Doctoral Thesis

Random Graph Models for Complex Systems

Author(s):
Peter, Ueli

Publication Date:
2014

Permanent Link:
https://doi.org/10.3929/ethz-a-010290422

Rights / License:
In Copyright - Non-Commercial Use Permitted
Random Graph Models for Complex Systems

A thesis submitted to attain the degree of
Dr. sc. ETH ZÜRICH

presented by
UELI PETER
Master of Science ETH in Computer Science
born 23.03.1982
citizen of Dinhard ZH

accepted on the recommendation of
Prof. Dr. Angelika Steger, examiner
Prof. Dr. Benjamin Doerr, co-examiner
Prof. Dr. Konstantinos Panagiotou, co-examiner

2014
# Contents

Abstract iii

Zusammenfassung vii

Acknowledgments xi

1 Introduction and Results 1

1.1 Our Results ........................................ 4
1.2 Organization of this Thesis .......................... 9

2 Notation 11

2.1 Asymptotic and Probabilistic Notation .......... 13
2.2 Probabilistic Tools .................................. 13

3 Background and Related Work 15

3.1 Properties of Large Graphs ...................... 15
3.2 Random Graph Models .............................. 20

4 Universality of Binomial Random Graphs 29

4.1 Results ............................................ 30
4.2 Graph-Theoretic Facts .............................. 31
4.3 Partitioning Lemmas ............................... 33
Abstract

Graphs are a fundamental tool to model large complex systems such as the brain, the internet, social networks, or the world economy. Simple generative random graph models are therefore important for many areas of research. In applied computer science they are used to measure the performance of algorithms for web mining and distributed computing. In theoretical computer science they are used as a powerful tool for the average case analysis of graph algorithms, and they can be helpful for the design of efficient algorithms which exploit the topology of real world networks. In other fields, such as economics, biology, statistical physics, and the social sciences, large graphs are essential for computer experiments. Moreover, suitable models are used by social scientists and economists to make predictions from measured data. In mathematics and theoretical computer science the extensive results on random graphs have brought new insights and methods, which contributed to a deeper understanding of stochastic processes.

The many areas of application make the requirements for a general random graph model hard to fulfill. A good model should capture the relevant features of the real-world system. The better these properties are captured the better are the predictions and the more meaningful are the results obtained from simulations and benchmarks. For a
proper evaluation of the accuracy of a random graph model the model needs to be analyzed rigorously. Simulations demonstrate the behavior of the model at a small scale, but, for an understanding of large instances in their whole parameter space, a rigorous analysis is unavoidable. Therefore, the model should also be simple enough that it is feasible to build a framework by studying the probability distribution of elementary events. For performance measurements and simulations it is necessary that algorithms exist which can quickly generate graphs according to the model’s distribution.

There has been a huge amount of work concerning the design and analysis of random graph models. For many applications it seems that the most successful models are those that are simple to state and analyze. It is probably for this reason that two of the best known models are the model of uniform random graphs by Erdős and Rényi, [ER60] and its close relative, the binomial random graph model by Gilbert [Gil59], which is often more accessible. In the universal random graph model every graph on \(n\) vertices and \(m\) edges is sampled with the same probability, while in the binomial random graph model a graph is constructed by drawing each possible edge uniformly at random with the same edge probability \(p\).

However, the structural properties of these abstract random graphs are very different from the properties that are observed in typical complex systems. It is therefore not very surprising that more sophisticated models have been suggested with the goal of capturing the salient features of complex systems. Unfortunately, they are either designed to reproduce one particular property and fail miserably for many others, or they are of such a high complexity that a rigorous analysis is infeasible.

After more than a decade of model design and evaluation, it seems that there is a fundamental trade-off between a models susceptibility for a mathematical analysis and its ability to reproduce the structure of real world networks. Naturally, the perfect model that succeeds in both parts might be just around the corner, but, since this scenario is rather unlikely, researchers are trying to push for progress in different directions. While a fair share of complex network researchers is working on the design of better, simpler, and more accurate models, statistical physicists apply their methods to the latest suggested models. Meanwhile, researchers from the algorithms and discrete mathematics community focus on the analysis of simple random graph models and on the development of tools that might be applicable to the more advanced models.
Abstract

In this thesis we follow this approach by presenting results for three
random graph models of very different nature and complexity.

The first main result in this thesis considers the model of binomial ran-
dom graphs. In this classical random graph model we search for the "universality threshold" of spanning subgraphs; that is the smallest edge
probability \( p \) for which a binomial random graph contains every graph
from a specific family as a subgraph. A result by Dellamonica, Koh-
hayakawa, Rödl and Rucińcki roughly states that a random graph on \( n \)
vertices with edge probability at least \( \Omega(\frac{n^{-1/\Delta}}{\Delta}) \) (omitting polylogarith-
mic terms) contains with high probability every spanning subgraph of
maximum degree at most \( \Delta \). We improve this result for a large class of
spanning subgraphs by showing that a random graph with edge prob-
ability at least \( \Omega(\frac{n^{-1/(2d)} \log^3 n}{d \log^3 n}) \) contains every subgraph of maximum
density \( d \). If the subgraphs additionally have no cycles of length at
most six, our result implies that they are already contained in a ran-
dom graph with edge probability \( \Omega(\frac{n^{-1/d}}{d \log^3 n}) \). Since the maximum
density of trees is smaller than two, this significantly improves a result
of Johannsen, Krivelevich and Samotij which states that a binomial ran-
dom graph with edge probability larger than \( \Omega(\frac{n^{-1/3} \log^2 n}{3 \log^2 n}) \) is universal
for the family of all spanning trees of constant maximum degree.

Next, we consider the behavior of the classical push broadcast process
on random graphs with a power-law degree sequence. Power-law degree
sequences have been observed in many real world networks. Therefore,
this model is more suitable as a model for complex systems. The goal
of the push process is to efficiently send a message from one initial ver-
tax to every vertex in the graph. It is a well known weakness of the
push process that it cannot efficiently (in polylogarithmic time) inform
all vertices of a power-law random graph. Our analysis shows that for
every \( \varepsilon > 0 \), with high probability, \( O(\log n) \) rounds are sufficient to
inform all but an \( \varepsilon \)-fraction of the vertices. This bound is best possi-
bile and it extends a result by Fountoulakis, Panagiotou and Sauerwald
who proved that for random graphs that have a power-law degree se-
quency with exponent \( \beta > 3 \) the more involved push-pull protocol needs
\( \Omega(\log n) \) rounds to inform all but \( \varepsilon n \) vertices with high probability. Our
result demonstrates that for power-law random graphs with exponent
\( \beta > 3 \) the addition of the pull mechanism does not (asymptotically) im-
prove the running time. This stands in contrast to power-law random
graphs with exponent \( 2 < \beta < 3 \) where push-pull is exponentially faster
than pull. While random graphs with a power-law degree sequence are
certainly the most interesting class of random graphs considered in this thesis, our results hold for a much larger class of sparse degree sequences.

