K-cores in random graphs

Vincent Wassenaar

TU Delft

K-cores in random graphs

by

Vincent Wassenaar

to obtain the degree of Bachelor of Science
at the Delft University of Technology.

Student number: 4251490
Project duration: April 18, 2022 – July 8, 2022
Thesis committee: Dr. J. Komjáthy, TU Delft, supervisor
Dr. J.G. Spandaw , TU Delft

This thesis is confidential and cannot be made public until August 31, 2022.

TU Delft

Abstract

A graph $G = (V, E)$is a mathematical model for a network with vertex set V and edge set E. A Random Graph model is a probabilistic graph. A Random Geometric Graph is a Random Graph were each vertex has a location in a space χ. We compare the Erdős-Rényi random graph, $G(n, p)$, to the Random Geometric Graph model, RGG(n, r) where, in general we use $r = c \cdot n^{-\frac{1}{d}}$, with dimension d. It is known that for $p = \frac{\lambda_k}{n}$ the k-core has a first-order phase transition in $G(n, p)$ where λ_k is the critical value for the k-core. The k-core is a global property of a graph. The k-core is the largest induced subgraph where each vertex has at least degree k. We suggest by simulations and a supportive proof that for the RGG-model a first-order phase transition not plausible. A inhomogeneous extension of the RGG-model with a vertex weight distribution T is a Geometric Inhomogeneous Random Graph model (GIRG). We also prove why some heavy-tailed (i.e. power-law) distributions almost surely have a k-core, when the amount of vertices v which has a weight $w_v > \sqrt{n}$ is greater than k. Furthermore, we rephrase from known literature how using a fixed equation for a branching process is a useful tool for analysing the existence of a k-core. In particular, the critical value for the 3-core is recovered using the probability of a binary tree embedding in branching processes, with the root having at least 3 children.

Lay abstract

Networks are vital to modern life. In this thesis we study networks and in particular the stability of networks using mathematical models called graphs. We investigate the behaviour of a well-known property, k-cores, as a telltale for stability of networks. Most importantly we compare this property between spatial networks, for example wireless transmission networks or electricity grids, and non-spacial, like a social network. The latter has a critical point where the stability drastically increase in terms of strong the network is interconnected. In spatial networks the stability grows proportional to the interconnectedness as one would suspect. In this thesis we try to understand this difference directly and via a model to describe network growth.

Preface

Due to renovating my house, the first two weeks were far from optimal. This meant I had to compensate in the weeks to come, which was challenging at times. Looking back I realised I have learned a lot, primarily on the topic of random graphs. So many new interesting topics I have read in the field of random graphs and branching processes, that I realise even more how little I have yet grasped in this exciting area of mathematics. However, I can confidently say that my scientific writing has improved during this thesis.

To that end, I would first and foremost like to thank Júlia Komjáthy not only for her patience, guidance and feedback as a mentor, but also for her very extensive knowledge in the field of random graphs. I learned much from explanations during our weekly sessions, both in the matter of probability as well as in properly mathematical writing. Not in the least, to write more objectively without accidental bashing respected researchers.

Moreover, I would like to thank Jeroen Spandaw for taking the time to review this thesis and immense helpful feedback during the colloquium part of the bachelor thesis.

Lastly, great support I have had throughout this thesis from my family and my partner, Maaike, in particular.

Vincent Wassenaar
Delft, July 2022

Contents

1 Introduction

1.1. Modern Networks

Networks are vital to modern life, which becomes especially visible when a network fails as mentioned by Strogatz, a mathematician from Cornell University [74]. He mentions a major power blackout in the US and Canada in 1996 due to a cascading series of failure, which provides insight in the underlying power grid structure [30].

Networks are thoroughly studied for a myriad purposes [32]. A few examples are presented. The 'small-world effect', which states how people are connected through a limited number of intermediates. This effect serves as a possible predictor of success, when collaborating with people [60]. Another example, linguists study of semantic networks, which provide topological insight and syntactic dependency trees [5, 39, 57]. Studying social networks is a third example. Graph theory is used how to predict unrest and group behavior when social groups break apart into smaller components after friendships end, based on the weight of ties [42]. Another social network example is a controversial paper, which describes how an obesity pandemic can be imitated by a time-evolving network, simulating population dynamics [28].

Traditionally networks are used to solve logistical problems and consequently algorithms have been developed. A few practical applications are: the shortest path algorithm for transportation optimization [64], cost effective cable networks using minimum spanning trees [76] and competitive currency stock-market trading using network cycles [67]. Moreover, humankind is reliant on computer networks to function properly. Hence greedy queuing algorithms for wireless network and switches are of interest [11]. In the 21st century the electricity grid, and our dependency of it, is of great importance. Therefore research on power grid robustness against cascading failures is another important example of network theory with practical applications [65, 78].

Networks can be gigantic, which makes them very difficult, if not impossible, to handle in a computational sense [53]. These gigantic networks are called 'large-scale networks'. Just trying all options (brute-force) to find an optimal strategy is folly in most cases. To illustrate this imagine all shipping containers ($\sim$15 million) to process in a major port like Rotterdam. To schedule these containers efficiently is an example of a task where efficient algorithms are vital. Other examples are, analysing the size of World Wide Web [2, 66, 71], the way the human brain regulates neuron interactions [21], molecular behaviour of hydration water around proteins [58] and large social networks [75]. These examples are a selection of a wide variety of papers in which large-scale network research is potent.

In 2001 Strogatz stated: "Researchers are only now beginning to unravel the structure and dynamics of complex networks" [74]. Many complex networks have been identified to are closely related to either the 'scale-free' or 'small-world' model. There is currently no precise technical definition of a *complex network*, but key markers have been identified and are widely accepted in the field [12, 59]:

1. The system consist of many interacting parts, i.e. large-scale.
2. The collective behavior supersedes the sum of the collective individual behaviour.
3. Dynamically evolve

Many examples listed are in fact complex systems; the brain, condensed matter systems, road traffic, the Internet (scale-free), Social networks (small-world). To study these phenomena simplified mathematical models are used.

1.1.1. Models of networks

In mathematics a representation of a network is called a '*graph*'. Both network and graph are used interchangeably in literature. Graphs are a mathematical model for representing networks, a form of abstraction which allows a wide collections of problems to be solved using knowledge and techniques acquired over the last decades. What makes a graph model powerful is the generalization and unification of problems when modelled as a graph. As illustrated with the examples, very different phenomena can be described by a graph. Therefore, efficient algorithms are universally applicable. A famous example is the use of Dijkstra's algorithm [29]. The software which uses graph theory to find efficient routes to complete all deliveries, is a product of developments in fast algorithms for graphs [22]. Hence most logistical operations, like trucking companies, supermarket chains, waste-collection services and food delivery companies use software based on these mathematical developments in graph theory.

Generally graphs will be represented an a list of points (*vertices* or *nodes*) and a list of connections between those points (*edges* or *links*). A drawing of a graph is a visual representation of these vertices and edges. An example is a metro network where the metro stations are the dots (*vertices*) and the lines between the stations are the connections (*edges*).

Figure 1.1: Visual representation of the metro network of Amsterdam

When formulating an adequate model for a network (graph), multiple difficulties for capturing reality into it arise. According to Strogatz six difficulties can be distinguished [74]:

1. Structural complexity and size of the network
2. Network evolution; Another difficulty is networks which evolve over time. Brain cells die and simultaneously new brain cells are created. New roads are build over time or old ones broadened. New relationships are formed and ended on daily basis within a social network. A routing network might turn on/off certain parts of a circuit. Importantly, to understand how the network evolves is a first step in order to predict its behaviour.
3. Connection diversity; The size of arteries of different sizes (*edge weight*) and flow direction (*edge direction*). Electricity circuits have charges (*edge sign* and *edge weight*)

4. Dynamical complexity
5. Node diversity; A computer chip has different type of components on its circuit. Brain cells are of various types. Crossroads or stations have different sizes and requirements.
6. Meta-complication; Mutual influences between the points above; Positive reinforcement of a neural connection is a good example. The connections get stronger, edge weights increase, if they are used more often. Another example are topology constraints. New power grid extensions are more common in highly dense urban areas.

It is important to stress that some real-world structures still lie beyond our current mathematical models. As with many areas in physics and mathematics simplifying certain processes or constraints are a key to still unlocking the basis underlying networks. When mathematicians try to model a phenomena, they try to find the simplest model possible which still reproduces the phenomena, whilst trying to filter which properties are crucial and which are irrelevant for the phenomena to occur.

Two sorts of graphs can be distinguished; deterministic and probabilistic. Deterministic graphs are those of which all information is fixed and known. Even though it may evolve over time, the way it evolves is governed by deterministic principles. A probabilistic graph however, is a probability distribution of a graph. The focus of this paper shall be on the latter. These probabilistic graphs are also referred to as *Random Graphs*.

1.1.2. Classical Random Graphs Models

Random graph theory was founded around 1960 by Paul Erdős and Alfréd Rényi as a consequence of probabilistic methods to solve graph problems. In 1959, Erdős proved the existence of graphs which has a *chromatic number*[1] of at least k and a girth[2] of at least g, under the condition that $k \geq 3, g \geq 3$. The idea of a non-constructive proof was a major breakthrough in this area of mathematics. This changed the way of looking at graphs and opened a new field; Random Graph Theory. They introduced the 'classical random graph' model [35], denoted $G(n, M)$, where n represents the number of vertices and M the number of edges. Moreover, $\mathscr{G}(n, M)$ stands for the collection of all possible graphs which consists of n vertices and M edges. The probability of picking a certain graph $G(n, M)$ out of the collection $\mathscr{G}(n, M)$ of $\binom{\binom{n}{2}}{M}$ possible graph is uniform. Note that edges are distinct and can be viewed as labelled.

A variant by Gilbert, $G(n, p)$, also fixes the number of vertices n, but each edge now has a probability p to be included in the graph or not [41]. The random graphs consisting of a vertex set V, with n vertices, consists of a induced subgraph of a complete graph on n vertices, where each edge is part of the subgraph with probability p.

Since the 1960's more than 10.000 papers have been published studying various properties of this classical Erdős-Rényi random graph model.

In 1980, B. Bollobás founded another model inspired by work of Bender and Canfield [9, 13]. Networks often do have multiple edges per vertex[15]. Some networks even have a fixed number of edges per vertex. A graph where every vertex has d edges connected to it vertex, is called 'd-regular'. This 'configuration model' is based on the notion of d-regular graphs. In this model every node starts out with d 'half-edges'. An example is shown in figure 1.2. At random half-edges are connected to form a matching, thus creating a random graph. The process of matching half-edges stops when there are no half-edges left to connect. Self-loops are allowed as well as multiple edges between two vertices. When these appear the graph becomes a 'multi-graph'. Other contributions to this model include asymptotic analysis on graph cycles, enumeration of matrices and labeled graphs [8, 81]. The configuration model has an extended literature, we refer the reader to the book of Hofstad on extensive treatment of the configuration model and its variation [45]. Expending on the configuration model, a degree sequence for vertices can also be used, where d_v is the degree for vertex v. Since early 2000, the configuration model is popular for real networks with inhomogeneous degree sequences [55]. Hence when modelling certain real-life network, the configuration model begin a multigraph can be advantageous, for example when modelling social media platform posts where the network is a multigraph with multiple edges between users and loops as users can post on their own page. Or it can be disadvantageous, if a simple graph is required. For the latter, simple graph variations of the configuration model exists, notably the

[1] The *chromatic number* is the minimum number for a graph to be coloured such that no adjacent vertices share the same colour
[2] The *girth* of a graph is the shortest length of a cycle.

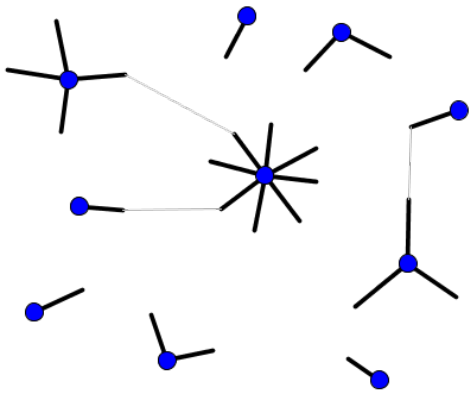


Figure 1.2: Visual representation of the configuration model. Vertices with half-edges. [63]

'erased configuration model' and 'repeated configuration model' [45]. Secondly, the number of users with 0 posts (i.e. vertex degree) does not align with the degree distribution of a classical random graph [45].

The field, ever branching and extending into discrete maths, computer science and even analysis has gained additional traction at the turn of the century. Different kind of random graph models are listed in a surveys by Dorogovtsev, Newton and Boccaletti [12, 32, 60]. With time and the advancement of computer one could verify how the classical model and configuration model compare to real-world data of large networks. Barabási and Albert showed that the World Wide Web did not adhere to this model [2] and the Faloutsos's showed similar results on Internet router networks in their famous paper [37]. In their paper the Faloutsos brothers showed how many real-world networks fit a log-log plot for their degree sequences. This verified the power-law of 'scale-free' networks as apposed to Gaussian bell-curves for the degree distribution.

1.1.3. Models for complex network

Another famous network model, published in the late 1990's, is the Watts-Strogatz (WS) model. Realizing that many real networks fall within a spectrum Watts and Strogatz came up with their network model in order to mimic this diversity in networks. Where on one end of the spectrum one has the d-regular graphs, which have no randomness, classical random graphs are on the other, being totally random. The Watts-Strogatz model introduces a parameter p to 'tune' between regularity ($p = 0$) and disorder ($p = 1$). Networks which can be produced by a rewiring process in the middle ground of the spectrum are coined 'small-world' graphs [79]. Small-world graph are characterized by short path lengths between vertices combined with high (local) clustering.

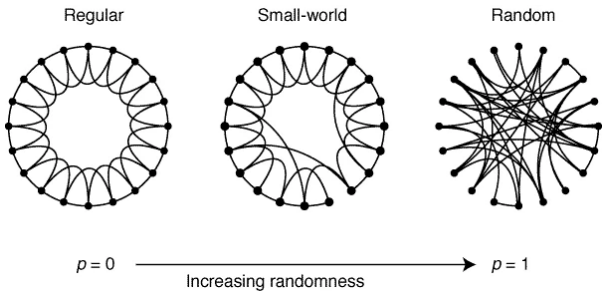


Figure 1.3: Rewiring procedure going clockwise, reconnect with probability p to another vertex uniformly. Duplicate edges are forbidden. with $n = 20$ and $d = 4$. [3]

There have been several variations since, most notably the Newman-Watts 'small-world' variation. In the Newman-Watts variation shortcuts are added between randomly chosen pairs of sites, but no connections are removed from the regular lattice [61]. Another adaptation, the Song-Wang adaptation with distance-dependent formulation, simplifies the generation of directed and undirected Watts-Strogatz model networks making it possible to derive exact expressions for the degree distribution and global

clustering coefficient [73]. Edges are drawn between pairs of vertices with a distance-based probability in stead of rewiring uniformly random. This adaptation also approaches a true classical random graph in the random limit case [52]. Moreover, the model allows a geometric embedding. Consider a discrete small-world model with a torus containing n vertices, then adding shortcuts uniformly between vertices. Distances in these discrete grids are linear.

Discoveries, for example by the Faloutsos brothers, led to formulation of another model, the preferential attachment model (PAM), by A.L. Barabási and R. Albert, also abbreviated as (BA-model) [6]. The standard preferential attachment model treats undirected graphs, while the directed preferential attachment graph model is also known as the Price's model after D. Price [70]. In this model the number of edges and vertices grow linearly in time. Competition or natural selection can stimulate growth and explained the behaviour of the existence of graphs generated with the preferential attachment model [10]. The preferential attachment model is prone to produce 'scale-free' networks [24]. Another example of this model is the economic reinforcement 'the-rich-get-richer', where having more resources enables more options to acquire even more resources at an increased rate [45].

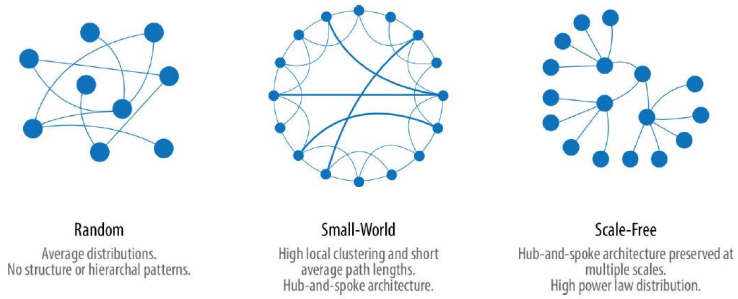


Figure 1.4: Different structural graphs. Left the 'classical random graph', middle a 'small-world' and at the right a 'scale-free' graph example. [3]

There has also been progress how these graphs interrelate and their benefits and disadvantages [1] A well-known fact is the vulnerability of scale-free networks to targeted attacks, which helps to analyse power grid failures [7, 25, 77, 82, 83]. Though scale-free networks have proven their worth, the universality of this graph model remains controversial [20].

1.2. Random Geometric Graphs

Random graphs also led to ideas closely related to statistical physics [31]. The dealing of atoms and molecules comes natural when looking at small-scale behaviour and network. The distance between atoms is related to the strength of the bonds between them. Thus is make sense to provide vertices with a geometrical property. Furthermore, in a dynamic graph, the location of a vertex can also change over time. A specific type of random graphs are random geometric graphs (RGGs), also known as random spatial graphs. In these RRGs a node is a point in a d-dimensional space. Besides physical dimensions, they represent various states of the underlying physics (e.g. temperature, pressure, activity). Others reasons for which spatial graphs are of interest are when the degrees are constraint by the physical space [12], like a limited number of streets that can cross an intersection [69], or radio towers where at a certain distance the signal becomes too weak. Wireless networks is an active field of application.

> *Suppose we immerse a large porous stone in a bucket of water. What is the probability the centre of the stone is wetted?* - Grimmett (1999) [43]

The percolation model, founded by Broadbent and Hammersley in 1957, considers a $\mathbb{Z}^2$ plane square lattice [19]. Furthermore p is the probability an edge between two neighbouring points on the lattice is *open* or *closed*. Related to the porous stone the edges are cavities in the stone through which water can pass. The stone itself can be modelled as a large, finite subset of this lattice. Many articles have been published on percolation in the sixty years, hence a cornerstone in Random Geometric Graphs. For further reading we refer to the book by Grimmit on 'Percolation' [43] and a surveys from Araujo et al. (2014) and Copin (2017) [4, 33].

Another geometric graph model uses the spatial Poisson point process to generate the vertex set. This process generates N points in the infinite plane $\mathbb{R}^2$ with a given density per finite area. The connectivity rule is defined as follows: Two points (vertices) are connected by an edge if their (Euclidean) distance is less then some threshold r. Other norms, e.g. l_1 or l_∞ can also be used. Important to note is that the number of vertices N, i.e. vertex set size, may vary for each realisation of this process.

Notably, Edgar Gilbert published about random planar networks as early as 1961. In this paper he defines a Poisson process to pick points from the infinite plane with density D points per unit area. The imposed edge condition is the Euclidean distance with radius R. He shows in [40] that there exists a critical value r_{crit} to estimate whether there exists an infinite component or not. Gilbert's research paper provides useful information on the size of the graph components.

