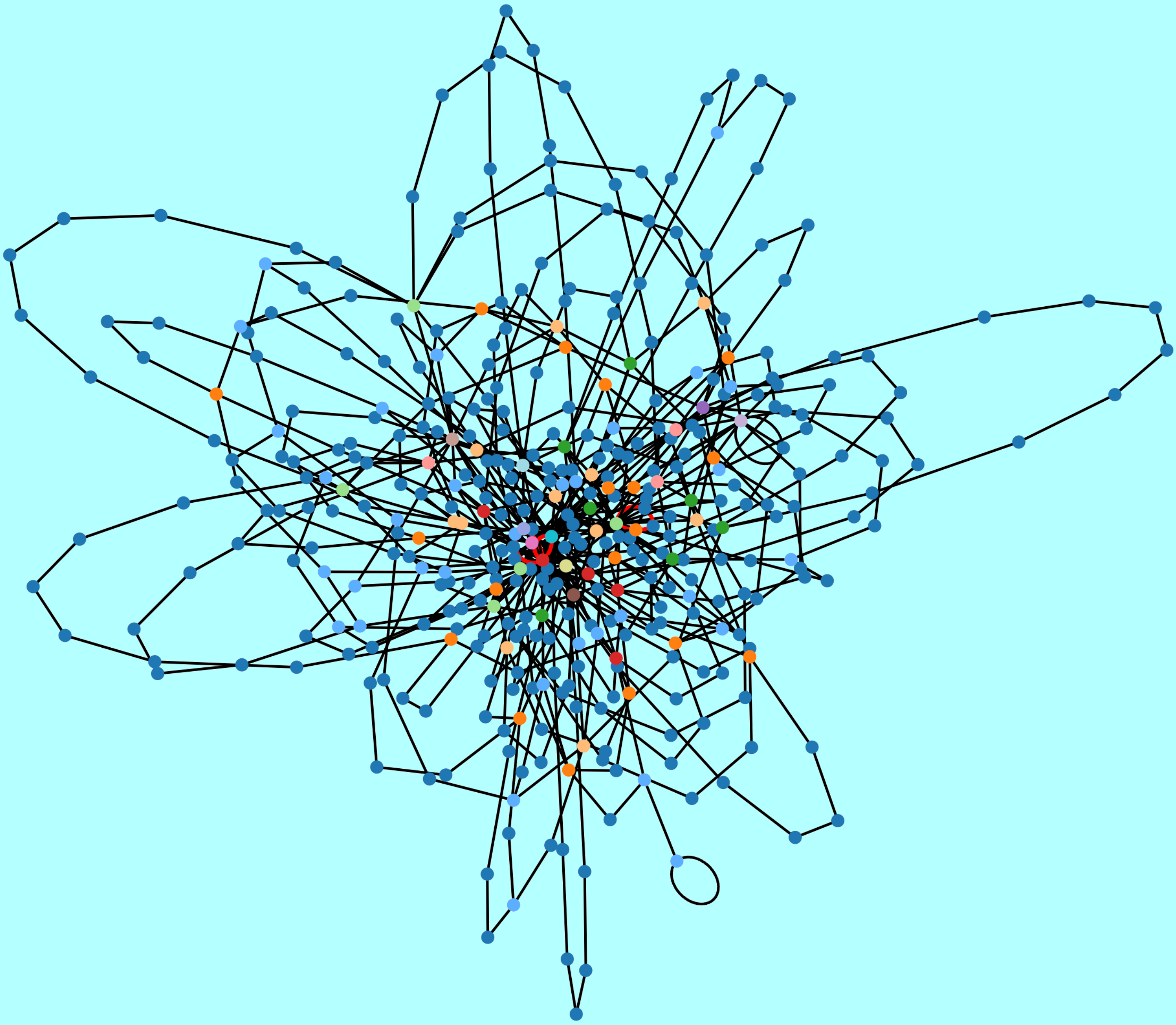
k(n)-cores
in the scale-free configuration model

Jesse Jaspers



TU Delft

k(n)-cores in the scale-free configuration model

Understanding the structure of a commonly used null model for scale-free networks

Jesse Jaspers

Student number: 5645867

Bachelor project thesis

BSc Applied Mathematics

EEMCS TU Delft

Supervisor: Júlia Komjáthy

Graduation committee: Anurag Bishnoi

Defense: 20 June 2024

Layman's summary

English version

In this research, we looked at a certain kind of network. This network is a commonly used way to simulate many real life networks. In the network we have a number of points. Then we determine how many connections we want to make to each point. Subsequently, we connect the points in a special random way that makes sure that every point has the desired number of connections. In this research, we attempt to find a group of points in the network, where every point has a minimum number of k connections to other points in the group. We have shown how k-values for which such a group exists depend on the number of points in the network. We hope to better understand the structure of the network by doing this.

Nederlandse versie (Dutch version)

In dit onderzoek bekijken we een bepaald soort netwerk, dat wordt gebruikt voor het nabootsen van veel netwerken die in de praktijk voorkomen. Het netwerk bestaat uit een aantal punten. Voor elk punt bepalen we hoeveel verbindingen we aan dit punt willen koppelen. Vervolgens maken we verbindingen in dit netwerk op een speciale willekeurige manier, zodat elk punt het gewenste aantal verbindingen heeft. In dit onderzoek, zoeken we een groep punten, waarbij elk punt in de groep minimaal k verbindingen heeft naar andere punten in de groep. Wij hebben laten zien hoe de waardes van k waarvoor zo een groep bestaat, afhangt van het aantal punten in het netwerk. Hiermee hopen we de structuur van het netwerk beter te begrijpen.

Summary

During this research, we investigate if there exists a $k(n)$-core in the scale-free configuration model, this is a commonly used null model to simulate networks. The scale-free configuration model produces a random graph, where the degree of every vertex is determined using a random variable. In this thesis, the discrete Pareto variable is used with parameter $\tau \in (2, 3)$. The k-core of a graph is the biggest induced subgraph where every vertex is connected to at least k edges in the subgraph. In this thesis, we let k be dependent on the number of vertices in a graph, this makes it a $k(n)$-core. By investigating whether $k(n)$-cores exist in graphs produced by the scale-free configuration model, we hope to to get a better understanding of the structure of these graphs.

In this thesis, the scale-free configuration model is modeled as a death process. This death process will produce the graph and its $k(n)$-core jointly. First, we removes all edges which cannot be part of the $k(n)$-core until the point where no such edges exist anymore. This is the moment where we reach the $k(n)$-core. Using this method, we prove that whenever the number of vertices is sufficiently high a $\log^{\alpha}(n)$-core exists with high probability for all constant $\alpha > 0$. For $\alpha < \frac{3-\tau}{8\tau}$, also a n^{α}-core exists with high probability when the number of vertices is sufficiently high. We also find a lower bound for the number of vertices and edges left in the cores.

Contents

Chapter 1

Introduction

A network is scale-free when the number of links to each node can be approximated by a power law distribution. This means that for D the distribution of the number of links to a node, it holds that $\mathbb{P}[D = x] \approx cx^{-\tau}$ for some constants $c > 0$ and $\tau > 0$ [17]. A lot of networks have this scale-free property. A commonly used null model to simulate scale-free networks is the scale-free configuration model. The configuration model forms a random graph such that the degrees of the vertices satisfy a given degree sequence. When these degrees are determined by a power law distribution, we call the configuration model scale-free. For every vertex, a number of half-edges is created that matches with the degree of the vertex. By creating a uniform random matching of the half-edges, the edges for the graph are created. In this thesis, we will attempt to better understand this commonly used null model.

First, we will discuss some networks that have been observed to be scale-free. Firstly, the network of websites linking to each other via hyperlinks has been observed to be scale-free [1]. Also the network of people connected via email is scale-free [8]. In addition, a network of scientific articles is scale-free [39]. In this networks, all articles are nodes and the links between articles are formed when one article cites the other one. Another example is the autonomous systems graph [10]. In this graph every node is an autonomous system, that is a group of routers which are under the same control [29]. When there is traffic flow between different autonomous systems, links are created between the nodes. Also in networks where the nodes are people, scale-free networks have been observed. The network of Hollywood actors who worked together has the scale-free property [3]. In Sweden it has even been observed that the network of people sleeping with each other is scale-free [25]. Scale-free networks also occur in biology. For 43 different organisms, a scale-free property was shown in the metabolism. It was shown that the metabolic reactions a molecule takes part in forms a scale-free network. In this network the molecules are the nodes and the links are formed when two molecules are in a biochemical reaction together [22]. Finally, the communication between different parts of the human brain has been observed to be scale-free [9]. Thus there are a lot of networks, for which the scale-free configuration model could form an interesting null model.

We would like to better understand the graphs that are formed by the scale-free configuration model. An important graph property is the k-core, this is the largest induced subgraph where every vertex is connected to at least k edges that are part of the subgraph. To analyze the structure of a graph, k-cores are really helpful. Core-decomposition is the main reason why k-cores are important. With core decomposition, for every vertex in the graph the highest value of k is determined such that the vertex is in the k-core [26]. This gives a way to create a hierarchy in a graph, where the highest k-core a vertex is in determines the hierarchy. The algorithm determines the importance of each vertex in a computation-efficient way. This makes it possible to quickly determine what the most important part of a network is, even when that network is large and complex. To visualize graphs, also an algorithm is used that is based on k-core decomposition [2]. The visualization shows important graph properties like the degrees of the vertices in the graph, the highest k-core a vertex is part of and whether a vertex has many neighbors in high k-cores. This can quickly give an overview of a graph, even for large graphs. An example of this visualization can be seen in Figure 1.1. Another important property of nodes in a k-core is that $k-1$ links can be removed and then the node is still connected to some node in the k-core. This gives information about the robustness of a network.

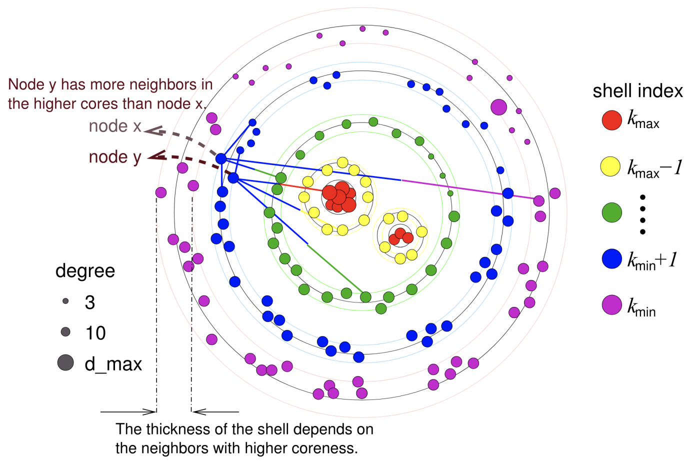


Figure 1.1: Visualization algorithm for graphs. The shell index is the highest k-core a node is a part of [2].

We will now discuss applied research in which k-cores are used to analyze networks. First we will look at research on online networks. To investigate the structure of the previously mentioned autonomous systems graph, k-cores have been used [27]. In this study it was also determined how the network of autonomous systems would react to power outages or a DDoS-attack. Also in research on social networks, k-cores have been used. It has for instance been investigated whether people remain engaged in a social network [38]. The research assumed that people only remained engaged when at least k of their friends were also active on the same platform, which meant this could be modeled as a k-core (as can be seen in Figure 1.2). Also the graph of all active Facebook users has been analyzed with the help of a k-core decomposition [34]. To do research on information spreading in networks, also k-cores are used [23].

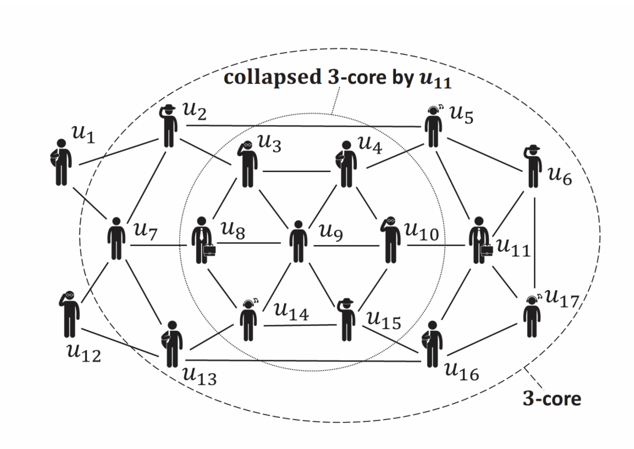
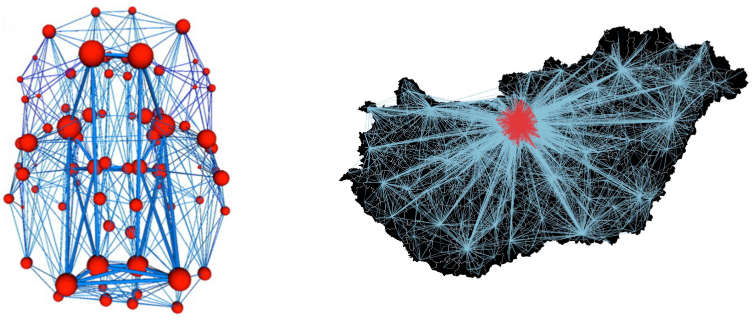


Figure 1.2: Network of friends that interact with each other on a social network. It is assumed that most people remain engaged when at least 3 friends are active on the platform. The outer circle contains all people who are likely to remain engaged, the inner circle contains the people who are likely to remain engaged after person u_{11} quits [38].

In biology and ecology k-cores are also used in research. A study on the structure of a human brain has been done using k-core decomposition [16], this can be seen in Figure 1.3a. In [30], the spread of an epidemic was investigated with the help of k-cores. The researchers used core decomposition to identify the towns in Hungary with most mutual commuting (see Figure 1.3b). Subsequently, they investigated the spread of the epidemic in two different situations. In the first case, the epidemic started only in the core, while in the second case the epidemic started uniformly across the country. This gave insight in the spread of the epidemic with different starting conditions. Lastly, we discuss the research done in mutualistic networks using k-core decomposition [14]. For 89 different mutualistic networks involving plant pollinators or seed dispersers it was analyzed which species were crucial for keeping an ecosystem alive. An overview of more studies which used k-cores can be found in [24].



(a) **(b)**

Figure 1.3: (a) Graph on connections between regions of a brain. Every node is a brain region. The thick blue lines are the edges in the core, while the thin lines are the other edges [16].
(b) Depiction of all commuting flows of more than 25 people in Hungary between towns with more than 1000 inhabitants. The core with most mutual commuting can be seen in red [30].

There has already been done research on k-cores in the configuration model. Most importantly, Janson & Luczak [20] have discovered conditions to check if a k-core exists with high probability when k is a constant. But first we will discuss research done by Fernholz & Ramachandran [11]. They determined the degree sequence for the scale-free configuration model with a special type of power law distribution, which is called the discrete Pareto distribution. For certain parameter values of this distribution and large enough number of vertices, they discovered that with high probability there would form a k-core for any positive constant k. In this thesis, we also use the discrete Pareto distribution. However, we shift focus to $k(n)$-cores. This means we let k be dependent on the number of vertices in the graph. We thus investigate if a k-core still exists when k scales with the number of vertices. By investigating $k(n)$-cores, we hope to get a better understanding of the highest k such that a k-core exists when we have n vertices. This is important for the previously explained core decomposition, which uses all different k values for which a k-core exists. Additionally, $k(n)$-cores could give a better understanding of how the highest k such that a k-core exists changes, when the number of nodes in a network changes.

In this thesis we elaborate on the ideas of Janson and Luczak [20], who formulated the configuration model as a death process. This death process builds the random graph and the k-core at the same time. In every step of the death process, two half-edges die to form an edge. First all edges outside the k-core are formed, then all remaining edges which are part of the k-core. This death process was used to obtain bounds on the number of edges left to make and the number of edges left that might go in the k-core. However, these bounds are not accurate enough to find a $k(n)$-core. Therefore, in thesis we will use empirical measures to obtain stricter bounds. This will allow us to extend the results to check whether a $k(n)$-core exists. This would also be an extension of the research from Fernholz & Ramachandran [11], who only looked at k-cores for constant k.

We will now discuss the structure of the rest of the thesis. First, in Chapter 2 we give background knowledge on $k(n)$-cores and the scale-free configuration model. In this chapter, we also state the main results of the thesis and explain how the configuration model can be modelled as a death process. Thereafter, in Chapter 3 we approximate the number of edges that could still be part of the $k(n)$-core after a certain time. Subsequently, in Chapter 4 we estimate the number of edges left in the death process after a certain time. Then in Chapter 5 we calculate the time when all remaining edges are part of the $k(n)$-core. In Chapter 6, we combine all previous knowledge to prove the main results. Finally, in Chapter 7 we look back at the research and do suggestions for further research.

Chapter 2

Preliminaries

In this thesis, we will look if a $k(n)$-core exists in a scale-free configuration model. But before we can look at this question, first some background knowledge is needed. First, in Section 2.1 it is explained what a $k(n)$-core is. Then, in Section 2.2 the discrete Pareto-distribution and the configuration model to create a random graph will be explained. Subsequently, in Section 2.3 the main results of this thesis will be stated. Furthermore, in Section 2.4 it is explained how the process of finding a $k(n)$-core in a configuration model can be modelled with the help of death processes. Finally, in Section 2.5 it will be explained how we can use this death process to find a $k(n)$-core.

2.1 The k(n)-cores

We will start this chapter by defining what a k-core is [20].

Definition 2.1.1 (k-core)**.** *Let G be a graph. For $k \in \mathbb{Z}_{>0}$, the k-core of G is the largest induced subgraph H_k of G where every vertex has a degree of at least k within the subgraph H_k.*

The k-core of a graph is unique and can be found by recursively removing vertices of degree less than k. It is important to update the degrees after every removal step since we also remove edges connected to the removed vertices. This can mean other vertices also drop below k edges. If there are no vertices left at all after every vertex with degree lower than k has been removed, the k-core is empty. If there are still vertices left, than the induced subgraph between these vertices forms the k-core. In the following example, the k-cores of a small graph are drawn.

Example 2.1.2. *In the figure below, all k-cores of a graph are drawn. The 3-core consists of the orange vertices. The 2-core is formed by the orange and green vertices. Lastly, the 1-core is formed by all vertices. For $k \geq 4$, the k-core is empty.*

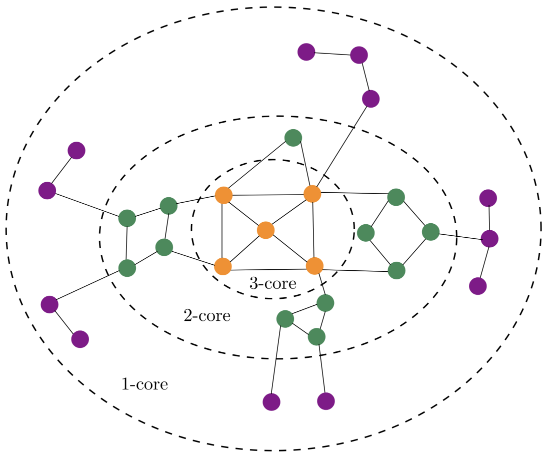


Figure 2.1: The k-cores of a graph [26].

We will now extend Definition 2.1.1 to be dependent on the number of vertices in the graph.

Definition 2.1.3 ($k(n)$-core)**.** *Let G be a graph with n vertices. Let $k(n)$ be a function, then the $k(n)$-core of G is the largest induced subgraph of G where every vertex has a degree of at least $k(n)$.*

Remark 2.1.4. *During this thesis, $k(n)$ is sometimes written as k_n for shorter notation.*

In the next section, we will introduce the configuration model that forms the random graphs we will investigate $k(n)$-cores in.

2.2 Configuration model

In this thesis we work with random graphs. Instead of the Erdős-Rényi models $G(n, p)$ and $G(n, m)$, we will create a random graph with a given degree distribution. Before we explain why it is interesting to look at $k(n)$-cores in this graph, a configuration model for this random graph will be defined .