Our last main result considers the recently suggested model of so-called random hyperbolic graphs. These random graphs have startling similarities with several real-world networks, in particular with the internet graph (the network formed by the routers and their physical connections). We analyze the degree sequence and the clustering coefficient of hyperbolic random graphs. In particular, we show that the degree sequence follows a power-law and that the clustering coefficient, that is the graph’s average of the probability that two neighbors of a vertex are connected, is of the same order as observed in real world systems. Furthermore, we study the compressibility of such random graphs. Compressibility is strongly related to the design of data structures that efficiently store and access large graphs, which is a fundamental topic in computer science. In a series of papers Boldi, Santini and Vigna provided efficient data structures that are able to store the internet graph using only a constant number of bits per edge. The classical random graph models are not able to reproduce this property since they all have a large entropy and therefore require at least \( \Omega(\log n) \) bits per edge (where \( n \) denotes the number of vertices in the graph). We analyze the performance of a basic compression scheme on random hyperbolic graphs and prove that it is able to store these graphs such that the number of bits per edge is constant with high probability.
Zusammenfassung


Die jeweiligen Anwendungsgebiete stellen unterschiedliche Anforderun-
Zusammenfassung


Im ersten Teil der Arbeit betrachten wir das klassische Gilbert Modell. Für eine gegebene Menge von Graphen $F$ bezeichnen wir einen Graphen genau dann als universell für $F$, wenn er jeden Graphen aus $F$ als Teilgraphen enthält. Wir widmen uns folgender Frage: „Für welche Kantendichte $p$ ist ein Zufallsgraph universell für eine gegebene Menge $F$?“. Dellamonica, Kohayakawa, Rödl und Rucińcki zeigten, dass ein Zufallsgraph universell für die Menge aller Graphen mit Maximumgrad höchstens $\Delta$ ist, solange $p = \Omega(n^{-1/\Delta})$ (wobei wir logarithmische Faktoren vernachlässigen). Wir zeigen, dass die Menge aller Graphen mit Dichte höchstens $d$ für $p = \omega(n^{-1/(2d)} \log^3 n)$ im Zufallsgraphen enthalten ist. Weiter folgt aus unserem Beweis, dass der Zufallsgraph bereits für $p = \omega(n^{-1/d} \log^3 n)$ universell für die Menge der Teilgraphen ist, welche zusätzlich keine kurzen Kreise enthalten. Da die Dichte von Bäumen kleiner als zwei ist, verbessern wir damit die beste bekannte Schranke von Johannsen, Krivelevich und Samotij für die Kantendichte, über welcher der Zufallsgraph universell für die Menge aller Spannbäume mit konstantem Maximumgrad ist.


Im letzten Kapitel untersuchen wir das erst kürzlich vorgeschlagene Modell der hyperbolischen Zufallsgraphen. Diese Zufallsgraphen haben verblüffend ähnliche Eigenschaften wie die aus Daten sozialer Netzwerke und dem World Wide Web gewonnenen Graphen. Zuerst bewei-
First I want to thank my advisor Angelika Steger for giving me the opportunity to write a thesis in here group. You gave me many valuable advices over the years. I hope that all my future bosses will be as supportive as you!

Next I want to thank Konstantinos Panagiotou. Not only for co-refering my thesis but also for guiding my first steps in research and writing, for inviting me to Saarbrücken and München, and for sharing his tabaco in Buchboden. It was always fun to work with you!

I also thank Benjamin Doerr for co-refering my thesis. I hope you did not suffer to much.

Furthermore, I thank Asaf Ferber for sharing his enthusiasm about random graphs. You brought the fun back into research. I wish you had appeared earlier. At this occasion I also want to thank my last two co-workers Rajko Nenadov and Nemanja Škorić: the spanners are the best!

Moreover I want to thank all my co-authors. Karl Bringmann, Asaf Ferber, Luca Gugelmann, Tomas Hruz, Fabian Kuhn, Ralph Keusch, Florian Meier, Torsten Mütze, Rajko Nenadov, Andreas Noever, Konstantinos Panagiotou, Nemanja Škorić, and Henning Thomas: I learned
from everyone of you!

The atmosphere in our group, especially at the many joint lunches on the roof, during the coffee breaks, and at our annual workshop in Buchboden was always very nice because of all the great members and associates of the CSA group: Reto, Dan, Kosta, Fabian, Crisse, Flo, Torsten, Thomas, Luca, Henning, Johannes, Rajko, Haffi, Frank, Andreas, Asaf, Ralph, Felix, and Nemanja. Thank you all!

I also want to thank the many friends I have outside academia. The balance they provided is very important for me. In particular, I want to thank Thomas, Jann, and Gäbe for proofreading parts of my thesis. Last but not least, I want to thank my family for supporting me through all the years. Mama, Papa, Christa, and Jürg, without you I would not be the man I am today!
Introduction and Results

Understanding large complex systems such as the world wide web, the world economy, or the human brain is not only important in science, but it strongly affects our every day life. A simple way of describing binary relationships among entities in a system is to draw a line between two entities if and only if they are related. In the language of graph theory, a branch of mathematics that dates back to Leonard Euler, entities are called vertices and the lines that connect them are called edges. It turns out that graphs are, despite their simple graphical representation, interesting objects and the study of their combinatorial structure is a fundamental topic in discrete mathematics. Moreover, graphs are used in almost every field of science to model and analyze relationships in complex systems.

The second ingredient for the design of models for complex systems is
randomness. It is merely a philosophical question whether randomness occurs in the dynamics of complex systems. But the interactions of the various entities are of such a high complexity that it can hardly be distinguished from randomness. This makes randomness a powerful tool for the design of random graph models for complex systems. A random graph model (also referred to as random graph process) is a formal description of how random experiments (e.g., coin tosses) can be used to create an instance from a class of graphs. This description can be a probability distribution from which graphs are sampled, or it can be a more sophisticated algorithmic description of how random experiments are used to construct a graph. Such a randomized algorithm then implicitly defines a probability distribution over all graphs.