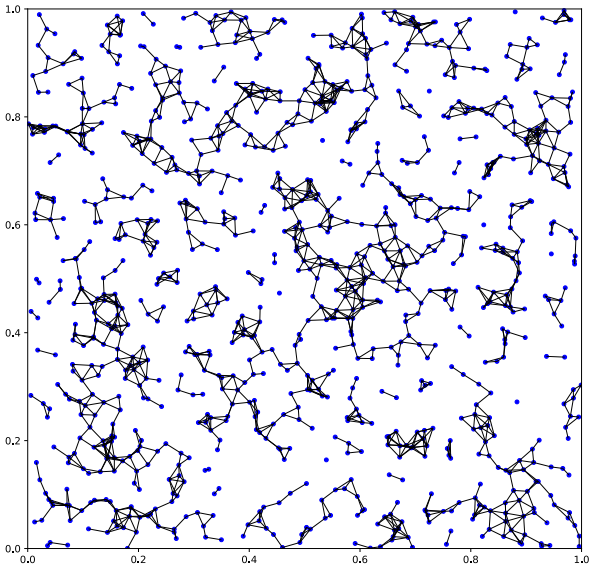


Figure 1.5: Visual of generated random spatial graph ($n = 250, r = 0.05$) on $[0, 1] \times [0, 1]$. Points were drawn uniformly on the grid. Edge condition used is the 2-norm with radius r. No wrapping on the borders

1.3. k-core property

Properties of graphs give insight into the structure and provide useful leads for finding better understanding and solving graphs. There are many properties of a graph or its elements (vertices, edges). As mentioned before, an important property is the 'degree' of a vertex, usually defined as the number of directly connected vertices to the given vertex. Consequently the distribution of degrees and the maximum degree are of interest. In a graph, the degree distribution informs one about the interconnectedness of the graph. Let $P(k)$ be the fraction of the number of vertices with degree k to the total number of vertices. Then for the classical random graph model the degree distribution has been shown to be Poisson in the limit,

$$P(k) = \frac{e^{-\lambda}\lambda^k}{k!} \tag{1.1}$$

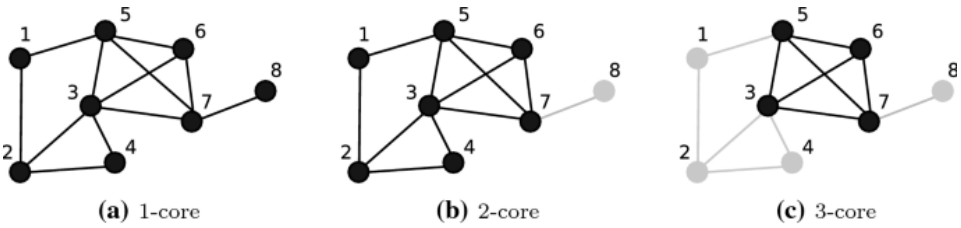
as the size tends to infinity [14].

For scale-free networks, the degree distribution can often be formulated as

$$P(k) \propto k^{-\gamma} \tag{1.2}$$

for sufficiently large values of k [24].

While having a large degree is a local property, it only depends on the number of neighbours a vertex has, connectivity of a graph is often measured by global characteristics. A commonly known



(a) 1-core **(b)** 2-core **(c)** 3-core

Figure 1.6: K-core decomposition of a graph ([54])

property based on the number of vertices is the existence of a 'giant' (i.e. linear sized) component, which is a global property of the graph. The notion of the k-core is also a global property and is used to measure the stability of this giant component in terms of the degrees of the vertices, where $k \geq 2$.

The 'k-core', see Definition 8 for the mathematical formulation, is a subgraph [3] where each vertex has at least k neighbours which have a degree of at least k, after trimming. Trimming is a process of removing vertices with a degree below k and update the degree of the remaining vertices. Hence the 'k-core' of a graph describes the interconnected subgraph of degree k. k-cores are like highways, their presence is important for process behaviour on the graphs; e.g. it facilitates quick information transmission. So there presence or lack of is of interest for determining the stability of the network.

Both the size of the largest component, denoted $\mathcal{C}_{\max}(n)$, and the size of the k-core, $|\text{Core}_k(n)|$ provide information of the interconnectedness of a graph G. The following result is known in literature for the classical Erdős-Rényi random graph model: the largest component, $C_{\max}(n)$, undergoes a phase-transition in the $G(n, \frac{p}{n})$ model. A small change in p_c results in the emergence of a giant component at the critical value p_c. This phase-transition for $C_{\max}(n)$ is second-order. A second-order phase transition is characterised by a continuous transition from one state to the other [16, 36].

For the percolation model, Gilbert showed that the infinite component itself can be very 'loosely' connected, thus not necessarily resulting in k-cores for higher values of $k > 2$.

In contrast to the largest component $C_{\max}$, the k-core size shows a first-order phase-transition in the Erdős-Rényi random graph model. This results in a discontinuity at the critical value, or a jump in a plot of the size compared to the values. The k-core property is desirable for communication networks for example, hence a threshold for when this property occurs is an important result. These communication networks, for example wireless networks like 5G, require stability which a k-core provides. If one (or more) of the transmitters fails, the network remains operational. Thus, the 'k-core' can be used to quantify the resilience of a network.

Using the Configuration model approach for the classical Random Graphs, $G(n, M)$ and $G(n, p)$, Luczak published in 2007 a paper called 'A Simple Solution To k-core Problem' [47], a simplified and elegant solution for a result by Pittel, Spencer and Wormald for the sudden emergence of a giant k-core component in random graph [68].

The goal of this thesis is to understand the behaviour of k-cores in various random graph models, and to contrast the theory with simulations. Moreover we give intuition to the k-core through the analogue of neighbourhoods in random graphs by Branching Processes. The outline of this thesis is as follows. Chapter 2 starts with basic definitions from graph theory. Then, Section 2.1.1, focuses on the classical random graph model and its extensions, both inhomogeneous and spatial. Next, we define the k-core property and study theorems regarding the phase transition of the k-core. In Section 2.4 and 2.5 we investigate if some well-known theorems on k-cores transfer to the geometric random graph model and its inhomogeneous variant respectively, by use of simulations. In Chapter 3, we describe branching processes as to study the existence and size of k-cores as well as giant components. Branching processes are close linked to existence of connected components in random graphs. In chapter 4 we discuss the results of our findings and research questions which remain open. Lastly, Chapter 5 summarises the main results of this thesis.

[3] A graph with a subset of vertices and edges. For a mathematical definition see 7

2

K-cores in Random Graphs

2.1. Random Graphs

There are multiple ways to define a graph. The following definition is suitable for undirected[1], simple[2] graphs.

Definition 1. *Let $G = (V, E)$ be a **graph**. Where $V = \{v_1, v_2, \dots, v_n\}$ is the set of vertices and $E \subseteq \binom{V}{2}$ the set of edges. To specify, let $u, v \in V$ be vertices, than $\{u, v\}$ is an edge (i.e. an unordered pair of two vertices).*

If not explicitly stated otherwise, $\#V = n$ is assumed to be finite. The basic vertex property is required for defining the k-core graph property later on.

Definition 2. *The **degree of a vertex** v, denoted $deg(v)$, is the number of edges which include v.*

$$d_v^{(G)} = deg^{(G)}(v) = \#\{e \in E : v \in e\} \tag{2.1}$$

The shorthand notation, d_v, will be used instead. Any graph G has a degree sequence.

Definition 3. *A **degree sequence** of G, is a sequence of vertex degrees, denoted*

$$\boldsymbol{d}^{(G)} = (d_1, d_2, \dots, d_n) \tag{2.2}$$

The superscript to specify the graph G can be omitted if it is clear which graph is considered. Moreover, when studying the structure of the graph the degree distribution is an important tool.

Definition 4. *Let $G = (V, E)$ be a graph, where $n = |V|$. The number of vertices with degree k is denoted by*

$$n_k = \#\{v \in V : d_v = k\} \tag{2.3}$$

*Now the **degree distribution** of a graph $G = (V, E)$ is a sequence $\boldsymbol{p}_n$ which list the proportion of vertices in the graph of degree k;*

$$\boldsymbol{p}_n^{(G)} = (p_n(1), p_n(2), \dots, p_n(n-1)) \tag{2.4}$$

where $p_n(k) = \frac{n_k}{n}$.

Since the $p_n(k)$'s are proportions the following consequence is obtained $\sum_{k=1}^{n-1} p_n(k) = 1$. The degree distribution is extensively studied for random graphs.

[1] Edges have no direction
[2] At most one edge exists between two vertices and no vertices has an edge to itself (loop)

2.1.1. Erdős-Rényi random graph model

The Erdős-Rényi random graph is defined by $G(n,p)$ in general. For limiting processes and scaling purposes, $G\left(n,\frac{\lambda}{n}\right)$ is studied with a fixed $\lambda > 0$, i.e. λ is not a function of n. Now in $G\left(n,\frac{\lambda}{n}\right)$ the degree of a vertex v, denoted d_v, becomes a random variable. Because the edge probabilities are independent and identically distributed, all are Bernoulli $\left(\frac{\lambda}{n}\right)$, the degree becomes a Binomial distribution. Thus in the Erdős-Rényi random graph model, it follows that

$$\mathbb{P}(d_v = k) = \text{Bin}\left(n-1, \frac{\lambda}{n}\right) \tag{2.5}$$

and the expectation of the degree is

$$\mathbb{E}[d_v] = \frac{n-1}{n} \cdot \lambda \tag{2.6}$$

Since λ is fixed, if $n \to \infty$ the expectation of the degree tends to λ. Furthermore by the Poisson Limit Theorem, under the conditions that, $p = \frac{\lambda}{n} \to 0$ as $n \to \infty$ and that $np = n \cdot \frac{\lambda}{n} \to c$ for some finite real-valued constant $c \in (0,\infty)$ as n tends to infinity, one has convergence in distribution

$$d_v \xrightarrow{d} \text{Poi}(\lambda) \tag{2.7}$$

Since λ was a fixed constant clearly $\frac{\lambda}{n} \to 0$ as $n \to \infty$. The second condition is also satisfied as $n \cdot \frac{\lambda(n)}{n} \to \lambda$, which is indeed a constant in $(0,\infty)$. We provide a proof for the geometric equivalent in Theorem 2.12. The Poisson distribution has a very thin tail, finding vertices with large degrees very unlikely. Thus this model is not 'scale-free' and does not correspond to many real-world networks which adhere to the power-law. Therefore, we present extension of this model where inhomogeneity is added to the degree of the vertices.

2.1.2. Inhomogeneous variations

Inhomogeneous extensions of the Erdős-Rényi random graph are a way to model these real-world phenomena however. The widely studied Chung-Lu model, proposed in [23], extends the classical Erdős-Rényi random graph model to satisfy the power-law under certain conditions. We now define the Chung-Lu model. Consider a given expected degree sequence $\mathbf{w} = (w_1, \dots, w_n)$, where each vertex v_i is assigned vertex weight w_i. Then redefine the edge probability p_{ij}, now inhomogeneous, to be

$$p_{ij} := \frac{w_i w_j}{\sum_{k=1}^{n} w_k} \tag{2.8}$$

for each pair of vertices $\{v_i, v_j\}$. The edge probabilities are still independent, given the weights. The Chung-Lu model allows i and j to be equal, creating the probability for loops at v_i with probability p_{ii}. To ensure $p_{ij} \in [0,1]$, Chung and Lu add the following assumption to the model

$$\max_{i \in \{1,\dots,n\}} w_i^2 < \sum_{k=1}^{n} w_k \tag{2.9}$$

We shall quickly prove why this assumption ensures $p_{ij} \in [0,1]$. Let $\mathbf{w}$ be a finite sequence of non-zero real numbers and let $w_i, w_j \in \mathbf{w}$. Since n is assumed finite, there exists a maximum w' in $\mathbf{w}$. Then by definition of a maximum one has

$$w_i \leq w' \text{ and } w_j \leq w'$$

Combined with Equation 2.9 it follows that

$$\frac{w_i w_j}{\sum_{k=1}^{n} w_k} \leq \frac{w'w'}{\sum_{k=1}^{n} w_k} < \frac{\sum_{k=1}^{n} w_k}{\sum_{k=1}^{n} w_k} = 1$$

Moreover, since $w_i \geq 0,\ \forall w_i \in \mathbf{w}$, also $\sum_{k=1}^{n} w_k \geq 0$. The ratio of two positive real numbers must be a positive real number. Excluding the case of isolated vertices only, i.e. $\mathbf{w} = (0, 0, \dots, 0)$, where $p_{ij} = 0$ by definition, indeed $p_{ij} \geq 0$. Hence $\forall w_i, w_j \in \mathbf{w} : p_{ij} \in [0, 1]$.

Similarly to the Chung-Lu model, Norros and Reittu defined an inhomogeneous model where the vertices are weighted. Norros en Reittu refer to these 'weights' as *capacities*, denoted Λ_i. Then $E(i, j)$ are the number of edges between vertices v_i and v_j, with

$$E(i, j) := \text{Poi}\left(\frac{\Lambda_i \Lambda_j}{\sum_{k=1}^{n} \Lambda_k}\right) \tag{2.10}$$

This results in a multi-graph, not only with loops as in the Chung-Lu model, but also with multiple edges between vertices. This proved useful for finding thresholds for a giant component in the multi-graphs generated by this component, mentioned in [62], similar to those found in the configuration model. Another, more general, inhomogeneous random graph model is proposed by Söderberg in [72], where between each pair of vertex types there is a given edge probability.

Definition 5. *Let $\mathbf{r} = (r_1, \dots, r_m)$ be a m-dimensional probability vector, i.e. $\sum_{j=i}^{m} r_j = 1$, where $m \in \mathbb{N}$. Moreover, consider a symmetric $M \times M$ matrix $\boldsymbol{A}$, with the elements $A_{uv} \in [0, 1] \forall u, v$. Then each vertex is independently assigned a type $i \in \{1, \dots, m\}$ with probability r_i and for each unordered pair of vertices $\{u, v\}$ the corresponding edge probability is A_{r_u, r_v}.*

Note that $m > n = |V|$ is allowed. It is possible though that some vertices are of the same type, which is clearly unavoidable if $m < n$. Hence the way to assign types is of importance. Bollobás, Janson and Riordan popularised this model, also called the 'Stochastic block-model', in their paper on thresholds in inhomogeneous random graphs [17].

2.1.3. Configuration model

As mentioned in the introduction, there are other random graph models. A generalized version of the classical graph model, the Configuration model by Bollobás in [13], must be mentioned, since many results have been proved using this generalized view. Also theorems have been proved in more concise way using the configuration model, like the k-core theorem in [47]. Another example is the Molloy and Reed paper on the size of the giant component of a random graph with a given degree sequence [56]. As opposed to the degree of a vertex to follow from the probability of edges, one can start with given degree sequence $\mathbf{d} = (d_v)_n$, similar to the inhomogeneous variations of the classical model, where n is the fixed number of vertices of the graph. Without loss of generality we assume $d_v \geq 1$. Consider half-edges labelled from 1 to $l_n = \sum_{v=1}^{n} d_v$. Note that l_n to be even is a necessary condition to end up with a *simple* graph. Half-edges are paired to form a single edge of the multigraph. The pairing of the half-edges called a configuration. Every half-edge is only paired once. The resulting graph is called the configuration model with degree sequence $\mathbf{d}$, denoted $\text{CM}_n(\mathbf{d})$. For further insight into the configuration model we refer the reader to Chapter 7 of Hofstad's Random Graph and Complex Networks book [45]. In this thesis we do not use the configuration model. Instead we shall focus on geometric random graphs.

2.2. Random Geometric Graphs

An extension to the classical regular graph is to include a spatial setting. The vertices now represent locations in a given metric space. There are multiple spatial graph models, like the hyperbolic geometric graph (HGG) by Krioukov et al in [50], the scale-free percolation model (SFP) and the random Geometric Graphs (RGG). For the SFP-model, we refer the reader to the publication and books of Hofstad et. al. [26, 44, 46]. In this report we shall focus on the RGG-model, and define it accordingly as many variations have been proposed.

Definition 6. *Let $G = (V, E)$ a graph and let (χ, d) be a metric space. Assume that $(X_i)_{i \in \{1, \dots, n\}}$ represent the vertices chosen i.i.d. from a measure μ on χ. Then the vertex set is $V = (X_1, X_2, \dots, X_n)$. Moreover, the edge set E is defined by*

$$E := \{\{i, j\} : d(X_i, X_j) \leq r\} \tag{2.11}$$

*with $r \in [0, \infty)$. Then G is called a **random geometric graph**, denoted RGG(n, r).*

In this report we shall focus on the d-dimensional hypertorus $\mathbb{T}^d$. More specifically we use a grid, $\chi = [0,1]^d \subseteq \mathbb{R}^d/\mathbb{Z}^d$, with periodic boundary conditions. The X_i are drawn uniformly from χ. In simulations, for convenience, a $[0,1] \times [0,1]$ grid is used which acts like a torus based on the edge conditions. If not mentioned otherwise, l_2 norm is used, denoted $\|\cdot\|_2$. For large calculations, sometimes the infinity norm $\|\cdot\|_\infty$ is used to speed up calculations and will be mentioned accordingly when applied. The edge condition $\|x_u - x_v\| \leq r$ is used by default. Instead of increasing the $[0,1]^d$ space to $[0,n]^d$ with a fixed radius r, we scale back to $[0,1]^d$ and set $r = c \cdot n^{-\frac{1}{d}}$. This ensures the proper scaling, with c a fixed constant similar to the λ in the classical Erdős-Rényi random graph model.

Similarly to the Erdős-Rényi random graph model, this model is homogeneous and can be regarded as a binomial geometrical random graph. In Figure 2.1 the result of generated samples of the empirical degree distribution of random geometrical graphs are compared to the Poisson distribution with the same constant. Even for $n = 250$ the fit is already plausible and remains so for higher values of n.

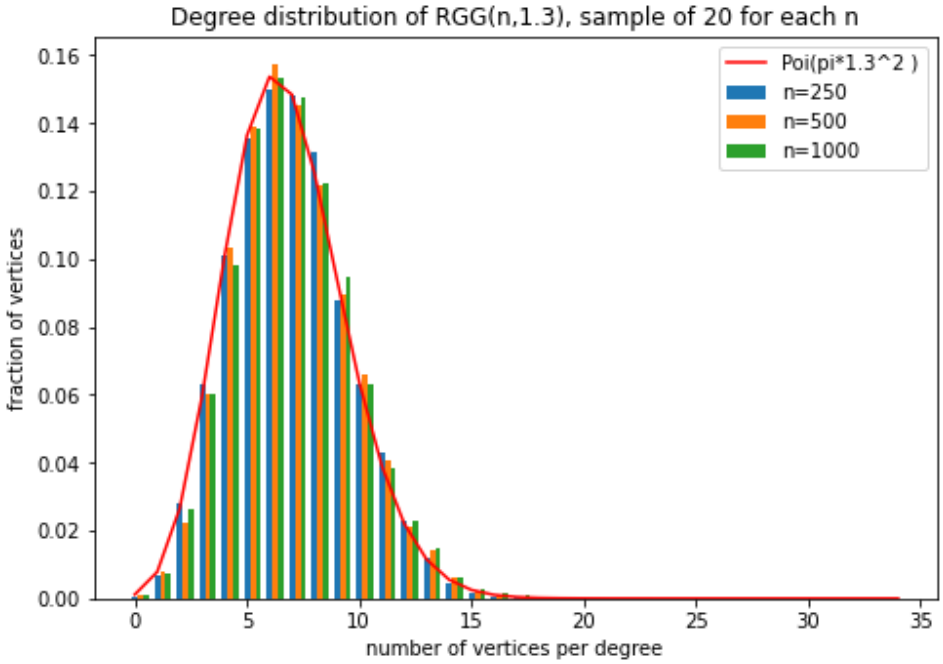


Figure 2.1: The theoretical degree distribution compared to empirical average of 20 samples of the RGG-model with the same parameter $c = 1.3$ for different sizes, $n = 250, 500, 1000$. The fit, proven to hold for $n \to \infty$, is already visible for relative low n.

To confirm this mathematically we shall prove the following theorem. The proof is my own, inspired by a similar proof for the Erdős-Rényi random graph model.

Theorem 1. *Let $G = (V, E)$ be a random geometric graph with $\chi = [0,1]^2$ acting like $\mathbb{T}^2$ and with norm $\|\cdot\|_2$. Let X_i be drawn uniformly from $[0,1]^2$. Furthermore, assume $r = c \cdot n^{-\frac{1}{2}}$ for some fixed $c > 0$. Let X_u be an arbitrary vertex in V, then it holds that*

$$deg^{(G)}(X_u) \xrightarrow{d} Poi(\lambda) \tag{2.12}$$

Proof. Since the points are drawn uniformly and independently we can consider a fixed point $x \in \chi$, the realisation of the location of vertex X_u. Moreover, we consider the circle $A := B(x, r)$ around x with radius r. Let y_i be the realisation of X_i. Then for each point $y_i \in \chi, i \neq u$, around x we obtain the following

$$\mathbb{P}(\|y - x\| \leq r) = \frac{\text{Area}(A)}{\text{Area}(\chi)} = \frac{\pi \cdot r^2}{1} \tag{2.13}$$

because the y_i are i.i.d. The probability if $y_i \in A$ we view as a random variable $Y_i \sim \text{Bernoulli}(\mathbb{P}(y_i \in A))$. Now for multiple events with a Bernoulli distribution to occur we shall use a Binomial random variable.