Definition 2.2.1 (Configuration model [20])**.** *Let $n \in \mathbb{Z}_{>0}$ and $(d_i)_1^n = (d_1, d_2, \ldots, d_n)$ a sequence of non-negative integers such that $\sum_{i=1}^n d_i$ is even. To create a graph where there are n vertices that have degrees $(d_i)_1^n$, d_i half edges are created for every vertex v_i. Then the edges of the graph are decided by taking a uniform matching of the half-edges.*

During this thesis we will need some conditions on the degree sequence $(d_i)_1^n$. We will now introduce these conditions, later this section we verify they are true for the degree sequence we want to use.

Condition 2.2.2 (Conditions on degree sequence [20]). *Let $(d_i)_1^n$ be a sequence of non-negative integers with $\sum_{i=1}^n d_i$ even. Then there needs to be a probability distribution such that $(p_r)_{r=0}^{\infty}$*
(i) $\frac{\#\{i:d_i=r\}}{n} \to p_r$ for every $r \geq 0$ as $n \to \infty$
(ii) For D with distribution $(p_r)_{r=0}^{\infty}$, it holds $M := \mathbb{E}[D] = \sum_{r=1}^{\infty} rp_r \in (0, \infty)$
(iii) $\frac{1}{n}\sum_{i=1}^n d_i \to M$ as $n \to \infty$

In this thesis, we will use the discrete Pareto distribution in the configuration model, since it is useful for modeling scale-free networks. We will now define the discrete Pareto distribution.

Definition 2.2.3 (Discrete Pareto distribution). *A variable X is discrete Pareto distributed with parameter τ if $\mathbb{P}[X = x] = \frac{1}{\zeta(\tau)}x^{-\tau}$ for $x \in \mathbb{Z}_{>0}$, where $\zeta(\tau) = \sum_{i=1}^{\infty} i^{-\tau}$ is the Riemann-Zeta function. We denote this as $X \sim dpareto(\tau)$.*

Remark 2.2.4. *The discrete Pareto distribution is also known as the Zipf distribution or the (Riemann) Zeta distribution [32].*

In this thesis, the discrete Pareto distribution will be used with $\tau \in (2, 3)$. The discrete Pareto distribution is a power-law distribution. For a power-law distribution, we have a slowly varying function $L(x)$, which means that for all $t > 0$

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1. \tag{2.1}$$

Then a variable X has a power-law distribution when [18, Definition 1.19]

$$\mathbb{P}[X > x] = L(x)x^{-(\tau-1)}. \tag{2.2}$$

This also means that for a constant $c > 0$ the following holds [18]:

$$\mathbb{P}[X = x] \approx c * x^{-\tau}. \tag{2.3}$$

Clearly this holds for the discrete Pareto distribution and therefore it is a power law distribution. A characteristic feature of a power law distribution is that a $\log$-$\log$ plot of the degrees against the proportion of vertices with that degree gives an approximately straight line. Using a logarithm on (2.3), namely gives the form:

$$\log(\mathbb{P}[X = x]) \approx c - \tau * \log(x). \tag{2.4}$$

Example 2.2.5. *To demonstrate the property from* (2.4)*, we make a* log*-*log *plot of a dpareto*(2.2)*-sample with* 500 *vertices.*

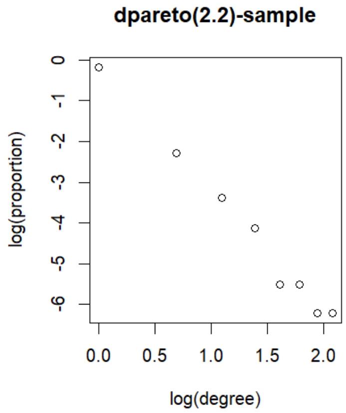


Figure 2.2: A log-log-plot of the degrees and proportion of vertices that have a degree of a dpareto(2.2)-sample with 500 vertices.

Let us check if we can satisfy Condition 2.2.2 with the discrete Pareto distribution, first we check if Condition 2.2.2(ii) holds for $X \sim$ dpareto(τ) with $\tau \in (2,3)$.

$$\begin{aligned}\mathbb{E}[X] &= \sum_{m=1}^{\infty} \frac{m}{\zeta(\tau)} m^{-\tau} \\ &= \frac{1}{\zeta(\tau)} \sum_{m=1}^{\infty} m^{-(\tau-1)}.\end{aligned} \tag{2.5}$$

Note that $\zeta(\tau-1) = \sum_{m=1}^{\infty} m^{-(\tau-1)}$, therefore we obtain

$$\mathbb{E}[X] = \frac{\zeta(\tau-1)}{\zeta(\tau)}. \tag{2.6}$$

For the Riemann-Zeta function it is known that $\zeta(x) < \infty$ for $x > 1$. It can also be seen that for all real x, $\zeta(x)$ is positive. Therefore for $\tau > 2$, both $\zeta(\tau) \in (0,\infty)$ and $\zeta(\tau-1) \in (0,\infty)$. Therefore, also the following equation holds.

$$\mathbb{E}[X] = \frac{\zeta(\tau-1)}{\zeta(\tau)} \in (0,\infty). \tag{2.7}$$

And thus Condition 2.2.2(ii) holds for a dpareto(τ) distribution with $\tau \in (2,3)$. Let us now check if we can satisfy the remaining conditions with a discrete Pareto sample. In case we sample the degrees d_i i.i.d. discrete Pareto, then it could be that the sum of the degrees is odd. To fix this, we use the approach from [17, Section 7.2]. If the sum of all d_i is odd, then we add 1 to d_n. We will now make a claim about this sampling.

Claim 2.2.6. *The explained sampling of the d_i satisfies conditions 2.2.2(i) and 2.2.2(iii).*

A proof of this claim can be found in [17, Section 7.2]. Therefore we can use this sampling of the discrete Pareto distribution to satisfy Condition 2.2.2.

Now we have shown that the scale free configuration model can be used with the discrete Pareto distribution, we will see why this is interesting. A network is called scale-free if the distribution of the vertex degrees can be well approximated by power laws. The reason that we work with the discrete Pareto distribution in this thesis, is that it is a simple distribution that follows a power law. We work with $\tau \in (2,3)$, since this parameter is commonly seen in scale-free networks [3]. The scale-free property is often seen in networks, as we already already discussed in the introduction. There has recently been discussion whether scale-free models are the best way to model many problems (see e.g. [5]). However, they are still seen as a useful modeling method [18].

As explained in the introduction, k-cores are an in important way to investigate the structure of a network. There has been previously done research on k-cores in random graphs where the vertices-degrees are dpareto(τ)-distributed for $\tau \in (2,3)$. Before we state a theorem from this research, we need to introduce a definition [20].

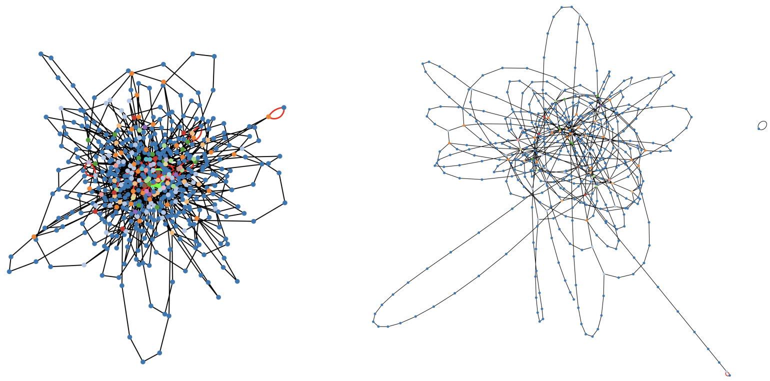
Definition 2.2.7 (whp)**.** *An event $\mathcal{E}_n$ holds with high probability (whp) if $\mathbb{P}\left[\mathcal{E}_n\right] \to 1$ as $n \to \infty$.*

Now we state a results that holds with high probability.

Theorem 2.2.8 (Existence k-cores [11])**.** *Take a random graph where the vertices-degrees are dpareto(τ)-distributed for $\tau \in (2,3)$. Then for arbitrary $k \in \mathbb{Z}_{\geq 0}$ and large enough number of vertices, there exists a k-core with high probability.*

To show that cores actually occur, we will now show an example of a graph produced in the scale-free configuration model with 500 vertices. We will also show this core does not form for a certain τ-value larger than 3.

Example 2.2.9. *We simulate the scale-free configuration model with 500 vertices, where the vertex degrees obey a power law. It can clearly be seen that for $\tau = 2.5$ a core exists, while for $\tau = 3.2$ no cohesive core arises.*



(a) $\tau = 2.5$ **(b)** $\tau = 3.2$

Figure 2.3: Simulation of the scale-free configuration model for 500 vertices, when the degrees are determined by a power law. The parameter τ is set on 2.5 and 3.2 respectively [13].

By Theorem 2.2.8 we now know that for constant $k \in \mathbb{Z}_{\geq 0}$, there exists a k-core for large enough graph size. Therefore in this thesis, we will shift focus to $k(n)$-cores. So it will be investigated if there exists a k-core where k scales with the number of vertices. Now, we have introduced the necessary background information, in the next section we will introduce the main results that will be derived in this thesis.

2.3 Main results

In this section, we state the results that will be proven during the rest of the thesis. The goal of the thesis is to discover whether a $k(n)$-core exists in the scale-free configuration model. To be able discuss the number of vertices and edges in the $k(n)$-core, we need one more definition.

Definition 2.3.1 (Asymptotic bounds). *y is $\Theta(x)$ if $\lim_{n\to\infty} \frac{y}{x} = a \in \mathbb{R} \setminus \{0\}$. Furthermore, $y < \Theta(x)$ means $\lim_{n\to\infty} \frac{y}{x} = 0$ and $y > \Theta(x)$ means $\lim_{n\to\infty} \frac{y}{x} = \infty$.*

Now we look at $k(n)$-cores for the case where $k(n)$ is polylogarithmic.

Theorem 2.3.2 (Polylogarithmic $k(n)$-cores)**.** *Consider the configuration model from Definition 2.2.1 for a degree sequence that satisfies the conditions in 2.2.2 with $(p_r)_{r=0}^{\infty}$ discrete Pareto distributed with parameter $\tau \in (2,3)$. Then for any constant $\alpha > 0$ there is a $\log^{\alpha}(n)$-core for all sufficiently large n whp. Furthermore, the number of vertices in the core is whp minimally*

$$\Theta\left(\frac{n}{(\log(n))^{\alpha(\tau-1)/(3-\tau)}}\right), \tag{2.8}$$

and the number of edges in the core is whp minimally

$$\Theta\left(\frac{n}{(\log(n))^{2\alpha(\tau-2)/(3-\tau)}}\right). \tag{2.9}$$

Now, we will look at a second form for $k(n)$. This time

Theorem 2.3.3 (Polynomial $k(n)$-cores)**.** *Consider the configuration model from Definition 2.2.1 for a degree sequence that satisfies the conditions in 2.2.2 with $(p_r)_{r=0}^{\infty}$ discrete Pareto distributed with parameter $\tau \in (2,3)$. Then for all $\alpha \in \left(0, \frac{3-\tau}{8\tau}\right)$ there is a n^{α}-core for all sufficiently large n with the number of vertices in it being whp minimally*

$$\Theta\left(n^{1-\alpha(\tau-1)/(3-\tau)}\right), \tag{2.10}$$

and the number of edges in the core is whp minimally

$$\Theta\left(n^{1-2\alpha(\tau-2)/(3-\tau)}\right). \tag{2.11}$$

It is unknown whether for the conditions in Theorem 2.3.3, a n^{α}-core also exists for $\alpha \geq \frac{3-\tau}{8\tau}$. The currently given values of α in this theorem are a consequence of a technical artifact in the proof. Now we have stated the main results, in the next section we turn to proof techniques by modeling the configuration model as a death process.

2.4 Death process

To analyze the configuration model, we would like to relate it to a different process which can be expressed into formulas easier. Janson & Luczak [20], have actually found a way to relate this configuration model to a death process. A death process is a special type of Markov process (see e.g. [15, Section 11.5]). It will now be explained how the configuration model can be modeled as a death process.

Remember that the configuration model from Definition 2.2.1 starts with a sequence of vertex degrees $(d_i)_1^n = (d_1, d_2, \dots, d_n)$. We start formulating the configuration model as a death process by creating n boxes, filled with $d_1, d_2, \dots d_n$ balls respectively. Every box represents a vertex in the graph, while the number of balls in a box is the number of edges that should be connected to the vertex. We call the balls in the box half-edges, and two half-edges will be connected every time to form the edges of the graph.

We actually want to build the graph and the k_n-core at the same time. This will be done by first creating all edges which are not part of the k_n-core. Once we have created these edges, there are two options. Either there are no balls left into all bins or there are still balls left in some bins. In case there are no balls left, there are no half-edges available anymore to form new edges. This would mean there are no edges left in the k_n-core and therefore the k_n-core is empty. However, if there are still balls in some bins, these half-edges will form the edges of the k_n-core. This would mean the k_n-core is non-empty.

For this algorithm it is important to know which edges cannot be part of the k_n-core. By Definition 2.1.1 a vertex can only be part of the k_n-core if it has degree at least k_n, otherwise it definitely cannot have degree k_n in an induced subgraph. Therefore all balls that are in bins that have less than k_n balls will not become part of the k_n-core anymore. We will call these balls light. Contrary, all balls that are in a bin that has at least k_n balls, are called heavy balls. During every step of the process, we remove one light ball. Then, a random ball (which can be either heavy or light) is chosen uniformly to match to the light ball. This chosen ball is also removed from the box it is in. The two chosen half-edges then form an edge, which is not part of the k_n-core. If the removal of the balls causes a bin to have less than k_n balls, the balls in this bin are now changed from heavy to light.

The process is terminated once we need to remove a light ball, but there are no light balls left. Since there are no light balls, there are no balls left in bins with less than k_n balls. The remaining bins with balls thus have at least k_n balls. By uniformly pairing the remaining half-edges, we thus create an induced subgraph where every vertex has degree at least k_n. Therefore these edges form a k_n-core by Definition 2.1.1. Note that during this process we made a uniform random pairing of the half-edges (as was asked by the configuration model), this was just done in a smart order. Now it is time to explain how this balls-into-bins method can be seen as a death process.

In the process that was just explained, we remove two balls every time. First, a light ball and then a random one. To get a death process we would like to reverse this order. Therefore we remove a single light ball first. After this is done, we will still remove two balls at a time. But now, we first remove a random ball and then a light ball. In a death process, at every step the index of the removed ball is also random [35]. Furthermore, in a death process with m individuals, the time until the next jump is exponentially distributed with rate m. To make the removal a death process, we therefore use the following trick. If there are m balls left after a step, we wait an $\exp(m)$-distributed time until we start the next step. After every step, we jump size 2 downwards (since we also remove a light ball).

We now have a death process, which is slightly different then the configuration model. The only difference is that we removed one ball at the start to switch the order of removing light and random balls. The impact of removing the first ball is negligible, as the number of balls tends to infinity when the number of boxes tends to infinity. However, switching the order causes one problem. We terminate the process once there are no light balls left when we need to choose one. During the last step, we do remove a random ball. We then also need to remove a light ball, because the death process jumps down by 2. But there are no light balls left, to fix this we set the number of light balls after the last step to -1. This has an interesting consequence. We know that the total number of balls is the the sum of the number of light and heavy balls. But since the number of light balls is negative, the number of heavy balls in the expression is bigger than the number of total balls.

Let us summarize the algorithm we have described:

Step 0. Remove one light ball

Step 1a. Wait an $\exp$ (#balls)-time to remove a random ball

Step 1b. Immediately remove a random light ball too. If this ball exists, go back to step 1a. If this ball does not exist, go to step 2.

Step 2. Set the number of light balls to -1.

Step 3. If balls still exist, pair two random balls. These balls form an edge in the k_n-core. Repeat this step until there are no balls left

Let us check whether the described algorithm actually follows the configuration model. For this, we state a claim.

Claim 2.4.1. *The described algorithm produces a uniform random matching.*

This claim holds by [17, Section 7.2]. Since we have a uniform random matching on the half-edges, we satisfy the definition of the configuration model (Definition 2.2.1). In the next section we will explain how we will use the described algorithm to prove the main results.

2.5 Arriving at k(n)-core

In the last section, we derived that the number of heavy balls is larger than the total number of balls, at the moment when we reach the k_n-core. Let $L(t)$ be the number of light balls at time t and $H(t)$ be the number of heavy balls at time t. Also let $B(t)$ be the total number of balls a time t. We want to know the first time $t \geq 0$ such that:

$$L(t) = B(t) - H(t) < 0. \tag{2.12}$$

Exact expression for $B(t)$ and $H(t)$ are hard to find. But let us rewrite the statement to

$$L(t) = \mathbb{E}[B(t)] - \mathbb{E}[H(t)] + B(t) - \mathbb{E}[B(t)] + \mathbb{E}[H(t)] - H(t) < 0. \tag{2.13}$$

Assume now that for large enough n, the following statement holds:

$$B(t) - \mathbb{E}[B(t)] + \mathbb{E}[H(t)] - H(t) < n^{\delta}. \tag{2.14}$$

If we use this assumption (2.14) in (2.13), then we see that

$$L(t) < \mathbb{E}[B(t)] - \mathbb{E}[H(t)] + n^{\delta}. \tag{2.15}$$

Therefore if we manage to prove that,

$$\mathbb{E}\left[B(t)\right] - \mathbb{E}\left[H(t)\right] + n^{\delta} \leq 0, \tag{2.16}$$

then by using (2.15), also (2.12) holds. By rewriting (2.16), we obtain the following goal:

$$\mathbb{E}\left[B(t)\right] - \mathbb{E}\left[H(t)\right] \leq -n^{\delta}. \tag{2.17}$$

If we can obtain a time t^* such that this equation holds, $H(t^*)$ gives an expression for the minimum number of edges in the $k(n)$-core. If this number of edges is positive, than we are sure that the $k(n)$-core cannot be empty. This would mean that the $k(n)$-exists. Obtaining this reaching time t^* will be done in Chapter 5. In Chapter 6, we find the minimum number of edges and vertices at this reaching time. However, we have made an assumption in (2.14) that needs to be satisfied. We split this assumption up into two statements that are sufficient to satisfy (2.14) . This are the following statements, that need to be satisfied for large enough n:

$$|B(t) - \mathbb{E}\left[B(t)\right]| < \frac{n^{\delta}}{2}, \tag{2.18}$$

$$|H(t) - \mathbb{E}\left[H(t)\right]| < \frac{n^{\delta}}{2}. \tag{2.19}$$

In Chapter 4, we will find out for which δ-values (2.18) is satisfied. But first in Chapter 3 we will find a value for δ such that the equation in (2.19) holds.

Chapter 3

Approximating heavy half-edges

In this chapter, we obtain a bound on the number of heavy half-edges $H(t)$ being close to its expectation. First, in Section 3.1 the expected number of heavy half edges is calculated. In Section 3.2 we also determine the expected number of vertices. Furthermore, in Section 3.3 we determine the desired bound on the number of heavy half-edges $H(t)$ being close to its expectation. Finally, in Section 3.4 we bound the expected number of heavy boxes being close to its expectation. This will allow us to say something about the number of vertices and edges in the $k(n)$-core

3.1 Expected number of heavy half-edges

In this section, we would like to calculate the expected number of heavy half-edges. Janson & Luczak [20], have already discovered that if a bin starts with $l \geq k_n$ balls and has $j \geq k_n$ balls at time t, then the number of balls at time t is Bin(l, e^{-t})-distributed. Note that the condition that there are at least $j \geq k_n$ balls at time t means that the bin is still heavy at time t. Since we remove maximum one heavy ball in every step of the death process, we know that the number of balls left is expected to be e^{-t} times its starting value. Since we choose the balls uniformly random, the chance that a ball in a heavy bin still exists at time t is e^{-t}, therefore the number of balls remaining in a heavy bin that started with $l \geq k_n$ balls and remained heavy is Bin(l, e^{-t}).