Random graph models that replicate the structure and dynamics of real world systems have an impact on various fields of research as they have the potential to provide several benefits. Firstly, they can give valuable insight into the dynamics of the system. In particular, they might help to define quantities for the analysis of the system, and for predictions of its future states. For this purpose, a heavy machinery of tools, the so called network analysis method, has been developed (see [Was94]). Secondly, in computer science, the knowledge we gain from studying random graph models might allow us to exploit structural properties of complex systems. This can lead to the design of more efficient algorithms for problems such as broadcasting, routing, and data mining. Last but not least, random graph models can provide algorithms to efficiently create instances of graphs that behave similar to graphs obtained from real world data. Those instances can be used as benchmarks for the design of algorithms as well as to verify models of dynamic processes such as the spreading of viruses in social networks.

The design of a suitable random graph model is an important first step in complex network research. But in order to profit from the previously mentioned benefits a rigorous analysis of the model is substantial to compare the model’s features to real world systems. Computer simulations might confirm the similarity of small instances, but, for a deeper understanding, the mathematical study of the random graph model is unavoidable. The true challenge in the design of random graph models therefore not only lies in finding a model that captures the critical properties, but in finding one that is also mathematically tractable in the sense that it is simple enough to be analyzed by known methods.

Surprisingly, it turns out that most of the graphs attained from real
world data have very similar structural properties. The web graph, which models the link structure between websites, is certainly one of the best studied large graphs since the understanding of its structure is of particular importance for the design of efficient algorithms in computer science. During the last twenty years, many properties and quantities of the web graph have been empirically studied. The most basic quantity for a vertex in a graph is its degree, which counts the number of edges that are adjacent to the vertex. In 1999, Faloutsos et al. [FFF99] observed that the degree sequence of the web graph follows a power-law, i.e. that the proportion of vertices with degree $k$ is of order $k^{-\delta}$ for some fixed exponent $2 \leq \delta \leq 3$. During the following decade, power-laws have been observed in many large real world networks. Another important feature of the web graph is clustering. The clustering coefficient measures the network’s average of the probability that two neighbors of a vertex are connected. It has been observed that the web graph, social networks and biological networks have a surprisingly large clustering coefficient [RB03].

Many random graph models have been proposed and some of them capture the aforementioned properties quite well. Unfortunately, it seems that we observe the emergence of a trade-off between the ability of a model to replicate these properties and its mathematical tractability. There are models which are simple enough for their structure to be analyzed using methods known from probability theory. In fact, simple random graph models became very popular in mathematics with books, journals, and conferences devoted only to the study of their combinatorial structure. However, because of their simplicity those models fail to reproduce the salient features of real world systems such as a power-law degree sequence and a strong clustering. On the other hand, the analysis of the more realistic models that are better in reproducing these features is often infeasible with today’s methods. For many models not even the degree sequence can be formally analyzed. It is because of this trade-off that researchers started to work on different models simultaneously. Some study simple models and develop tools that might be applicable to more complex models, while others attack the more sophisticated models.
Chapter 1. Introduction and Results

1.1 Our Results

In this thesis we start with the very simple model of binomial random graphs for which we study the containment of spanning subgraphs. Then, we increase the complexity of the model slightly by considering random graphs with a given degree sequence on which we study the performance of a rumor spreading process. Finally, we attack a model that is much more sophisticated, and prove that its basic properties, such as degree sequence and clustering, are a good fit for real world systems.

1.1.1 Universality of Binomial Random Graphs

Perhaps the simplest way of creating a random graph with $n$ vertices and $m$ edges is to choose an element uniformly at random from the set of all such graphs. This model of uniform random graphs was introduced by Paul Erdős and Alfréd Rényi in 1959 [ER59]. Edgar Gilbert suggested in the same year the binomial random graph model $G(n,p)$ [Gil59], in which a graph on $n$ vertices is created by choosing each of the $\binom{n}{2}$ possible edges independently at random with a fixed edge probability $p := p(n)$. The edge probability $p$ and the edge ratio $m/\binom{n}{2}$ both control the density of the random graph, and it is therefore not very hard to see that, for many natural properties, the two models are equivalent if $p = m/\binom{n}{2}$ and $n$ is large enough. The first question when studying a property of binomial random graphs is to determine for which values of $p$ the graph typically has this property. A property is monotone increasing if the addition of edges in a graph can not destroy it. One particularly nice phenomena of binomial random graphs is that there exists, for many properties $Q$, a so called threshold function $\bar{p} := \bar{p}(n)$ such that

$$\lim_{n \to \infty} \Pr[G(n,p) \in Q] = \begin{cases} 0 & \text{if } p \ll \bar{p}, \\ 1 & \text{if } p \gg \bar{p}. \end{cases}$$

Bollobás and Thomason [BT87] proved the existence of a threshold for every monotone increasing property. In the last fifty years, binomial random graphs have been intensively studied. The thresholds for many interesting properties have been determined and several books have been published on that topic. It is therefore fair to say that binomial random graphs are not only an important tool in combinatorics, but are its own branch of discrete mathematics with conferences and journals dedicated to the matter. We discuss some known results that are of importance.
Part of the richness of binomial random graphs is that, even though they have been studied for more than five decades, there are still many interesting open problems. This constantly requires the development of new methods and techniques. One of these problems is the containment of specific graphs in binomial random graphs, and in particular the existence of so called spanning subgraphs which contain all the vertices of the random graph. We say that a graph $G$ is universal for a family of graphs $\mathcal{F}$, if $G$ contains a copy of every graph $F \in \mathcal{F}$. Before we proceed let us introduce two large families of graphs for which we will study the universality threshold. Let $\mathcal{H}(n, \Delta, d)$ be the family of all graphs on $n$ vertices with maximum degree at most $\Delta$ and with density at most $d$, where the density of a graph $G$ (denoted by $d(G)$) is defined as

$$d(G) := \max \left\{ \frac{2|E(H)|}{|V(H)|} : H \subseteq G \right\}.$$ 

Furthermore, let $\mathcal{T}(n, \Delta)$ be the family of all forests on $n$ vertices with maximum degree at most $\Delta$.