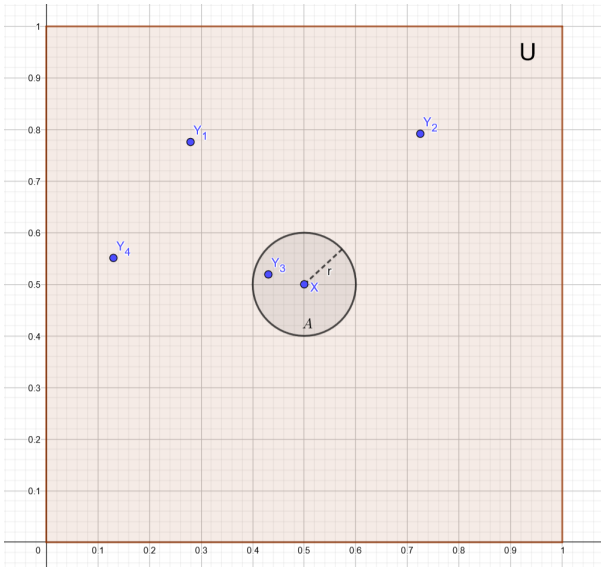


Figure 2.2: Example of points drawn uniformly around centered point X, with a radius r. Points inside the circle count as a neighbour of X.

Let Y be the number of vertices y_i within circle A. An example of this process is shown in Figure 2.2. Thus $\deg^{(G)}(x) \sim \text{Bin}(n-1, \pi \cdot r^2)$. Using $r = c \cdot n^{-\frac{1}{2}}$ we have

$$\begin{aligned}
\mathbb{P}(\deg^{(G)}(x) = i) &= \binom{n-1}{i}(\pi \cdot r^2)^i \cdot (1 - \pi r^2)^{n-1-i} \\
&= \binom{n-1}{i}\left(\pi\left(c \cdot n^{-\frac{1}{2}}\right)^2\right)^i \cdot \left(1 - \pi\left(c \cdot n^{-\frac{1}{2}}\right)^2\right)^{n-1-i} \\
&= \binom{n-1}{i}\left(\frac{\pi \cdot c^2}{n}\right)^i \left(1 - \left(\frac{c^2 \cdot \pi}{n}\right)\right)^{n-1-i}
\end{aligned}$$

Now remark that $n \cdot \frac{\pi \cdot c^2}{n} = \pi \cdot c^2 \to \pi c^2$ as $n \to \infty$. Then applying the limit it follows that

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(\deg^{(G)}(x) = i) &= \lim_{n\to\infty} \binom{n-1}{i}\left(\frac{\pi \cdot c^2}{n}\right)^i \left(1 - \left(\frac{c^2 \cdot \pi}{n}\right)\right)^{n-1-i} \\
&= \lim_{n\to\infty} \frac{(n-1)!}{i!(n-1-i)!} \frac{(\pi \cdot c^2)^i}{n^i} \left(1 - \left(\frac{c^2 \cdot \pi}{n}\right)\right)^{n-1-i} \\
&= \frac{(\pi \cdot c^2)^i}{i!} \lim_{n\to\infty} \frac{(n-1)!}{(n-1-i)! \cdot n^i} \left(1 - \left(\frac{c^2 \cdot \pi}{n}\right)\right)^{n} \left(1 - \left(\frac{c^2 \cdot \pi}{n}\right)\right)^{-(i+1)} \\
&= \frac{(\pi \cdot c^2)^i}{i!} \lim_{n\to\infty} \frac{(n-1) \cdot (n-2) \cdot \ldots \cdot (n-i)}{n^i} \left(1 - \left(\frac{c^2 \cdot \pi}{n}\right)\right)^{n} \left(1 - \left(\frac{c^2 \cdot \pi}{n}\right)\right)^{-(i+1)}
\end{aligned}$$

To compute this limit, first we conclude that each piece is bounded. This allows us to compute the limit of each factor. For large n it holds that $\frac{(n-1)(n-2)\ldots(n-i)}{n^i} \to 1$ since

$$\begin{aligned}
\frac{(n-1)(n-2)\ldots(n-i)}{n^i} &= \frac{n^i}{+} \frac{\sum_{k=1}^{i}(-k) \cdot n^{i-1}}{n^i} + \ldots + \frac{\prod_{k=1}^{i}(-k)}{n^i} \\
&= 1 + \frac{\sum_{k=1}^{i}(-k)}{n} + \ldots + \frac{\prod_{k=1}^{i}(-k)}{n^i}
\end{aligned}$$

by expending the factors. So all terms except the first tend to zero for $n \to \infty$.

In the second factor we recall the limit for the exponential.

$$\lim_{n\to\infty}\left(1-\left(\frac{c^2\cdot\pi}{n}\right)\right)^n = \lim_{n\to\infty}\left(1+\left(\frac{-c^2\cdot\pi}{n}\right)\right)^n = e^{-c^2\cdot\pi}$$

For the last factor notice that i is fixed. So $\lim_{n\to\infty}\frac{c^2\cdot\pi}{n}=0$ and therefore

$$\lim_{n\to\infty}\left(1-\left(\frac{c^2\cdot\pi}{n}\right)\right)^{-(i+1)} = \left(1-\lim_{n\to\infty}\left(\frac{c^2\cdot\pi}{n}\right)\right)^{-(i+1)} = (1-0)^{-(i+1)} = 1$$

In all, the limit simplifies to

$$\lim_{n\to\infty}\mathbb{P}(\deg^{(G)}(x)=i) = \frac{(\pi\cdot c^2)^i}{i!}\cdot 1\cdot e^{-(c^2\cdot\pi)}\cdot 1 \overset{\text{def}}{=} \mathbb{P}(\text{Poi}(\pi c^2)=i)$$

Thus, renaming $\lambda=\pi c^2$ for convenience, we conclude that the degree of any fixed vertex converges to a Poisson distribution. □

This means we just proved the convergence in distribution for the degree of a vertex locally. A much stronger result would be if this would also hold globally. Therefore we now prove the following theorem

Theorem 2. *Let $G=(V,E)$ be a random geometric graph with $\chi=[0,1]^2$ acting like $\mathbb{T}^2$ and with norm $\|\cdot\|_2$. Let X_i be drawn uniformly from $[0,1]^2$. Furthermore, assume $r=c\cdot n^{-\frac{1}{2}}$ for some fixed $0<c<\infty$. Given $p_n^{(G)}(i)=\frac{\sum_{j=1}^n \mathbb{1}_{deg(v_j)=i}}{n}$, then*

$$p_n^{(G)}(i)\to\mathbb{P}(Poi(\lambda)=i)\ \forall i\in\mathbb{N} \tag{2.14}$$

***a.s.** and hence $\boldsymbol{p}_n^{(G)}\xrightarrow{d}Poi(\lambda)$.*

Proof. First, let $d=2$ without loss of generality and let $i\in\mathbb{N}$ be an arbitrary fixed constant. The idea of the proof is to find $p_n(i)$ to be arbitrarily close to $\mathbb{E}[p_n(i)]$ and $\mathbb{E}[p_n(i)]$ to be arbitrarily close to $\lambda_i=\mathbb{P}(\text{Poi}(\lambda)=i)$. The latter we find using the first moment. It follows that

$$\mathbb{E}[p_n^{(G)}(i)] = \mathbb{E}\left[\frac{\sum_{j=1}^n \mathbb{1}_{\deg(v_j)=i}}{n}\right] = \frac{1}{n}\sum_{j=1}^n \mathbb{E}[\mathbb{1}_{\deg(v_j)=i}]$$

Now the expectation of the indicator function is the probability of it occurring, hence

$$\mathbb{E}[p_n^{(G)}(i)] = \frac{1}{n}\sum_{j=1}^n \mathbb{P}(\deg(v_j)=i)$$

Due to the symmetry, we argue that these probabilities equal a priori. Therefore we have

$$\frac{1}{n}\sum_{j=1}^n \mathbb{P}(\deg(v_j)=i) = \frac{1}{n}\sum_{j=1}^n \mathbb{P}(\deg(v_1)=i) = \frac{1}{n}\cdot n\mathbb{P}(\deg(v_1)=i)$$

Then from Theorem 1 it follows that

$$\mathbb{E}[p_n^{(G)}(i)] = \mathbb{P}(\deg(v_1)=i)\overset{n\to\infty}{\to}\mathbb{P}(\text{Poi}(\lambda)=i) \tag{2.15}$$

Let $\epsilon > 0$ be arbitrary. Furthermore, we write $X = p_n(i)$ for convenience. Now we consider the following probability

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq \epsilon\right) \tag{2.16}$$

. Clearly we want this probability to be zero, for $p_n(i)$ to be arbitrarily close to $\mathbb{E}[p_n(i)]$. Making use of Chebyshev's inequality we derive

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq \epsilon\right) \leq \frac{\mathrm{Var}(X)}{\epsilon^2}$$

So we want to prove that $\mathrm{Var}(X) \to 0$ as $n \to \infty$. Since $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$ we have

$$\begin{aligned}\mathrm{Var}\left(\frac{\sum_{j=1}^{n} \mathbb{1}_{\deg(v_j)=i}}{n}\right) &= \frac{1}{n^2}\mathrm{Cov}\left(\sum_{j=1}^{n} \mathbb{1}_{\deg(v_j)=i}, \sum_{j=1}^{n} \mathbb{1}_{\deg(v_j)=i}\right) \\ &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n} \mathrm{Cov}\left(\mathbb{1}_{\deg(v_j)=i}, \mathbb{1}_{\deg(v_k)=i}\right)\end{aligned}$$

using the distributive property of the covariance. We split up the double sum in a diagonal sum and use the property that $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ to sum triangular twice. Hence

$$\begin{aligned}&= \frac{1}{n^2}\left(\sum_{j=1}^{n}\left[\mathrm{Cov}(\mathbb{1}_{\deg(v_j)=i}, \mathbb{1}_{\deg(v_j)=i})\right] + 2\cdot\sum_{j=1}^{n}\sum_{k=j+1}^{n} \mathrm{Cov}\left(\mathbb{1}_{\deg(v_j)=i}, \mathbb{1}_{\deg(v_k)=i}\right)\right) \\ &= \frac{1}{n^2}\sum_{j=1}^{n}\left[\mathrm{Cov}(\mathbb{1}_{\deg(v_j)=i}, \mathbb{1}_{\deg(v_j)=i})\right] + \frac{2}{n^2}\cdot\sum_{j=1}^{n}\sum_{k=j+1}^{n} \mathrm{Cov}\left(\mathbb{1}_{\deg(v_j)=i}, \mathbb{1}_{\deg(v_k)=i}\right)\end{aligned}$$

We recall that $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$ and then we simplify the first term to find

$$\frac{1}{n^2}\sum_{j=1}^{n}\left[\mathrm{Cov}(\mathbb{1}_{\deg(v_j)=i}, \mathbb{1}_{\deg(v_j)=i})\right] = \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{Var}\left(\mathbb{1}_{\deg(v_j)=i}\right) \tag{2.17}$$

Now we use the fact that the $\mathbb{1}_{\dots} \sim \mathrm{Bernoulli}(p)$ for some p and $\mathrm{Var}(\mathrm{Bernoulli}(p)) = p(1-p) \leq 1$. Therefore

$$\begin{aligned}\frac{1}{n^2}\sum_{j=1}^{n}\mathrm{Var}\left(\mathbb{1}_{\deg(v_j)=i}\right) &\leq \frac{1}{n^2}\sum_{j=1}^{n} 1 \\ &= \frac{n}{n^2} = \frac{1}{n} \to 0 \quad (n \to \infty)\end{aligned}$$

So we have proven the first term indeed tends to zero. The second term requires us to write out the Cov terms using the definition $\mathrm{Cov}(X, Y) \overset{\text{def}}{=} \mathbb{E}[X \cdot Y] - \mathbb{E}[X]\mathbb{E}[Y]$. Thus the second term becomes

$$\frac{2}{n^2}\cdot\sum_{j=1}^{n}\sum_{k=j+1}^{n}\left(\mathbb{P}(\deg(v_j) = i, \deg(v_k) = i) - \mathbb{P}(\deg(v_j) = i)\mathbb{P}(\deg(v_k) = i)\right)$$

Next we split the the probability using the law of total probability for the cases that v_j and v_k have overlap or not. We note

$$\begin{aligned}\mathbb{P}(\deg(v_j) = i, \deg(v_k) = i) = \mathbb{P}(\deg(v_j) = i, \deg(v_k) = i \mid v_j \leftrightarrow v_k)\mathbb{P}(v_j \leftrightarrow v_k) \\ +\mathbb{P}(\deg(v_j) = i, \deg(v_k) = i \mid v_j \nleftrightarrow v_k)\mathbb{P}(v_j \nleftrightarrow v_k)\end{aligned}$$

In the case that they do not have overlap we can actually view them as independent events and find

$$\mathbb{P}(\deg(v_j)=i,\deg(v_k)=i)=\mathbb{P}(\deg(v_j)=i,\deg(v_k)=i\mid v_j\leftrightarrow v_k)O\left(\frac{1}{n}\right)$$
$$+\mathbb{P}(\deg(v_j)=i,\deg(v_k)=i\mid v_j\nleftrightarrow v_k)O\left(1-\frac{1}{n}\right)$$

as $n\to\infty$. Moreover, the double sum we view as a lower triangle, which we can also write as $\frac{(n-1)(n-1)}{2}$. Using the symmetry on the v_i we find

$$\frac{2(n-1)^2}{2n^2}\left(\mathbb{P}(\deg(v_j)=i,\deg(v_k)=i\mid v_j\nleftrightarrow v_k)-\mathbb{P}(\deg(v_j)=i)\mathbb{P}(\deg(v_k)=i)\right)=0$$

The first term can be described with a Multinomial distribution with $n-2$ trials, $(i,i,n-2-2i)$ successes, and probability vector $\mathbf{p}=\left(\frac{\pi c^2}{n},\frac{\pi c^2}{n},1-\frac{2\pi c^2}{n}\right)$, the second term is a Binomial $\left(n-1,\frac{\pi c^2}{n}\right)$ distribution squared. Moreover, the double sum we view as a lower triangle, which we can also write as $\frac{(n-1)(n-1)}{2}$. Using the symmetry on the v_j,v_k we find

$$\mathbb{P}\left(\text{Multinomial}(n-2,\frac{2\pi c^2}{n},\frac{2\pi c^2}{n},1-\frac{2\pi c^2}{n})=(i,i,n-2-2i)\right)\to\mathbb{P}(\text{Poi}(\lambda)=k)^2 \tag{2.18}$$

but also

$$\mathbb{P}\left(\text{Bin}(n-1,\frac{2\pi c^2}{n})=i\right)^2\to\mathbb{P}(\text{Poi}(\lambda)=k)^2 \tag{2.19}$$

for $n\to\infty$. Thus for large enough values of n the terms approximately cancel out and the variance tends to zero. Therefore, now that we have shown the variance to tends to zero as n tends to infinity we conclude that

$$\mathbb{P}\left(|X-\mathbb{E}[X]|\geq\epsilon\right)\overset{n\to\infty}{\to}0 \tag{2.20}$$

and therefore $|X-\mathbb{E}[X]|<\epsilon$ as $n\to\infty$. Then finally we can write

$$\begin{aligned}|p_n^{(G)}(i)-\mathbb{P}(\text{Poi}(\lambda)=i)|&=|p_n^{(G)}(i)-\mathbb{E}[p_n^{(G)}(i)]+\mathbb{E}[p_n^{(G)}(i)]-\mathbb{P}(\text{Poi}(\lambda)=i)|\\&\leq|p_n^{(G)}(i)-\mathbb{E}[p_n^{(G)}(i)]|+|\mathbb{E}[p_n^{(G)}(i)]-\mathbb{P}(\text{Poi}(\lambda)=i)|\\&=\epsilon+\epsilon=2\epsilon\end{aligned}$$

Since $\epsilon>0$ was chosen arbitrary it can be arbitrarily small, thus proving that indeed $p_n^{(G)}(i)\overset{n\to\infty}{\to}\mathbb{P}(\text{Poi}(\lambda)=i)$ for all finite $i\in\mathbb{N}$. This implies that $\mathbf{p}_n^{(G)}(i)\overset{d}{\to}\text{Poi}(\lambda)$, concluding the proof. □

The proof is work of my own, heavily inspired by a sketch of my supervisor Júlia Komjáthy.
We note that the important link between the $RGG(n,r)$ model and the $G(n,p)$ to be in the form of

$$\boxed{\lambda=\pi c^2} \tag{2.21}$$

for dimension $d=2$. As mentioned in Section 2.1.2 another extension to the classical random graph is inhomogeneity. We shall now procede to define a inhomogeneous extension to the random geometric graph.

2.2.1. Geometric Inhomogeneous Random Graphs

A random graph model with both the location extension as well as the inhomogeneous edge condition variation can be defined, called **GIRG** (Geometric Inhomogeneous Random Graph). A natural way to extend the RGG-model is to vary the radius for each pair of vertices

$$\|x-y\|\leq f(r_x,r_y) \tag{2.22}$$

where $f : [0,\infty)^2 \to [0,\infty)$ is a function 'join' the radii, for example $f(r_x, r_y) = \max(r_x, r_y)$. Different functions of f can result in drastic different behaviour. However, in order to best compare a RGG-model and its GIRG extension we prefer to use a weight based approach. Henceforth, a joint measure $\mu \otimes W^{\mathbb{P}}$ is used, where W is the distribution of vertex weights. One can think of the weight of a vertex as the the 'fitness' of the vertex, i.e. its ability to form many edges. Moreover, the joint measure is independent. We define $\tilde{V} = (V, W_V)$ as the realisation of the stochastic variable. One can also view W as adding another dimension and this results in double randomness; first a realisation in the vertex location and then a realisation in their weights. For example let u be a random vertex in the GIRG-model, then x_u is the realisation of X_u and w_u of W_u, where

$$(x_u, w_u) \in \tilde{V} \tag{2.23}$$

Specifically, $\chi \subseteq \mathbb{R}^d$ is assumed with $d \geq 1$. Let $(X_i, W_i)_{i\in\{1,\dots,n\}}$ be the vertices as a pair of its location and weight, where X_i is a point in χ. The X_i's are chosen i.i.d. from some measure μ on χ. Moreover each vertex i has a weight $W_i \in [0,\infty)$ chosen i.i.d. as well, but with possibly another distribution. The realisations of X_i and W_i are denoted x_i and w_i respectively. The edges are conditionally independent with given the x_i and w_i. One example of an connection probability used in a GIRG-model is

$$\mathbb{P}(\{u,v\} \in E \big| (x_u, w_u), (w_v, w_v) \in \tilde{V}) = \begin{cases} \min\left(1, \frac{w_u \cdot w_v \cdot r}{\|x_u - x_v\|_2}\right) & \text{if} X_u \neq X_v \\ 0 & \text{otherwise} \end{cases} \tag{2.24}$$

where the edge probability both depends on the location as well as the vertex weight. If the edge condition contains the corrector term $\sum_{i=0}^{n} w_i$ in the denominator it is a Chung-Lu GIRG adaptation. Using this sum is qualitatively similar since $\sum W_i = n \cdot \mathbb{E}[W] \pm \sqrt{n}$ by the Law of Large numbers.

In this chapter many different random graph models have been mentioned and defined. Now we shall focus on the properties in these random graph models, where most of our efforts are towards the k-core property.

2.3. K-core property

Before we continue with k-core graph property, we first shall define a subgraph. Like most structures in mathematics, substructures can be defined. Most fundamentally a set and its subset. For a graph a subgraph can be defined.

Definition 7. *Let $G = (V, E)$ be a graph. Then $H = (V', E')$ is a subgraph of G if $V \subseteq V'$ and $E' \subseteq E$, such that every $e \in E'$ must consist of vertices from V'.*[3]

With the induced subgraph in mind, the definition of the k-core can be written in a compact form.