Remark 3.1.1. *Once the number of balls in the bin drops below k_n, all the balls in the bin become light. Thus there are now two possible chances at every step of the death process for the ball to be chosen. Therefore the number of balls is not Bin(l, e^{-t})-distributed anymore after a bin becomes light. However, since we only care about the number of heavy balls and boxes, this does not cause a problem.*

Using the property that the number of balls in heavy boxes is Bin(l, e^{-t})-distributed, we will now slightly alter the results obtained by Janson & Luczak [20], about the expected number of heavy half-edges. First, we define

$$p_l(n) := \frac{\#\{i \in \{1, \ldots, n\} : d_i = l\}}{n}, \tag{3.1}$$

such that $n_l := np_l$ is the number of bins that start with l balls. By Condition 2.2.2(i), it is known that $p_l(n) \to p_l$ as $n \to \infty$. This also means that for large n, we have $c_1 p_l \le p_l(n) \le c_2 p_l$ for constants $c_1 < 1$ and $c_2 > 1$. Let $X_l(t) \sim$ Bin(l, e^{-t}). We define

$$h\left(e^{-t}\right) := \mathbb{E}\left[H(t)\right] = \sum_{l=k_n}^{\infty} p_l(n) n \mathbb{E}\left[X_l(t) \mathbb{1}_{\{X_l(t) \ge k_n\}}\right]. \tag{3.2}$$

Let us write out the expectation in this sum to get

$$h\left(e^{-t}\right) = \sum_{l=k_n}^{\infty} p_l(n) n \sum_{r=k_n}^{l} r \mathbb{P}\left[X_l(t) = r\right]. \tag{3.3}$$

We have now found an expression for the expected number of heavy balls left at time t, since these heavy balls represent the number of heavy half-edges at this time. If we take a the time in the death process where the $k(n)$-core is reached, then we have the expected number of half-edges left to form the $k(n)$-core. Therefore $\frac{1}{2} h\left(e^{-t}\right)$ is the expected number of edges in the $k(n)$-core. We will now briefly look at the expected number of vertices.

3.2 Expected number of heavy boxes

In the previous section, the expected number of heavy half balls left at time t was calculated. To get the expected number of vertices, we need to calculate how many heavy boxes there are left at the time when we reach the $k(n)$-core. Therefore let us find an expression of the number of heavy boxes expected a time t. Note that in (3.3), we expressed the expected number of heavy balls by calculating the expected number of heavy boxes with j balls left. And then this was multiplied with a factor j to get the number of balls. However, if we remove this factor, we get the number of boxes. Therefore we let $X_l(t) \sim$ Bin(l, e^{-t}). Then gives us the following expression if we denote the number of boxes by $V(t)$:

$$v(e^{-t}) := \mathbb{E}\left[V(t)\right] = \sum_{l=k_n}^{\infty} p_l(n) n \sum_{r=k_n}^{l} \mathbb{P}\left[X_l(t) = r\right]. \tag{3.4}$$

Let us also write out the function in a more compact form, this is

$$v(e^{-t}) = \sum_{l=k_n}^{\infty} p_l(n) n \mathbb{P}\left[X_l(t) \ge k_n\right]. \tag{3.5}$$

In the next section, we will obtain a bound on the number of heavy half-edges being close to its expectation. From this, a bound on the number of heavy boxes being close to its expectation will follow in Section 3.4.

3.3 Heavy half-edges close to expectation

We want to get a bound on the number of half-edges being close to its expectation. We define the number of boxes with r balls at time t as $U_r(t)$. In the article by Janson & Luczak [20, Lemma 4.4], the following result is already shown:

$$\mathbb{P}\left[\sup_{t\geq 0}\left|h\left(e^{-t}\right)-\sum_{r=k_n}^{\infty} rU_r(t)\right|>n\right]\to 0. \tag{3.6}$$

We would like to obtain a better bound then n. To achieve this, we state a result on empirical measures. This result is a strengthened version of the Dvoretzky–Kiefer–Wolfowitz inequality [7].

Theorem 3.3.1 (Convergence empirical measure to distribution function [28])**.** *Let $x_1, x_2, \dots, x_n$ be an i.i.d. sample of a distribution with continuous distribution function $F(x)$. Define the empirical distribution function as following:*

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i\leq x\}}. \tag{3.7}$$

Then the following statement holds for $\lambda > 0$:

$$\mathbb{P}\left[\sup_{x\geq 0}\sqrt{n}\left|F_n(x)-F(x)\right|>\lambda\right]\leq 2\exp\left(-2\lambda^2\right). \tag{3.8}$$

We will now use this result to prove the main theorem of this section.

Theorem 3.3.2 (whp bound on heavy balls)**.** *For large enough graph size n and $\hat{\beta} = \frac{3\tau+2}{4\tau}$, the following equation holds for the function h from (3.2):*

$$\mathbb{P}\left[\sup_{t\geq 0}\left|h\left(e^{-t}\right)-H(t)\right|\geq \frac{n^{\hat{\beta}}}{2}\right]\to 0. \tag{3.9}$$

Proof. We begin this proof by writing out $h\left(e^{-t}\right)$ and $H(t)$ to get that we need to prove that

$$\mathbb{P}\left[\sup_{t\geq 0}\left|\sum_{l=k_n}^{\infty} np_l(n)\sum_{r=k_n}^{l} r\mathbb{P}\left[X_l(t)=r\right]-\sum_{r=k_n}^{\infty} rU_r(t)\right|\geq \frac{n^{\hat{\beta}}}{2}\right]\to 0. \tag{3.10}$$

To make notation easier, binomial chances are denoted in the following way,

$$m_{lr}(t) := \mathbb{P}\left[X_l(t)=r\right]=\binom{l}{r}\left(e^{-t}\right)^r\left(1-e^{-t}\right)^{l-r}. \tag{3.11}$$

We will also define $U_{lr}(t)$ as the number of boxes that have l balls at the start and $l\leq r$ balls left at time t. We use this to note that

$$\sum_{r=k_n}^{\infty} rU_r(t)=\sum_{r=k_n}^{\infty}\sum_{l=r}^{\infty} rU_{lr}(t)=\sum_{l=k_n}^{\infty}\sum_{r=k_n}^{l} rU_{lr}(t). \tag{3.12}$$

We use this to rewrite (3.10) to

$$\mathbb{P}\left[\sup_{t\geq 0}\left|\sum_{l=k_n}^{\infty} np_l(n)\sum_{r=k_n}^{l} rm_{lr}(t)-\sum_{l=k_n}^{\infty}\sum_{r=k_n}^{l} rU_{lr}(t)\right|\geq \frac{n^{\hat{\beta}}}{2}\right]. \tag{3.13}$$

We would like to split the expression into multiple parts. After this, we show that each of these parts is smaller than $\Theta\left(n^{\hat{\beta}}\right)$ with probability tending to 1. This would mean that also the sum of these parts is smaller than $\frac{n^{\hat{\beta}}}{2}$ with probability tending to 1. To be able to split up the expression, we first define

$$\begin{aligned} H_L(t) &:= \sum_{l=k_n}^{L_n} \sum_{r=k_n}^{l} r U_{lr}(t), \\ h_L\left(e^{-t}\right) &:= \sum_{l=k_n}^{L_n} n p_l(n) \sum_{r=k_n}^{l} r m_{lr}(t). \end{aligned} \tag{3.14}$$

The absolute value in (3.9) can now be written as

$$\left| h\left(e^{-t}\right) - H(t) + h_L\left(e^{-t}\right) - h_L\left(e^{-t}\right) + H_L(t) - H_L(t) \right| \tag{3.15}$$

We would like to re-order this equation in a different way. After this is done, the triangle inequality is applied to divide the expression into three different parts.

$$\begin{aligned} &\left| \left(h_L\left(e^{-t}\right) - H_L(t)\right) + \left(H_L(t) - H(t)\right) + \left(h\left(e^{-t}\right) - h_L\left(e^{-t}\right)\right) \right| \\ &\leq \left| h_L\left(e^{-t}\right) - H_L(t) \right| + \left| H(t) - H_L(t) \right| + \left| h_L\left(e^{-t}\right) - h\left(e^{-t}\right) \right|. \end{aligned} \tag{3.16}$$

The goal of the rest of this proof is to bound these three absolute values. We want to discover if they are all smaller than $\Theta\left(n^{\hat{\beta}}\right)$ with probability tending to 1 for smartly chosen $\hat{\beta}$ and L_n. First the focus will lie on $\left|H(t) - H_L(t)\right|$:

$$\begin{aligned} \left|H(t) - H_L(t)\right| = \sum_{L_n}^{\infty} \sum_{r=k_n}^{l} r U_{lr}(t) &\leq \sum_{L_n}^{\infty} l \sum_{r=k_n}^{l} U_{lr}(t) \\ &\leq \sum_{L_n}^{\infty} l n p_l(n). \end{aligned} \tag{3.17}$$

Let us now work out a second term from (3.16),

$$\begin{aligned} \left| h_L\left(e^{-t}\right) - h\left(e^{-t}\right) \right| &= \sum_{l=L_n}^{\infty} n p_l(n) \sum_{r=k_n}^{l} r m_{lr}(t) \\ &\leq \sum_{l=L_n}^{\infty} l n p_l(n). \end{aligned} \tag{3.18}$$

We see that the same upper bound can be used for the terms in (3.17) and (3.18). Let us see if we can approximate this upper bound by using an integral. Since $p_l(n) \leq c_2 l^{-\tau}$ for $c_2 > 1$ and large enough n, it holds

$$\sum_{l=L_n}^{\infty} l n p_l(n) \leq c_2 n \sum_{l=L_n}^{\infty} l^{1-\tau} \leq c_2 n \int_{L_n-1}^{\infty} u^{1-\tau} du. \tag{3.19}$$

We can calculate this integral, where it is important to notice that $2 - \tau < 0$:

$$c_2 n \int_{L_n-1}^{\infty} u^{1-\tau} du = c_2 n \left[\frac{1}{2-\tau} u^{2-\tau} \right]_{L_n-1}^{\infty} = \frac{c_2 n}{\tau - 2} \left(L_n - 1\right)^{2-\tau}. \tag{3.20}$$

We have now created upper bounds for two terms in (3.16). We will choose L_n in a convenient way later this proof, to make sure that these terms are below $\Theta\left(n^{\hat{\beta}}\right)$. Consequently, we get the following condition that needs to be satisfied:

$$\frac{c_2 n}{\tau-2}\left(L_n-1\right)^{2-\tau}<\Theta\left(n^{\beta}\right). \tag{3.21}$$

Before choosing $\hat{\beta}$ and L_n, we will first investigate what is need to bounded the remaining term in (3.16). This term is

$$\begin{aligned}\left|h_L\left(e^{-t}\right)-H_L(t)\right| &=\sup_{t\geq 0}\left|\sum_{l=k_n}^{L_n} np_l(n)\sum_{r=k_n}^{l} rm_{lr}(t)-\sum_{l=k_n}^{L_n}\sum_{r=k_n}^{l} rU_{lr}(t)\right| \\ &=\sup_{t\geq 0}\left|\sum_{l=k_n}^{L_n}\sum_{r=k_n}^{l} r\left(np_l(n)m_{lr}(t)-U_{lr}(t)\right)\right|.\end{aligned} \tag{3.22}$$

By taking the sum out, we obtain a larger value. This gives the following upper bound,

$$\sum_{l=k_n}^{L_n}\sum_{r=k_n}^{l} r\sup_{t\geq 0}\left|np_l(n)m_{lr}(t)-U_{lr}(t)\right|. \tag{3.23}$$

We would like to bound all terms in the following way for a certain function $\lambda(n,l)$,

$$\sup_{t\geq 0}\left|n_l m_{lr}(t)-U_{lr}(t)\right|\leq\sqrt{n_l}\lambda(n,l). \tag{3.24}$$

This would give the following upper bound for (3.23),

$$\sum_{l=k_n}^{L_n}\sum_{r=k_n}^{l} r\sup_{t\geq 0}\left|np_l(n)m_{lr}(t)-U_{lr}(t)\right|\leq\sum_{l=k_n}^{L_n}\sum_{r=k_n}^{l} r\sqrt{n_l}\lambda(n,l). \tag{3.25}$$

If it is true that we can use this upper bound, then we have the following condition that is sufficient for convergence of the remaining term in (3.16):

$$\sum_{l=k_n}^{L_n}\sum_{r=k_n}^{l} r\sqrt{n_l}\lambda(n,l)<\Theta\left(n^{\hat{\beta}}\right). \tag{3.26}$$

We will use Theorem 3.3.1 to see if this bound holds with probability tending to 1. First, we note that $\left|n_l m_{lr}(t)-U_{lr}(t)\right|$ can be rewritten to

$$\begin{aligned}&\left|\left(\sum_{s=0}^{r} n_l m_{ls}(t)-\sum_{s=0}^{r-1} n_l m_{ls}(t)\right)+\left(\sum_{s=0}^{r-1} U_{ls}(t)-\sum_{s=0}^{r} U_{ls}(t)\right)\right| \\ &=\left|\left(\sum_{s=0}^{r} n_l m_{ls}(t)-\sum_{s=0}^{r} U_{ls}(t)\right)+\left(\sum_{s=0}^{r-1} U_{ls}(t)-\sum_{s=0}^{r-1} n_l m_{ls}(t)\right)\right|.\end{aligned} \tag{3.27}$$

For convenient notation, we denote

$$\varepsilon_1 := \left| \sum_{s=0}^{r} (n_l m_{ls}(t) - U_{ls}(t)) \right|, \tag{3.28}$$

$$\varepsilon_2 := \left| \sum_{s=0}^{r-1} (U_{ls}(t) - n_l m_{ls}(t)) \right|. \tag{3.29}$$

Applying the triangle inequality then gives:

$$\left| n_l m_{lr}(t) - U_{lr}(t) \right| \leq \varepsilon_1 + \varepsilon_2. \tag{3.30}$$

We will now see ε_1 and ε_2 can be estimated using Theorem 3.3.1. For this theorem, we need an i.i.d. sample. It is known that all boxes that start with l balls are $\text{Bin}(l, e^t)$ distributed. Let us order all n_l boxes that start with l balls, by assigning an index $i \in \{1, 2, \ldots, n_l\}$ to each box. We will look at the moment when the j-th ball is removed from a box. Now let $T_{ij}^{(l)}$ be the time that the j-th ball is removed from the box with index i. Let us compare the empirical distribution function of $T_{ij}^{(l)}$ to the cumulative distribution function of $\text{Bin}(l, e^t)$. By Theorem 3.3.1, it holds that

$$\mathbb{P}\left[\sup_{t \geq 0} \sqrt{n_l} \left| \frac{\#\{i : T_{ij}^{(l)} \leq t\}}{n_l} - \sum_{r=0}^{l-j} m_{lr}(t) \right| > \lambda(n, l) \right] \leq 2e^{-2\lambda(n,l)^2}. \tag{3.31}$$

Note that in all boxes where $T_{ij}^{(l)} \leq t$, at least j balls have been removed. This means that these are exactly the boxes with maximum $l - j$ balls left, therefore we get

$$\#\{i : T_{ij}^{(l)} \leq t\} = \sum_{s=0}^{l-j} U_{ls}(t). \tag{3.32}$$

Let us plug this expression into (3.31) to get

$$\mathbb{P}\left[\sup_{t \geq 0} \sqrt{n_l} \left| \frac{\sum_{s=0}^{l-j} U_{ls}(t)}{n_l} - \sum_{r=0}^{l-j} m_{lr}(t) \right| > \lambda(n, l) \right] \leq 2e^{-2\lambda(n,l)^2}. \tag{3.33}$$

By multiplying both sides in the equation in the probability with $\sqrt{n_l}$, we can rewrite this to

$$\mathbb{P}\left[\sup_{t \geq 0} \left| \sum_{s=0}^{l-j} (U_{ls}(t) - n_l m_{lr}(t)) \right| > \sqrt{n_l} \lambda(n, l) \right] \leq 2e^{-2\lambda(n,l)^2}. \tag{3.34}$$

This means, we have obtained the form that was presented in (3.30) by choosing $j = l - r$ for ε_1 and $j = l - r + 1$ for ε_2. We would like this upper bound to hold with probability tending to one. Let us look at the probability that the upper bound from (3.24) does not hold for one of the terms. If for a term, it holds that

$$\sup_{t \geq 0} \left| n_l m_{lr}(t) - U_{lr}(t) \right| > \sqrt{n_l} \lambda(n, l). \tag{3.35}$$

Then it must also happen that

$$\sup_{t\geq 0}\left|\varepsilon_1\right| > \frac{\sqrt{n_l}}{2}\lambda(n,l) \text{ or } \sup_{t\geq 0}\left|\varepsilon_2\right| > \frac{\sqrt{n_l}}{2}\lambda(n,l). \tag{3.36}$$

Therefore we bound

$$\begin{aligned}&\mathbb{P}\left[\sup_{t\geq 0}\left|n_l m_{lr}(t) - U_{lr}(t)\right| > \sqrt{n_l}\lambda(n,l)\right] \\ &\leq \mathbb{P}\left[\sup_{t\geq 0}\left|\varepsilon_1\right| > \frac{\sqrt{n_l}}{2}\lambda(n,l)\right] + \mathbb{P}\left[\sup_{t\geq 0}\left|\varepsilon_2\right| > \frac{\sqrt{n_l}}{2}\lambda(n,l)\right].\end{aligned} \tag{3.37}$$

Using the bounds obtained in 3.34, it is therefore derived that

$$\mathbb{P}\left[\sup_{t\geq 0}\left|n_l m_{lr}(t) - U_{lr}(t)\right| > \sqrt{n_l}\lambda(n,l)\right] \leq 4e^{-\frac{1}{2}\lambda(n,l)^2}. \tag{3.38}$$

Now, we can bound the chance that the upper bounds from (3.25) are invalid by

$$\sum_{l=k_n}^{L_n}\sum_{r=k_n}^{l}\mathbb{P}\left[\sup_{t\geq 0}\left|n_l m_{lr}(t) - U_{lr}(t)\right| > \lambda(n,l)\right] < \sum_{l=k_n}^{L_n}\sum_{r=k_n}^{l} 4e^{-\frac{1}{2}\lambda(n,l)^2}. \tag{3.39}$$

We can only use the upper bound if the probability it is invalid tends to 0. This is the last condition that we need to satisfy, together with the conditions mentioned previously in equations (3.21) and (3.26). We will now choose the value for $\hat{\beta}$ that was already stated in the theorem:

$$\hat{\beta} = \frac{3\tau + 2}{4\tau}. \tag{3.40}$$

Now, we will also choose the following values for L_n and $\lambda(n,l)$:

$$\begin{aligned}L_n &= n^{1/(2\tau)}, \\ \lambda(n,l) &= \left(c_2 n l^{-\tau}\right)^{(\tau-2)/(2\tau)}.\end{aligned} \tag{3.41}$$

We will show that for these choices of $\hat{\beta}$, L_n and $\lambda(n,l)$ the three given conditions are satisfied. First, we will see if (3.39) tends to zero. Note that the biggest term in this equation is the one where $l = L_n$ (for r arbitrary), therefore

$$\sum_{l=k_n}^{L_n}\sum_{r=k_n}^{l} 4e^{-\frac{1}{2}\lambda(n,l)^2} \leq 4L_n^2 e^{-\frac{1}{2}\lambda(n,L_n)^2}. \tag{3.42}$$

Since exponents grow faster than polynomials, this equation converges to zero if

$$\frac{1}{2}\lambda(n,l)^2 = \frac{1}{2}\left(c_2 n L_n^{-\tau}\right)^{(\tau-2)/\tau} \to \infty. \tag{3.43}$$

Since c_2 and the power $\frac{\tau-2}{\tau}$ are positive, this tends to ∞ as long as $nLn^{-\tau}$ tends to ∞. Therefore we have

$$nL_n^{-\tau} = nn^{-\tau/(2\tau)} = n^{1/2} \to \infty. \tag{3.44}$$

So (3.39) indeed goes to zero, which means the first condition is satisfied. Now, let us check if the condition from (3.21) holds, we plug in the value for L_n to get

$$\frac{n}{\tau-2}\left(n^{1/(2\tau)}-1\right)^{2-\tau} < \Theta(n^{\hat{\beta}}). \tag{3.45}$$