It is well known that binomial random graphs are disconnected for $p = o(\log/n)$. The threshold for being connected clearly provides a lower bound on the threshold for containing any spanning tree. In fact, it is widely believed that the threshold for $\mathcal{T}(n, \Delta)$-universality is $\Theta(\log n/n)$ for every fixed $\Delta$. The best previously known upper bound for universality of bounded degree spanning trees, due to Johannsen, Krivelevich, and Samotij, was $O(\Delta^8 n^{-1/3} \log^2 n)$ [JKS13]. For the family $\mathcal{H}(n, \Delta, d)$ we discuss in Section [3] that $p$ must be at least $\Omega(n^{-\frac{2}{\Delta+1}})$ to be $\mathcal{H}(n, \Delta, d)$-universal. The best known upper bound for this class of graphs is $O(n^{-1/\Delta} \log^{1/\Delta} n)$ for $\Delta > 2$. This upper bound was obtained by Dellamonica, Kohayakawa, Rödl, and Rucińcki in [DJKRR12].

Thus, there remain large gaps between upper and lower bounds for general graphs and trees.

In this thesis we prove an upper bound of $p = O(n^{-1/(2d)} \log^3 n)$ on the $\mathcal{H}(n, \Delta, d)$-universality threshold. This improves the upper bound in [DJKRR12] if $d < \Delta/2$. If we consider only the subfamily of $\mathcal{H}(n, \Delta, d)$, that consists of graphs without short cycles, we can show that these graphs are already contained in the random graph for $p = O(n^{-1/d} \log^3 n)$. Since the density of a tree is smaller than two and since trees are cycle free, our result improves the result of Johannsen, Krivelevich, and Samotij to $p = O(n^{-1/2} \log^3 n)$. 

for this thesis in Chapter [3]
The complexity of the subgraph containment problem can be increased by randomly coloring the edges of the random graph, and by asking for the containment of a specifically colored subgraph. To that matter, let $c$ be a positive integer and let $G_c(n, p)$ denote the random graph model in which every edge of a $G(n, p)$ is colored uniformly at random with a color from $[c]$. For a given graph $H$ we say that an edge colored graph $G$ contains a rainbow copy of $H$ if $G$ contains a copy of $H$ with all the edges colored in distinct colors. If we fix a graph $H$ and an integer $c$, such that $c$ is at least the number of edges in $H$, we can ask for the threshold of $G_c(n, p)$ for containing a rainbow copy of $H$. Thus, the problem of finding the threshold has two dimensions: we can change the probability $p$ and the number of colors $c$. In [FL14], Frieze and Loh showed that for every $\varepsilon > 0$, every $p \geq (1 + \varepsilon) \log n/n$, and for $c = n + o(n)$ a typical member of $G_c(n, p)$ contains a rainbow Hamilton cycle (a cycle that passes through all the vertices of the graph). This is almost optimal in both parameters since the lower bound on $p$ is essentially necessary for $G(n, p)$ to contain a Hamilton cycle, and since at least $n$ colors are necessary for a Hamilton cycle to be rainbow. Recently, Bal and Frieze [BF13] managed to improve this further to exactly $n$ colors but a required density of $C \log n/n$ for a large constant $C$.

We prove that for any $H \in \mathcal{H}(n, \Delta, d)$ and any $\alpha > 0$ the threshold for containing a rainbow copy of $H$ is at most $n^{-1/d} \log^{5/d} n$, given that $c \geq (1 + \alpha)|E(H)|$. It seems that the bounds on $c$ and on $p$ in our result are far from optimal for any fixed graph. On the other hand, the strength of our result is that it holds for a very large class of graphs.

All results on binomial random graphs are joint work with Asaf Ferber and Rajko Nenadov.

1.1.2 Rumor Spreading in Sparse Random Graphs

Binomial random graphs are a simple and powerful tool for combinatorics. Unfortunately, their degree sequence follows a binomial distribution, and the degree of every vertex is therefore close to the average degree of the graph. This is very different from the power-law degree sequences observed in real world networks where vertices of very large degree exist.

A reasonable approach to overcome this weakness of the model is to fix a degree sequence and to choose uniformly at random from all the graphs
that have this degree sequence. Note that requiring a certain degree sequence is a stronger restriction than requiring a certain number of edges (as it is done in the uniform random graph model) and the model of random graphs with a given degree sequence is therefore a refinement of the uniform random graph model. The study of random graphs with a given degree sequence dates back to [Bol80] and [BC78], but the focus at that time was mainly on random regular graphs. During the first decade of the twentieth century, researchers started to consider random graphs with a fixed power-law distributed degree sequence (see Section 3 for a discussion of the corresponding results).

The study of so-called rumor spreading processes is popular in the randomized algorithms and the distributed systems community. Methods and results from this area have many applications ranging from the design of efficient broadcast algorithms to the development of vaccination strategies, that prevent the spreading of diseases. The often considered push model and its symmetric equivalent pull are important building blocks of such rumor spreading processes. In the push model every informed vertex chooses in every round one of its neighbors uniformly at random and sends the rumor to the chosen neighbor. In the pull model, conversely, every non-informed vertex chooses in every round one of its neighbors uniformly at random and receives the rumor if the chosen neighbor is informed at the time. The combination of push and pull is called push-pull and it is known to be very efficient on many graph classes.

The behavior of push on binomial random graphs has been analyzed by Fountoulakis, Huber and Panagiotou in [FHP10], where they showed that, on a sufficiently dense random graph on \( n \) vertices, push typically succeeds in informing all vertices in \( O(\log n) \) rounds. Random graphs with power-law degree sequences usually contain many vertices of small degree that are only connected to vertices of very large degree (i.e. of degree \( n^{\Omega(1)} \)). It follows that the push protocol needs with high probability \( n^{\Omega(1)} \) rounds to inform all these vertices. For the push-pull protocol, Fountoulakis, Panagiotou and Sauerwald [FPS12] showed that the time needed to spread the information to at least an \((1 - \epsilon)\)-fraction of the vertices of a random power-law graph\(^1\) with exponent \( 2 < \beta < 3 \) is with high probability of order \( O(\log \log n) \), while if \( \beta > 3 \) it is with high probability \( \Omega(\log n) \). We follow this approach of considering the

\(^1\)They use the ACL model [CL04], which is slightly different than choosing from the uniform distribution over all graphs with a fixed degree sequence.
running time until the push algorithm informs at least a \((1 - \varepsilon)\)-fraction of the vertices of a graph. In particular, we show that on power-law random multigraphs the push model suffices to inform almost surely a \((1 - \varepsilon)\)-fraction of the vertices (for every \(\varepsilon > 0\)) in logarithmic time. Our result holds for a much larger class of sparse random multigraphs that contains random regular graphs, random bounded degree graphs, and many more, and generalizes for most of those classes (unfortunately not for power-law random graphs with \(\beta < 3\)) to the corresponding class of random (simple) graphs. Note that for push the number of informed vertices can at most double with every round. Thus, \(\Omega(\log n)\) is a lower bound for every graph, and our result therefore implies that, while for power-law random graphs with exponent \(2 < \beta < 3\) push-pull is exponentially faster than push (at least in the ACL model, see [FPS12]), for \(\beta > 3\), the pull mechanism does not lead to an (asymptotical) improvement.