Definition 8. *Let $G = (V, E)$ be a graph. The **k-core** of G, denoted Core_k, is the maximal induced subgraph H of G where $\forall v \in \text{Core}_k^{(G)} : \deg^{(H)}(v) \geq k$. Moreover, the k-**shell** of a graph are those vertices which are in the k-core vertex set but not in the $(k+1)$-core vertex set.*

Maximal means that there is no other subgraph which also meets the degree condition, whilst having a bigger vertex set. Important to note is that the k-core of a graph does not need to be one component, but can consists of multiple components (clusters), as can be seen in the example in Figure 2.3. In the figure, three clusters of red points, all part of the 4-core, can be seen. Alternative definitions exist which require the k-core to be a *connected* subgraph. This means that in case of Figure 2.3 the largest k-core component is refered to as 'the k-core' according to this alternative definition. We shall only refer to the k-core as given by Definition 8 in this thesis.

The size as well as the existence of a k-core are studied extensively for some random graph models [49]. Upper and lower bounds in terms of elements of G are important to consider both from a deterministic point of view as well as in the random graph setting. We provide a sketch of a simple proof for lower bound of edges for the existence of a k-core. Furthermore, we shall write out a full proof for an upper bound for the existence of the k-core, also in terms of the number of edges in G. For both bounds we use the definition of a *clique* in a graph, so we shall state this first.

[3] To clarify $E' \subseteq \binom{V'}{2}$, since in this paper we avoid edges in the subgraph edge set where the endpoints are not in the subgraph vertex set. For more information these alternative concepts search for 'half edges' and 'loose edges'

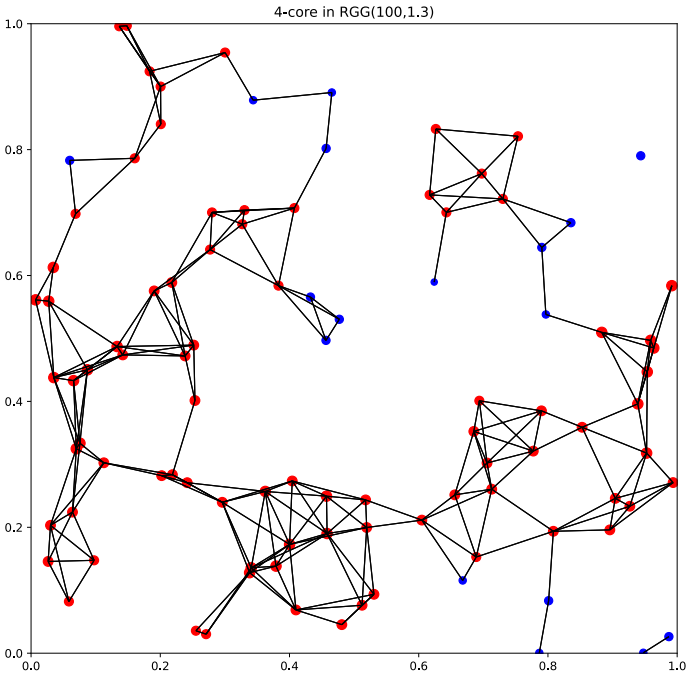


Figure 2.3: The k-core in red consisting of two (connected) components. Visual of generated random geometric graph (RGG), with $n = 100$. Points were drawn uniformly on the grid. Edge condition used is the infinity norm with radius $r = c \cdot n^{-\frac{1}{2}}$, since the dimension $d = 2$.

Definition 9. *A clique of size m of a graph $G = (V, E)$ is a set of vertices $C_m \subseteq V$ such that*

$$C_m := \{u_1, u_2, \dots, u_m \in V : \{u_i, u_j\} \in E \forall i, j \in \{1, \dots, m\}\} \tag{2.25}$$

meaning that every vertex in the clique is connected to any other vertex in the clique.

Now we continue with the lower and upper bounds for the k-core in terms of edges in a graph G. The first bound is the following: "What is the minimum number of edges in a graph with n vertices to have a k-core?". To deduce this answer, we first note the graph must contain a *clique* of size k and $n - k$ isolated vertices. Therefore the minimum number of edges is precisely those edges needed for the clique; $\frac{k(k-1)}{2}$.

The second edge case result for k-cores in graphs is the following.

Theorem 3. *Let $G = (V, E)$ be a simple graph with n vertices. Then the maximum number of edges $e(n, k)$ in G with a empty k-core is given by*

$$e(n, k) = n(k - 1) - \frac{k(k - 1)}{2} \tag{2.26}$$

where $n \geq k \geq 2$.

Now we proof Theorem 3.

Proof. First we remark that a clique of size $k + 1$ results in a k-core. We shall use this argument repeatedly throughout this proof. Hence, the biggest clique in G can be at most of size k, which is C_k. Any u in the clique is connected to $k - 1$ other vertices in the clique, by definition. Now every vertex not in C_k can be connected to at most $k - 1$ vertices from the clique C_k without creating a clique of size $k + 1$ which would create a k-core. Now, assume every vertex $v \in V \setminus C_k$ is connected to $k - 1$ vertices from C_k only. Now we claim that we already have found the maximum number of edges without a k-core.

Claim 1. *A graph $G = (V, E)$ with a clique C_k and where every vertex not in the clique is only connected to $k-1$ vertices from the clique is a maximal graph, in terms of edges, with no k-core. Maximal in the sense that adding any edge will result in a k-core.*

If the claim is correct any additional edge would create a k-core. To prove this claim we use case analysis. Suppose we now add an additional edge $\{u, v\}$ between two arbitrary vertices in G, then there are two cases to be distinguished.

1. An edge between two vertices both **not** in C_k
2. An edge between a vertex in C_k and a vertex **not** in C_k.

In the first case the edge is between two arbitrary vertices $u, v \in V \setminus C_k$. By assumption both u and v are connected to $k-1$ arbitrary vertices from C_k. This implies they have either $k-2$ or $k-1$ shared adjacent vertices in C_k. First assume u and v share precisely $k-2$ edges in C_k, say $w_1, w_2, \dots, w_{k-2} \in C_k$. Then it holds that all vertices in the set $\Gamma_1 = \{u, v\} \cup C_k$ have at least k edges. Thus the vertices in Γ_1 form a k-core. Alternatively we consider u and v share $k-1$ adjacent vertices in C_k, say $s_1, s_2, \dots, s_{k-1}$. Now we note that the set $\Gamma_2 = \{u, v, s_1, s_2, \dots, s_{k-1}\}$ has at least k edges for each element (vertex). So Γ_2 forms a k-core since all vertices in Γ_2 is a connected component and each vertex in Γ_2 has at least k edges.

In the second case, given $C_k = \{w_1, w_2, \dots w_k\}$, we have an arbitrary vertex $u \notin C_k$ such that $\{u, w_i\}$ is the additional edge added to G, for some i. Since u was already connected to $k-1$ vertices from C_k this newly added edge must be formed with the only vertex not yet connected to u, since G is assumed a simple graph. Now u is connected to any vertex in C_k, meaning $C_{k+1} = \{u\} \cup C_k$. Since C_{k+1} is a k-core the second case of the claim is now proven.

Both cases lead to the existence of a k-core in G, hence the claim is true. Now from the construction it follows that the number of edges for the k vertices in the clique is

$$e(C_K) = \frac{k(k-1)}{2} \tag{2.27}$$

Moreover, the number of edges for the $(n-k)$ vertices not in C_k is $(n-k)(k-1)$. Therefore we find the maximum number of edges to be

$$\begin{aligned} e(n,k) &= (n-k)(k-1) + \frac{k(k-1)}{2} \\ &= n(k-1) - k(k-1) + \frac{k(k-1)}{2} \\ &= n(k-1) - \frac{k(k-1)}{2} \end{aligned}$$

which is the required equation from the theorem, thus finishing the proof. □

The proof is a personal work. Another, more common, way of proving this theorem requires induction. Theorem 3 provides an upper bound on the emergence of a k-core in terms of the number of edges. However, this bound is not very useful for random graphs as probability for the appearance of the k-core can be shown to be almost surely 1 for large values of n. Therefore we would like to formulate the upper bound in terms of parameter(s) of a random graph model, for example, in terms of λ for $G\left(n, \frac{\lambda}{n}\right)$. We refer to the first appearance of the k-core in a random graph as the *threshold*. In the next section we explore this k-core threshold.

2.3.1. K-core thresholds

A brief remark on the notation regarding asymptotic behaviour. An event which occurs with a probability tending to 1 as $n \to \infty$ is said occur 'almost surely' (**a.s.**) or in other words to hold 'with high probability' (**whp**).

The k-core property in the classical random graph $G(n,p)$ has been well studied. Back in 1987, Luczak's paper on the size of the k-core in $G(n,p)$ establishes the discontinuous phase transition of the k-core size. He defines $v(k;n,p)$, we refer to as $v(\text{Core}_k^{(G)})$ or $|\text{Core}_k^{(G)}|$, to be the random variable representing the number of vertices in the k-core of the $G(n,p)$ graph [51]. Important to remark is that p is dependent on n, such that $p = p(n) = \frac{c}{n}$, where c is the average degree.

Theorem 4. *Let G be a realisation of the classical random graph model $G(n,p(n))$. If $k \geq 3$, then **a.s.** either $v(Core_k^{(G)}) = 0$ or $v(Core_k^{(G)}) \geq 0.0002n$.*

This phenomena was confirmed by Pittel et al: "When a k-core appears for the first time it is very likely to be giant, ...". The model $G\left(n, \frac{\lambda}{n}\right)$ was used by Pittel, Spencer and Wormald to derive a similar result in [68]. Different than in [51], a critical value is provided for every $k \geq 3$ using a clever construction. Given $\mu > 0$ and $j \in \mathbb{N}$ and let $\psi_j := \mathbb{P}(\text{Poi}(\mu) \geq j)$. Further define

$$\lambda_k = \min_{\mu>0} \frac{\mu}{\psi_{k-1}(\mu)} \tag{2.28}$$

as the threshold value for the appearance of the k-core. In Figure 2.4, the function $\frac{\mu}{\psi_{k-1}(\mu)}$ is plotted for various μ. The threshold value λ_k is visible as the minimum λ for $k = 3$, $k = 4$ and $k = 5$. For $k = 3$ at $\mu \approx 1.81$ we find $\lambda_k \approx 3.35$. For $\lambda > \lambda_k$ define $\mu_k(\lambda)$ to be the largest solution to the equation

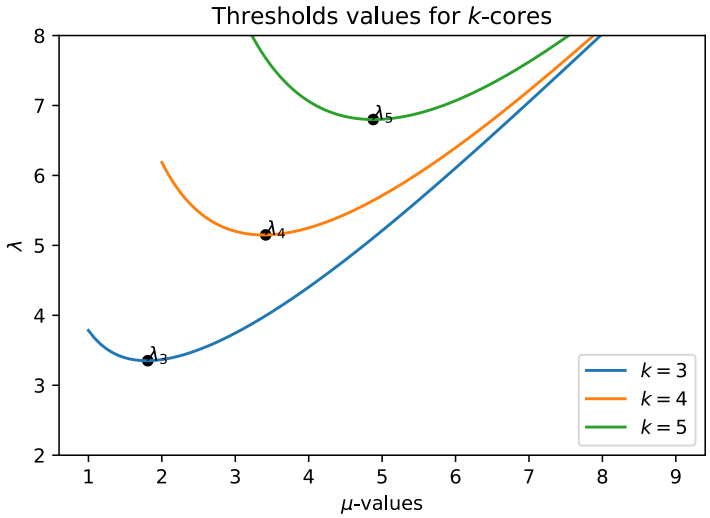


Figure 2.4: The critical points for the k-cores in $G\left(n, \frac{\lambda}{n}\right)$, with $k = 3, 4, 5$, are marked by the the threshold value λ_k. Values of μ on the left of this threshold value fall into the subcritical regime and those the right into the supercritical regime. For λ_3 the critical value is around $\mu \approx 3.55$

$\frac{\mu}{\psi_{k-1}(\mu)} = \lambda$. We shall now state the theorem by Pittel et al. as mentioned in [47]

Theorem 5. *Consider the random graph $G\left(n, \frac{\lambda}{n}\right)$, where $\lambda > 0$ is fixed. Let $k \geq 2$ be fixed and let $Core_k = Core_k(n,\lambda)$ be the k-core of $G\left(n, \frac{\lambda}{n}\right)$. Then*

*(i) If $\lambda < \lambda_k$ and $k \geq 3$, then $Core_k$ is empty **whp**.*

*(ii) If $\lambda > \lambda_k$, then **whp** $Core_k$ is non-empty, and $\frac{v(Core_k)}{n} \xrightarrow{p} \psi_k(\mu_k(\lambda)), e(Core_k) \xrightarrow{p} \mu_k(\lambda)\psi_{k-1}(\mu_k(\lambda))/2 = \mu_k(\lambda)^2/(2\lambda)$.*

The same results hold for the random graph $G(n,m)$, for any sequence $m = m(n)$ with $\frac{2m}{n} \to \lambda$.

We note that λ_k is the critical threshold value for the emergence of the k-core in $G\left(n, \frac{\lambda}{n}\right)$. The relative size of the k-core is now explicitly expressed.

In [47] Janson and Luczak study the k-core of a random (multi)graph, using the configuration model, for a degree sequence **d**. A breakthrough in this field of research is the fact that Janson and Luczak recover the result from Pittel et al by an alternative proof, stating "unlike, ..., we do not use differential equations, but rely solely on the convergence of empirical distributions of independent random variables."

2.4. K-cores in Random Geometric Graphs

Given the λ_k threshold for the homogeneous $G(n,p)$, for example $\lambda_3 \approx 3.35$, we can derive the radius amplification factor for our RGG, which is also homogeneous. By Equation 2.21 it follows that

$$\lambda = \pi c^2 \tag{2.29}$$

$$c = \sqrt{\frac{\lambda}{\pi}} \tag{2.30}$$

Denote r_3 the radius amplification factor r_f for the critical value λ_3. Then $r_3 \approx 1.03$. It would be desirable to transfer the theorem of Pittel et al to the realm of the RGG.

For multiple values of r_f the proportion of the vertices in the k-core are calculated by simulation. For each r_f a sample of 25 simulations ran. These samples were based on $n \in \{250, 750, 2500\}$. The largest simulation ran for $n = 50000$ produces similar result but took to long to run multiple times, also taking up close to 32Gb of RAM memory.

Overall the dispersion of the samples decreased with the increase of n. This is contrary to what is expected. The variance should increase with n if there is a first-order phase transition around at critical value for r_f More closely around r_3, simulations also including $n = 7500$ are performed. Hence the

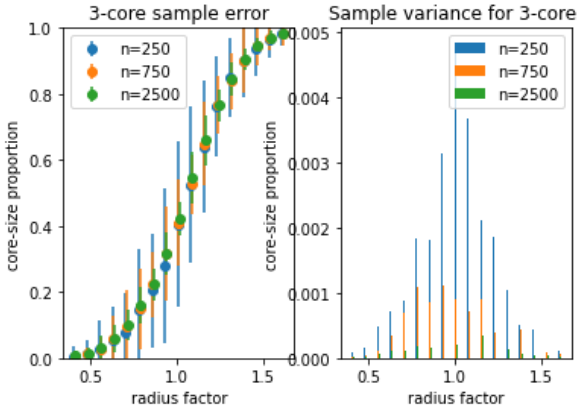


Figure 2.5: Left: Proportion of vertices in the 3-core plotted against the radius factor. The dispersion of the sample of 25 simulations annotated with the vertical bar and the dot represents the sample mean. Right: The sample variance for each sample of 25 simulations.

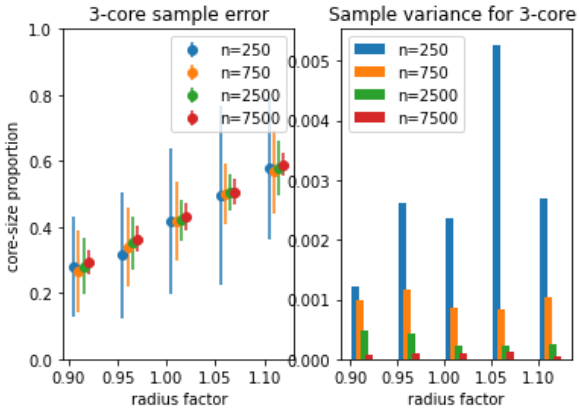
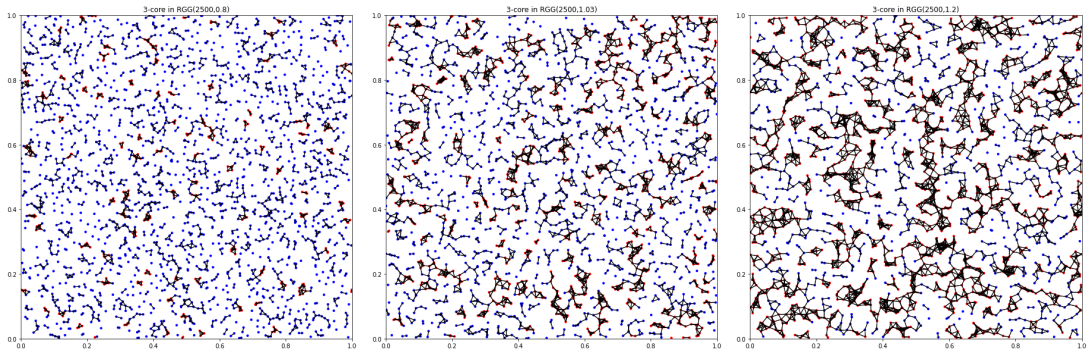


Figure 2.6: Left: Proportion of vertices in the 3-core plotted against the radius factor. The dispersion of the sample of 25 simulations annotated with the vertical bar and the dot represents the sample mean. Right: The sample variance for each sample of 25 simulations.

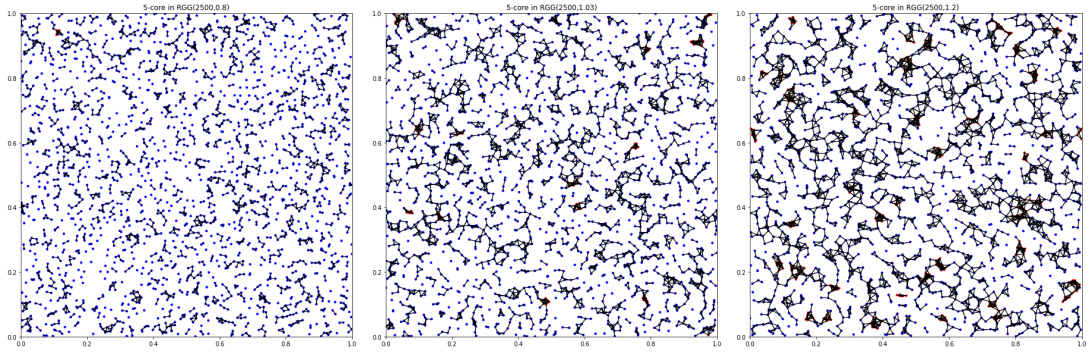
simulations suggest there is no first-order phase transition, as is the case in $G\left(n, \frac{\lambda}{n}\right)$. Simulations of



(a) Realisation of RGG(2500,0.8), where $|\text{Core}_3|/n \approx 0.1648$

(b) Realisation of RGG(2500,1.03), where $|\text{Core}_3|/n \approx 0.4428$

(c) Realisation of RGG(2500,1.2), where $|\text{Core}_3|/n \approx 0.69$



(d) Realisation of RGG(2500,0.8), where $|\text{Core}_5|/n \approx 0.0024$

(e) Realisation of RGG(2500,1.03), where $|\text{Core}_5|/n \approx 0.0376$

(f) Realisation of RGG(2500,1.2), where $|\text{Core}_5|/n \approx 0.1392$

Figure 2.7: Examples of the growth behaviour of the k-core for realisations of the RGG-model. The k-core vertices are marked in red and the other vertices in blue. Shown for $k = 3$ in the top row and $k = 5$ in the bottom row.

the RGG-model show the k-core grows in 'patches', i.e. clusters, for both $k = 3$ as $k = 5$; see Figure 2.7. Meaning, the k-core appears to exist even in proportions below the $0.27n$ threshold.