We can reduce this statement by using that Θ only looks at the highest order in an equation, this yields the following statement:

$$n^{1+(2-\tau)/(2\tau)} < \Theta(n^{\hat{\beta}}). \tag{3.46}$$

This is the case if

$$1+\frac{2-\tau}{2\tau} < \hat{\beta} = \frac{3\tau+2}{4\tau}. \tag{3.47}$$

And by rewriting the left-hand side, indeed we see that since $\tau > 2$ it holds that

$$\frac{4\tau}{4\tau}+\frac{4-2\tau}{4\tau} = \frac{2\tau+4}{4\tau} < \frac{3\tau+2}{4\tau}. \tag{3.48}$$

It is only left to check that condition (3.26) holds. We will first simplify the double sum by using that $r \leq l$ to obtain

$$\sum_{l=k_n}^{L_n}\sum_{r=k_n}^{l} r\sqrt{n_l}\lambda(n,l) \leq \sum_{l=k_n}^{L_n} l^2\sqrt{n_l}\lambda(n,l). \tag{3.49}$$

Now we bound n_l and write out $\lambda(n,l)$ to get

$$\begin{aligned}\sum_{l=k_n}^{L_n} l^2\sqrt{n_l}\lambda(n,l) &\leq \sum_{l=k_n}^{L_n} l^2\sqrt{c_2 n l^{-\tau}}\lambda(n,l)\\ &= \sqrt{c_2}\sum_{l=k_n}^{L_n} l^2 n^{1/2} l^{-\tau/2}\left(nl^{-\tau}\right)^{(\tau-2)/(2\tau)}.\end{aligned} \tag{3.50}$$

By simplifying the exponents, we can rewrite this to

$$\sqrt{c_2}n^{(\tau-1)/\tau}\sum_{l=k_n}^{L_n} l^{3-\tau}. \tag{3.51}$$

Let us estimate this expression using an integral, this gives

$$\sqrt{c_2}n^{(\tau-1)/\tau}\sum_{l=k_n}^{L_n} l^{3-\tau} \leq \sqrt{c_2}n^{(\tau-1)/\tau}\int_{l=0}^{L_n+1} u^{3-\tau}du. \tag{3.52}$$

Calculating this integral gives

$$\sqrt{c_2}n^{(\tau-1)/\tau}\left[\frac{1}{4-\tau}u^{4-\tau}\right]_0^{Ln+1} = \sqrt{c_2}\frac{n^{(\tau-1)/\tau}}{4-\tau}\left(L_n+1\right)^{4-\tau}. \tag{3.53}$$

Plugging in the chosen value of L_n gives

$$\sqrt{c_2}\frac{n^{(\tau-1)/\tau}}{4-\tau}\left(n^{1/(2\tau)}+1\right)^{4-\tau}. \tag{3.54}$$

Remember, we want this expression to be smaller than $\Theta\left(n^{\hat{\beta}}\right)$. Therefore, we can forget the $+1$ in the brackets, since it does not change the order of the equation. We satisfy the asked upper bound of $\Theta\left(n^{\hat{\beta}}\right)$ if

$$\frac{\tau-1}{\tau}+\frac{4-\tau}{2\tau}<\hat{\beta}=\frac{3\tau+2}{4\tau}. \tag{3.55}$$

Rewriting the right-hand side of the equation gives:

$$\frac{4\tau-4}{4\tau}+\frac{8-2\tau}{4\tau}=\frac{2\tau+4}{4\tau}<\frac{3\tau+2}{4\tau}. \tag{3.56}$$

Which means the third condition is also satisfied. Therefore we can conclude that (3.10) holds, which finishes the proof. □

Remark 3.3.3. *It could be possible that the theorem still works for a smaller β, by setting different L_n and $\lambda(n,l)$ in (3.42) that still satisfy the three conditions in the proof.*

Now we have created a whp bound on the number of heavy balls, in the next section we will use this result to get a whp bound on the number of heavy boxes.

3.4 Heavy boxes close to expectation

With the proof of the previous theorem, we can also say something about the number of heavy boxes. This is necessary, because we want to say something about the number of vertices in the $k(n)$-core. Note that we can denote the number of boxes that are still heavy at time t in the following way:

$$V(t)=\sum_{r=k_n}^{\infty} U_r(t). \tag{3.57}$$

Now we can prove the following result on the number of vertices.

Corollary 3.4.1 (whp bound on heavy boxes)**.** *For large enough graph size n and $\hat{\beta}=\frac{3\tau+2}{4\tau}$, the following holds for the function v from* (3.4) *and k_n the core that needs to be found:*

$$\mathbb{P}\left[\sup_{t\geq 0}\left|v\left(e^{-t}\right)-V(t)\right|\geq\frac{n^{\hat{\beta}}}{k_n}\right]\to 0. \tag{3.58}$$

We can prove this corollary by altering the proof of Theorem 3.3.1. A fully worked-out proof can be found in Appendix A. Now we have derived whp bounds on the number of heavy half-edges and heavy boxes alive in the death process, in the next chapter we will shift focus to obtain bounds on the number of total balls.

Chapter 4

Approximating death process

The goal of this chapter is to get a good bound on the number of balls alive in the death process. First, in Section 4.1 the death process will be related to a hitting time. In Section 4.3 a bound on this hitting time is created using martingales. Therefore in Section 4.2 some background knowledge about martingales is explained. Then in Section 4.4 the bound on the hitting time will be used to obtain a bound on the number of balls alive. Finally, in Section 4.5 we obtain a different bound by using the information on convergence to empirical measures, that was obtained in the previous chapter.

4.1 Relating to hitting time

In the article by Janson & Luczak [20, Lemma 4.3], it was already revealed that

$$\mathbb{P}\left[\left|X^{(n)}(t) - Mne^{-\gamma t}\right| \geq \varepsilon n\right] \to 0. \tag{4.1}$$

Here $Mne^{-\gamma t}$ is the expected number of individuals alive in a death process at time t [37]. Therefore if we take $\gamma = 2$, this is equal to the following statement:

$$\mathbb{P}\left[|B(t) - \mathbb{E}\left[B(t)\right]| \geq \varepsilon n\right] \to 0. \tag{4.2}$$

It will now be investigated if a stricter bound than ϵn can be found. For this inspiration is taken from [19, Claim 3.6]. Before stating this claim, first we introduce a hitting time on a pure death process,

$$\mathcal{T}_x = \min\{t : X^{(n)}(t) \leq x\}. \tag{4.3}$$

Now let us state the discussed transfer lemma.

Lemma 4.1.1 (Concentration of a death process and its hitting times [19])**.** *Let $\left(X^{(n)}(t)\right)_{t\geq 0}$ be a pure death process for each $n \in \mathbb{N}$ with initial condition $a_n = X^{(n)}(0) \to \infty$ as $n \to \infty$. Let f: $[0, \infty) \to [0, 1]$ be a function that is strictly decreasing, $f(0) = 1$, f is continuous and $f^{(-1)}$ is continuous. Then the following statements are equivalent:*

(i) for any $t_0 < \infty$, $\sup\limits_{t\leq t_0}|\frac{X^{(n)}(t)}{a_n} - f(t)| \xrightarrow{\mathbb{P}} 0$

(ii) for any $c_0 \in (0, 1)$, $\sup\limits_{c\geq c_0}|\mathcal{T}_{c\cdot a_n} - f^{(-1)}(c)| \xrightarrow{\mathbb{P}} 0$

We would like to investigate for which $\psi(n) = n^{-\beta}$ the following statement holds:

$$\mathbb{P}\left[\sup_{y \geq x} |\mathcal{T}_y - \mathbb{E}\left[\mathcal{T}_y\right]| > \psi(n)\right] \to 0. \tag{4.4}$$

In the rest of this chapter, we will shorten $\psi(n)$ to ψ to lighten the notation. Note that at a death process, all states that are reached are of the form $Mn - \gamma i$. For all other values of y, the hitting time is equal to the first $Mn - \gamma i$ below y. Therefore we can discretize the supremum. Let us look at which value the death process is when we reach the hitting time of y. For this we need

$$y \geq Mn - \gamma * i, \tag{4.5}$$

let us re-order this equation to obtain

$$i \geq \frac{Mn - y}{\gamma}. \tag{4.6}$$

Therefore we obtain the hitting time of y at $i = \lceil \frac{Mn-y}{\gamma} \rceil$. The value in $y \geq x$ that will last be reached is x, since the death process has a decreasing number of half-edges. And therefore we can discretize the supremum to obtain

$$\mathbb{P}\left[\max_{\{Mn, Mn-\gamma, \ldots, Mn-\gamma\lceil \frac{Mn-x}{\gamma} \rceil\}} |\mathcal{T}_y - \mathbb{E}\left[\mathcal{T}_y\right]| > \psi(n)\right] \to 0. \tag{4.7}$$

We can actually give an expression for $\mathcal{T}_y$. In (4.6) it was already discovered that there need to be done $\lceil \frac{Mn-y}{\gamma} \rceil$ steps before y is reached. At the first step there are still Mn vertices, so the time until the first death is exponentially distributed with rate Mn. After i steps, there are $Mn - \gamma * i$ vertices left so the time until the next death is exponentially distributed with rate $Mn - \gamma i$. Let $E(j)$ be an exponentially distributed variable with rate j. Then:

$$\mathcal{T}_y = \sum_{i=1}^{\lceil \frac{Mn-y}{\gamma} \rceil} E(Mn - \gamma(i-1)). \tag{4.8}$$

By using that exponential variables are closed under scaling by a positive factor [12], we can rewrite this equation to the following statement where all E_i are i.i.d $\exp(1)$-variables,

$$\mathcal{T}_y = \sum_{i=1}^{\lceil \frac{Mn-y}{\gamma} \rceil} \frac{E_i}{Mn - \gamma(i-1)}. \tag{4.9}$$

Let us calculate the expectation of $\mathcal{T}_y$ as well. Then by linearity of the expectation

$$\begin{aligned} \mathbb{E}\left[\mathcal{T}_y\right] &= \mathbb{E}\left[\sum_{i=1}^{\lceil \frac{Mn-y}{\gamma} \rceil} \frac{E_i}{Mn - \gamma(i-1)}\right] \\ &= \sum_{i=1}^{\lceil \frac{Mn-y}{\gamma} \rceil} \mathbb{E}\left[\frac{E_i}{Mn - \gamma(i-1)}\right]. \end{aligned} \tag{4.10}$$

By using that $\exp(1)$-distributed variables have expectation 1, we can simplify this to

$$\mathbb{E}\left[\mathcal{T}_y\right] = \sum_{i=1}^{\lceil \frac{Mn-y}{\gamma} \rceil} \frac{1}{Mn - \gamma(i-1)}. \tag{4.11}$$

It is hard to give a nice expression for this summation, therefore we will derive an upper and lower bound. These bounds will be used in the proofs of the transfer lemma later this chapter. First, we derive a lower bound. We can bound the sum of the previous equation by an integral. Note that the terms of the summation are positive and increasing (as long as $y \geq \gamma$, which will always be the case in this thesis for large n). Thus:

$$\begin{aligned} \mathbb{E}\left[\mathcal{T}_y\right] &\geq \int_0^{-1+\lceil \frac{Mn-y}{\gamma} \rceil} \frac{1}{Mn - \gamma(u-1)}\, du \\ &\geq \int_0^{-1+\frac{Mn-y}{\gamma}} \frac{1}{Mn - \gamma(u-1)}\, du. \end{aligned} \tag{4.12}$$

We can calculate this integral:

$$\begin{aligned} \left[-\frac{1}{\gamma} \log\left(Mn - \gamma(u-1)\right) \right]_0^{-1+\frac{Mn-y}{\gamma}} &= \frac{1}{\gamma} \left(\log(Mn+\gamma) - \log(y+2\gamma) \right) \\ &= \frac{1}{\gamma} \log\left(\frac{Mn+\gamma}{y+2\gamma} \right). \end{aligned} \tag{4.13}$$

Let us also derive an upper bound for the expectation:

$$\begin{aligned} \mathbb{E}\left[\mathcal{T}_y\right] &\leq \int_1^{1+\lceil \frac{Mn-y}{\gamma} \rceil} \frac{1}{Mn - \gamma(u-1)}\, du \\ &\leq \int_1^{2+\frac{Mn-y}{\gamma}} \frac{1}{Mn - \gamma(u-1)}\, du. \end{aligned} \tag{4.14}$$

We can calculate this integral:

$$\begin{aligned} \left[-\frac{1}{\gamma} \log\left(Mn - \gamma(u-1)\right) \right]_1^{2+\frac{Mn-y}{\gamma}} &= \frac{1}{\gamma} \left(\log(Mn-\gamma) - \log(y) \right) \\ &= \frac{1}{\gamma} \log\left(\frac{Mn-\gamma}{y} \right). \end{aligned} \tag{4.15}$$

In the next section, Doob's maximal inequality will be introduced. This inequality will be useful in Section 4.3 to see for which values of $\psi(n)$ (4.7) holds.

4.2 Martingales

In this section, an inequality will be introduced that will give a bound on $|\mathcal{T}_x - \mathbb{E}[\mathcal{T}_x]|$ in the next section. This inequality will make use of martingales, therefore first we define what a martingale is.

Definition 4.2.1 (Martingale [31]). *A stochastic process X_n for $n \in \{1, 2, \dots\}$ is a martingale if for all n, the following conditions are satisfied*
(i) $\mathbb{E}[|X_n|] < \infty$
(ii) $\mathbb{E}[X_{n+1}|X_n, \dots, X_1] = X_n$

For the desired inequality, we actually need a submartingale. This concept will now be introduced.

Definition 4.2.2 (Submartingale [31]). *Let X_n and Y_n for $n \in \{1, 2, \dots\}$ be stochastic processes. Then X_n is a submartingale with respect to Y_n if for all Y_n*
(i) $\mathbb{E}[\max\{X_n, 0\}] < \infty$
(ii) $\mathbb{E}[X_{n+1}|Y_n, \dots, Y_1] \geq X_n$
(iii) X_n is a function of $Y_0, \dots, Y_n$

Now let us prove there is an equivalent definition, which is nicer to work with in the special case that $X_n = \sum Y_n$.

Lemma 4.2.3 (Submartingale for series). *If $X_n = \sum_{i=1}^{n} Y_i$ is a stochastic process for $n \in \{1, 2, \dots\}$. Then X_n is a submartingale if the following conditions are satisfied:*
(I) $\mathbb{E}[\max\{X_n, 0\}] < \infty$
(II) $\mathbb{E}[X_{n+1}|X_n, \dots, X_1] \geq X_n$

Proof. It will be shown that X_n satisfies the conditions from Definition 4.2.2. Firstly, X_n is clearly a function from $Y_1, \dots, Y_n$, so condition (iii) in Definition 4.2.2 is satisfied. Furthermore, condition (I) guarantees condition (i) also holds. Now it is left to prove condition (ii), this is done by showing that the information in $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ is equal. If $Y_1, \dots, Y_n$ is known, then we can calculate X_j using $X_j = \sum_{i=1}^{j} Y_i$. Contrary, if $X_i, \dots, X_n$ is known, then we can calculate Y_j by using $Y_j = X_j - X_{j-1}$ (unless $j = 1$, then $Y_j = X_j$). Therefore $\mathbb{E}[X_{n+1}|X_n, \dots, X_1] = \mathbb{E}[X_{n+1}|Y_n, \dots, Y_1]$. This means condition (II) implies condition (ii) in Definition 4.2.2. Therefore all three conditions in this definition are satisfied and thus X_n is a Martingale. □

Now the useful inequality, for which martingales were introduced will be stated.

Lemma 4.2.4 (Doob's maximal inequality [31]). *Let X_n be a nonnegative submartingale, then for any $\psi > 0$, $\mathbb{P}\left[\max_{1 \leq j \leq n} X_j > \psi\right] \leq \frac{\mathbb{E}[X_n]}{\psi}$.*

We will use this lemma in the next section, to bound the hitting time $\mathcal{T}_x$.

4.3 Bounding hitting time

Let us define the following sequence, which we saw in Section 4.1 is related to $\mathcal{T}_x - \mathbb{E}[\mathcal{T}_x]$.

$$\begin{aligned} N_l &= \sum_{i=1}^{l} \left(\frac{E_i}{Mn - \gamma(i-1)} - \mathbb{E}\left[\frac{E_i}{Mn - \gamma(i-1)} \right] \right) \\ &= \sum_{i=1}^{l} \frac{E_i - 1}{Mn - \gamma(i-1)}. \end{aligned} \tag{4.16}$$

This summation is designed such that every term in the sum has expectation 0. Note that for $l = \lceil \frac{Mn-y}{\gamma} \rceil$, we have $N_l = \mathcal{T}_y - \mathbb{E}[\mathcal{T}_y]$ by equations (4.9) and (4.11). We would like to use Doob's maximal inequality (Lemma 4.2.4) to get a bound on N_l. Remember it was shown in Section 4.1, that we can discretize the supremum ((4.4) is equal to (4.7)). If it can be shown that N_l^2 is a submartingale, then for $T = \lceil \frac{Mn-x}{\gamma} \rceil$ we can use Doob's maximal inequality to get:

$$\begin{aligned} \mathbb{P}\left[\sup_{y \geq x} |\mathcal{T}_y - \mathbb{E}[\mathcal{T}_y]| > \psi \right] &= \mathbb{P}\left[\max_{0 \leq l \leq T} |N_l| > \psi \right] \\ &= \mathbb{P}\left[\max_{0 \leq l \leq T} N_l^2 > \psi^2 \right] \leq \frac{\mathbb{E}[N_T^2]}{\psi^2}. \end{aligned} \tag{4.17}$$

Let us verify N_l^2 is indeed a submartingale.

Theorem 4.3.1. *The sequence* $N_l^2 = \left(\sum_{i=1}^{l} \frac{E_i - 1}{Mn - \gamma(i-1)} \right)^2$ *is a submartingale.*

We can prove this theorem by checking the conditions from Lemma 4.2.3, a full proof can be found in Appendix A. Now it is known that N_l^2 is a submartingale, we can use Doob's maximal inequality in the way of (4.17). This equality will be used to prove a convergence theorem

Theorem 4.3.2 (Concentration hitting time)**.** *For* $\frac{1}{\psi^2} < \Theta(x)$*, it holds that* $\mathbb{P}\left[\sup_{y \geq x} |\mathcal{T}_y - \mathbb{E}[\mathcal{T}_y]| > \psi \right] \to 0.$

Proof. First, let us recall (4.17), which stated that

$$\begin{aligned} \mathbb{P}\left[\sup_{y \geq x} |\mathcal{T}_y - \mathbb{E}[\mathcal{T}_y]| > \psi \right] &= \mathbb{P}\left[\max_{0 \leq l \leq T} |N_l| > \psi \right] \\ &= \mathbb{P}\left[\max_{0 \leq l \leq T} N_l^2 > \psi^2 \right] \leq \frac{\mathbb{E}[N_T^2]}{\psi^2}. \end{aligned} \tag{4.18}$$

We can use Doob's maximal inequality here, since in Theorem 4.3.1 it was shown that N_l^2 is a submartingale. Let us derive a bound on $\mathbb{E}[N_T^2]$ by using the integral bound for positive increasing functions it is achieved:

$$\begin{aligned} \sum_{i=1}^{\lceil \frac{Mn-x}{\gamma} \rceil} \frac{1}{(Mn - \gamma(i-1))^2} &\leq \int_{2}^{1 + \lceil \frac{Mn-x}{\gamma} \rceil} \frac{1}{(Mn - \gamma(u-1))^2} \, du \\ &\leq \int_{1}^{2 + \frac{Mn-x}{\gamma}} \frac{1}{(Mn - \gamma(u-1))^2} \, du. \end{aligned} \tag{4.19}$$

Now, let us use the substitution $v = Mn - \gamma(u-1)$. This gives the following expression for the integral:

$$-\frac{1}{\gamma}\int_{Mn}^{x-\gamma}\frac{1}{v^2}\,dv = \frac{1}{\gamma}\left[\frac{1}{v}\right]_{Mn}^{x-\gamma} = \frac{1}{\gamma(x-\gamma)} - \frac{1}{\gamma Mn}. \tag{4.20}$$

Naturally, we can drop $-\frac{1}{\gamma Mn}$ to get a slightly worse inequality, doing this and then using (4.17). We obtain that:

$$\mathbb{P}\left[\sup_{y\geq x}|\mathcal{T}_y - \mathbb{E}\left[\mathcal{T}_y\right]| > \psi\right] \leq \frac{1}{\gamma\psi^2(x-\gamma)}. \tag{4.21}$$

By the assumptions: $\frac{1}{\psi^2} < \Theta(x)$, this means that

$$\lim_{n\to\infty}\frac{1/\psi^2}{x} = \lim_{n\to\infty}\frac{1}{x\psi^2} = 0. \tag{4.22}$$

Now, since x scales with n and γ is just a constant, $x-\gamma$ will scale with n in the same order as x. Furthermore, multiplying the fraction with a positive constant will not change the limit going to 0. Therefore

$$\lim_{n\to\infty}\frac{1}{\gamma\psi^2(x-\gamma)} = 0. \tag{4.23}$$

And therefore, we obtain that

$$\mathbb{P}\left[\sup_{y\geq x}|\mathcal{T}_y - \mathbb{E}\left[\mathcal{T}_y\right] > \psi|\right] \leq \frac{1}{\gamma\psi^2(x-\gamma)} \to 0. \tag{4.24}$$

□

Remark 4.3.3. *If this theorem holds for a certain x-value, then for $x' = x - n^{1-\beta}$ this still works as long as $n^{1-\beta} < x$. Because then*

$$\Theta\left(x'\right) = \Theta\left(x - n^{1-\beta}\right) = \Theta\left(x\right). \tag{4.25}$$

Now, a special case will be considered where x is set useful for finding a $\log^{\alpha}(n)$-core.