This is joint work with Florian Meier.

1.1.3 Random Hyperbolic Graphs: Degree Sequence, Clustering and Compressibility

The model of random hyperbolic graphs was recently introduced by Papadopoulos, Krioukov, Boguñá and Vahdat [PKBnV10]. Papadopoulos et al. argued empirically and by some preliminary mathematical analysis that the resulting graphs have many of the desired properties for models of large real-world graphs, such as a high clustering coefficient and a degree sequence that follows a power-law. By computing explicitly a maximum likelihood fit of the Internet graph they demonstrated impressively that this model is adequate for reproducing the structure of the Internet with high accuracy.

Considering the experiments just described, it seems at least fair to say that random hyperbolic graphs provide an attractive model that has a high potential of being suitable for describing the characteristics of real-world networks. Moreover, a simple formulation and a strong affinity to random geometric graphs indicate that this model might be mathematically tractable. In this thesis we prove that this is indeed the case. First, we show that the expected degree sequence indeed follows a power-law across all scales, i.e. even up to the maximum degree. We then prove a constant lower bound on the clustering coefficient of hyperbolic random graphs which confirms the claimed high clustering. Note that in
the seminal papers [PKBnV10] and [KPK+10] the degree distribution was also considered, however, only for constant degrees and without any error guarantees. In addition, we prove small probabilities for large deviations, i.e. we show that sampling from this distribution returns with high probability a graph with the desired properties, which is crucial for validating experimental results. We also compute tight bounds for the average and maximum degree that hold with high probability.

Finally, we study the compressibility of random hyperbolic graphs. It has been discovered that the web graph can be stored by using only an average of three bits per link. The classical models for the web graph fail in reproducing this property as their entropy is so large that every data structure needs at least $\Omega(\log n)$ bit per link. We analyze the performance of a basic compression scheme on random hyperbolic graphs and prove that it is able to store them using only a constant number of bits per edge in expectation and with high probability.

All these results except the ones on compressibility are joint work with Luca Gugelmann, Konstantinos Panagiotou, and Ralph Keusch. The result on compressibility was obtained with Karl Bringmann and Konstantinos Panagiotou.

1.2 Organization of this Thesis

In Chapter 2 we describe the notation that we use in this thesis, and explain the necessary tools from probability theory. A brief discussion of the history of the different random graph models as well as the considered properties follows in Chapter 3. Chapter 4 is devoted to the proof of our bounds on the universality threshold of binomial random graphs. Then, in Chapter 5 we analyze the behavior of push on sparse random graphs with a given degree distribution. Finally, in Chapter 6 we prove all our results on random hyperbolic graphs.
In this chapter we introduce the basic notation that will be used throughout this thesis.

**Basic Notation**

Our graph-theoretic notation is standard and follows that of [Wes01]. We usually denote a graph or multigraph by $G$ or $M$ and we use the following notation for a graph $G$, sets of vertices $U$ and $W$, a vertex $v$, and an integer $k$.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V(G)$</td>
<td>vertex-set of $G$</td>
</tr>
<tr>
<td>$E(G)$</td>
<td>edge-set of $G$</td>
</tr>
<tr>
<td>$v(G)$</td>
<td>number of vertices of $G$, $v(G) :=</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>$e(G)$</td>
<td>number of edges of $G$, $e(G) :=</td>
</tr>
</tbody>
</table>
| $N(v), N_G(v)$  | neighborhood of a vertex $v \in V(G)$  
$N_G(v) := \{u \in V(G) : \{u, v\} \in E(G)\}$                                                                                       |
| $\text{deg}(v), \text{deg}_G(v)$ | degree of a vertex $v$ in $G$, $\text{deg}_G(v) := |N_G(v)|$  
maximum degree of $G$, $\Delta(G) := \max_{v \in V(G)} \text{deg}_G(v)$                                                             |
| $\bar{k}(G)$   | the average degree of $G$, $\bar{k}(G) := 2e(G)/v(G)$                                                                                     |
| $d(G)$          | the maximum density of $G$  
$d(G) := \max \left\{ \frac{2e(H)}{v(H)} : H \subseteq G \right\}$                                                                         |
| $E(U), E_G(U)$  | all the edges of $G$ with both endpoints in $U$  
$E_G(U) := \{\{u, w\} \in E(G) : u, v \in U\}$                                                                                      |
| $e(U), e_G(U)$  | number of edges in $E(U)$, $e(U) := |E(U)|$                                                                                             |
| $E(U,W), E_G(U,W)$ | all the edges of $G$ with one endpoint in $U$ and one in $W$  
$E_G(U,W) := \{\{u, w\} \in E(G) : u \in U, w \in W\}$                                                                 |
| $E(v,U), E_G(v,U)$ | all the edges of $G$ with one endpoint being $v$ and one endpoint in $U$  
$E_G(v,U) := \{\{v, u\} \in E(G) : u \in U\}$                                                                                  |
| $e(v,U), e_G(v,U)$ | number of edges in $E(v,U)$, $e(v,U) := |E(v,U)|$                                                                                       |
| $D_k(G)$        | the set of vertices of degree exactly $k$  
$D_k(G) := \{v \in V(G) : \text{deg}_G(v) = k\}$                                                                                   |
| $D_{\geq k}(G), D_{\leq k}(G)$ | the number of vertices of degree at least or at most $k$  
$D_{\geq k}(G) := \{v \in V(G) : \text{deg}_G(v) \geq k\}$  
$D_{\leq k}(G) := \{v \in V(G) : \text{deg}_G(v) \leq k\}$                                                                 |
| $D_{> k}(G), D_{< k}(G)$ | the number of vertices of degree strictly larger or smaller than $k$  
$D_{> k}(G) := \{v \in V(G) : \text{deg}_G(v) > k\}$  
$D_{< k}(G) := \{v \in V(G) : \text{deg}_G(v) < k\}$                                                                 |

We omit the graph $G$ whenever it is clear from the context what graph we mean, and write e.g. $D_k$ instead of $D_k(G)$.