Due to the geometric embedding, the k-cores appear to behave similar in fashion to the C_{max}, but a first order phase-transition we could not manage to find in simulations, except for the trivial $r_f = 0$. We conclude this from several types of simulations. Figure 2.10, shows simulation samples with error margins, along with the variance of the samples. These simulations show less variation for larger values of n, and support the idea of only a trivial jump at $r_f = 0$ for large values of n. This suggests that the randomness within the geometric embedding ensures clustering due to the symmetry of the homogeneous conditions, hence small cliques appear. Those small cliques add to the total proportion of the k-core. To illustrate this behaviour, we think of a $\mathbb{R}^2$, an infinite grid, with vertices on all of $\mathbb{Z}^2$. Consider points x and y to be connected if $\|x-y\|_\infty < 1$. Now we let the points deviate from their integer valued coordinates by some $\epsilon > 0$ at random. For example $x = (2, 3)$ becomes $x' = (2.1, 2.89)$. We started without any connections, but clearly now many connections will occur due to the randomness of the deviation. Surely triangles or cliques of size 4 will be formed. With this illustration in mind, we now give a possible explanation for the lack of the first-order phase-transition. Consider the following scenario. Let u, v, w vertices, then for $G(n,p)$ we have

$$\mathbb{P}\left(\{u,w\} \in G(n,p) \| \{u,v\} \in G(n,p) \wedge \{v,w\} \in G(n,p)\right) = \mathbb{P}(\{u,w\} \in G(n,p)) = \frac{\lambda}{n} \tag{2.31}$$

since edge probabilities are independent. In the RGG model however, this is not the case.

Lemma 6. *Let $G = (V,E)$ be an instance of the RGG-model such that $r = \lambda n^{-\frac{1}{d}}$. Then, in general, $\mathbb{P}\left(\{u,w\} \in E \middle| \{u,v\} \in E \wedge \{v,w\} \in E\right) \neq \frac{\lambda}{n}$.*

Proof. Let $\text{RGG}(n,r) = (V,E)$ be a random geometric graph on $\mathbb{T}^2$ with the edge probability $\mathbb{P}(\{v_1,v_2\} \in E) = \|v_1 - v_2\|_2 \leq r, \quad \forall v_1, v_2 \in V$. Now suppose $\{u,v\},\{v,w\} \in E$. Let u have location (u_x, u_y) and without loss of generality assert $u_x = u_y = 0$. Now since $\{u,v\} \in E$ clearly $\|u - v\|_2 \leq r$.

In the extreme case $\|u - v\|_2 = r$. This means that v lies on the edge of the circle of connection of u and vice versa. Denote the area of connection of u with A and the area of connection for v with B. Now the circle of connection of v can be split into two parts

1. The part outside; $D := B \setminus A$

2. The overlap; $C := B \cap A$

as shown in Figure 2.8.

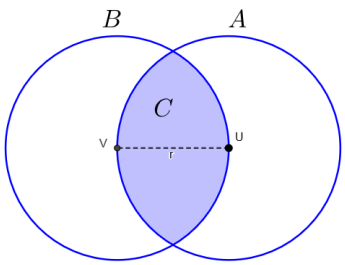


Figure 2.8: Intersection area of circles of connection between two vertices which are connected at the end.

Now by assumption $\{v, w\} \in E$, so either $w \in (B \setminus A)$ or $w \in B \cap A$. Meaning that, since w is connected to v,

$$\mathbb{P}(\{u, w\} \in E \,|\, (\{u, v\} \in E \wedge \{v, w\} \in E)) \geq \frac{\text{Area}(C)}{\text{Area}(B)} \tag{2.32}$$

where $\text{Area}(B) = \pi r^2$. Furthermore, C is has the shape of a symmetric lens. Since the total distance between u and v is $\|u - v\| = r$, the halfway point is $\frac{r}{2}$. Now translate and rotate everything such that $u = \left(0, \frac{-r}{2}\right)$ and $v = \left(0, \frac{r}{2}\right)$. Then the distance still remains r. Now the upper half of the lens is the part above the horizontal axis. Using basic algebraic geometry this upper lens area can be described as

$$C_{\text{upper}} = \int_{-\frac{\sqrt{3}r}{2}}^{\frac{\sqrt{3}r}{2}} \sqrt{r^2 - x^2} - \frac{r}{2} dx \tag{2.33}$$

To solve this substitute for $x = r\sin(s)$ on the interval $[s_1, s_2]$. Use $dx = r\cos(s)ds$ to write

$$\begin{aligned} C_{\text{upper}} &= \int_{s_1}^{s_2} r\sqrt{1 - \sin(s)^2} - \frac{r}{2} \cdot r\cos(s) ds \\ &= \int_{s_1}^{s_2} r\cos(s)^2 - \frac{r^2\cos(s)}{2} ds \end{aligned}$$

Using the double angle trigonometry identity $\cos(s)^2 = \frac{1}{2} + \frac{1}{2}\cos(2s)$ we rewrite to

$$\begin{aligned}
C_{\text{upper}} &= \int_{s_1}^{s_2} \frac{r}{2} + \frac{r}{2}\cos(2s) - \frac{r^2\cos(s)}{2} ds \\
&= \left[\frac{r}{2}s + \frac{r}{4}\sin(2s) - \frac{r^2\sin(s)}{2}\right]_{s_1}^{s_2} \\
&= \left[\frac{r}{2}\arcsin\left(\frac{x}{r}\right) + \frac{r}{4}\cdot 2\frac{x}{r}\cdot\cos(s) - \frac{rx}{2}\right]_{x=-\frac{\sqrt{3}r}{2}}^{x=\frac{\sqrt{3}r}{2}} \\
&= \left[\frac{r}{2}\arcsin\left(\frac{x}{r}\right) + \frac{r}{4}\cdot 2\frac{x}{r}\cdot\sqrt{1-\sin(s)^2} - \frac{rx}{2}\right]_{x=-\frac{\sqrt{3}r}{2}}^{x=\frac{\sqrt{3}r}{2}} \\
&= \left[\frac{r}{2}\arcsin\left(\frac{x}{r}\right) + \frac{x}{2}\cdot\sqrt{1-\frac{x^2}{r^2}} - \frac{rx}{2}\right]_{x=-\frac{\sqrt{3}r}{2}}^{x=\frac{\sqrt{3}r}{2}} \\
&= \frac{1}{2}r^2\left(\frac{2\pi}{3} - \sqrt{3}2\right)
\end{aligned}$$

Here we used back substitution and trigonometry identities. Thus

$$\text{Area}(C) = r^2\left(\frac{2\pi}{3} - \sqrt{3}2\right) \tag{2.34}$$

because of the symmetry of the lens shaped region C. The same result could also have been obtained using ordinary geometry for circle intersections [80]. Using the result (Equation 2.34) and the fact that $\text{Area}(B) = \pi r^2$ in Equation 2.32 we obtain

$$\mathbb{P}(\{u,w\} \in E \,|\, (\{u,v\} \in E \wedge \{v,w\} \in E)) \geq \frac{r^2\left(\frac{2\pi}{3} - \sqrt{3}2\right)}{\pi r^2} \approx 0.39 \tag{2.35}$$

Clearly, in general, $\frac{\lambda}{n} \not\geq 0.39$. Moreover, it does not even depend on n. Intuitively this means the presence of triangles, or clusters in general, are guaranteed for large values of n. To really bring home this point we provide a comparison between the homogeneous Erdős-Rényi random graph model and the RGG counterpart. Figure 2.9 shows, on the left, the first-order phase transition for the 3-core in the Erdős-Rényi random graph model at the critical $\lambda_3 \approx 3.35$. On the right the 3-core proportion, $\frac{|\text{Core}_k|}{n}$, is computed by simulations for the same values of λ. To overcome this locality preference in the RGG model, we shall explore an inhomogeneous extension of this model in Section 2.5. But before we do, we first analysis of the k-core growth for the RGG-model further. □

For $C_{\max}$, there is a second-order phase transition in the $G(n,p)$ model, compared to the first-order phase-transition for the Core_k. The question if the thresholds for the appearance of a giant component transfers from the $G(n,p)$-model to the RRG-model remains. Due to the geometrical cluster forming we added the connected variant of the k-core. Let C_i be the components of graph G. Define the connected k-core by

$$\text{Core}_k^* := \max_i \{C_i \cap \text{Core}_k\} \tag{2.36}$$

For the connected k-core it seems plausible to find a first-order phase transition in its size, close to the $C_{\max}$ threshold. In Figure 2.11, we compare the largest component, the k-core and the connected k-core. Around $r_f = 1.2$, presumably a critical threshold, a jump in the size of the largest component is witnessed. This ensures the difference between the k-core and its connected version is (almost fully) nullified.

To conclude, in this section we compared the geometric random graph model to the homogeneous Erdős-Rényi random graph model. We shall continue to do the same for the inhomogeneous variant.

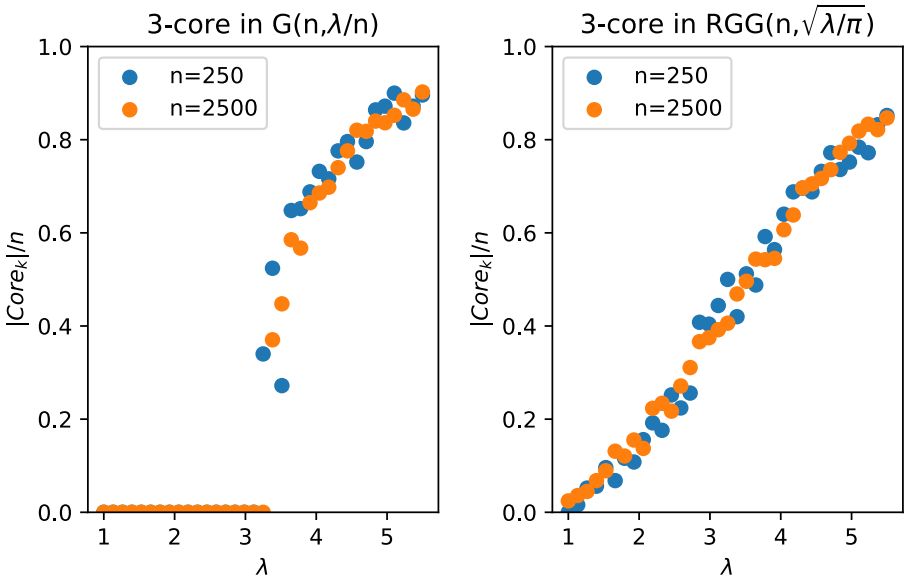
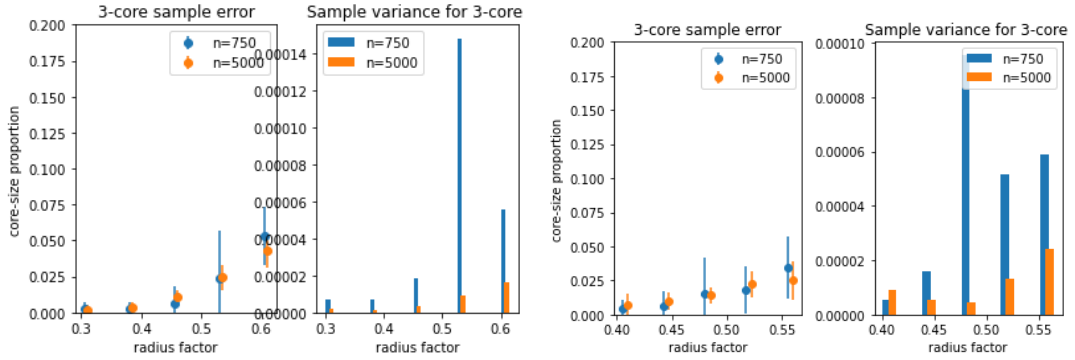


Figure 2.9: Comparison of the 3-core growth in terms of λ for the ER-model (left) and RGG-model (right), based on sample of 35 values for λ each. The ER-model clearly shows the first-order transition, while the 3-core in the RGG-model increases roughly linear throughout.



(a) Simulation samples of 25 for lower values of r_f do not indicate a phase-transition

(b) No apparent first emergence of the k-core. Simulation samples of 25 on the specific r_f region with small proportions.

Figure 2.10: Simulations suggest no first-order phase transition occurs for k-cores in the RGG-model.

2.5. K-cores in Geometric Inhomogeneous Random Graphs

While vertices in the RGG-model are limited by their location and the general radius r, in the GIRG-model the weight of the vertex results in different phenomena. 'Hub'-like behaviour is observed in these GIRG realisations if the weight distribution is heavy-tailed, as described by Bringmann et al in [18]. This is confirmed by simulations with power-law probability distributions, like the Pareto distribution. In Figure 2.12, realisations are shown of both the GIRG-model with $W \sim \text{Poi}^+(2)$ and $W \sim \text{Pareto}(2)$. Here Poi^+ indicates the zero-truncated Poisson distribution with probability density function

$$\mathbb{P}(\text{Poi}^+(\lambda) = k) = \frac{e^{-\lambda}\lambda^k}{k!\left(1 - e^{-\lambda}\right)} \tag{2.37}$$

The truncated Poisson is used to avoid weights of zero, which would make those vertex nonreceptive with respect to any other vertex, i.e. an isolated vertex.

The question arises how the k-core behaves in the GIRG-model with a vertex weight Pareto(λ) distribution. In Figure 2.13 the k-core proportion is plotted for $W \sim \text{Pareto}(3)$. We find a continuous growth for the size of the k-core for this setup of the GIRG-model. Decreasing the heaviness of the tail

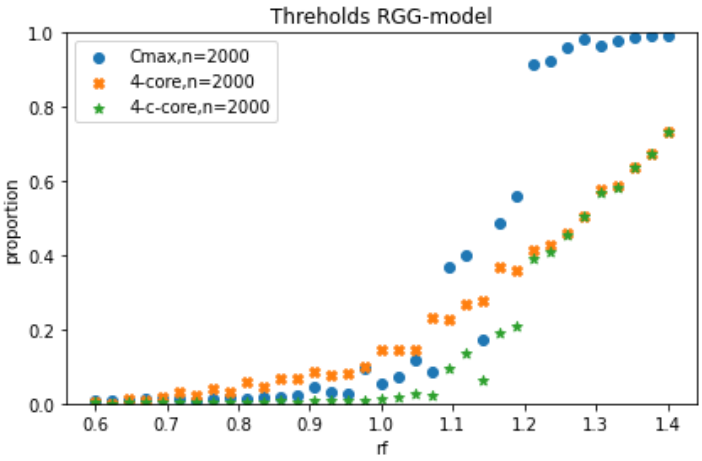
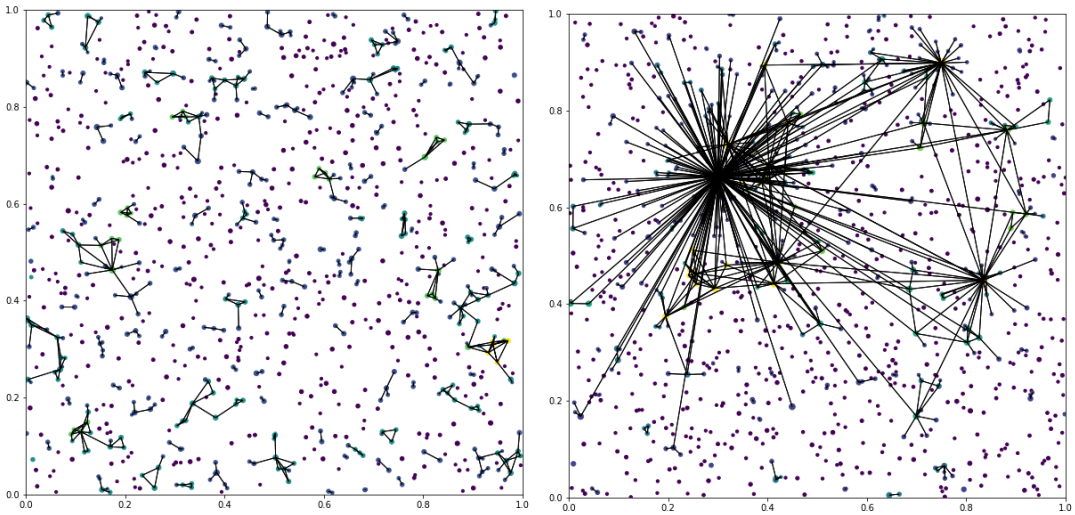


Figure 2.11: Comparison of $C_{\max}$, Core$_k$ and the Core$_k^*$



(a) Realisation of the GIRG-model with a Poi$^+(2)$ weight distribution and $r_f = 0.08$

(b) Realisation of the GIRG-model with a Pareto(2) weight distribution and $r_f = 0.08$

Figure 2.12: Comparison between a realisation of a light- and heavy-tailed weight distributed GIRG-model.

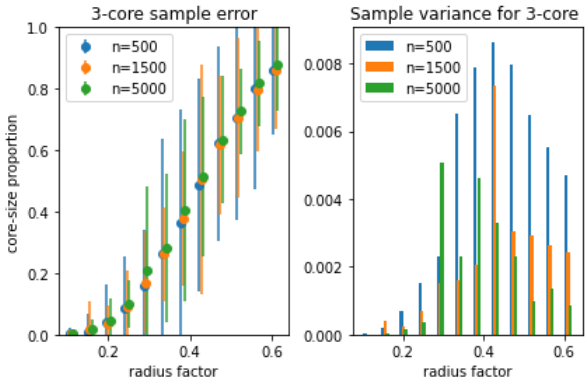
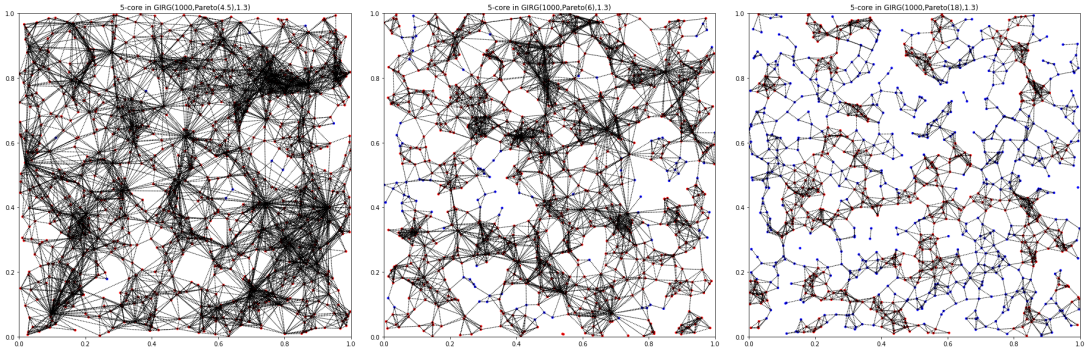


Figure 2.13: Left: Proportion of vertices in the 3-core plotted against the radius factor. The dispersion of the sample of 25 simulations annotated with the vertical bar and the dot represents the sample mean. Right: The sample variance for each sample of 25 simulations.

of the Pareto distribution we find the 'hub' behaviour to lessen and the cluster. Moreover, the radius amplification factor influences the cluster sizes. To illustrate this, Figure 2.14 shows the GIRG-model with the Pareto distribution for the vertex weight with different values for the parameter of the distribution. This motivates us to start looking for a threshold in the 'hub' behaviour and its influence on the k-



(a) Realisation of the GIRG-model with a Pareto(4.5) weight distribution and $r_f = 1.3$

(b) Realisation of the GIRG-model with a Pareto(6) weight distribution and $r_f = 1.3$

(c) Realisation of the GIRG-model with a Pareto(18) weight distribution and $r_f = 1.3$

Figure 2.14: Comparison between realisations of Pareto weight distributed vertices in the GIRG-model with different scale parameter values. The 'hub'-like behavior decreases when the tail becomes less heavy. Edges which wrap around are not drawn to enhance readability.

core. Using simulations we found the following phenomena. We found a combination of the Pareto parameter and radius amplification factor which varies drastically in k-core proportion. Simulations with Pareto(1.4) and $r_f = 0.08$ both $\frac{|\text{Core}_5|}{n} \approx 1$ and $\frac{|\text{Core}_5|}{n} \approx 0$ have been realised. In Figure 2.15 the 5-core proportion based on samples of 25 simulations are presented for different values of b for Pareto(b), with $r_f = 0.08$. Moreover, the variance slightly increases with the size of the graph around $b = 1.4$.