Corollary 4.3.4 (Concentration hitting time for polylogarithmic $k(n)$)**.** *If $x = \tilde{c}\frac{n}{\log^{\omega}(n)}$ for some constants $\omega > 0$ and $\tilde{c} > 0$, then for $\beta < \frac{1}{2}$ and large n:*

$$\mathbb{P}\left[\sup_{y\geq x-n^{1-\beta}}|\mathcal{T}_y - \mathbb{E}\left[\mathcal{T}_y\right]| > n^{-\beta}\right] \to 0. \tag{4.26}$$

Proof. In this proof, Theorem 4.3.2 will be applied with $x = \tilde{c}\frac{n}{\log^{\omega}(n)}$ and $\psi = n^{-\beta}$. After this, it will be shown that the convergence still holds for $x' = \tilde{c}\frac{n}{\log^{\omega}(n)} - n^{1-\beta}$. By Theorem 4.3.2, it is known that the convergence holds for

$$\frac{1}{n^{-2\beta}} = n^{2\beta} < \Theta\left(\tilde{c}\frac{n}{\log^{\omega}(n)}\right) = \Theta\left(\frac{n}{\log^{\omega}(n)}\right). \tag{4.27}$$

Let us write this statement out using the definition of theta, it is necessary that

$$\lim_{n\to\infty} \frac{n^{2\beta}}{n/\log^{\omega}(n)} = \lim_{n\to\infty} \frac{\log^{\omega}(n)}{n^{1-2\beta}} \to 0. \tag{4.28}$$

Since positive powers of n converge faster than logarithms, this will indeed converge if $1-2\beta > 0$. Therefore we obtain that convergence indeed holds for $\beta < \frac{1}{2}$. Now it is clear that for $0 < \beta < \frac{1}{2}$ it holds that $n^{1-\beta} < \frac{n}{\log^{\omega}(n)}$, so by remark 4.3.3 the convergence still works for $x' = \frac{n}{\log^{\omega}(n)} - n^{1-\beta}$ as well. ☐

Now another special case will be considered where x is set useful for finding a n^{α}-core.

Corollary 4.3.5 (Concentration hitting time for polynomial $k(n)$). *If $x = \tilde{c}n^{\omega}$ for constants $\omega > \frac{2}{3}$ and $\tilde{c} > 0$, then for $\beta < \frac{\omega}{2}$ and large n it holds that:*

$$\mathbb{P}\left[\sup_{y \geq x - n^{1-\beta}} |\mathcal{T}_y - \mathbb{E}\left[\mathcal{T}_y\right]| > n^{-\beta}\right] \to 0. \tag{4.29}$$

Proof. We apply Theorem 4.3.2 with $x' = \tilde{c}n^{\omega} - n^{1-\beta}$ and $\psi = n^{-\beta}$. We need x to be positive, therefore it is needed that: $\omega > 1 - \beta$. Assume that this is indeed the case, then to get convergence in Theorem 4.3.2, we also need that

$$\frac{1}{n^{-2\beta}} = n^{2\beta} < \Theta\left(\tilde{c}n^{\omega} - n^{1-\beta}\right) = \Theta\left(n^{\omega}\right). \tag{4.30}$$

This statement is true for $\omega > 2\beta$. By combining the statements $\omega > 2\beta$ and $\omega > 1 - \beta$, we see that

$$3\omega > 2\beta + 2(1-\beta) = 2 \tag{4.31}$$

Consequently the convergence holds for $\omega > \frac{2}{3}$ and $\beta < \frac{\omega}{2}$. ☐

In the next section, we will transfer the derived bounds on the hitting time to bounds on the number of balls alive.

4.4 Transfer theorem

In the previous section, the concentration of the hitting times in the death process was observed. It is now time to create an adjusted version of the transfer lemma 4.1.1, to show convergence of the number of half-edges as well.

Theorem 4.4.1 (Transfer theorem). *For $t^* = \frac{1}{\gamma}\log\left(\frac{Mn}{x}\right)$ and $\epsilon > 0$, if it holds that*

$$\mathbb{P}\left[\sup_{y \geq x - n^{1-\beta}} |\mathcal{T}_y - \mathbb{E}\left[\mathcal{T}_y\right]| > n^{-\beta-\epsilon}\right] \to 0, \tag{4.32}$$

then also

$$\mathbb{P}\left[\sup_{t \leq t^*} |X^{(n)}(t) - Mne^{-\gamma t}| > n^{1-\beta}\right] \to 0. \tag{4.33}$$

Proof. To start this proof, the following events are defined:

$$\begin{aligned} \mathcal{E}_1(t) &= \sup_{t \leq t^*} |X^{(n)}(t) - Mne^{-\gamma t}| > n^{1-\beta}, \\ \mathcal{E}_2(t) &= \sup_{y \geq x} |\mathcal{T}_y - \mathbb{E}\left[\mathcal{T}_y\right]| > n^{-\beta-\epsilon}. \end{aligned} \tag{4.34}$$

During this proof, it will be shown that if $\mathcal{E}_1(t)$ holds, then also $\mathcal{E}_2(t)$ holds. From this, it follows that the chance that $\mathcal{E}_2(t)$ holds is bigger than the chance that $\mathcal{E}_1(t)$ holds. From the assumption $\mathbb{P}\left[\mathcal{E}_2(t)\right] \to 0$, it then also follows that $\mathbb{P}\left[\mathcal{E}_1(t)\right] \to 0$. Let us start this proof by assuming that $\mathcal{E}_1(t)$ holds, then for some $t \leq t^*$:

$$|X^{(n)}(t) - Mne^{-\gamma t}| > n^{1-\beta}. \tag{4.35}$$

There are two possible cases for $\mathcal{E}_1(t)$ to hold:

$$\begin{aligned} (i) &: X^{(n)}(t) > Mne^{-\gamma t} + n^{1-\beta}, \\ (ii) &: X^{(n)}(t) < Mne^{-\gamma t} - n^{1-\beta}. \end{aligned} \tag{4.36}$$

Case (i): Let us start by setting

$$y = Mne^{-\gamma t} + n^{1-\beta}. \tag{4.37}$$

It holds $\mathcal{T}_y > t$. By rewriting (4.37), we obtain the following expression for t:

$$t = -\frac{1}{\gamma}\log\left(\frac{y - n^{1-\beta}}{Mn}\right). \tag{4.38}$$

Using this equation, let us derive a lower bound for $\mathcal{T}_y - \mathbb{E}\left[\mathcal{T}_y\right]$,

$$\begin{aligned} \mathcal{T}_y - \mathbb{E}\left[\mathcal{T}_y\right] &> t - \mathbb{E}\left[\mathcal{T}_y\right] \\ &= -\frac{1}{\gamma}\log\left(\frac{y - n^{1-\beta}}{Mn}\right) - \mathbb{E}\left[\mathcal{T}_y\right]. \end{aligned} \tag{4.39}$$

Now, let us use the lower bound on $\mathbb{E}\left[\mathcal{T}_y\right]$ that was obtained in (4.15), to get

$$\mathcal{T}_y - \mathbb{E}\left[\mathcal{T}_y\right] > -\frac{1}{\gamma}\log\left(\frac{y - n^{1-\beta}}{Mn}\right) - \frac{1}{\gamma}\log\left(\frac{Mn - \gamma}{y}\right). \tag{4.40}$$

Rewriting the logarithms gives

$$\begin{aligned} &\frac{1}{\gamma}\left(\log(y) - \log\left(y - n^{1-\beta}\right) + \log(Mn) - \log(Mn - \gamma)\right) \\ >&\frac{1}{\gamma}\left(\log(y) - \log\left(y - n^{1-\beta}\right)\right) = \frac{1}{\gamma}\log\left(\frac{y}{y - n^{1-\beta}}\right). \end{aligned} \tag{4.41}$$

Reversing the fracture gives the following result:

$$-\frac{1}{\gamma}\log\left(\frac{y - n^{1-\beta}}{y}\right) = -\frac{1}{\gamma}\log\left(1 - \frac{-n^{1-\beta}}{y}\right). \tag{4.42}$$

Now let us look at the Taylor series of this logarithm to see that, this logarithm is equal to

$$\begin{aligned} -\frac{1}{\gamma}\left(\frac{-n^{1-\beta}}{y} + \mathcal{O}\left(\left(\frac{-n^{1-\beta}}{y}\right)^2\right)\right) &= \Theta\left(\frac{n^{1-\beta}}{y}\right) \\ &> \Theta\left(n^{-\beta-\epsilon}\right). \end{aligned} \tag{4.43}$$

Therefore, during equations (4.39) to (4.43) it was derived that for the chosen y,

$$|\mathcal{T}_y - \mathbb{E}[\mathcal{T}_y]| \geq \mathcal{T}_y - \mathbb{E}[\mathcal{T}_y] > n^{-\beta-\epsilon}. \tag{4.44}$$

Case (ii): Start by setting

$$y = Mne^{-\gamma t} - n^{1-\beta}. \tag{4.45}$$

Now it holds that $\mathcal{T}_y \leq t$. Rewriting y, gives the following expression for t:

$$t = -\frac{1}{\gamma}\log\left(\frac{y+n^{1-\beta}}{Mn}\right). \tag{4.46}$$

Using this expression of t, it follows

$$\begin{aligned}\mathbb{E}[\mathcal{T}_y] - \mathcal{T}_y &\geq \mathbb{E}[\mathcal{T}_y] - t \\ &= \mathbb{E}[\mathcal{T}_y] + \frac{1}{\gamma}\log\left(\frac{y+n^{1-\beta}}{Mn}\right).\end{aligned} \tag{4.47}$$

By using the lower bound on the expectation from (4.13), we obtain that

$$\mathbb{E}[\mathcal{T}_y] - \mathcal{T}_y \geq \frac{1}{\gamma}\log\left(\frac{y+n^{1-\beta}}{Mn}\right) + \frac{1}{\gamma}\log\left(\frac{Mn+\gamma}{y+2\gamma}\right). \tag{4.48}$$

Rewriting the logarithms gives

$$\begin{aligned}&\frac{1}{\gamma}\left(\log\left(y+n^{1-\beta}\right) - \log(y+2\gamma) + \log(Mn+\gamma) - \log(Mn)\right) \\ >&\frac{1}{\gamma}\left(\log\left(y+n^{1-\beta}\right) - \log(y+2\gamma)\right) = \frac{1}{\gamma}\log\left(\frac{y+n^{1-\beta}}{y+2\gamma}\right).\end{aligned} \tag{4.49}$$

We can rewrite this logarithm by using a Taylor series,

$$\begin{aligned}\frac{1}{\gamma}\log\left(\frac{y+n^{1-\beta}}{y+2\gamma}\right) &= \frac{1}{\gamma}\log\left(1+\frac{n^{1-\beta}-2\gamma}{y+2\gamma}\right) \\ &= \frac{1}{\gamma}\frac{n^{1-\beta}-2\gamma}{y+2\gamma} + \mathcal{O}\left(\left(\frac{n^{1-\beta}-2\gamma}{y+2\gamma}\right)^2\right).\end{aligned} \tag{4.50}$$

Let us approximate this expression, by determining the order of the equation

$$\Theta\left(\frac{n^{1-\beta}-2\gamma}{y+2\gamma}\right) = \Theta\left(\frac{n^{1-\beta}}{y}\right) > \Theta\left(n^{-\beta-\epsilon}\right). \tag{4.51}$$

So also in the second case for the chosen y

$$|\mathcal{T}_y - \mathbb{E}[\mathcal{T}_y]| \geq \mathbb{E}[\mathcal{T}_y] - \mathcal{T}_y > n^{-\beta-\epsilon}. \tag{4.52}$$

Let us check if for the chosen y-values, $y \geq x - n^{1-\beta}$. If this is the case, the chosen y-value is in the supremum of $\mathcal{E}_2(t)$. Note that the y from case (ii) is smaller than the one in case (i), so only checking the y from case (ii) suffices:

$$\begin{aligned}y &= Mne^{-\gamma t} - n^{1-\beta} \\ &\geq Mne^{-\gamma t^*} - n^{1-\beta}.\end{aligned} \tag{4.53}$$

By filling in the expression, for t^*, we obtain that

$$\begin{aligned} y &\geq Mne^{-\log\left(\frac{Mn}{x}\right)} - n^{1-\beta} \\ &= \frac{Mn}{Mn/x} - n^{1-\beta}. \end{aligned} \tag{4.54}$$

So we derive that

$$y \geq x - n^{1-\beta}, \tag{4.55}$$

so indeed we have found a value in the supremum of $\mathcal{E}_2(t)$. Therefore in both cases it is now known that

$$\sup_{y \geq x - n^{1-\beta}} |\mathcal{T}_y - \mathbb{E}\left[\mathcal{T}_y\right]| > n^{-\beta-\epsilon}. \tag{4.56}$$

should also hold. Therefore we can conclude that if $\mathcal{E}_1(t)$ holds, also $\mathcal{E}_2(t)$ holds. Consequently, we conclude that:

$$\mathbb{P}\left[\mathcal{E}_1(t)\right] \leq \mathbb{P}\left[\mathcal{E}_2(t)\right] \to 0. \tag{4.57}$$

□

Let us look at a specific choice for x which can be helpful to find a $\log^\alpha(n)$-core.

Corollary 4.4.2 (Concentration of death process for polylogarithmic $k(n)$)**.** *For $\omega > 0$ and $c > 0$ constants, take*

$$t^* = \log\left(\hat{c}\log^{\omega/2}(n)\right). \tag{4.58}$$

Then it holds for any $\upsilon > 0$ that

$$\mathbb{P}\left[\sup_{t \leq t^*} |X^{(n)}(t) - Mne^{-2t}| > \frac{n^{1/2+\upsilon}}{2}\right] \to 0. \tag{4.59}$$

This corollary follows from combining Theorem 4.4.1 and Corollary 4.3.4. A full proof can be found in Appendix A. We will now also consider a special value of x, which is convenient for discovering a n^α-core.

Corollary 4.4.3 (Concentration of death process for polynomial $k(n)$)**.** *For constants $\tilde{c} > 0$ and $\omega > \frac{2}{3}$, set*

$$t^* = \log\left(\tilde{c}n^{(1-\omega)/2}\right). \tag{4.60}$$

Then for $\beta < \frac{\omega}{2}$ it holds that:

$$\mathbb{P}\left[\sup_{t \leq t^*} |X^{(n)}(t) - Mne^{-2t}| > \frac{n^{1-\beta}}{2}\right] \to 0. \tag{4.61}$$

This corollary follows from combining Theorem 4.4.1 and Corollary 4.3.5. A full proof can be found in Appendix A. In the next section, we will use empirical measures to derive another bound on the number of balls alive.

4.5 Empirical measure approach

In the previous sections in this chapter, we have used a transfer theorem to bound the number of balls being close to its expectation. However, it is also possible to use the previously stated Theorem 3.3.1. We will now investigate whether this gives a better bound. First, we will look how a death process with death rate 1 converges to its distribution function. Thereafter, we will couple this result to a death process with death rate 2.

Lemma 4.5.1 (Convergence rate 1 death process to distribution function). *Let $(X(t))_{t\geq 0}$ be a pure death process with death rate 1 with initial condition $X(0) = a_n$. Then for $\epsilon > 0$, the following result holds:*

$$\mathbb{P}\left[\sup_{t\geq 0}\left|X(t) - a_n e^{-t}\right| > \sqrt{a_n} n^{\epsilon}\right] \to 0. \tag{4.62}$$

Proof. $X(t)$ starts with a_n balls alive, let us give every ball an index $i \in \{1, 2, \dots a_n\}$. We know ball i will die after T_i time, where $T_i \sim \exp(1)$. Let $F_{a_n}(t)$ be the empirical distribution function of all T_i. The amount of balls alive at time t can be expressed as

$$\begin{aligned} X(t) = \sum_{i=1}^{a_n} \mathbb{1}_{\{T_i > t\}} &= a_n - \sum_{i=1}^{a_n} \mathbb{1}_{\{T_i \leq t\}} \\ &= a_n - a_n F_{a_n}(t). \end{aligned} \tag{4.63}$$

Now let $F(t)$ be the distribution function of an $\exp(1)$-distributed variable. Then $a_n F(t)$ represents the expected number of balls that have already died. Therefore we can use the expected number of balls still alive [20] to deduce that

$$a_n F(t) = a_n - \mathbb{E}\left[X(t)\right] = a_n - a_n e^{-t}. \tag{4.64}$$

We will now apply Theorem 3.3.1 to obtain (4.62). By Theorem 3.3.1, it is known that

$$\mathbb{P}\left[\sup_{t\geq 0}\sqrt{a_n}\left|F_{a_n}(t) - F(t)\right| > \lambda\right] \leq 2\exp\left(-2\lambda^2\right). \tag{4.65}$$

By multiplying both sides of the equation in the probability with $\sqrt{a_n}$, we obtain the equivalent expression

$$\mathbb{P}\left[\sup_{t\geq 0}\left|a_n F_{a_n}(t) - a_n F(t)\right| > \sqrt{a_n}\lambda\right] \leq 2\exp\left(-2\lambda^2\right). \tag{4.66}$$

Now let us slightly rewrite this statement, so we can use (4.63) and (4.64).

$$\mathbb{P}\left[\sup_{t\geq 0}\left|\left(a_n F_{a_n}(t) - a_n\right) + \left(a_n - a_n F(t)\right)\right| > \sqrt{a_n}\lambda\right] \leq 2\exp\left(-2\lambda^2\right). \tag{4.67}$$

By applying (4.63) and (4.64), we now get:

$$\mathbb{P}\left[\sup_{t\geq 0}\left|-X(t) + a_n e^{-t}\right| > \sqrt{a_n}\lambda\right] \leq 2\exp\left(-2\lambda^2\right). \tag{4.68}$$

By setting $\lambda = n^\epsilon$, we obtain that it indeed holds that

$$\mathbb{P}\left[\sup_{t\geq 0}\left|X(t) - a_n e^{-t}\right| > \sqrt{a_n}n^\epsilon\right] \leq 2\exp\left(-2n^{2\epsilon}\right) \to 0. \tag{4.69}$$

□

This lemma will now help to give a bound on $B(t)$ being close to its expectation in the next theorem.