We use $\log n$ for the base two logarithm and $\ln n$ for the natural logarithm.
2.1 Asymptotic and Probabilistic Notation

We use the Landau symbols $O$, $\Omega$, $\Theta$, $o$, and $\omega$ to describe the asymptotic behavior of functions. We write $a \ll b$ (respectively $a \gg b$) to express that $a = o(b)$ (respectively $a = \omega(a)$).

We abbreviate with high probability by $w.h.p$ and use it whenever an event holds with probability $1 - o(1)$. Sometimes, we write $u.a.r.$ for uniformly at random and we use the notation $e \in_r U$ when we draw $e$ uniformly at random from a set $U$.

2.2 Probabilistic Tools

In our proofs we apply at several occasions the well known Chernoff-Hoeffding bound (see e.g. Chapter 1 of [DP09]) for sums of independent Bernoulli random variables. Sometimes we need these bounds in the following slightly dependent setting. The proof of the following theorem follows directly from the Chernoff bound (see for example Problem 1.7 in [DP09]).

**Theorem 2.1.** Let $p, q \in [0, 1]$ and let $X_1, \ldots, X_n \in \{0, 1\}$ be $n$ indicator variables and $X := \sum_{i=1}^n X_i$. If for each $1 \leq i \leq n$

$$E[X_i|X_1, \ldots, X_{i-1}] \geq p \quad \text{and} \quad E[X_i|X_1, \ldots, X_{i-1}] \leq q,$$

then it holds for every $0 < \varepsilon < 1$ that

$$\Pr[X \geq (1 + \varepsilon) nq] \leq e^{-nq \varepsilon^2 / 3} \quad \text{and} \quad \Pr[X \leq (1 - \varepsilon) np] \leq e^{-np \varepsilon^2 / 2}.$$

We will use the following large deviation inequality. Let $f$ be a function on the random variables $X_1, \ldots, X_n$ that take values in some set $A_i$. We say that $f$ is Lipschitz with coefficients $c_1 \ldots c_n$ and bad event $B$ if for all $x, y \in A_i$

$$|E[f|X_1, \ldots, X_{i-1}, X_i = x, B] - E[f|X_1, \ldots, X_{i-1}, X_i = y, \overline{B}]| \leq c_i.$$

(We denote by $\overline{B}$ the complement of $B$.) Then the following estimates are true.

**Theorem 2.2** (Theorem 7.1 in [DP09]). Let $f$ be a function of $n$ independent random variables $X_1, \ldots, X_n$, each $X_i$ taking values in a set...
$A_i$, such that $E[f]$ is bounded. Assume that
\[
m \leq f(X_1, \ldots, X_n) \leq M.
\]

Let $B$ any event, and let $c_i$ be the maximum effect of $f$ assuming the complement $\overline{B}$ of $B$:
\[
\max_{x, y \in A_i} |E[f \mid X_1, \ldots, X_{i-1}, X_i = x, \overline{B}] - E[f \mid X_1, \ldots, X_{i-1}, X_i = y, \overline{B}]| \leq c_i.
\]

Then
\[
Pr[f > E[f] + t + (M - m) Pr[B]] \leq e^{-2t^2/\sum i c_i^2} + Pr[B]
\]
and
\[
Pr[f < E[f] - t - (M - m) Pr[B]] \leq e^{-t^2/\sum i c_i^2} + Pr[B].
\]
Chapter 3

Background and Related Work

In this thesis we consider binomial random graphs, random graphs with a given degree sequence, and random hyperbolic graphs. For each of the three models, we analyze a different set of properties. In this chapter we first describe in Section 3.1 all the properties that are of importance for this thesis. Then, in Section 3.2 we give a short summary of the background and related work to the considered properties and the three random graph models.

3.1 Properties of Large Graphs

The common approach to complex network research is as follows. In a first step, empirical researchers measure data in the real world and identify the system’s characteristic properties. The goal of the second
The elementary quantities of a graph \( G = (V, E) \) are the number of vertices \(|V|\) and the number of edges \(|E|\). However, the study of more sophisticated parameters has a long history, which we discuss in the following paragraphs. This summary is by no means complete, it only reflects the background for this thesis. We refer the reader to [New03] and [DM02] for a more complete summary on the matter. Moreover, we recommend [BF02] for an entertaining book on the research history of complex networks.

Since the 60’s, the study of networks of various kinds has grown into a significant research area. One of the initiators in this field, the sociologist Stanley Milgram, investigated the network that is obtained from the relationships among people [Mil67, TM69]. In his work he discovered what is nowadays known as the small-world phenomenon, which postulates that the distance between two random people is on average between five and six. Outside the context of social networks this has become synonymous to a graph with a comparatively low diameter, and nowadays many networks are known to possess this property [AJB99, ASBS00, WS98].

It was further observed that a person can only maintain a limited number of contacts and that social networks are therefore sparse in the sense that the average degree

\[
\bar{k} := \frac{2|E|}{|V|}
\]

is bounded by a constant. The number of edges is therefore linear in the number of vertices, i.e.

\[|E| = O(|V|).\]

This feature can be observed in many complex networks and sparsity is therefore a general assumption for such graphs.

Another property that is found in many networks addresses the degree
sequence, i.e. the sequence $D_k$ for $0 \leq k \leq |V| - 1$, where

$$D_k := |\{v \in V | \deg(v) = k\}|.$$

In a celebrated paper Faloutsos et. al. [FFF99] studied the degree sequence of the web graph. They observed that the web graph exhibits a so-called scale-free nature: the degree sequence follows approximately a power-law distribution, which means that

$$D_k/|V| \approx k^{-\delta},$$

for a constant power-law exponent $\delta$. This sets such a network dramatically apart from e.g. a typical uniform or binomial random graph, and stirred significant interest in exploring the causes of this phenomenon. From today’s viewpoint, it is well-known that many graphs have a heavy-tailed degree distribution, which may be close to a power-law, a log-normal, or a combination of these distributions (see [Mit04] and references therein).

A third distinctive feature of large real-world graphs is the appearance of clustering [NP03,SBn06,WS98]. The network’s average of the probability that two neighbors of a random vertex are also directly connected is called the clustering coefficient. Measured clustering coefficients of social networks are typically tens of percent, and similar values have been measured for many other networks including technological and biological ones.