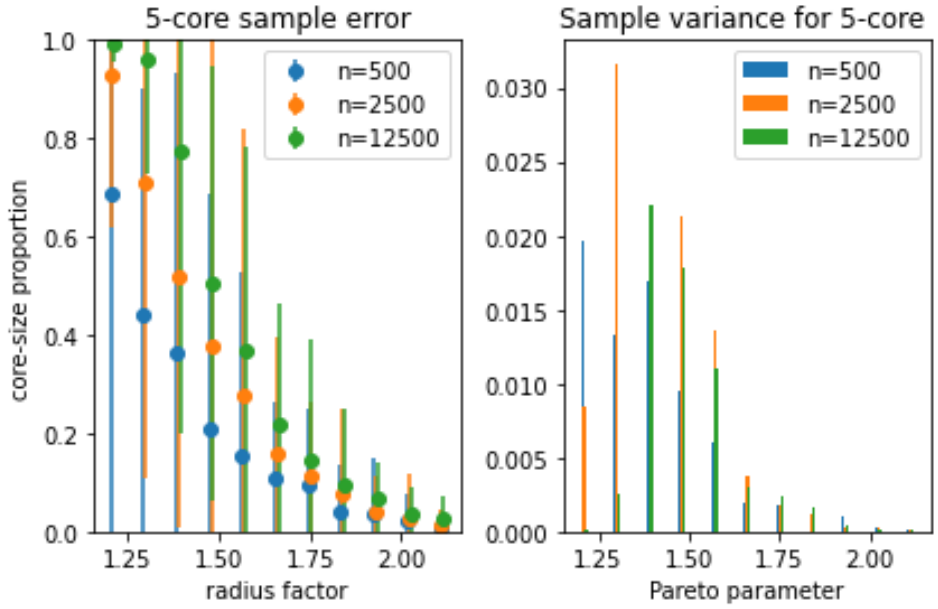


Figure 2.15

For some distribution it is possible to derive a bound for the k-core in terms of super-heavy weighted vertices. In other words, if a certain distribution produces more than k super-heavy weighted vertices, almost surely a k-core exists. Next section we shall prove a statement closely related to this statement of k-core existence in the GIRG-model.

2.5.1. K-cores embedded in complete subgraphs

In this section we shall prove why the GIRG-model for some distributions is guaranteed to have a k-core **a.s.** The intuition behind the proof is that the super-spreaders from a clique of size at least $k+1$, which implies the existence of the k-core.

Theorem 7. *Consider the GIRG-model on the torus $\mathbb{T}^d = [0,1]^d$ with dimension d and the Euclidean norm, and vertex weight distribution T. Let $G = (\tilde{V}, E)$ be a random graph in this GIRG-model with connection probability*

$$\mathbb{P}(\{u,v\} \in E \big| (x_u, w_u), (w_v, w_v) \in \tilde{V}) = \begin{cases} \min\left(1, \frac{w_u \cdot w_v \cdot r^d}{\|x_u - x_v\|_2}\right) & \text{if } X_u \neq X_v \\ 0 & \text{otherwise} \end{cases} \tag{2.38}$$

*where u and v are vertices in G. Now assume that T is a heavy-tailed distribution such that $|\{u \in \tilde{V} : w_u \geq \sqrt{n}\}| > k$, $\forall k \in \mathbb{N}$ for large values of n. Then graph G with the has a k-core **w.h.p.** if the radius amplification factor $c > \sqrt[d]{\frac{1}{2}\sqrt{d}}$*

Proof. First, let $k \in \mathbb{N}$ and $d \in \mathbb{N}$ be arbitrary fixed constants. Let $G = (\tilde{V}, E)$ be a random graph the GIRG-model on the torus $\mathbb{T}^d = [0,1]^d$ with dimension d and the Euclidean norm, and vertex weight distribution T. Now we define the set S by

$$S := \{u \in \tilde{V} : w_u \geq \sqrt{n}\} \tag{2.39}$$

Now by assumption T is heavy-tailed such that $|S| \geq k$ as $n \to \infty$. Therefore, we consider the following probability that vertices $u, v \in S,\ u \neq v$, arbitrarily chosen, are connected. We find that

$$\begin{aligned} \mathbb{P}(\{u,v\} \in E) &= \min\left(1, \frac{w_u \cdot w_v \cdot r^d}{\|x_u - x_v\|_2}\right) \\ &= \min\left(1, \frac{w_u \cdot w_v \cdot \left(cn^{-\frac{1}{d}}\right)^d}{\|x_u - x_v\|_2}\right) \end{aligned}$$

Since the largest distance two vertices can be apart in $\mathbb{T}^d$ is at most $\sqrt{d \cdot \left(\frac{1}{2}\right)^2} = \frac{1}{2}\sqrt{d}$. We find that $\|x_u - x_v\|_2 \leq \frac{1}{2}\sqrt{d}$. Moreover, we have that $w_u, w_v \geq \sqrt{n}$, because $u, v \in S$. Then it follows that

$$\begin{aligned} \mathbb{P}(\{u,v\} \in E) &\geq \min\left(1, \frac{w_u \cdot w_v \cdot \frac{c}{n}}{\frac{1}{2}\sqrt{d}}\right) \\ &\geq \min\left(1, \frac{\sqrt{n} \cdot \sqrt{n} \cdot c^d}{\frac{n}{2}\sqrt{d}}\right) \\ &= \min\left(1, \frac{nc^d}{\frac{n}{2}\sqrt{d}}\right) \\ &= \min\left(1, \frac{2c^d}{\sqrt{d}}\right) \end{aligned}$$

By assumption $c > \sqrt[d]{\frac{1}{2}\sqrt{d}}$ and we find

$$\mathbb{P}(\{u,v\} \in E) \geq \min(1,1) = 1 \tag{2.40}$$

Since $u, v \in S$ were chosen arbitrarily this holds for all vertices in S, thus S forms a clique of size $|S| > k$. The clique of size greater than k implies the existence of a k-core. □

The condition on c is mild as $c(d) \to 1$ for $d \to \infty$. We remark that an alternative version of the proof includes the condition $c = 1$ and to move the dimension formula to the condition on the weights in the set S. Important to note that this proof does not provide a tight lower bound on the weights however. Many interesting question remain on k-cores in the GIRG-model. For example, the size of the k-core can vary widely. Even thought the super-heavy vertices form a clique, it does not imply any other vertices are connected few or more of the super-heavy vertices, due to the multiplications of weights. Although very interesting, a more in-depth analysis of the k-core size in the GIRG-model is out of the scope of this thesis. Thus, the next section we briefly shall discuss thresholds for inhomogeneous random graphs in general.

2.5.2. Thresholds in Inhomogeneous Random Graphs

In the paper 'the phase transition inhomogeneous random graph' by Bollobás, Janson and Riordan they state in particular a necessary and sufficient condition for the existence of a giant component in a inhomogeneous random graph under a weak assumption [17]. Their model is an extension to the inhomogeneous random graph model by Söderberg [72]. Due to the generality of the model this is an important result. The paper contains is lengthy and highly technical and contains many results. The paper itself is out of the scope of this thesis. Bollobás et al state: "... many properties of the model can be determined, in particular the critical point of the phase transition, and the size of the giant component above the transition. We do this by relating our random graphs to branching processes, which are much easier to analyze". This is our motivation to focus on the basics of branching processes in the next chapter in order to better understand thresholds.

3

Branching processes for k-cores

In this chapter first a branching process is defined. Next an example is provided using the branching progress. Moreover, some key results regarding 'extinction' of a tree branching process are discussed. Then, in Section 3.2, the k-core existence is viewed as a branching process, recovering results earlier stated about k-core thresholds. Lastly these theoretical results are matched with empirical simulations.

3.1. Branching process

The emergence of the giant component behaviour can be explained through a branching process [17]. Similarly the phase transition of the k-core can be described [48]. To understand the branching process the mathematical object 'tree' must be introduced first.

3.1.1. Trees

Trees are an important mathematical model to study phenomena. Trees can be viewed a graph with no cycles.

Definition 10. *A **tree** is an undirected, connected, acyclic graph. A **forest** is a graph containing tree components only.*

An important observation follows from the definition: there is a unique path between any two vertices in a tree. For the reader unfamiliar with paths in a graph, we provide some definitions. Let $G = (V, E)$ be a graph. A **walk** w is a sequence of alternating vertices and edges, so $w = (v_0, e_0, v_1, e_1, \dots, e_n, v_n)$, beginning and ending with a vertex. Important to note the following condition must hold $\forall i \in \{0, \dots, n\}$: $e_i \in E \wedge v_i \in e_i \wedge v_{i+1} \in e_i$ to make the walk sensible. Otherwise nonsensical walks could be constructed. A **path** from u to v is a walk where $v_0 = u$ and $v_n = v$ in which every vertex appears at most once.

In the subcritical regime, $\lambda < 1$, in the Erdős-Rényi random graph model w.h.p. all components are small and are cycle-free [45].

3.1.2. Extinction of the branching process

Now it is time to properly define the Galton-Watson branching process. First a root vertex is taken. The children of the root are generated by a discrete random distribution. These children form the first generation. Formally

Definition 11. *Let $X_{n,i}$ be a double infinity array of i.i.d. discrete random variables. Here $X_{n,i} \sim X$ represent the number of offspring of vertex i in the nth generation. Let $X_{n,i} \sim X, \forall n, i \geq 0$, where X is the off-spring distribution. Hence, one assumes $X_{n,1} \geq 0$. Let Z_n be the number of vertices in the n-th generation. By assumption $Z_0 = 1$.*

Moreover Z_{n+1} can be formulated by recursion relation since it solely consists of children of the previous generation.

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i} \tag{3.1}$$

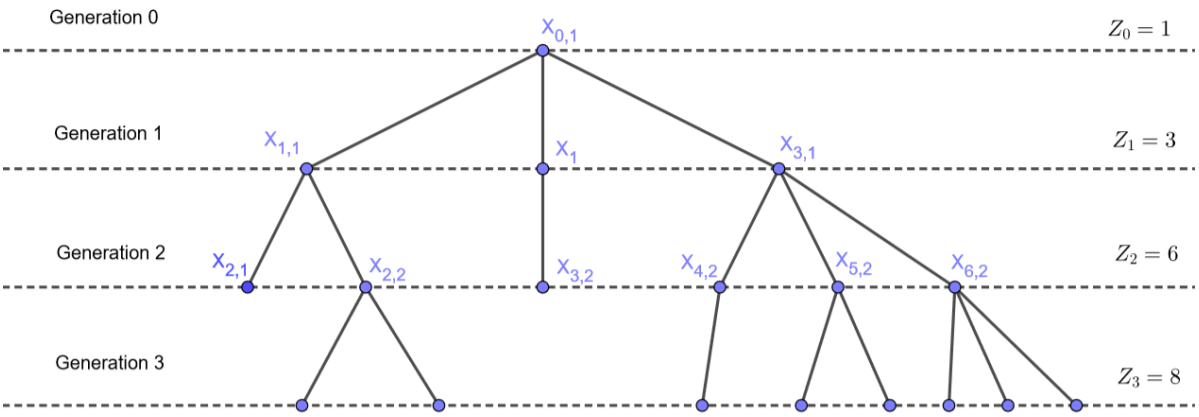


Figure 3.1: Realisation of three generations of a branching process.

In Figure 3.1, the realisation of the first generations of the branching process is visualised.

The **extinction probability** is formulated as the probability of the event that the n-th generation has no children.

$$\eta := \mathbb{P}(\exists n < \infty : Z_n = 0) \tag{3.2}$$

Not surprisingly, when $\mathbb{E}[X] < 1$, $\eta = 1$. Since, on average, each generation has less children than parents, in the end the branching process will stop. The probability that the population survives up to generation n is exponentially small according to Theorem (8).

Theorem 8. *Let $0 \leq n < \infty$ fixed. Let $\mu = \mathbb{E}[X]$ and assume $\mu < 1$. Then*

$$\mathbb{P}(Z_n > 0) \leq \mu^n \tag{3.3}$$

Proof. Let $0 \leq n < \infty$ fixed and assume $\mathbb{E}[X] = \mu < 1$. Since Z_n is assumed integer we find

$$\mathbb{P}(Z_n > 0) = \mathbb{P}(Z_n \geq 1) \leq \frac{\mathbb{E}[Z_n]}{1}$$

using Markov's Inequality. Now to compute the expectation of Z_n we note that the $X_{n,i} \sim X$ are all i.i.d. for each generation. Then from Wald's equation it follows that

$$\mathbb{E}[Z_n] = \mathbb{E}[X] \cdot \mathbb{E}[Z_{n-1}]$$

using the fact that $Z_n = \sum_{i=0}^{Z_{n-1}} X_{n-1,i}$ where Z_{n-1} is a random variable. Unfolding this recursive expression and using the default $Z_0 = 1$, one now has

$$\mathbb{E}[Z_n] = \mathbb{E}[X]^n \cdot \mathbb{E}[Z_0] = \mu^n \cdot 1$$

Combining this with the result from Markov's inequality we conclude

$$\mathbb{P}(Z_n > 0) \leq \mu^n$$

□

Thus indeed, if $\mu = \mathbb{E}[X] < 1$, then $\eta = 1$ as $\mu^n \to 0$ as $n \to \infty$. If $\mu > 1$ however, $\eta < 1$ as $n \to \infty$. To find a bound for this probability, one should look at the *survival probability* ξ in stead of the 'extinction probability' η. Being mutually exclusive it follows that $\xi = 1 - \eta$ and by definition one has

$$\xi = \mathbb{P}(Z_n > 0 \forall n \geq 0)$$

For the branching process to 'survive', the corresponding tree must keep growing over time. The total number of vertices in the tree can be described as the total number of children $T = \sum_n^\infty Z_{n+1}$, effectively

summing the generations [34]. T is also referred to as the *total progeny*. Then equivalently one could say the branching process survives is T increases indefinitely.

This brings us to another way to assess the survival of the tree branching process is a breadth-first approach. This view enables to view each vertex independently, where the recursion can be reduced to a sum of i.i.d. random variables.

Definition 12. *Let $X'_1, X'_2, ... \sim X_{1,1}$. Then the random-walk formulation of the branching process is recursively described by*

$$S'_0 = 1 \tag{3.4}$$
$$S'_i = S'_{i-1} - 1 = X'_1 + ... + X'_i - (i-1) \tag{3.5}$$

where S'_i is the branching number.

The branching number depicts the number of 'active' vertices which still branch and have not been visited. After a vertex has been visited the total 'active' vertices is decreased by 1 and the sum of the children is added to the 'active' count. So, when each vertex has precisely 1 child, the tree continues forever, as we would expect. If, however $\mathbb{E}[X] < 1$, the branching process will stop at some point, matching the earlier branching process result. This stopping point can be formulated as

$$T' := \inf\{t : S'_t = 0\} \tag{3.6}$$

i.e. the point where no more 'active' vertices, and thus none can produce off-spring for further branching. This brings us to the bound for the extinction probability with $\mu > 1$.

Theorem 9. *Let X be the offspring distribution with $\mu = \mathbb{E}[X] > 1$. Let T denote the total progeny. Let $k \in \mathbb{N}$ finite, then*

$$\mathbb{P}(T < \infty) \leq \frac{1}{1 - e^{-I}} \tag{3.7}$$

with exponential rate I defined as $I := \sup_{t \leq 0} \left(t - \log \mathbb{E}[e^{tX}]\right)$

Now all the preliminary knowledge required has been to relate the branching process of trees to the existence of the k-core.

3.2. Tree embedded k-cores

When a vertex is picked and observed locally, they neighbours can be modelled as children and the local portion of the graph as a tree [27]. Disregarded are any edges between 'children' of the same generation. Take C_3, a cycle of three vertices, then when picking a starting vertex at random, the tree either looks like a root with two children ór a tree with one child which also has one child. In this thesis we shall use the 'neighbours first' mapping. For this we define the neighbourhood of a vertex v.

Definition 13. *Let $G = (V, E)$ be a graph. Let $v \in V$ be a vertex. Then the **neighbourhood of a vertex** v is given by*

$$N(v) = \{u \in V : u \neq v \wedge \{u, v\} \in E\} \tag{3.8}$$

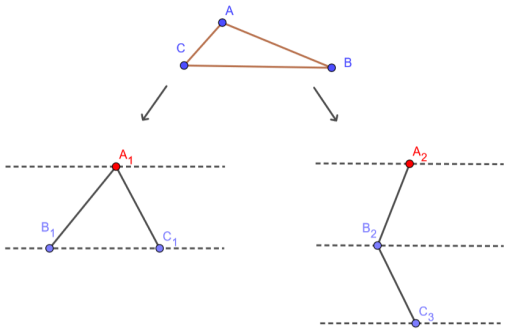
*Moreover let U be a subset of vertices, i.e. $U \subseteq V$, then the **neighbourhood of a vertex set** U is given by*

$$N(U) = \{w \in V : w \notin U \wedge [\exists u \in U : \{w, u\} \in E\}] \tag{3.9}$$

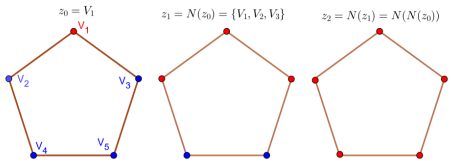
In other words, any vertex you can reach from v by traversing only one edge is in the neighbourhood of v.

Now the existence of k-core is similar to a branching process which contains $k-1$-ary tree.

Assume the k-core exists in a random graph G. Now let v_{root} be a random vertex in the k-core, then surely it must hold that $|N(v_{\text{root}})| \geq k$, by definition, since it is part of the k-core. Now if G is K_k, i.e. a clique of size k, then the process stops as there are no vertices left to explore. Therefore, let us assume that G is a large graph but not extremely interconnected. Now each 'child' must have at least $k-1$ children itself, i.e. the neighbourhood must be at least size $k-1$. This leads to the general idea that if the total tree has a $k-1$-ary subtree it is very probable the k-core exists. Using the random branching process in this way, established in [68], Fernholz and Ramachandran in [38] find the following.



(a) The tree mapping of the cyclic graph C_3. Left the 'neighbourhood' mapping, i.e. breadth-first, and right a depth-first mapping



(b) Example of a tree mapping of the cyclic graph C_5. Each generation equals another neighbourhood of the previous generation. The red vertices have been already visited and have been added to a generation. Blue vertices have yet to be added.

Theorem 10. *Let $k \geq 3$ be a fixed integer. Let G be a Chung-Lu graph with a degree distribution $\mathbf{p}_n$ such that $p_i \propto i^{-\beta}$ for some $\beta \geq 0$. Then the following holds*

1. *If $\beta \geq 3$, there is no giant k-core w.h.p.*
2. *If $2 < \beta < 3$, there is a giant k-core w.h.p.*

To compare the branching process more directly with the existence of the k-core the $k-1$-ary tree critical threshold is shown to be equal to the one found by Pittel et al in [68]. Consider a random branching process with a tree of n generations. Now we define the *root* of the tree to be 'good' if at least $d-1$ children are 'good'. A child is defined 'good' if has at least $d-1$ children which are 'good'.

Definition 14. *Given a random branching process with $X_{n,i} \sim X$. Denote the probability of a child in the tree containing a d-ary tree for n generation, defined recursively, by*

$$\zeta(d-1, n-1) = \sum_{i=0}^{\infty} \mathbb{P}(X=i)\mathbb{P}\left(\mathit{Bin}\left(i, \zeta(d-1, n-2)\right) \geq d-1\right) \tag{3.10}$$

Now the probability of the root can also be formally written down

$$\mathbb{P}(v_{\text{root}} \text{ is } \textit{good}) = \sum_{i=0}^{\infty} \mathbb{P}(X_0 = i)\mathbb{P}(\text{at least } d-1 \text{ children are 'good'}) \tag{3.11}$$

$$= \sum_{i=0}^{\infty} \mathbb{P}(X_0 = i)\mathbb{P}\left(\text{Bin}\left(i, \zeta(d-1, n-1)\right) \geq d\right) \tag{3.12}$$

using the law of total probability, where $\forall n, i \geq 0 : X_{n,i} \sim X$.