Theorem 4.5.2 (Concentration bound of total balls)**.** *Let $(B(t))_{t\geq 0}$ be a death process with rate 2 with initial condition $B(0) = Mn$. Then for $\upsilon > 0$, we get*

$$\mathbb{P}\left[\sup_{t\geq 0}\left|B(t) - Mne^{-2t}\right| > \frac{n^{1/2+\upsilon}}{2}\right] \to 0. \tag{4.70}$$

Proof. $B(t)$ is a death process with rate 2, we would like a rate 1 death process. This would mean we can apply Lemma 4.5.1. We will scale $B(t)$ to a different Markov process. We will divide both t and $B(t)$ by 2. This gives a Markov process $X^{(a)}(t)$ that jumps down size 1 with the waiting time until the next event being $\exp\left(\frac{1}{X^{(a)}(t)}\right)$ [20]. The variable a stands for the starting value, which is

$$a = \frac{B(0)}{2} = \frac{Mn}{2}. \tag{4.71}$$

This is almost a rate 1 death process, the only difference is that $X^{(a)}(t)$ is not integer if $B(0)$ is odd. To relate $X^{(a)}(t)$ to a rate 1 death process, we turn to the coupling principle applied in the proof of [20, Lemma 4.3]. Here $X^{\lceil a\rceil}(t)$ is defined with

$$X^{\lceil a\rceil}(0) = \lceil a\rceil. \tag{4.72}$$

Now we couple $X^{(a)}(t)$ and $X^{\lceil a\rceil}(t)$ such that both jump whenever the smaller does. Since $X^{\lceil a\rceil}(t)$ starts at an integer value, it is a rate 1 death process. For all t, it holds that:

$$\left|X^{(a)}(t) - X^{\lceil a\rceil}(t)\right| < 1. \tag{4.73}$$

Now we let $a = \frac{Mn}{2}$, we can apply Lemma 4.5.1 on $X^{\lceil a\rceil}(t)$ to get for $\epsilon > 0$

$$\mathbb{P}\left[\sup_{t\geq 0}\left|X^{\lceil a\rceil}(t) - \left\lceil\frac{Mn}{2}\right\rceil e^{-t}\right| > \sqrt{\left\lceil\frac{Mn}{2}\right\rceil}n^\epsilon\right] \to 0. \tag{4.74}$$

We instead would like to bound

$$\left|X^{(a)}(t) - \frac{Mn}{2}e^{-t}\right| = \left|X^{(a)}(t) - ae^{-t}\right|. \tag{4.75}$$

Note that we can rewrite this absolute value in the following way:

$$\left|\left(X^{\lceil a\rceil}(t) - \lceil a\rceil e^{-t}\right) + \left(X^{(a)}(t) - X^{\lceil a\rceil}(t)\right) + \left(\lceil a\rceil - a\right)e^{-t}\right|. \tag{4.76}$$

By the triangle inequality, we can bound this with the absolute value of the three terms. We can bound the second term by (4.73). It is also clear that the third term is smaller than 1, since both $e^{-t} < 1$ and $\lceil a\rceil - a < 1$. Therefore:

$$\left|X^{(a)}(t) - \frac{Mn}{2}e^{-t}\right| < 2 + \left|X^{\lceil a\rceil}(t) - \left\lceil\frac{Mn}{2}\right\rceil e^{-t}\right|. \tag{4.77}$$

But then we can use (4.74) to get

$$\mathbb{P}\left[\sup_{t\geq 0}\left|X^{(a)}(t)-\frac{Mn}{2}e^{-t}\right|>2+\sqrt{\left\lceil\frac{Mn}{2}\right\rceil n^{\epsilon}}\right]\to 0. \tag{4.78}$$

This means that it is also true that

$$\mathbb{P}\left[\sup_{t\geq 0}\left|X^{(a)}(t)-\frac{Mn}{2}e^{-t}\right|>2+\sqrt{\left\lceil\frac{M}{2}\right\rceil n^{1/2+\epsilon}}\right]\to 0. \tag{4.79}$$

We now scale back the process by multiplying $X^{a}(t)$ and t by 2, this gives

$$\mathbb{P}\left[\sup_{2t\geq 0}\left|B(t)-Mne^{-2t}\right|>4+2\sqrt{\left\lceil\frac{M}{2}\right\rceil n^{1/2+\epsilon}}\right]\to 0. \tag{4.80}$$

We can drop the multiplication with 2 in the supremum, since it makes no difference. Now this means that

$$\mathbb{P}\left[\sup_{t\geq 0}\left|B(t)-Mne^{-2t}\right|>\frac{n^{1/2+2\epsilon}}{2}\right]\to 0. \tag{4.81}$$

By setting $\epsilon=\frac{\upsilon}{2}>0$, we obtain that

$$\mathbb{P}\left[\sup_{t\geq 0}\left|B(t)-Mne^{-2t}\right|>\frac{n^{1/2+\upsilon}}{2}\right]\to 0 \tag{4.82}$$

□

We will now compare the results from Theorem 4.5.2 to the results obtained in Section 4.4. In Theorem 4.5.2, we obtained a bound for $t\geq 0$. Where the results in Corollaries 4.3.4 and 4.4.3 only worked for $t\leq t^{*}$. If these corollaries give a stricter bound, they could still be preferred. However, Corollary 4.3.4 bounds by $n^{\hat{\beta}}$ for $\hat{\beta}>\frac{1}{2}$. This is the same result as obtained in Theorem 4.5.2. If we look at Corollary 4.4.3 instead we need $\hat{\beta}>1-\frac{\omega}{2}\geq\frac{1}{2}$, so we will never get a better result than from Theorem 4.5.2. Therefore Theorem 4.5.2 is an improvement of previous obtained results, since it gives a bound at least as good, but for all $t\geq 0$. Thus this theorem will be used for results later this thesis. In the next chapter, we will work on finding a time t^{*} such that the $k(n)$-core has been reached.

Chapter 5

Finding reaching time

In this chapter, we will find the time when the balls-into-boxes algorithm reaches the $k(n)$-core. This will be done by proving one main theorem in Section 5.1.

5.1 Reaching k(n)-core

The goal of this chapter is to find a time t^* as small as possible that satisfies (2.17), which means that we have reached the $k(n)$-core. Using the expectations of $B(t)$ and $H(t)$, that were obtained in (3.2) and (4.1), we can rewrite (2.17) to

$$Mne^{-2t} - h\left(e^{-t}\right) \leq -n^{\delta}. \tag{5.1}$$

By substituting $p = e^{-t}$, we obtain that

$$Mnp^2 - h\left(p\right) \leq -n^{\delta}. \tag{5.2}$$

Note that e^{-t} is a decreasing function with $e^0 = 1$. Since we tried $t \geq 0$ as small as possible such that (5.1) holds, this will now change to $p \leq 1$ as large as possible in (5.5). We will now state a theorem about a p that satisfies this equation.

Theorem 5.1.1 (Reaching time $k(n)$-core)**.** *Assume that for some $\delta < 1$ and k_n it holds that*

$$n^{(\delta-1)/2} < \Theta\left(k_n^{(2-\tau)/(3-\tau)}\right). \tag{5.3}$$

Then for large enough n there also exists

$$p = \Theta\left(k_n^{(2-\tau)/(3-\tau)}\right), \tag{5.4}$$

that for the function $h(p)$ from (3.2) satisfies the equation

$$Mnp^2 - h\left(p\right) \leq -n^{\delta}. \tag{5.5}$$

Proof. We start this proof by rewriting the equation that needs to be satisfied. First, let us write out $h(p)$ using (3.2). For $X_l(p) \sim \text{Bin}(l, p)$, we get

$$Mnp^2 - \sum_{l=k_n}^{\infty} p_l(n) n \mathbb{E}\left[X_l(p) \mathbb{1}_{\{X_l(p) \geq k_n\}}\right] \leq -n^{\delta}. \tag{5.6}$$

Let us divide both sides by n and re-order the equation to get:

$$Mp^2 + n^{\delta-1} \leq \sum_{l=k_n}^{\infty} p_l(n)\mathbb{E}\left[X_l(p)\mathbb{1}_{\{X_l(p)\geq k_n\}}\right]. \tag{5.7}$$

From now on, we will make the statement slightly stricter by using a lower bound on the right-hand side a few times to make the calculations easier. This will mean that the p from this theorem is not certainly the largest solution to (5.5), but the difference with the largest solution will be small. The maximum degree in the scale-free configuration model is n^ϵ for $\epsilon = \frac{1}{\tau-1}$ whp [18, Theorem 7.13]. Therefore we will give a slightly stricter bound by letting the sum not go to ∞ anymore. The upper bound will be set to n^ε instead, this gives

$$Mp^2 + n^{\delta-1} \leq \sum_{l=k_n}^{n^\varepsilon} p_l(n)\mathbb{E}\left[X_l(p)\mathbb{1}_{\{X_l(p)\geq k_n\}}\right]. \tag{5.8}$$

Now, it is time to use the distribution of the discrete Pareto variable. We know that $p_l = \frac{1}{\zeta(\tau)}l^{-\tau}$ and by Condition 2.2.2(i) also $p_l(n) \to p_l$ for $l \leq n^\varepsilon$. Therefore, we know that for large n, $p_l(n) \geq \hat{c}p_l$ for all constants $\hat{c} \in (0,1)$. By using this information, again the bound in the inequality becomes slightly stricter, we now get

$$Mp^2 + n^{\delta-1} \leq \sum_{l=k_n}^{n^\varepsilon} \frac{\hat{c}}{\zeta(\tau)}l^{-\tau}\mathbb{E}\left[X_l(p)\mathbb{1}_{\{X_l(p)\geq k_n\}}\right]. \tag{5.9}$$

The next questions is how to bound the expectation. Naturally, we can note that

$$\begin{aligned}\mathbb{E}\left[X_l(p)\mathbb{1}_{\{X_l(p)\geq k_n\}}\right] &\geq \mathbb{E}\left[X_l(p)\right] - \mathbb{E}\left[X_l(p)\mathbb{1}_{\{X_l(p)\leq k_n\}}\right] \\ &= lp - \sum_{i=0}^{k_n} i\mathbb{P}\left[X_l(p) = i\right].\end{aligned} \tag{5.10}$$

Let us rewrite the second part of the equation:

$$\begin{aligned}\sum_{i=0}^{k_n} i\mathbb{P}\left[X_l(p) = i\right] &\leq \sum_{i=0}^{k_n} k_n\mathbb{P}\left[X_l(p) = i\right] \\ &= k_n\mathbb{P}\left[X_l(p) \leq k_n\right].\end{aligned} \tag{5.11}$$

Assume that we have $lp \geq 2k_n$. Then

$$\mathbb{P}\left[X_l(p) \leq k_n\right] \leq \mathbb{P}\left[X_l(p) \leq \frac{1}{2}lp\right]. \tag{5.12}$$

Now we can use a Chernoff bound on the probability [33] to get

$$\begin{aligned}\mathbb{P}\left[X_l(p) \leq \frac{1}{2}lp\right] &\leq \exp\left(-\frac{1}{2}\left(\frac{1}{2}\right)^2 lp\right) \\ &\leq \exp\left(-\frac{1}{2}\left(\frac{1}{2}\right)^2 2k_n\right) = \exp\left(-\frac{k_n}{4}\right).\end{aligned} \tag{5.13}$$

Using this in (5.10), we obtain that for $lp \geq 2k_n$, it holds that

$$\mathbb{E}\left[X_l(p)\mathbb{1}_{\{X_l(p)\geq k_n\}}\right] \geq lp - k_n e^{-k_n/4}. \tag{5.14}$$

To use the Chernoff bound, we made the assumption that $lp \geq 2k_n$. Since $p \leq 1$, this also means that $l \geq 2k_n > k_n$, therefore it is true that

$$\sum_{l=k_n}^{n^{\varepsilon}} \frac{\hat{c}}{\zeta(\tau)} l^{-\tau} \mathbb{E}\left[X_l(p)\mathbb{1}_{\{X_l(p)\geq k_n\}}\right] \geq \sum_{l=\lceil \frac{2}{p}k_n \rceil}^{n^{\varepsilon}} \frac{\hat{c}}{\zeta(\tau)} l^{-\tau}\left(lp - k_n e^{-k_n/4}\right). \tag{5.15}$$

Thus (5.9) can be made stricter to obtain:

$$\begin{aligned} Mp^2 + n^{\delta-1} &\leq \sum_{l=\lceil \frac{2}{p}k_n \rceil}^{n^{\varepsilon}} \frac{\hat{c}}{\zeta(\tau)} l^{-\tau}\left(lp - k_n e^{-k_n/4}\right) \\ &= \frac{\hat{c}}{\zeta(\tau)} p \sum_{l=\lceil \frac{2}{p}k_n \rceil}^{n^{\varepsilon}} l^{1-\tau} - \sum_{l=\lceil \frac{2}{p}k_n \rceil}^{n^{\varepsilon}} \frac{\hat{c}}{\zeta(\tau)} l^{-\tau} k_n e^{-k_n/4}. \end{aligned} \tag{5.16}$$

It can be seen that the second term in this equation is a lot smaller, since it has a lower power of l. Therefore this term will eventually go below $(1-\hat{c})$ times the leading term. Therefore, we can again create a stricter inequality by setting

$$Mp^2 + n^{\delta-1} \leq \frac{\hat{c}^2 p}{\zeta(\tau)} \sum_{l=\lceil \frac{2}{p}k_n \rceil}^{n^{\varepsilon}} l^{1-\tau}. \tag{5.17}$$

We are working with $\tau \in (2,3)$. Therefore, $l^{1-\tau}$ is a decreasing function in terms of l. Therefore, let us bound the sum using an integral bound for decreasing positive terms to get

$$\begin{aligned} \sum_{l=\lceil \frac{2}{p}k_n \rceil}^{n^{\varepsilon}} l^{1-\tau} &\geq \int_{\lceil \frac{2}{p}k_n \rceil}^{n^{\varepsilon}} u^{1-\tau} du \\ &\geq \int_{\frac{2}{p}k_n+1}^{n^{\varepsilon}} u^{1-\tau} du. \end{aligned} \tag{5.18}$$

Calculating this integral gives:

$$\left[\frac{1}{2-\tau}u^{2-\tau}\right]_{\frac{2}{p}k_n+1}^{n^{\varepsilon}} = \frac{1}{2-\tau}\left(n^{\varepsilon}\right)^{2-\tau} - \frac{1}{2-\tau}\left(\frac{2k_n}{p}+1\right)^{2-\tau}. \tag{5.19}$$

We are using a discrete Pareto distribution with $\tau \in (2,3)$, therefore $\frac{1}{2-\tau} < 0$, so the second term is actually the positive one. Also the power of u is negative because of this, so $u = \frac{2}{p}k_n + 1$ gives a larger value than $u = n^{\varepsilon}$. And this difference becomes bigger to eventually go below $(1-\hat{c})$ times the leading term for large n. Therefore we conclude that

$$\sum_{l=\lceil \frac{2}{p}k_n \rceil}^{n^{\varepsilon}} l^{1-\tau} \geq \hat{c}\frac{1}{\tau-2}\left(\frac{2k_n}{p}+1\right)^{2-\tau}. \tag{5.20}$$

We thus again get a stricter inequality, namely:

$$
\begin{aligned}
Mp^2 + n^{\delta-1} &\leq \frac{\hat{c}^3 p}{\zeta(\tau)} \frac{1}{\tau-2} \left(\frac{2k_n}{p} + 1\right)^{2-\tau} \\
&= p \frac{\hat{c}^3}{\zeta(\tau)(\tau-2)} \left(\frac{2k_n + 2p}{p}\right)^{2-\tau}.
\end{aligned}
\tag{5.21}
$$

Now let us re-order the p-powers in the equation to get:

$$
Mp^{3-\tau} + n^{\delta-1} p^{1-\tau} \leq \frac{\hat{c}^3}{\zeta(\tau)(\tau-2)} (2k_n + 2p)^{2-\tau}. \tag{5.22}
$$

We would like to have $Mp^{3-\tau}$ as leading term on the left-hand side of the equation, this is the case if:

$$
n^{\delta-1} p^{1-\tau} < \Theta\left(Mp^{3-\tau}\right) = \Theta\left(p^{3-\tau}\right). \tag{5.23}
$$

By rewriting this equation, we see that we need:

$$
n^{(\delta-1)/2} < \Theta(p). \tag{5.24}
$$

If this is indeed the case, then we again use that this term goes below $(1-\hat{c})$ times the leading term eventually. Then (5.22) is definitely true for

$$
Mp^{3-\tau} \leq \frac{\hat{c}^4}{\zeta(\tau)(\tau-2)} (2k_n + 2p)^{2-\tau}. \tag{5.25}
$$

Since $2k_n$ is far bigger than $2p$ we can get a slightly stricter statement, by multiplying k_n with a constant bigger than 1. We choose $2-\hat{c}$ as this constant. This gives the following statement.

$$
Mp^{3-\tau} \leq \frac{\hat{c}^4}{\zeta(\tau)(\tau-2)} (2(2-\hat{c}))^{2-\tau} k_n^{2-\tau}. \tag{5.26}
$$

We can rewrite this to

$$
p \leq \left(\frac{\hat{c}^4}{M\zeta(\tau)(\tau-2)} (2(2-\hat{c}))^{2-\tau}\right)^{1/(3-\tau)} k_n^{(2-\tau)/(3-\tau)}. \tag{5.27}
$$

So if $n^{(\delta-1)/2} < \Theta(p)$, then we know that (5.5) holds for a certain p, such that

$$
p = \Theta\left(k_n^{(2-\tau)/(3-\tau)}\right). \tag{5.28}
$$

For this value of p, the assumption that $n^{(\delta-1)/2} < \Theta(p)$ becomes

$$
n^{(\delta-1)/2} < \Theta\left(k_n^{(2-\tau)/(3-\tau)}\right). \tag{5.29}
$$

And this is exactly why this was stated as a condition in the theorem. Thus we can conclude that if this condition holds, the given value for p satisfies the equation in (5.5). $\square$

In the next chapter, we will use the theorem from this section to prove the main results of this thesis.

Chapter 6

Finding the cores

With the information gathered in the previous chapters, we prove the main results that were stated in Section 2.3. First, in Section 6.1 we will prove Theorem 2.3.2 about the existence of $\log^\alpha(n)$-cores. Then, in Section 6.2 we will prove Theorem 2.3.3 about the existence n^α-cores.

6.1 Logarithmic core

We will now prove Theorem 2.3.2, to show a $\log^\alpha(n)$-core exists for all fixed $\alpha > 0$.

Proof of Theorem 2.3.2. In Section 2.4, we created an algorithm that follows the configuration model from Definition 2.2.1. This algorithm reaches the $k(n)$-core if the number of heavy balls is larger than the number of total balls, since the number of total balls is set to -1 at the moment we reach the $k(n)$-core. In Section 2.5 it was derived that this is the case if (2.17), (2.18) and (2.19) hold for some $\delta > 0$. We will now derive a $\delta > 0$ such that these equations hold, with help of theorems from the previous chapters.