The high values observed in clustering coefficients of complex networks already suggests that their local structure is not purely random, since a sparse random graph has a clustering coefficient of order

$$O\left(\frac{|E|}{(|V|^2)}\right) = O(1/|V|),$$

if the probability for two vertices to be connected does not depend on whether they share a common neighbor. The clustering coefficient, as well as all the previously mentioned graph quantities, describes only the local structure around vertices in a graph. An interesting global property is the compressibility of a graph. Since the web graph is very large (the indexed web contains at least 1.7 billion pages), the question of how efficient it can be stored is not only of theoretical interest but also of practical importance. In a series of papers [BV04,BV05,BSV09], Boldi, Santini and Vigna demonstrated that the web graph can be stored using
an average of only three bits per edge. The main idea of their compression scheme is to use the lexicographical ordering of the website’s urls and to encode for every edge only the distance between the two web sites in this ordering. This approach is very efficient since most of the links are between web sites that are close in the lexicographical ordering.

Shannon’s source coding theorem (see [Mac03] for a formal statement) implies that the entropy of a random graph model gives a lower bound on the expected length of the encoding that a compression scheme generates. The entropy of a random variable measures the amount of randomness that is involved in determining the random variable’s value. A fair coin toss has entropy exactly 1, and a general random variable $X$, on a set of values $\Omega$, has entropy

$$H(X) = \sum_{x \in \Omega} \Pr[X = x] \log_2 \Pr[X = x]. \quad (3.1)$$

The observation of Boldi, Santini and Vigna thus suggests that a model for the web graph should have an entropy of $\Theta(|V|)$, which corresponds to only $\Theta(|V|)$ fair coin tosses. This suggests that the link structure of the web graph is not very random at all, as there are $\Theta(|V|^2)$ potential links.

### 3.1.1 Rumor Spreading

The behavior of stochastic processes on a graph is another interesting global property. A class of stochastic processes that has been intensively studied on various graph classes are rumor spreading processes, which emulate the distribution of an information in a graph. A particularly prominent rumor spreading process is the so called push process. In the push process (or protocol/model) on a graph $G$ with $n$ vertices, initially (in round 0) only one arbitrary vertex knows the rumor. In each round $i \geq 1$, every vertex that knows the rumor informs uniformly at random one of its neighbors. The push process has a counter part called pull process, where in every round, every uninformed vertex asks a random neighbor for the rumor. Moreover, the straight forward combination of push and pull is called push-pull. Both push and pull where introduced in [DGH+87], but the push protocol was already studied in [FG85].

Frieze and Grimmet showed in [FG85] that the push process informs w.h.p all vertices of the complete graph $K_n$ in $(1+o(1)) \log_2 n + \ln n$, but only very recently, Doerr and Künnemann [DK14] managed to tighten
the error term up to a constant. For general graphs, the running time of push-pull strongly depends on whether the graph contains a subset of vertices of bad expansion. A measure that serves well in quantifying such "bottleneck" structures is the conductance of a graph. The conductance is defined as

\[ \varphi(G) := \min_{S \subseteq V, \text{vol}(S) \leq |E|} \frac{|E(S, V \setminus S)|}{\text{vol}(s)}, \]

where the volume of a set \( S \subseteq V \) is defined as \( \text{vol}(S) := \sum_{v \in S} \text{deg}(v) \).

Giakkoupis \cite{Gia11} improved on a result of Chierichetti, Lattanzi and Panconesi \cite{CLP10a, CLP10b} by showing that the push-pull protocol informs w.h.p every vertex of a graph \( G \) in \( O(\log |V|/\varphi(G)) \) rounds. A construction described in \cite{CLP10a} shows that this upper bound of \( O(\log |V|/\varphi(G)) \) is optimal in \( |V| \) and \( \varphi(G) \).

### 3.1.2 Subgraph Containment and Universality

An embedding of a graph \( H \) into a graph \( G \) is an injective function

\[ f : V(H) \rightarrow V(G), \]

which satisfies that for all \( u, v \in V(H) \):

\[ \{u, v\} \in E(H) \implies \{f(u), f(v)\} \in E(G). \]

We say that \( G \) contains \( H \) as a subgraph if and only if such an embedding exists, and we say that \( G \) is the host graph. We also consider the containment problem for families of graphs, where a family \( \mathcal{H} \) is simply a set of graphs. A graph \( G \) is universal for a family \( \mathcal{H} \) (we write \( G \) is \( \mathcal{H} \)-universal) if \( G \) contains every graph \( H \in \mathcal{H} \) as a subgraph. The construction (explicit and/or randomized) of sparse universal graphs for various families received a considerable amount of attention (see \cite{ABH+08, AA02, AC07, ACK+00, AKS07, BCE+82, BCS11, BCLR89, BPTW10, Cap10, CG83, DJKRR12, JKS13, KL13}).

We consider families of graphs that are spanning in the sense that all graphs contained in the family have the same number of vertices as the host graph. Two very large families of spanning graphs are of particular interest. First, the family \( \mathcal{H}(n, \Delta, d) \) of graphs on \( n \) vertices with maximum degree at most \( \Delta \) and with maximum density at most \( d \), where
the maximum density of a graph $G$ is defined as

$$d(G) = \max \left\{ \frac{2|E(H)|}{|V(H)|} : H \subseteq G \right\}.$$ 

And second, the family $\mathcal{T}(n, \Delta)$ of all forests on $n$ vertices with maximum degree at most $\Delta$.

### 3.2 Random Graph Models

The random graph models studied in this thesis can be distinguished into two categories. Binomial random graphs and random graphs with a given degree sequence are rather artificial random graph models, and therefore mainly of interest for the discrete mathematics community. We first discuss their background in the following two subsections. On the other hand, the model of random hyperbolic graphs has been suggested by a group of physicists, and, as it seems that it explains many behaviors of real world networks, it is of interest for any researcher who studies complex systems. The model of random hyperbolic graphs is not the first model that was motivated by real world networks. There are many other random graph models for complex systems. In Section 3.2.3 we discuss some of these models, before we summarize the short history of random hyperbolic graphs.

#### 3.2.1 Uniform and Binomial Random Graphs

In Section 1.1.1 we explained that uniform and binomial random graphs behave identically for many properties. This holds in particular for all the properties considered in this thesis. We therefore focus on the binomial random graph model $G(n, p)$ on $n$ vertices in which every pair of vertices forms an edge with probability $p$ independently at random. There are several books on binomial and uniform random graphs (most prominently [JLR11] and [Bol98]). We only discuss the work that is strongly related to this thesis.