3.2.1. Fixed point equation for k-core tree embedding

If one assumes there exists a stable for ζ in the branching process with 'good' vertices, then surprisingly at first, such equation can be solved for some distributions of X. This Now for the case that $n \to \infty$ we assume $\zeta(d-1, n-k) \to \zeta, \forall k \in \mathbb{N}$. The reasoning here is that simply every child here has a comparable subtree, i.e. landing at some random child in the tree provides equal information. Under the assumption $\zeta(d-1, n-k) \to \zeta$, Equation (3.10) can be rewritten as

$$\zeta = \sum_{i=0}^{\infty} \mathbb{P}(X=i) \sum_{j=d-1}^{i} \binom{i}{j} \zeta^j (1-\zeta)^{i-j} \tag{3.13}$$

using the probability density function of the Binomial distribution.

In order to solve for ζ, probability generating functions will be used. By definition the generating function for a random variable X is

$$G_X(s) = \sum_{i=0}^{\infty} \mathbb{P}(X = i)s^i = \mathbb{E}(z^X), \quad s \in [0, 1] \tag{3.14}$$

Often the radius of convergence is greater than $[0, 1]$, but for this derivation this suffices. Since we are within the radius of convergence, we are safe to use term-wise differentiation with respect to z. Therefore we find the following derivatives of the generating function

$$G'_X(s) = \sum_{i=0}^{\infty} \mathbb{P}(X = i)is^{i-1} \tag{3.15}$$

$$G_X^{(j)}(s) = \sum_{i=0}^{\infty} \mathbb{P}(X = i)i(i-1)(i-2)\dots(i-j+1)s^{i-j}, \quad j < i \tag{3.16}$$

First, however, we must modify Equation (3.13). Therefore we write out

$$\begin{aligned}\zeta &= \sum_{i=0}^{\infty} \mathbb{P}(X = i) \sum_{j=d-1}^{i} \binom{i}{j} \zeta^j (1-\zeta)^{i-j} \\ &= \sum_{i=0}^{\infty} \mathbb{P}(X = i) \sum_{j=d-1}^{i} \frac{i!}{(i-j)!j!} \zeta^j (1-\zeta)^{i-j}\end{aligned}$$

Because of our assumptions on ζ we count this double sum in a different way, allowing the interchanging of the sums. The first terms of i are zero until $i = j = d - 1$. So we start counting for $j = d - 1$, hence

$$\begin{aligned}\zeta &= \sum_{j=d-1}^{\infty} \mathbb{P}(X = i) \sum_{i=j}^{\infty} \frac{i!}{(i-j)!j!} \zeta^j (1-\zeta)^{i-j} \\ &= \sum_{j=d-1}^{\infty} \sum_{i=j}^{\infty} \mathbb{P}(X = i) \frac{i!}{(i-j)!j!} (1-\zeta)^{i-j} \\ &= \sum_{j=d-1}^{\infty} \frac{\zeta^j}{j!} \sum_{i=j}^{\infty} \mathbb{P}(X = i)i(i-1)\dots(i-j+1)(1-\zeta)^{i-j}\end{aligned}$$

rearranging the terms solely dependent on j and cancelling factorial factors. Now recall Equation (3.16) and substitute this derivative result of the generating function to find

$$\zeta = \sum_{j=d-1}^{\infty} \frac{\zeta^j}{j!} G_X^{(j)}(1-\zeta) \tag{3.17}$$

$$\tag{3.18}$$

Since ζ is a probability we have $\zeta \in [0, 1]$. So the following change of variable is allowed:

$$v = 1 - \zeta, \quad v \in [0, 1] \tag{3.19}$$

So when substituting ζ for $1 - v$ we find

$$1 - v = \sum_{j=d-1}^{\infty} \frac{(1-v)^j}{j!} G_X^{(j)}(v) \tag{3.20}$$

Moreover, in the case of examining the existence of a k-core, it makes sense to assume $X \sim \text{Poi}(\lambda)$. Then the generating function can be written down explicitly;

$$\begin{aligned} G_{\text{Poi}(\lambda)}(v) = G_{\text{Poi}(\lambda)}(v) &= \sum_{k=0}^{\infty} \mathbb{P}\left(\text{Poi}(\lambda) = k\right) v^k \\ &= \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} v^k \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda v)^k}{k!} \\ &= e^{-\lambda} e^{\lambda v} = e^{-\lambda(1-v)} \end{aligned}$$

where in the last step the definition of the exponential is used. Differentiate with respect to v to find

$$G^{(j)}_{\text{Poi}(\lambda)}(v) = \lambda^j e^{-\lambda(1-v)} \tag{3.21}$$

Now, we use the expression above in Equation (3.20). Then one has

$$1 - v = \sum_{j=d-1}^{\infty} \frac{(1-v)^j}{j!} \lambda^j e^{-\lambda(1-v)} \tag{3.22}$$

Furthermore, we recall $\zeta = 1 - v$ and write

$$\begin{aligned} \zeta &= e^{-\lambda\zeta} \sum_{j=d-1}^{\infty} \frac{(\lambda\zeta)^j}{j!} \\ &= e^{-\lambda\zeta} \left(e^{\lambda\zeta} - \sum_{j=0}^{d-2} \frac{(\lambda\zeta)^j}{j!} \right) \\ &= 1 - e^{-\lambda\zeta} \sum_{j=0}^{d-2} \frac{(\lambda\zeta)^j}{j!} \end{aligned} \tag{3.23}$$

In particular, for $k = 3$, we found

$$\zeta = 1 - e^{-\lambda\zeta}\left(1 + \lambda\zeta\right) \Leftrightarrow \tag{3.24}$$

$$e^{-\lambda\zeta} = \frac{1-\zeta}{1+\lambda\zeta} \tag{3.25}$$

Recall that, from Figure 2.4 it is known that $\lambda_3 \approx 3.35$ is the critical 3-core existence threshold. For $k = 3$ we observe, in Figure 3.3, that $\forall \lambda < \lambda_3$ Equation (3.24) has no solutions within the interval $[0, 1]$. There a unique solution at $\lambda = \lambda_3$ and are multiple solutions for $\lambda > \lambda_3$. This corresponds precisely to results found earlier using in Chapter 2. Solving for ζ with a given λ value enables us to calculate the probability of when the *root* of the tree is 'good'. Let ζ_λ be a solution to the equation, then

$$\mathbb{P}(v_{\text{root}} \text{ is } \textit{good}) = \sum_{i=0}^{\infty} \mathbb{P}(\text{Poi}(\lambda) = i)\mathbb{P}\left(\text{Bin}\left(i, \zeta_\lambda\right) \geq d\right) \tag{3.26}$$

Therefore we can conclude that for $\lambda < \lambda_3$,

$$\mathbb{P}(\text{the branching process has a 3-ary tree embedding}) = 0 \tag{3.27}$$

while for $\lambda = \lambda_3$

$$\mathbb{P}(\text{the branching process has a 3-ary tree embedding}) > 0 \tag{3.28}$$

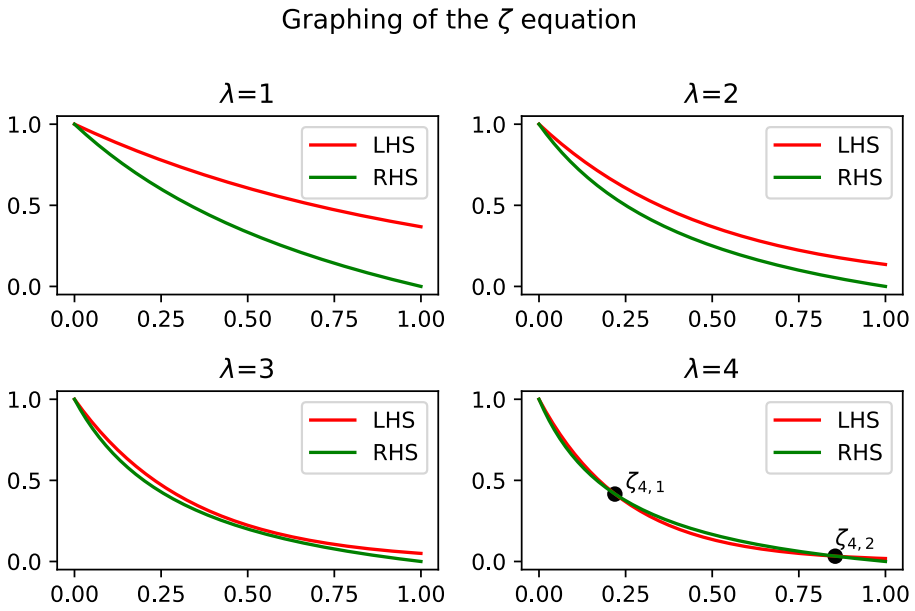


Figure 3.3: The left-handside (LHS, $e^{-\lambda\zeta}$) and right-handside (RHS, $\frac{1-\zeta}{1+\lambda\zeta}$) of Equation (3.24) plotted for different values of □with $k = 3$. There are no intersection points on the interval $[0, 1]$ for $\lambda < \lambda_3$.

hence there is a discontinuity of the cure $P_\lambda()$ plotted against λ, which causes the first-order phase transition we observed in Chapter 2 and know from literature. Moreover we conclude that this is different to actual survival of the branching process itself, which at criticality $\mathbb{E}[X] = 1$, under the obvious assumption $\mathbb{P}(X = 1) < 1$, has a survival probability $\zeta = 0$. When $\mathbb{E}[X] > 1$, there is a non-trivial solution and the ζ is continuous function of $\mathbb{X}$. Hence for the branching process itself a second-order phase transition occurs. With that distinction we conclude this chapter on the branching processes for finding thresholds in random graphs.

4

Discussion

While much is known about the $G(n,p)$ model due to great efforts of the last few decades, both the RGG and GIRG model prove to be fundamentally different. There are many angles to view the graph structures in these models.

Interestingly, no first-order phase-transitions could be detected for the k-core in the RGG-model. These phase-transitions were inspired by the theorems for the $G(n,p)$ model. A possible explanation was given for the conditional probability which, using arguments based on the geometric embedding of the RGG-model. Further research is possible regards a second-order phase-transition for the proportion of the *connected* k-core.

Though, it is probable to find phase-transitions for the GIRG-model as it versatile with the combination of the distribution parameter and the radius amplification factor. For heavy-tailed distributions further research into such transitions would be of interest.

Recovering the result by Pittel et al for the $G(n,p)$-model enables us to link the random branching process to the k-core of the $G(n,p)$-model. Since branching processes form an important way of reasoning about graph structures, further research of such processes for a geometric embedding might be insightful. Such geometric branching process has much in common with a geometric preferential attachment model. A root based geometric proces might also be of interest. Where for each generation using an offspring distribution X inside the, possible weight dependent, connectivity circle are drawn. One could think of this as migration for expanding population with limited resources per area.

While not extensively mentioned, the configuration model, is an important model to view random graphs with a given degree sequence. Much has been written about this model and maybe a geometric interpretation of this model, where half-edges can only connect randomly to other half-edges inside the area of overlap between connectivity circles. Most likely, this will result in clusters only. Extending this geometric configuration model with a weight distribution might allow for powerful results from the configuration model to be transferred to the geometric realm.

5

Conclusion

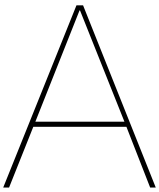
The main goal of the thesis is to understand the behaviour of k-cores in various random graph models, and to contrast the theory with simulations. Moreover to give intuition to the k-core through the analogue of neighbourhoods in random graphs by Branching Processes.

We were able to reproduce the k-core threshold with simulations of the Erdős-Rényi random graph model, as stated by Pittel et al in Theorem 5. The simulations match the expected first-order phase transition in this model. Next we compared this to a Random Geometric Graph model (RRG), which extends the Erdős-Rényi random graph model into a spatial setting. First we proved the degree distribution of the RGG-model to be Poisson and matched this with a fitting of simulation samples. Moreover, we found a cluster-like behaviour in the RGG-model, which results in a small size k-core even before a giant component exists. Furthermore, we gave a proof as a possible explanation why these clusters are more likely in the spatial setting. Notably, no second-order phase transition could be observed in the RGG-model.

Furthermore, we added an inhomogeneous extension to the RGG-model, called the Geometric Inhomogeneous Random Graph model (GIRG), based on vertex weights using a vertex weight distribution. We confirmed with simulations that this results in 'hub'-like behaviour where the center of hub is a vertex with a heavy weight, i.e. far above the average weight. We state it plausible for a GIRG-model with a heavy-tailed vertex distribution to find a first-order phase transition for the k-core in the GIRG-model, with a threshold based on the vertex weight distribution parameters combined with a radius amplification factor to scale properly.

Also we proved that for a heavy-tailed vertex weight distribution in the GIRG-model, under certain conditions, ensures a k-core with high probability as the number of vertices is sufficiently large. For light-tailed distributions the simulations are comparable to the RGG-model, where a small sized k-core exists.

In Chapter 3, we studied the k-core existence in homogeneous graphs using the branching processes. First we recover some known preliminaries. In particular the threshold for when a branching process stops (extinction) or continues for ever (survival). Furthermore, we investigate the embedding of a k-ary tree in a branching process tree. Using a fixed point equation method we find a k-ary tree embedding first occurs at the threshold stated by Pittel et al for k-cores in the Erdős-Rényi random graph model. Thus showing the power of reasoning with branching processes in general and for large network property thresholds in particular.



Python code

Only the most important functions have been listed. Many more code was used to investigate behaviour of Random Graph models or to plot figures, but can be easily replicated or is of minor importance.

```
#  *  coding: utf 8  *
"""
Updated on Monday 04 July at 202220:52:21

@author: Vincent Wassenaar
"""
import numpy as np
from numpy import linalg
from scipy.stats import poisson
from scipy.stats import pareto
from scipy.spatial import distance
import math
import matplotlib.pyplot as plt
from matplotlib.collections import LineCollection
import copy

#inserts in order of coordinates in order of dimension
#Example (1,2), (1,3),(2,1),(2,3),(3,1)
def sample_zero_truncated_poisson(rate):
        u = np.random.uniform(np.exp( rate), 1)
        t =  np.log(u)
        return 1 + np.random.poisson(rate   t)

#Function to generate a GIRG model graph
def createRandomWeightedGraph(n,rf, b=3):
    #create the scaled radius
    r_n = rf*(n**( 1/2))
    #drawn n points from [0,1]^2 uniformly
    p = np.random.rand(n,2)

    ###
    # Alternative code for creating a random weighted graph
    # using a zero truncated poission distribution.
    # w = []
    # for i in range(n):
    #     w.append(sample_zero_truncated_poisson(b))
    # w = np.add(0.1,np.random.poisson(b, n))
```

```
38      ###
39
40      #create a vertex weight sample from the Pareto
          ↪ distribution with parameter b
41      w = pareto.rvs(b, size=n)
42      #create an empty Adjacency matrix
43      adjMatrix = np.zeros((n,n))
44
45      for i in range(n 1):#calculate for each vertex the
          ↪ distance to the other vertices
46          ps = p[i,None]
47          adj_row = np.zeros((1,int(n i 1)))
48          #check the distance 9 times, due to the border
              ↪ wrapping, for each shifted copy of the other
              ↪ vertices
49          for xoff in [1,0,1]:
50              for yoff in [1,0,1]:
51                  jo=ps.copy()
52                  jo[:,0]+=xoff
53                  jo[:,1]+=yoff
54                  adj_row = np.add(adj_row,np.where(
55                      np.divide(distance.cdist(jo,p[ (n i 1):]),
                            ↪ w[ (n i 1):]*w[i])<r_n,1,0))
56          #fill the upper right triangle of the Adjacency
              ↪ matrices
57          adjMatrix[i, (n i 1):] = np.where(adj_row>0,1,0)
58      #the Adjacency matrix should be symmetrical
59      retMatrix = np.add(adjMatrix, adjMatrix.transpose())
60      #return the vertex location, the Adjacency matrix and the
          ↪ vertex weights.
61      return (p, retMatrix, w)
62
63  #Function to generate a graph for the default RGG model
64  def createRandomGraph(n,rf):
65      r_n = rf*(n**( 1/2))
66      p = np.random.rand(n,2)
67      adjMatrix = np.zeros((n,n))
68      for i in range(n 1):
69          ps = p[i,None]
70          adj_row = np.zeros((1,int(n i 1)))
71          for xoff in [1,0,1]:
72              for yoff in [1,0,1]:
73                  jo=ps.copy()
74                  jo[:,0]+=xoff
75                  jo[:,1]+=yoff
76                  adj_row = np.add(adj_row,np.where(distance.
                        ↪ cdist(jo,p[ (n i 1):])<r_n,1,0))
77          adjMatrix[i, (n i 1):] = np.where(adj_row>0,1,0)
78      retMatrix = np.add(adjMatrix, adjMatrix.transpose())
79      return (p, retMatrix, None)
80
81  #Function to find the k core shells of a graph.
82  #Note that the k core is simply the union of all shells
      ↪ greater than or equal to k.
83  #Time complexity is O(n*log(n))
84  def getCoreShells(vertices, edgeMatrix):
```

```
85      degrees = edgeMatrix.sum(axis=0)
86      coreShells = [[]]
87      zeros = np.where(degrees==0)[0].tolist()
88      if len(zeros) >0:
89          coreShells[0] = zeros
90      degrees[degrees == 0] =  1
91
92      i=1
93      while (degrees.sum()!= 1*len(degrees)):
94          if len(coreShells)<=i:
95              coreShells.append([])
96          #get all points where the degree is below i
97          depletables = np.argwhere((degrees <= i) &(degrees>=0)
              ↪ )
98          #depletables are those points who should still be
              ↪ explored but now fail to have a degree equal to
              ↪ the current shells.
99          if len(depletables)==0:
100             i+=1
101             continue
102         for v in depletables:
103             degrees[v[0]] =  1
104             coreShells[i].append(v[0])
105             #get column to decrease all connections to the
                  ↪ vertex which is explored
106             vrow = np.nonzero(edgeMatrix[v[0]])[0]
107             for r in vrow:
108                 edgeMatrix[v[0],r]  =1
109                 edgeMatrix[r,v[0]]  =1
110                 degrees[r] =1
111     return coreShells
112
113 #Helper function to visualise created Random Graphs
114 def visualizeRandomGraph(vertices, edgeMatrix, rf, w=None):
115     if w is None:
116         w = np.repeat(1,len(vertices))
117     r_n = rf*len(vertices)**( 1/2)
118     fig, ax = plt.subplots(figsize=(10,10))
119     ax.set_xlim([0,1])
120     ax.set_ylim([0,1])
121
122     lineMatrix = copy.deepcopy(edgeMatrix)
123     #compute the degrees
124     coreShells = getCoreShells(vertices,lineMatrix)
125
126     xpos = []
127     ypos = []
128     nsize = []
129     ncol = []
130     for i in range(len(coreShells)):
131         for j in coreShells[i]:
132             xpos.append(vertices[j][0])
133             ypos.append(vertices[j][1])
134             nsize.append(5+5*w[j])  #the size of vertex is
                  ↪ modified depending on its weight
135             ncol.append(i)          #each k shells has its
```

```
                ↪ only coloured vertices
136     node_collection = ax.scatter(xpos,ypos,s=nsize,c=ncol)  #
          ↪ draw the vertices on the torus
137     #draws only the edges which do not boundary wrap, to
          ↪ enhance readability.
138     edge_pos = np.asarray([(vertices[e[0]], vertices[e[1]])
          ↪ for e in np.transpose(np.nonzero(edgeMatrix)) if
          ↪ linalg.norm(np.subtract(vertices[e[0]],vertices[e
          ↪ [1]]))<=r_n*w[e[0]]*w[e[1]]])
139     edge_collection = LineCollection(
140         edge_pos,
141         colors='black',
142         linewidths=1,
143         antialiaseds=(1,)
144     )
145     edge_collection.set_zorder(1)  # edges are drawn behind
          ↪ vertices
146     ax.add_collection(edge_collection)
147
148     fig.show()
149     plt.savefig("core shells color.svg", format="svg")
150
151 #Helper function to view the k core more easily.
152 def visualizeKcore(vertices, edgeMatrix, k, rf,b):
153     r_n = rf*len(vertices)**( 1/2)
154     fig, ax = plt.subplots(figsize=(10,10))
155     ax.set_xlim([0,1])
156     ax.set_ylim([0,1])
157
158     lineMatrix = copy.deepcopy(edgeMatrix)
159     #compute the degrees
160     coreShells = getCoreShells(vertices,lineMatrix)
161     print(np.sum([len(coreShells[i]) for i in range(k,len(
          ↪ coreShells))])/len(vertices))
162     #print('cores',coreShells)
163     xpos = []
164     ypos = []
165     nsize = []
166     ncol = []
167     col = 'blue'
168     for i in range(len(coreShells)):
169         if i>=k:
170             col = 'red'
171         for j in coreShells[i]:
172             xpos.append(vertices[j][0])
173             ypos.append(vertices[j][1])
174             nsize.append(10)
175             ncol.append(col)
176     node_collection = ax.scatter(xpos,ypos,s=nsize,c=ncol)
177     edge_pos = np.asarray([(vertices[e[0]], vertices[e[1]])
          ↪ for e in np.transpose(np.nonzero(edgeMatrix)) if
          ↪ linalg.norm(np.subtract(vertices[e[0]],vertices[e
          ↪ [1]]))<=r_n])
178     edge_collection = LineCollection(
179         edge_pos,
180         colors='black',
```

```
181         linestyle='dotted',
182         linewidths=1,
183         antialiaseds=(1,)
184     )
185     edge_collection.set_zorder(1)  # edges go behind nodes
186     ax.add_collection(edge_collection)
187     ax.set_title(str(k)+' core in GIRG('+str(len(vertices))+',
        ↪ Pareto('+str(b)+'),'+str(rf)+')')
188     fig.show()
189     plt.savefig("multi comp 2.svg", format="svg")
```

Bibliography

[1] Jarle Aarstad, Håvard Ness, and Sven Haugland. “In what ways are small-world and scale-free networks interrelated?” In: Feb. 2013, pp. 1483–1487. ISBN: 978-1-4673-4567-5. DOI: 10.1109/ICIT.2013.6505891.