First, we will see when (2.17) holds, this will be determined by Theorem 5.1.1. This theorem has a condition that $n^{(\delta-1)/2} < \Theta\left(k_n^{(2-\tau)/(3-\tau)}\right)$, which we will now check. We fix $\alpha > 0$. By filling in $k(n) = \log^\alpha(n)$, we get:

$$n^{(\delta-1)/2} < \Theta\left((\log(n))^{\alpha(2-\tau)/(3-\tau)}\right). \tag{6.1}$$

This means that by writing out the definition of Θ (2.3.1), we get that it needs to be shown that

$$\frac{n^{(\delta-1)/2}}{(\log(n))^{\alpha(2-\tau)/(3-\tau)}} \to 0. \tag{6.2}$$

By using that $2 - \tau < 0$ for $\tau \in (2, 3)$, we rewrite this to:

$$\frac{(\log(n))^{\alpha(\tau-2)/(3-\tau)}}{n^{(1-\delta)/2}} \to 0. \tag{6.3}$$

Since positive powers of n increase faster than positive powers of logarithms, this statement is true for all $\delta < 1$. Therefore Theorem 5.1.1 tells us that (5.2) is satisfied for certain constant $c > 0$ and

$$p = c\,(\log(n))^{\alpha(2-\tau)/(3-\tau)}\,. \tag{6.4}$$

Therefore, also for $\delta < 1$ (2.17) is satisfied at the time t^* with $p = e^{-t^*}$. Now let us check what is the smallest δ for which we know that (2.18) and (2.19) hold whp. By Theorem 4.5.2 we know for $\delta > \frac{1}{2}$ (2.18) is satisfied whp. Also by Theorem 3.3.2, (2.19) holds for $\delta \geq \frac{3\tau+2}{4\tau}$ whp. Note that

$$\frac{3\tau+2}{4\tau} > \frac{3\tau}{4\tau} > \frac{1}{2}. \tag{6.5}$$

Therefore we set $\delta = \frac{3\tau+2}{4\tau}$, for which it indeed holds that $\delta < 1$. Therefore we know that the $\log^\alpha(n)$-core has been reached at t^* whp. If there are still edges left at t^*, then a $\log^\alpha(n)$-core exists. Otherwise the core is empty.

We will now derive lower bounds for the number of vertices and edges in the $\log^\alpha(n)$-core, this are the number of edges and vertices left at t^*. With the help of (2.17) and (2.19), that we have satisfied whp, we see that

$$H(t^*) > \mathbb{E}\left[H(t^*)\right] - \frac{n^\delta}{2} \geq \mathbb{E}\left[B(t^*)\right] + n^\delta - \frac{n^\delta}{2} > \mathbb{E}\left[B(t^*)\right]. \tag{6.6}$$

The expected number of balls left at t^* is:

$$\begin{aligned}\mathbb{E}\left[B(t^*)\right] &= Mnp^2 = c^2 Mn\,(\log(n))^{2\alpha(2-\tau)/(3-\tau)} \\ &= \Theta\left(\frac{n}{(\log(n))^{2\alpha(\tau-2)/(3-\tau)}}\right).\end{aligned} \tag{6.7}$$

So then by (6.6) we know that at time t^* there are at least this many half-edges left. The number of half-edges is twice the number of edges. Therefore the Θ-expression in (6.7) is whp a lower bound for the number of edges in the $\log^\alpha(n)$-core. Therefore we know an $\log^\alpha(n)$-core exists whp.

Let us now look at the minimum number of vertices in the $\log^\alpha(n)$-core, then we know by Theorem 3.4.1 that for the function $v(e^{-t})$ with $k(n) = \log^\alpha(n)$ from (3.4) it holds whp that for the value of p from (6.4)

$$V(t^*) > v(e^{-t^*}) - \frac{n^{(3\tau+2)/(4\tau)}}{\log^\alpha(n)} = v(p) - \frac{n^{(3\tau+2)/(4\tau)}}{\log^\alpha(n)}. \tag{6.8}$$

We can create a lower bound for this equation with the help of Chernoff bounds, like we did in the proof of Theorem 5.1.1. Because of large overlap with calculations in this proof, we work out this lower bound on $V(t^*)$ in Appendix B. In Corollary B.0.2 we derive that whp

$$V(t^*) \geq \Theta\left(\frac{n}{(\log(n))^{\alpha(\tau-1)/(3-\tau)}}\right). \tag{6.9}$$

□

Now we know a $\log^{\alpha(n)}$-core exists for all constant $\alpha > 0$, we will examine whether a n^α-core exists in the next section.

6.2 Polynomial core

We will now prove Theorem 2.3.3, to show a n^α-core exists for all $\alpha < \frac{3-\tau}{8\tau}$.

Proof of Theorem 2.3.3. We use the algorithm that was created in Section 2.4, that follows the configuration model from Definition 2.2.1. When the number of heavy balls is larger than the number of total balls, this algorithm reaches the $k(n)$-core. In Section 2.5 it was derived that this is the case if (2.17), (2.18) and (2.19) hold for some $\delta > 0$. We will now derive a $\delta > 0$ such that these equations hold, with help of theorems from the previous chapters.

To satisfy (2.17), Theorem 5.1.1 will be used. This theorem has the condition that $n^{(\delta-1)/2} < \Theta\left(k_n^{(2-\tau)/(3-\tau)}\right)$. We fix a constant $\alpha > 0$, then this becomes the following for $k_n = n^\alpha$:

$$n^{(\delta-1)/2} < \Theta\left((n^\alpha)^{(2-\tau)/(3-\tau)}\right). \tag{6.10}$$

By using definition of Θ (2.3.1) we get, that we need to have:

$$\frac{n^{(\delta-1)/2}}{(n^\alpha)^{(2-\tau)/(3-\tau)}} \to 0. \tag{6.11}$$

This is the case if

$$\frac{\delta-1}{2} - \alpha\frac{(2-\tau)}{(3-\tau)} < 0. \tag{6.12}$$

We rewrite this to conclude that for satisfying Theorem 5.1.1, we need

$$\alpha < \frac{(\delta-1)(3-\tau)}{2(2-\tau)} = \frac{(1-\delta)(3-\tau)}{2(\tau-2)}. \tag{6.13}$$

Theorem 5.1.1 then tells us that (5.2) is satisfied for certain constant $c > 0$ and

$$p = cn^{\alpha(2-\tau)/(3-\tau)}. \tag{6.14}$$

Therefore, also for α and δ that satisfy (6.13), (2.17) is satisfied at the time t^* with $p = e^{-t^*}$. Next, we will investigate which δ we need to choose to satisfy (2.18) and (2.19) whp. By Theorem 3.3.2, (2.19) holds whp for $\delta \geq \frac{3\tau+2}{4\tau}$. Also by Theorem 4.5.2, (2.18) holds whp for $\delta > \frac{1}{2}$. By (6.5), we conclude that the best possible choice is $\delta = \frac{3\tau+2}{4\tau}$. Therefore we conclude that we have reached the n^α-core at t^* if

$$\begin{aligned}\alpha < \frac{(1-\frac{3\tau+2}{4\tau})(3-\tau)}{2(\tau-2)} &= \frac{(\frac{\tau-2}{4\tau})(3-\tau)}{2(\tau-2)} \\ &= \frac{3-\tau}{8\tau}.\end{aligned} \tag{6.15}$$

This is the reason, why the theorem only holds for $\alpha < \frac{3-\tau}{8\tau}$, for the rest of the proof we use this constraint on α. If there are still heavy balls left at t^*, then a n^α-core exists whp. Otherwise it is possible that the core is empty. We estimate the minimum number of heavy balls left using (6.6). The expected number of balls left at t^* is

$$\mathbb{E}\left[B(t^*)\right] = Mnp^2 = c^2Mn\left(n^{2\alpha(2-\tau)/(3-\tau)}\right). \tag{6.16}$$

Therefore, whp the minimum number of edges in the n^α-core is

$$\Theta\left(n^{1-2\alpha(\tau-2)/(3-\tau)}\right). \tag{6.17}$$

Thus a n^α-core exists whp. Let us now determine a lower bound on the number of vertices in the n^α-core, then we know by Theorem 3.4.1 that for the function $v(e^{-t})$ from (3.4) with $k(n) = n^\alpha$ it holds whp that for the value of p from (6.14)

$$V(t^*) \geq v(p) - n^{-\alpha+(3\tau+2)/(4\tau)}. \tag{6.18}$$

We work out a lower bound for this equation in Appendix B. In Corollary B.0.3, we obtain the result that whp the number of vertices in the n^α-core is minimally

$$\Theta\left(n^{1-\alpha(\tau-1)/(3-\tau)}\right). \tag{6.19}$$

□

We have now proved the main results of this thesis. In the next chapter, we will look back at the research and look at possible future research opportunities.

Chapter 7

Conclusion & Discussion

During this research, we investigated if there exists a $k(n)$-core in the scale-free configuration model. The scale-free configuration model is a useful null-model for scale-free networks. Scale-free networks occur in many branches of science. To analyze the structure of a network, k-cores have already been an important tool. In this research, we shifted focus to $k(n)$-cores where k is dependent on the number of vertices. By investigating $k(n)$-cores, we hope to get a better understanding of the highest k such that a k-core exists when we have n vertices. Additionally, we hope $k(n)$-cores can give a better understanding of how the highest k such that a k-core exists changes, when the number of nodes in a network changes.

In this research we chose to work with degrees that are dpareto(τ)-distributed with $\tau \in (2,3)$. First in Chapter 2 we stated the main results on existence of $\log^{\alpha}(n)$-cores and n^{α}-cores in graphs formed by the scale-free configuration model. In this chapter, we also explained how we can express the configuration model as a death process. We found an expression for the total number of edges left to make (all balls) and the number of edges left which could be part of the $k(n)$-core (heavy balls). Approximation of the number of heavy edges left at a time was done in Chapter 3, after this the number of edges was approximated in Chapter 4. Then, in Chapter 5 we determined the time when all remaining edges in the process will be part of the $k(n)$-core. Finally, in Chapter 6 we proved the main results on existence of $\log^{\alpha}(n)$-cores and n^{α}-cores. We found that for all $\alpha > 0$ a $\log^{\alpha}(n)$-core exists whp, while for n^{α} we only know a core exists whp for: $\alpha < \frac{3-\tau}{8\tau}$. We also gave a lower bound for the number of vertices and edges left in the cores. In conclusion, we managed to prove that there exist $k(n)$-cores in the scale-free configuration model whp.

There are chances to extend the research done in this thesis. In this thesis only lower bounds were given for the number of vertices and edges in the $k(n)$-cores. For k-cores, Janson & Luczak have managed to create a central limit theorem for the number of vertices and edges [21]. It would be interesting to see whether this central limit theorem could be extended to $k(n)$-cores.

Other future research could be to investigate if a $k(n)$-core exists on a different graph type. Research on the existence of k-cores has already been done on random r-uniform n-vertex hypergraphs [6]. In this graph type, there are n vertices and every edge consists of r vertices (an example can be seen in Figure 7.1). The amount of edges that a vertex is part of, is still called the degree. In [6], a construct algorithm is discussed to generate uniform a hypergraph that satisfies a degree sequence. They also form the edges in an order such that first everything outside the k-core is formed. It would be interesting to see if it is possible to formulate this construct algorithm as a death process and attempt to find a $k(n)$-core this way. Additionally, it could be investigated whether $k(n)$-cores exist in random r-partite hypergraphs. This type of hypergraph consists of r groups with n vertices each, while every edge consists of one vertex out of each group (which means this is a special type of r-uniform hypergraph). In [4] the existence of a k-core in a random r-partite hypergraph was already linked to the existence of a k-core in a r-uniform hypergraph. So it would be interesting to see if this is still the case for $k(n)$-cores.

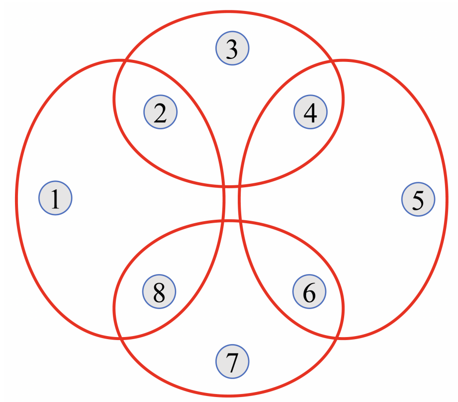


Figure 7.1: An example of a 3-regular hypergraph with 8 vertices and 4 edges [36].

It would also be interesting to see how the obtained theoretical results hold up in simulations. In this thesis, we worked with the situation that n tends to infinity. This means that it is not sure whether the $k(n)$-cores exist for a finite number of vertices. Simulating random graphs of different sizes with the help of the configuration model, could help shed light on the applicability of the results.

Bibliography

[1] R. Albert, H. Jeong, and A.-L. Barabási. “Diameter of the world-wide web”. In: *nature* 401.6749 (1999), pp. 130–131.

[2] J. Alvarez-Hamelin, L. Dall’Asta, A. Barrat, and A. Vespignani. “Large scale networks fingerprinting and visualization using the k-core decomposition”. In: *Advances in neural information processing systems* 18 (2005).

[3] A.-L. Barabási and E. Bonabeau. “Scale-free networks”. In: *Scientific american* 288.5 (2003), pp. 60–69.

[4] F. C. Botelho, N. Wormald, and N. Ziviani. “Cores of random r-partite hypergraphs”. In: *Information Processing Letters* 112.8 (2012), pp. 314–319. ISSN: 0020-0190. DOI: https://doi.org/10.1016/j.ipl.2011.10.017. URL: https://www.sciencedirect.com/science/article/pii/S0020019011002985.

[5] A. Broido and A. Clauset. “Scale-free networks are rare”. In: *Nature Communications* 10 (Mar. 2019). DOI: 10.1038/s41467-019-08746-5.

[6] C. Cooper. “The cores of random hypergraphs with a given degree sequence”. In: *Random Structures & Algorithms* 25.4 (2004), pp. 353–375.

[7] A. Dvoretzky, J. Kiefer, and J. Wolfowitz. “Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator”. In: *The Annals of Mathematical Statistics* 27.3 (1956), pp. 642–669. DOI: 10.1214/aoms/1177728174. URL: https://doi.org/10.1214/aoms/1177728174.

[8] H. Ebel, L.-I. Mielsch, and S. Bornholdt. “Scale-free topology of e-mail networks”. In: *Physical review E* 66.3 (2002), p. 035103.

[9] V. M. Eguiluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian. “Scale-free brain functional networks”. In: *Physical review letters* 94.1 (2005), p. 018102.

[10] M. Faloutsos, P. Faloutsos, and C. Faloutsos. “On power-law relationships of the Internet topology”. In: *SIGCOMM Comput. Commun. Rev.* 29.4 (Aug. 1999), pp. 251–262. ISSN: 0146-4833. DOI: 10.1145/316194.316229. URL: https://doi.org/10.1145/316194.316229.

[11] D. Fernholz and V. Ramachandran. “The giant k-core of a random graph with a specified degree sequence”. In: *manuscript, UT-Austin* (2003).

[12] P. Finke and Z. Zhu. *Exponential distribution*. 2019. URL: https://www.math.nagoya-u.ac.jp/~richard/teaching/s2019/Finke_Zhu.pdf.

[13] G. Fitzner. *Configuration model, degrees with power law distribution*. URL: https://www.networkpages.nl/CustomMedia/Animations/RandomGraph/CM/PowerLaw_DegreeDist.html.

[14] J. García-Algarra, J. M. Pastor, J. M. Iriondo, and J. Galeano. "Ranking of critical species to preserve the functionality of mutualistic networks using the k-core decomposition". In: *PeerJ* 5 (2017), e3321.

[15] G. Grimmett and D. J. A. Welsh. *Probability: An Introduction*. Oxford University Press, 2014. ISBN: 9780198709961. URL: https://books.google.nl/books?id=wIUSDAAAQBAJ.

[16] M. P. van den Heuvel and O. Sporns. "Rich-club organization of the human connectome". In: *Journal of Neuroscience* 31.44 (2011), pp. 15775–15786.

[17] R. van der Hofstad. *Random Graphs and Complex Networks*. 2017. URL: https://www.win.tue.nl/~rhofstad/NotesRGCN.pdf.

[18] R. van der Hofstad. *Random Graphs and Complex Networks*. 2021. URL: https://www.win.tue.nl/~rhofstad/NotesRGCNII_04_12_2021.pdf.

[19] R. van der Hofstad, J. Komjáthy, and V. Vadon. *Phase transition in random intersection graphs with communities*. 2019. arXiv: 1905.06253 [math.PR].

[20] S. Janson and M. J. Luczak. *A simple solution to the k-core problem*. 2005. arXiv: math/0508453 [math.CO].

[21] S. Janson and M. J. Luczak. "Asymptotic normality of the k-core in random graphs". In: (2008).

[22] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. "The large-scale organization of metabolic networks". In: *Nature* 407.6804 (2000), pp. 651–654.

[23] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. Eugene Stanley, and H. A. Makse. "Identification of influential spreaders in complex networks". In: *Nature Physics* 6 (Aug. 2010), pp. 888–893. URL: https://www.nature.com/articles/nphys1746#citeas.

[24] Y.-X. Kong, G.-Y. Shi, R.-J. Wu, and Y.-C. Zhang. "k-core: Theories and applications". In: *Physics Reports* 832 (2019). k-core: Theories and Applications, pp. 1–32. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2019.10.004. URL: https://www.sciencedirect.com/science/article/pii/S037015731930328X.

[25] F. Liljeros, C. R. Edling, H. Eugene Stanley, Y. Aberg, and L. A. Nunes Amaral. "Distributions of number of sexual partnerships have power law decaying tails and finite variance". In: *arXiv preprint cond-mat/0305528* (2003).

[26] F. D. Malliaros, C. Giatsidis, A. N. Papadopoulos, and M. Vazirgiannis. "The core decomposition of networks: Theory, algorithms and applications". In: *The VLDB Journal* 29.1 (2020), pp. 61–92.

[27] G. Marco and M. Patrignani. "Dynamic Analysis of the Autonomous System Graph". In: 2004. URL: https://api.semanticscholar.org/CorpusID:3042216.

[28] P. Massart. "The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality". In: *The Annals of Probability* 18.3 (1990), pp. 1269–1283. DOI: 10.1214/aop/1176990746. URL: https://doi.org/10.1214/aop/1176990746.

[29] NIST. *Autonomous System (AS)*. URL: https://csrc.nist.gov/glossary/term/autonomous_system#:~:text=Definitions%3A,operating%20the%20same%20routing%20policy.&text=An%20Autonomous%20System%20specifies%20a,network%20addresses%20to%20the%20Internet.

[30] G. Ódor, D. Czifra, J. Komjáthy, L. Lovász, and M. Karsai. "Switchover phenomenon induced by epidemic seeding on geometric networks". In: *Proceedings of the National Academy of Sciences* 118.41 (2021).

[31] S.K. and H. M. Taylor. "Chapter 6 - Martingales". In: *A First Course in Stochastic Processes (Second Edition)*. Ed. by S. Karlin and H. M. Taylor. Second Edition. Boston: Academic Press, 1975, pp. 238–339. ISBN: 978-0-08-057041-9. DOI: https://doi.org/10.1016/B978-0-08-057041-9.50010-6. URL: https://www.sciencedirect.com/science/article/pii/B9780080570419500106.

[32] Statistics How To. *Zeta Distribution (Zipf Distribution)*. URL: https://www.statisticshowto.com/zeta-distribution-zipf/.

[33] A. Tsun. *Chapter 6. Concentration Inequalities*. URL: https://courses.cs.washington.edu/courses/cse312/22wi/files/student_drive/6.2.pdf.

[34] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow. *The Anatomy of the Facebook Social Graph*. 2011. arXiv: 1111.4503 [cs.SI].

[35] Wikipedia. *Exponential distribution*. 2024. URL: https://en.wikipedia.org/wiki/Exponential_distribution#Distribution_of_the_minimum_of_exponential_random_variables.

[36] X.-J. Xu, S. He, and L.J. Zhang. "Dynamics of the threshold model on hypergraphs". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 32 (Feb. 2022), p. 023125. DOI: 10.1063/5.0075667.

[37] B. Xue. *Birth-Death Process*. 2022. URL: https://cmp.phys.ufl.edu/PHZ4710/files/unit3/birth-death.html.

[38] F. Zhang, Y. Zhang, L. Qin, W. Zhang, and X. Lin. "Finding Critical Users for Social Network Engagement: The Collapsed k-Core Problem". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 31.1 (Feb. 2017). DOI: 10.1609/aaai.v31i1.10482. URL: https://ojs.aaai.org/index.php/AAAI/article/view/10482.

[39] X. Zhong and H. Liang. "On the Scale-Free Property of Citation Networks: An Empirical Study". In: *Companion Proceedings of the ACM on Web Conference 2024*. 2024, pp. 541–544.

Appendix A

Proofs

In this appendix, we will give proofs for four theorems and corollaries from Chapters 3 and 4. This are Corollary 3.4.1, Theorem 4.3.1, Corollary 4.4.2 and Corollary 4.4.3.

Proof of Corollary 3.4.1. To prove this statement, we slightly alter the proof of Theorem 3.3.1. Similarly to (3.14), we define

$$\begin{aligned} V_L(t) &:= \sum_{l=k_n}^{L_n} \sum_{r=k_n}^{l} U_{lr}(t) \\ v_L(e^{-t}) &:= \sum_{l=k_n}^{L_n} np_l(n) \sum_{r=k_n}^{l} m_{lr}(t). \end{aligned} \tag{A.1}$$

In Theorem 3.3.1 it was proven (starting from (3.25)) that whp

$$\sum_{l=k_n}^{L_n} \sum_{r=k_n}^{l} r\sup_{t\geq 0} \left| (np_l(n)m_{lr}(t) - U_{lr}(t)) \right| < \Theta\left(n^{\hat{\beta}}\right). \tag{A.2}$$

Since $r \geq k_n$, also

$$k_n \sum_{l=k_n}^{L_n} \sum_{r=k_n}^{l} \sup_{t\geq 0} \left| (np_l(n)m_{lr}(t) - U_{lr}(t)) \right| < \Theta\left(n^{\hat{\beta}}\right). \tag{A.3}$$

And therefore we know whp that

$$\sum_{l=k_n}^{L_n} \sum_{r=k_n}^{l} \sup_{t\geq 0} \left| (np_l(n)m_{lr}(t) - U_{lr}(t)) \right| < \Theta\left(\frac{n^{\hat{\beta}}}{k_n}\right). \tag{A.4}$$

And this gives a bound on $|V_L(t) - v_L(e^{-t})|$. It was also shown in the proof of Theorem 3.3.1 (starting from (3.17)) that

$$\sum_{L_n}^{\infty} \sum_{r=k_n}^{l} rU_{lr}(t) < \Theta\left(n^{\hat{\beta}}\right). \tag{A.5}$$

By using that $r \geq k_n$ again, we obtain

$$\sum_{L_n}^{\infty} \sum_{r=k_n}^{l} U_{lr}(t) < \Theta\left(\frac{n^{\hat{\beta}}}{k_n}\right). \tag{A.6}$$

This gives a bound on $|V(t) - V_L(t)|$. Lastly, it was shown (starting from (3.18)) that

$$\sum_{l=L_n}^{\infty} np_l(n) \sum_{r=k_n}^{l} rm_{lr}(t) < \Theta\left(n^{\beta}\right). \tag{A.7}$$

So by using $r \geq k_n$ for a third time, also

$$\sum_{l=L_n}^{\infty} np_l(n) \sum_{r=k_n}^{l} m_{lr}(t) < \Theta\left(\frac{n^{\beta}}{k_n}\right). \tag{A.8}$$

Which also gives us a bound on $|v_L(e^{-t}) - v(e^{-t})|$. From the three obtained bounds, (3.58) follows analogously to the proof of Theorem 3.3.1. □

Proof of Theorem 4.3.1. During this proof, the conditions in Lemma 4.2.3 will be verified. First, let us verify that $\mathbb{E}[N_l^2|N_{l-1}^2, \ldots, N_1^2] \geq N_{l-1}^2$. For this, first N_l^2 is written out:

$$\mathbb{E}[N_l^2|N_{l-1}^2, \ldots, N_1^2] = \mathbb{E}\left[\left(\sum_{i=1}^{l} \frac{E_i - 1}{Mn - \gamma(i-1)}\right)^2 \middle| N_{l-1}^2, \ldots, N_1^2\right]. \tag{A.9}$$

To lighten the notation, let us write $c = Mn - \gamma(l-1)$. Using this notation, we can rewrite the sum as

$$\begin{aligned}\left(\sum_{i=1}^{l} \frac{E_i - 1}{Mn - \gamma(i-1)}\right)^2 &= \left(N_{l-1} + \frac{E_l - 1}{c}\right)^2 \\ &= N_{l-1}^2 + 2N_{l-1}\left(\frac{E_l - 1}{c}\right) + \left(\frac{E_l - 1}{c}\right)^2.\end{aligned} \tag{A.10}$$

By filling in (A.10) into (A.9), we obtain that

$$\mathbb{E}\left[N_{l-1}^2 + 2N_{l-1}\left(\frac{E_l - 1}{c}\right) + \left(\frac{E_l - 1}{c}\right)^2 \middle| N_{l-1}^2, \ldots, N_1^2\right]. \tag{A.11}$$

By linearity of the expectation, we can split this expectation into three separate parts. Let us now simplify these three parts. It is known that $\frac{E_l-1}{c}$ is independent of other exponential variables and therefore also independent of previous N_i^2-values. Therefore

$$\mathbb{E}\left[\left(\frac{E_l - 1}{c}\right)^2 \middle| N_{l-1}^2, \ldots, N_1^2\right] = \mathbb{E}\left[\left(\frac{E_l - 1}{c}\right)^2\right] \geq 0. \tag{A.12}$$

By using that N_{l-1}^2 is already known in the conditional expectation, it is also known that

$$\mathbb{E}\left[N_{l-1}^2 \middle| N_{l-1}^2, \ldots, N_1^2\right] = N_{l-1}^2. \tag{A.13}$$

By again using that $\frac{E_l-1}{c}$ is independent of other exponential variables, and by using multiplying rules for independent variables in the expectation we obtain that

$$\mathbb{E}\left[2N_{l-1}\left(\frac{E_l - 1}{c}\right) \middle| N_{l-1}^2, \ldots, N_1^2\right] = 2\mathbb{E}\left[\frac{E_l - 1}{c}\right] N_{l-1} = 0. \tag{A.14}$$

By filling in equations (A.12), (A.13) and (A.14) into (A.11), we obtain that:

$$\mathbb{E}[N_l^2 | N_{l-1}^2, \ldots, N_1^2] = N_{l-1}^2 + 0 + \mathbb{E}\left[\left(\frac{E_l - 1}{c}\right)^2\right] \geq N_{l-1}^2. \tag{A.15}$$

This means that condition (II) of Lemma 4.2.3 is satisfied. Now let us verify that also $\mathbb{E}\left[\max\{N_l^2, 0\}\right] < \infty$. Since N_l^2 is always nonnegative, this statement reduces to $\mathbb{E}\left[N_l^2\right] < \infty$. Notice that $E[N_l] = 0$, as it was designed this way in (4.16). Therefore we can use that

$$\begin{aligned} \mathbb{E}\left[N_l^2\right] &= \mathbb{E}\left[N_l^2\right] - \mathbb{E}\left[N_l\right]^2 = \mathrm{Var}[N_l] \\ &= \mathrm{Var}\left(\sum_{i=1}^{l} \frac{E_i - 1}{Mn - \gamma(i-1)}\right). \end{aligned} \tag{A.16}$$

Using the summation rules for variance of independent random variables and recalling that all E_i are $\exp(1)$ distributed, we can rewrite this to:

$$\begin{aligned} \sum_{i=1}^{l} \mathrm{Var}\left(\frac{E_i}{Mn - \gamma(i-1)}\right) &= \sum_{i=1}^{l} \frac{1}{(Mn - \gamma(i-1))^2} \mathrm{Var}\left(E_i\right) \\ &= \sum_{i=1}^{l} \frac{1}{(Mn - \gamma(i-1))^2}. \end{aligned} \tag{A.17}$$

And this sum is finite for every chosen value of l. Therefore also condition (I) in Lemma 4.2.3 is satisfied and thus N_l^2 is a submartingale. □

Proof of Corollary 4.4.2. Firstly, we rewrite t^* to a form that can be used in Theorem 4.4.1. We first take a factor $\frac{1}{2}$ out of the exponent in the logarithm:

$$t^* = \log\left(\hat{c}\log^{\omega/2}(n)\right) = \frac{1}{2}\log\left(\hat{c}^2 \log^{\omega}(n)\right). \tag{A.18}$$

Now we rewrite the term in the outer logarithm to

$$t^* = \frac{1}{2}\log\left(\hat{c}^2 \frac{n}{\frac{n}{\log^{\omega}(n)}}\right) = \frac{1}{2}\log\left(\frac{Mn}{\frac{Mn}{\hat{c}^2 \log^{\omega}(n)}}\right). \tag{A.19}$$

We now have the desired form for Theorem 4.4.1 by setting $\gamma = 2$ and

$$x = \frac{M}{\hat{c}^2} \frac{n}{\log^{\omega}(n)}. \tag{A.20}$$

We can use this form of x in Corollary 4.3.4. In this corollary, it was shown that for $\beta \in (0, \frac{1}{2})$. We set $\beta = \frac{1}{2} - \upsilon$, we have to assume that $\upsilon < \frac{1}{2}$ for this. We will explain the situation when $\upsilon \geq \frac{1}{2}$ at the end of the proof. Now by Corollary 4.3.4,

$$\mathbb{P}\left[\sup_{y \geq x - n^{1-\beta}} \left|\mathcal{T}_y - \mathbb{E}\left[\mathcal{T}_y\right]\right| > n^{-\beta}\right] \to 0. \tag{A.21}$$

To be able to use Theorem 4.4.1, we want $\beta + \epsilon \in (0, \frac{1}{2})$ to hold as well for a certain $\epsilon > 0$. To achieve this we set

$$\epsilon = \frac{1/2 - \beta}{2} > 0. \tag{A.22}$$

But then by Corollary 4.3.4 it holds that

$$\mathbb{P}\left[\sup_{y\geq x-n^{1-\beta}}|\mathcal{T}_y-\mathbb{E}\left[\mathcal{T}_y\right]|>n^{-\beta-\epsilon}\right]\to 0. \tag{A.23}$$

This gives us the possibility to use Theorem 4.4.1, which yields,

$$\mathbb{P}\left[\sup_{t\leq t^*}|X^{(n)}(t)-Mne^{-2t}|>n^{1-\beta}\right]\to 0. \tag{A.24}$$

By filling in β, we derive

$$\mathbb{P}\left[\sup_{t\leq t^*}|X^{(n)}(t)-Mne^{-2t}|>n^{1/2+\upsilon}\right]\to 0. \tag{A.25}$$

The result also holds if we take $\hat{\upsilon}=\frac{\upsilon}{2}$, since this is also bigger than 0. But this means that the convergence still holds for a slightly lower order of n. Therefore the result definitely stays valid if $n^{1/2+\upsilon}$ is replaced by $\frac{1}{2}n^{1/2+\upsilon}$. Lastly, we have only shown convergence using Theorem 4.4.1 for $\upsilon<\frac{1}{2}$. But this result still holds for all $\upsilon>0$ since this is a weaker result. □

Proof of Corollary 4.4.3. First we rewrite t^* to a form which can be used in Theorem 4.4.1

$$t^*=log\left(\tilde{c}n^{(1-\omega)/2}\right)=\frac{1}{2}\log\left(\tilde{c}n^{1-\omega}\right)=\frac{1}{2}\log\left(\frac{Mn}{\frac{M}{\tilde{c}}n^{\omega}}\right). \tag{A.26}$$

This is the desired form, when setting $\gamma=2$ and

$$x=\frac{M}{\tilde{c}}n^{\omega}. \tag{A.27}$$

We can use this form of x in Corollary 4.3.5, in this corollary it was shown that for $\omega>\frac{2}{3}$ and $\beta<\frac{\omega}{2}$,

$$\mathbb{P}\left[\sup_{y\geq x-n^{1-\beta}}|\mathcal{T}_y-\mathbb{E}\left[\mathcal{T}_y\right]|>n^{-\beta}\right]\to 0. \tag{A.28}$$

By setting $\epsilon=\frac{1}{2}\left(\frac{\omega}{2}-\beta\right)$, also $\beta+\epsilon<\frac{\omega}{2}$ and thus the statement still applies for $n^{-\beta-\epsilon}$ instead of $n^{-\beta}$. But then by taking $\gamma=2$, Theorem 4.4.1 almost shows that (4.61) holds. The only difference is the factor $\frac{1}{2}$. But since we can take a larger $\beta<\frac{\omega}{2}$, for which the equation holds this factor does not cause a problem. □

Appendix B

Lower bounds number of vertices

In this appendix, we will prove a lower bound on the number of vertices that is used in Chapter 6. We want to determine a lower bound on $v(p)$ from (3.4), since this terms occurs in both (6.8) and (6.18) for certain $k(n)$. Later this chapter, we choose $k(n)$ suitable for (6.8) and (6.18), but first we will use general $k(n)$.

Theorem B.0.1 (Lower bound $v(p)$). *For $v(p)$ from* (3.4)*, it holds that*

$$v(p) \geq \Theta\left(n\left(\frac{k_n}{p}\right)^{1-\tau}\right). \tag{B.1}$$

Proof. By using the expression from (3.5) and the knowledge that by Condition 2.2.2 $p_l(n) > \hat{c}p_l$ for $\hat{c} < 1$, we see that

$$v\left(p\right) \geq n \sum_{l=\lceil k_n \rceil}^{n^\epsilon} \frac{\hat{c}}{\zeta(\tau) l^\tau} \mathbb{P}\left[X_l(p) \geq k_n\right]. \tag{B.2}$$

To bound the probabilities in the sum, we first note that

$$\begin{aligned} \mathbb{P}\left[X_l(p) \geq k_n\right] &= 1 - \mathbb{P}\left[X_l(p) < k_n\right] \\ &\geq 1 - \mathbb{P}\left[X_l(p) \leq k_n\right]. \end{aligned} \tag{B.3}$$

By using the Chernoff bound from (5.13), we obtain that for $lp \geq 2k_n$, it holds that

$$\mathbb{P}\left[X_l(p) \geq k_n\right] \geq 1 - e^{-k_n/4}. \tag{B.4}$$

Since $p \leq 1$, this also means that $l \geq 2k_n > k_n$. But then by using the Chernoff bound, we get:

$$n \sum_{l=\lceil k_n \rceil}^{n^\epsilon} \frac{\hat{c}}{\zeta(\tau) l^\tau} \mathbb{P}\left[X_l(p) \geq k_n\right] \geq n \sum_{l=\lceil \frac{2}{p} k_n \rceil}^{n^\varepsilon} \frac{\hat{c}}{\zeta(\tau)} l^{-\tau}\left(1 - e^{-k_n/4}\right). \tag{B.5}$$

Since $e^{-k_n/4}$ goes to zero as $n \to \infty$, eventually $\left(1 - e^{-k_n/4}\right) > \hat{c}$, but then

$$n \frac{\hat{c}^2}{\zeta(\tau)} \sum_{l=\lceil \frac{2}{p} k_n \rceil}^{n^\varepsilon} l^{-\tau}. \tag{B.6}$$

We are working with $\tau \in (2, 3)$. Therefore, $l^{-\tau}$ is a decreasing function in terms of l. Therefore, let us bound the sum using an integral bound for decreasing positive terms to get

$$\begin{aligned} \sum_{l=\lceil \frac{2}{p} k_n \rceil}^{n^\varepsilon} l^{-\tau} &\geq \int_{\lceil \frac{2}{p} k_n \rceil}^{n^\varepsilon} u^{-\tau} du \\ &\geq \int_{\frac{2}{p} k_n + 1}^{n^\varepsilon} u^{-\tau} du. \end{aligned} \tag{B.7}$$

Calculating this integral gives:

$$\left[\frac{1}{1-\tau} u^{1-\tau}\right]_{\frac{2}{p} k_n + 1}^{n^\varepsilon} = \frac{1}{1-\tau} \left(n^\varepsilon\right)^{1-\tau} - \frac{1}{1-\tau} \left(\frac{2}{p} k_n + 1\right)^{1-\tau}. \tag{B.8}$$

We are using a discrete Pareto distribution with $\tau \in (2, 3)$, therefore $\frac{1}{1-\tau} < 0$, so the second term is actually the positive one. Also the power of u is negative because of this, so $u = \frac{2k}{p} + 1$ gives a larger value than $u = n^\varepsilon$. And this difference becomes bigger to eventually go below $(1 - \hat{c})$ times the leading term for large n. Therefore we conclude that

$$v(p) \geq n \frac{\hat{c}^3}{\zeta(\tau)(\tau - 1)} \left(\frac{2}{p} k_n + 1\right)^{1-\tau} = \Theta\left(n \left(\frac{k_n}{p}\right)^{1-\tau}\right). \tag{B.9}$$

□

We will now use this theorem for $k_n = \log^\alpha(n)$, to obtain a lower bound for (6.8).

Corollary B.0.2 (Lower bound number of vertices in $\log^\alpha(n)$-core)**.** *For $v(p)$ from (3.4) and $p = c\left(\log(n)\right)^{\alpha(2-\tau)/(3-\tau)}$ for $\alpha > 0$ and $c > 0$ constant, it holds for sufficiently large n that*

$$v(p) - \frac{n^{(3\tau+2)/(4\tau)}}{\log^\alpha(n)} \geq \Theta\left(\frac{n}{\left(\log(n)\right)^{\alpha(\tau-1)/(3-\tau)}}\right). \tag{B.10}$$

Proof. Fix $\alpha > 0$ constant, using Theorem B.0.1 for $k_n = \log^\alpha(n)$ yields

$$\begin{aligned} v(p) &\geq \Theta\left(n \left(\left(\log(n)\right)^{\alpha(1-(2-\tau)/(3-\tau))}\right)^{1-\tau}\right) \\ &= \Theta\left(\frac{n}{\left(\log(n)\right)^{\alpha(\tau-1)/(3-\tau)}}\right). \end{aligned} \tag{B.11}$$

It holds that $v(p)$ has a larger power of n than

$$\frac{n^{(3\tau+2)/(4\tau)}}{\log^\alpha(n)}. \tag{B.12}$$

Therefore for sufficiently large n, $v(p)$ is the leading term in the left-hand side of (B.10). Thus it follows that (B.10) holds. □

We will now look at the case $k(n) = n^\alpha$, to also derive a lower bound for (6.18).

Corollary B.0.3 (Lower bound number of vertices in n^α-core)**.** *For $v(p)$ from (3.4) and $p = cn^{\alpha(2-\tau)/(3-\tau)}$ with constant $\alpha \in (0, \frac{3-\tau}{8\tau})$ and $c > 0$, it holds for sufficiently large n that*

$$v(p) - n^{-\alpha+(3\tau+2)/(4\tau)} \geq \Theta\left(n^{1-\alpha(\tau-1)/(3-\tau)}\right). \tag{B.13}$$

Proof. Fix $\alpha > 0$ constant, using Theorem B.0.1 for $k_n = n^\alpha$ yields

$$\begin{aligned} v(e^{-t}) &\geq \Theta\left(n\left(n^{\alpha(1-(2-\tau)/(3-\tau))}\right)^{1-\tau}\right) \\ &= \Theta\left(n^{1-\alpha(\tau-1)/(3-\tau)}\right). \end{aligned} \tag{B.14}$$

For $v(p)$ to be the leading term on the left-hand side of (B.13), we need

$$1 - \frac{\alpha(\tau-1)}{(3-\tau)} > -\alpha + \frac{3\tau+2}{4\tau} \tag{B.15}$$

Re-ordering this equation gives

$$\alpha\left(\frac{2(2-\tau)}{3-\tau}\right) > \frac{2-\tau}{4\tau}. \tag{B.16}$$

And then we conclude that we need

$$\alpha < \frac{3-\tau}{8\tau}, \tag{B.17}$$

which is true for the values we have discovered a n^α-core exists. So $v(p)$ is the leading term on the left-hand side of (B.13). Thus (B.13) holds. $\square$

Remark B.0.4. *This corollary does actually give a result that grows as the number of vertices grows, since*

$$1 - \alpha\frac{\tau-1}{3-\tau} > 1 - \frac{\tau-1}{8\tau} = \frac{7\tau+1}{8\tau} > 0. \tag{B.18}$$