[2] Réka Albert, Hawoong Jeong, and Albert-László Barabási. “Diameter of the world-wide web”. In: *nature* 401.6749 (1999), pp. 130–131.

[3] George Anadiotis. *Graph shapes*. 2020. URL: https://pbs.twimg.com/media/EXfeTc9WoAA6b_u?format=jpg&name=medium (visited on 05/19/2022).

[4] N Araújo et al. “Recent advances and open challenges in percolation”. In: *The European Physical Journal Special Topics* 223.11 (2014), pp. 2307–2321.

[5] Michael E. Bales and Stephen B. Johnson. “Graph theoretic modeling of large-scale semantic networks”. In: *Journal of Biomedical Informatics* 39.4 (2006), pp. 451–464. ISSN: 1532-0464. DOI: https://doi.org/10.1016/j.jbi.2005.10.007. URL: https://www.sciencedirect.com/science/article/pii/S1532046405001103.

[6] Albert-László Barabási and Réka Albert. “Emergence of Scaling in Random Networks”. In: *Science* 286.5439 (1999), pp. 509–512. DOI: 10.1126/science.286.5439.509. eprint: https://www.science.org/doi/pdf/10.1126/science.286.5439.509. URL: https://www.science.org/doi/abs/10.1126/science.286.5439.509.

[7] Albert-László Barabási and Eric Bonabeau. “Scale-free networks”. In: *Scientific american* 288.5 (2003), pp. 60–69.

[8] A. Bekessy, P. Bekessy, and Janos Komlos. “Asymptotic enumeration of regular matrices”. In: *Studia Scientiarum Mathematicarum Hungarica* 7 (Jan. 1972).

[9] Edward A Bender and E.Rodney Canfield. “The asymptotic number of labeled graphs with given degree sequences”. In: *Journal of Combinatorial Theory, Series A* 24.3 (1978), pp. 296–307. ISSN: 0097-3165. DOI: https://doi.org/10.1016/0097-3165(78)90059-6. URL: https://www.sciencedirect.com/science/article/pii/0097316578900596.

[10] N. Berger et al. “Competition-Induced Preferential Attachment”. In: *Automata, Languages and Programming*. Ed. by Josep Díaz et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 208–221. ISBN: 978-3-540-27836-8.

[11] Berk Birand et al. “Analyzing the Performance of Greedy Maximal Scheduling via Local Pooling and Graph Theory”. In: *IEEE/ACM Transactions on Networking* 20.1 (2012), pp. 163–176. DOI: 10.1109/TNET.2011.2157831.

[12] S. Boccaletti et al. “Complex networks: Structure and dynamics”. In: *Physics Reports* 424.4 (2006), pp. 175–308. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2005.10.009. URL: https://www.sciencedirect.com/science/article/pii/S037015730500462X.

[13] Béla Bollobás. “A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs”. In: *European Journal of Combinatorics* 1.4 (1980), pp. 311–316. ISSN: 0195-6698. DOI: https://doi.org/10.1016/S0195-6698(80)80030-8. URL: https://www.sciencedirect.com/science/article/pii/S0195669880800308.

[14] Béla Bollobás. *Random Graphs*. 1st ed. Academic Press, 1985.

[15] Béla Bollobás. *Random Graphs*. 2nd ed. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001. DOI: 10.1017/CBO9780511814068.

[16] Béla Bollobás. “The evolution of random graphs”. In: *Transactions of the American Mathematical Society* 286.1 (1984), pp. 257–274.

[17] Béla Bollobás, Svante Janson, and Oliver Riordan. “The phase transition in inhomogeneous random graphs”. In: *Random Structures and Algorithms* 31.1 (2007), pp. 3–122. DOI: 10.1002/rsa.20168. URL: https://doi.org/10.1002%5C%2Frsa.20168.

[18] Karl Bringmann, Ralph Keusch, and Johannes Lengler. “Geometric inhomogeneous random graphs”. In: *Theoretical Computer Science* 760 (2019), pp. 35–54.

[19] Simon R Broadbent and John M Hammersley. “Percolation processes: I. Crystals and mazes”. In: *Mathematical proceedings of the Cambridge philosophical society*. Vol. 53. 3. Cambridge University Press. 1957, pp. 629–641.

[20] Anna D. Broido and Aaron Clauset. “Scale-free networks are rare”. In: *Nature Communications* 10.1 (Mar. 2019). DOI: 10.1038/s41467-019-08746-5. URL: https://doi.org/10.1038%5C%2Fs41467-019-08746-5.

[21] György Buzsáki. “Large-scale recording of neuronal ensembles”. In: *Nature neuroscience* 7.5 (2004), pp. 446–451.

[22] Roderick van Cann, Slinger Jansen, and Sjaak Brinkkemper. “Ortec”. In: *Software Business Start-up Memories*. Springer, 2013, pp. 112–120.

[23] Fan Chung and Linyuan Lu. “Connected components in random graphs with given expected degree sequences”. In: *Annals of combinatorics* 6.2 (2002), pp. 125–145.

[24] Reuven Cohen and Shlomo Havlin. *Complex networks: structure, robustness and function*. Cambridge university press, 2010.

[25] Paolo Crucitti et al. “Efficiency of scale-free networks: error and attack tolerance”. In: *Physica A: Statistical Mechanics and its Applications* 320 (2003), pp. 622–642.

[26] Maria Deijfen, Remco van der Hofstad, and Gerard Hooghiemstra. “Scale-free percolation”. In: (2011).

[27] F. M. Dekking. “Branching Processes That Grow Faster Than Binary Splitting”. In: *The American Mathematical Monthly* 98.8 (1991), pp. 728–731. ISSN: 00029890, 19300972. URL: http://www.jstor.org/stable/2324424 (visited on 06/23/2022).

[28] Jacques Demongeot, Mariem Jelassi, and Carla Taramasco. “From susceptibility to frailty in social networks: The case of obesity”. In: *Mathematical Population Studies* 24 (Oct. 2017), pp. 219–245. DOI: 10.1080/08898480.2017.1348718.

[29] E.W. Dijkstra. “A note on two problems in connexion with graphs”. In: *Numerische Mathematik* 1 (1959), pp. 269–271. ISSN: 0029-599X. DOI: 10.1007/BF01386390.

[30] *Disturbance Report for the Power System Outage that Occurred on the Western Interconnection on August 10th, 1996 at 1548 PAST*. Tech. rep. Western System Coordinating Counsil, 1996. URL: http://www.wscc.com.

[31] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. “Critical phenomena in complex networks”. In: *Rev. Mod. Phys.* 80 (4 Oct. 2008), pp. 1275–1335. DOI: 10.1103/RevModPhys.80.1275. URL: https://link.aps.org/doi/10.1103/RevModPhys.80.1275.

[32] S.N. Dorogovtsev et al. *Evolution of Networks: From Biological Nets to the Internet and WWW*. OUP Oxford, 2003. ISBN: 9780198515906. URL: https://books.google.nl/books?id=2isSDAAAQBAJ.

[33] Hugo Duminil-Copin. “Sixty years of percolation”. In: (2017). DOI: 10.48550/ARXIV.1712.04651. URL: https://arxiv.org/abs/1712.04651.

[34] Meyer Dwass. “The Total Progeny in a Branching Process and a Related Random Walk”. In: *Journal of Applied Probability* 6.3 (1969), pp. 682–686. ISSN: 00219002. URL: http://www.jstor.org/stable/3212112 (visited on 06/25/2022).

[35] P. Erdös and A. Rényi. “On Random Graphs I”. In: *Publicationes Mathematicae Debrecen* 6 (1959), p. 290.

[36] Paul Erdős, Alfréd Rényi, et al. “On the evolution of random graphs”. In: *Publ. Math. Inst. Hung. Acad. Sci* 5.1 (1960), pp. 17–60.

[37] Michalis Faloutsos, Petros Faloutsos, and Christos 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.

[38] Daniel Fernholz and Vijaya Ramachandran. “The giant k-core of a random graph with a specified degree sequence”. In: (2003).

[39] Ramon Ferrer i Cancho, Ricard V. Solé, and Reinhard Köhler. “Patterns in syntactic dependency networks”. In: *Phys. Rev. E* 69 (5 May 2004), p. 051915. DOI: 10.1103/PhysRevE.69.051915. URL: https://link.aps.org/doi/10.1103/PhysRevE.69.051915.

[40] E.N. Gilbert. “Random Plane Networks”. In: *Journal of the Society for Industrial and Applied Mathematics* 9 (1961), pp. 533–543. DOI: 10.1007/s00778-019-00587-4. URL: https://doi.org/10.1007/s00778-019-00587-4.

[41] Edgar N Gilbert. “Random graphs”. In: *The Annals of Mathematical Statistics* 30.4 (1959), pp. 1141–1144.

[42] M. Girvan and M. E. J. Newman. “Community structure in social and biological networks”. In: *Proceedings of the National Academy of Sciences* 99.12 (2002), pp. 7821–7826. DOI: 10.1073/pnas.122653799. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.122653799. URL: https://www.pnas.org/doi/abs/10.1073/pnas.122653799.

[43] Geoffrey Grimmett. “What is Percolation?” In: *Percolation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 1–31. ISBN: 978-3-662-03981-6. DOI: 10.1007/978-3-662-03981-6_1. URL: https://doi.org/10.1007/978-3-662-03981-6_1.

[44] Remco van der Hofstad and Julia Komjathy. “Explosion and distances in scale-free percolation”. In: *arXiv preprint arXiv:1706.02597* (2017).

[45] Remco van der Hofstad. “Configuration Model”. In: *Random Graphs and Complex Networks*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2016, pp. 216–255. DOI: 10.1017/9781316779422.010.

[46] van der Hofstad. *Random Graphs and Complex Networks*. Vol. 2. Apr. 2022. URL: https://www.win.tue.nl/~rhofstad/NotesRGCNII_colleagues_25_04_2022.pdf.

[47] Svante Janson and Malwina Luczak. *A simple solution to the k-core problem*. 2005. DOI: 10.48550/ARXIV.MATH/0508453. URL: https://arxiv.org/abs/math/0508453.

[48] Svante Janson, Andrzej Rucinski, and Tomasz Luczak. *Random graphs*. John Wiley & Sons, 2000.

[49] Yi-Xiu Kong et al. “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.

[50] Dmitri Krioukov et al. “Hyperbolic geometry of complex networks”. In: *Physical Review E* 82.3 (Sept. 2010). DOI: 10.1103/physreve.82.036106. URL: https://doi.org/10.1103%5C%2Fphysreve.82.036106.

[51] Tomasz Łuczak. “Size and connectivity of the k-core of a random graph”. In: *Discrete Mathematics* 91.1 (1991), pp. 61–68. ISSN: 0012-365X. DOI: https://doi.org/10.1016/0012-365X(91)90162-U. URL: https://www.sciencedirect.com/science/article/pii/0012365X9190162U.

[52] Benjamin F. Maier. “Generalization of the small-world effect on a model approaching the Erdős–Rényi random graph”. In: *Scientific Reports* 9.1 (June 2019). DOI: 10.1038/s41598-019-45576-3. URL: https://doi.org/10.1038%5C%2Fs41598-019-45576-3.

[53] Grzegorz Malewicz et al. “Pregel: A System for Large-Scale Graph Processing”. In: *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*. SIGMOD ’10. Indianapolis, Indiana, USA: Association for Computing Machinery, 2010, pp. 135–146. ISBN: 9781450300322. DOI: 10.1145/1807167.1807184. URL: https://doi.org/10.1145/1807167.1807184.

[54] F.D. Malliaros et al. “The core decomposition of networks: theory, algorithms and applications”. In: *The VLDB Journal* 29 (2020), pp. 61–92.

[55] Michael Molloy and Bruce Reed. “A critical point for random graphs with a given degree sequence”. In: *Random structures & algorithms* 6.2-3 (1995), pp. 161–180.

[56] Michael Molloy and Bruce Reed. “The size of the giant component of a random graph with a given degree sequence”. In: *Combinatorics, probability and computing* 7.3 (1998), pp. 295–305.

[57] Adilson E. Motter et al. “Topology of the conceptual network of language”. In: *Physical Review E* 65.6 (June 2002). DOI: 10.1103/physreve.65.065102. URL: https://doi.org/10.1103%5C%2Fphysreve.65.065102.

[58] M Nakasako. “Large-scale networks of hydration water molecules around proteins investigated by cryogenic X-ray crystallography”. In: *Cellular and molecular biology (Noisy-le-Grand, France)* 47.5 (July 2001), pp. 767–790. ISSN: 0145-5680. URL: http://europepmc.org/abstract/MED/11728092.

[59] M. E. J. Newman. “Resource Letter CS–1: Complex Systems”. In: *American Journal of Physics* 79.8 (Aug. 2011), pp. 800–810. DOI: 10.1119/1.3590372. URL: https://doi.org/10.1119%5C%2F1.3590372.

[60] M.E.J. Newman. “The Structure and Function of Complex Networks”. In: *Computer Physics Communications* 147 (Mar. 2003), pp. 40–45. DOI: 10.1016/S0010-4655(02)00201-1.

[61] M.E.J. Newman and D.J. Watts. “Renormalization group analysis of the small-world network model”. In: *Physics Letters A* 263.4-6 (Dec. 1999), pp. 341–346. DOI: 10.1016/s0375-9601(99)00757-4. URL: https://doi.org/10.1016%5C%2Fs0375-9601%5C%2899%5C%2900757-4.

[62] Ilkka Norros and Hannu Reittu. “On a conditionally Poissonian graph process”. In: *Advances in Applied Probability* 38.1 (2006), pp. 59–75. DOI: 10.1239/aap/1143936140.

[63] D.Q. Nykamp. *Generating networks with a desired degree distribution.* 2020. URL: https://mathinsight.org/generating_networks_desired_degree_distribution (visited on 05/19/2022).

[64] Stefano Pallottino and Maria Grazia Scutellà. “Shortest Path Algorithms In Transportation Models: Classical and Innovative Aspects”. In: *Equilibrium and Advanced Transportation Modelling*. Ed. by Patrice Marcotte and Sang Nguyen. Boston, MA: Springer US, 1998, pp. 245–281. ISBN: 978-1-4615-5757-9. DOI: 10.1007/978-1-4615-5757-9_11. URL: https://doi.org/10.1007/978-1-4615-5757-9_11.

[65] Marzieh Parandehgheibi and Eytan Modiano. “Robustness of interdependent networks: The case of communication networks and the power grid”. In: *2013 IEEE Global Communications Conference (GLOBECOM)*. IEEE. 2013, pp. 2164–2169.

[66] Romualdo Pastor-Satorras and Alessandro Vespignani. *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press, 2004. DOI: 10.1017/CBO9780511610905.

[67] Krzysztof Piasecki and Michał Dominik Stasiak. “The Forex Trading System for Speculation with Constant Magnitude of Unit Return”. In: *Mathematics* 7.7 (2019). ISSN: 2227-7390. DOI: 10.3390/math7070623. URL: https://www.mdpi.com/2227-7390/7/7/623.

[68] Boris G. Pittel, Joel H. Spencer, and Nicholas C. Wormald. “Sudden Emergence of a Giant k-Core in a Random Graph”. In: *J. Comb. Theory, Ser. B* 67 (1996), pp. 111–151.

[69] Sergio Porta, Paolo Crucitti, and Vito Latora. “The network analysis of urban streets: A dual approach”. In: *Physica A: Statistical Mechanics and its Applications* 369.2 (2006), pp. 853–866.

[70] Derek De Solla Price. “A general theory of bibliometric and other cumulative advantage processes”. In: *Journal of the American Society for Information Science* 27.5 (1976), pp. 292–306. DOI: https://doi.org/10.1002/asi.4630270505. eprint: https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.4630270505. URL: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.4630270505.

[71] Usha Nandini Raghavan, Réka Albert, and Soundar Kumara. “Near linear time algorithm to detect community structures in large-scale networks”. In: *Phys. Rev. E* 76 (3 Sept. 2007), p. 036106. DOI: 10.1103/PhysRevE.76.036106. URL: https://link.aps.org/doi/10.1103/PhysRevE.76.036106.

[72] Bo Söderberg. “General formalism for inhomogeneous random graphs”. In: *Physical Review E* 66.6 (Dec. 2002). DOI: 10.1103/physreve.66.066121. URL: https://doi.org/10.1103%5C%2Fphysreve.66.066121.

[73] H Francis Song and Xiao-Jing Wang. “Simple, distance-dependent formulation of the Watts-Strogatz model for directed and undirected small-world networks”. In: *Physical Review E* 90.6 (2014), p. 062801.

[74] Steven Strogatz. “Strogatz, S.H.: Exploring Complex Networks. Nature 410, 268”. In: *Nature* 410 (Apr. 2001), pp. 268–76. DOI: 10.1038/35065725.

[75] Jie Tang et al. “Social influence analysis in large-scale networks”. In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2009, pp. 807–816.

[76] Shalu Wadhwa. “Analysis of a network design problem”. In: (2000).

[77] Jianwei Wang et al. “Attack vulnerability of scale-free networks due to cascading failures”. In: *Physica A: Statistical Mechanics and its Applications* 387.26 (2008), pp. 6671–6678. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2008.08.037. URL: https://www.sciencedirect.com/science/article/pii/S0378437108007541.

[78] Xiangrong Wang et al. “A network approach for power grid robustness against cascading failures”. In: *2015 7th international workshop on reliable networks design and modeling (RNDM)*. IEEE. 2015, pp. 208–214.

[79] Duncan J Watts and Steven H Strogatz. “Collective dynamics of ‘small-world’networks”. In: *nature* 393.6684 (1998), pp. 440–442.

[80] E.W. Weisstein. *Circle-Circle Intersection.* 2017. URL: https://mathworld.wolfram.com/Circle-CircleIntersection.html (visited on 06/23/2022).

[81] Nicholas C. Wormald. “The asymptotic distribution of short cycles in random regular graphs”. In: *Journal of Combinatorial Theory, Series B* 31.2 (1981), pp. 168–182. ISSN: 0095-8956. DOI: https://doi.org/10.1016/S0095-8956(81)80022-6. URL: https://www.sciencedirect.com/science/article/pii/S0095895681800226.

[82] Jichang Zhao and Ke Xu. “Enhancing the robustness of scale-free networks”. In: *Journal of Physics A: Mathematical and Theoretical* 42.19 (2009), p. 195003.

[83] Liang Zhao, Kwangho Park, and Ying-Cheng Lai. “Attack vulnerability of scale-free networks due to cascading breakdown”. In: *Physical review E* 70.3 (2004), p. 035101.