As every edge is determined by its own independent random experiment, the elementary properties of binomial random graphs are very easy to obtain. It follows from Theorem 2.1 that the number of edges in a sparse binomial random graph is w.h.p $(1 \pm o(1))p(n)$. Moreover, one can prove (see Theorem 3.1 in [Bol98]) that the degree sequence follows
a Poisson distribution, which means that for $k$ sufficiently close to $np$ and every fixed $r$
\[
\Pr[N_k = r] \sim \frac{e^{-\lambda_k} \lambda_k^r}{r!},
\]
where $\lambda_k = n \Pr[\text{Bin}(n-1, p) = k]$. The expected clustering coefficient is clearly $p$, as sharing a common neighbor does not change the probability for two vertices to be connected. Therefore, we observe that the clustering coefficient is \textit{negligible} (it tends to zero) for the sparse regime $p = \Theta(1/n)$.

Binomial random graphs go through a structural evolution when $p$ moves from 0 (the empty graph) to 1 (the complete graph). For very small values of $p$ the graph contains w.h.p only small components and isolated vertices. In their groundbreaking paper [ER60] from 1960, Erdős and Rényi showed that binomial random graphs undergo a so called phase transition. More precisely, they showed that for every constant $\varepsilon > 0$, as long as $p \leq (1 - \varepsilon)/n$, w.h.p every component in the random graph contains at most $O(\log n)$ vertices, while for $p \geq (1+\varepsilon)/n$ there exists w.h.p a unique \textit{giant} component with $\Omega(n)$ vertices. The giant component grows when $p$ further increases, and as soon as the last isolated vertex vanishes (at $p = (\ln n + \ln \ln n + \omega(1))/n$) the random graph is w.h.p connected.

If the maximum degree $\Delta(H)$ of a graph $H$ is large, the threshold for containing a copy of $H$ is bounded by the event of containing a vertex of degree $\Delta(H)$. Therefore, the containment of large subgraphs is mostly considered for subgraphs of bounded maximum degree. Alon, Krivelevich, and Sudakov proved in [AKS07] the existence of a constant $c = c(\Delta, \varepsilon)$ for every fixed $\Delta > 0$ and $0 < \varepsilon < 1$, such that a typical member of $\mathcal{G}(n, c/n)$ is $\mathcal{T}((1 - \varepsilon)n, \Delta)$-universal. The constant $c$ in their result was further improved in [BCPS10]. Later on, Balogh, Csaba and Samotij showed in [BCS11] that $\mathcal{G}(n, c/n)$ is w.h.p $\mathcal{T}((1 - \varepsilon)n, \Delta)$-universal even if an adversary is allowed to delete almost half of the edges attached to any vertex. All these results confirm that every bounded degree almost spanning tree appears even before the random graph is connected.

The containment of connected spanning subgraphs clearly requires that the host graph is connected, and therefore that $p \geq \ln n/n$. Perhaps the most studied connected spanning subgraph of a random graph is the Hamilton cycle, a cycle through all vertices of the graph. A graph with minimum degree at most 1 can not contain such a cycle, and similarly
as the connectability threshold, the threshold for the containment of 
a Hamilton cycle is exactly the threshold for having minimum degree 
two. This threshold was proven to be $p = (\ln n + 2 \ln \ln n + \omega(1))/n$ 
by Komlós and Szemerédi in [KS83], and independently by Bollobás 
in [Bol84]. For other special cases of spanning trees like the complete 
binary tree or the comb it is also known that the threshold is $O(\log n/n)$ 
[HKM12, KLW14]. A more general (but not tight) result was presented 
by Krivelevich in [Kri10], where he proved that every fixed $T \in \mathcal{T}(n, \Delta)$ 
is w.h.p contained in $G(n, p)$ if $np \geq n \log \n + n\varepsilon$ for some $\varepsilon > 0$. 
For a bounded degree spanning tree $T$ that has linearly many leaves or 
a bare path (a path that contains only vertices of degree two) of linear 
length, Hefetz, Krivelevich and Szabó showed in [HKS12] that the tree 
is w.h.p contained in the random graph for $p \geq (1 + o(1)) \log n/n$.

Note that for large families $\mathcal{F}$ the statement ”$G(n, p)$ contains every fixed 
graph in $\mathcal{F}$ w.h.p” is a substantially weaker statement than ”$G(n, p)$ is 
w.h.p $\mathcal{F}$-universal”. For example if $|\mathcal{F}| = \Omega(n)$ (which is not large 
at all, most of the families which are considered are of exponential 
size), then even if every fixed member of $\mathcal{F}$ appears only with prob-
bility $1/\sqrt{n}$, it could still be the case that always at least one element 
of $\mathcal{F}$ is contained in the random graph. As it turns out, universal-
ity for spanning tree is much more challenging than universality of al-
most spanning trees. In fact, before the results presented in this thesis 
had been obtained, the best bound for $G(n, p)$ to be $\mathcal{T}(n, \Delta)$-universal 
was $p = \omega(\Delta n^{-1/3} \log^2 n)$, due to Johannsen, Krivelevich and Samotij 
[JKS13]. Only recently, Richard Montgomery [Mon14] managed to im-
prove this upper bound to $p = \Omega(\log^2 n/n)$, which is almost tight.

The family $\mathcal{H}(n, \Delta, \Delta)$ contains the so called clique factor, a graph that 
consists only of disjoint copies of $K_{\Delta+1}$ (and maybe some spare vertices). 
It is a celebrated result of Johansson, Kahn and Vu [JKV08] that for 
$\Delta > 2$ the threshold for the appearance of such a subgraph in $G(n, p)$ is 
\[ p = \Omega(n^{-\Delta+1}/(\Delta+1)(\log n)^{1/(\Delta+1)}). \]

This gives a lower bound on the threshold for $\mathcal{H}(n, \Delta, \Delta)$-universality. 
The best known upper bound is $O(n^{-1/\Delta} \log^{1/\Delta} n)$. This bound was 
proven by Dellamonica, Kohayakawa, Rödl and Ruciński in [DJKRR12] 
for $\Delta > 2$. For $\Delta = 2$ the same bound was obtained by Kim and Lee in 
[KL13].

Regarding the behavior of push on binomial random graphs, Fou-
toulakis, Huber, and Panagiotou generalized the result of Frieze and
3.2. Random Graph Models

Grimmet (FG85) by showing that push informs w.h.p all vertices in $O(\log n)$ rounds if $p = \omega(\log n/n)$ [FHP10]. For smaller values of $p$, the random graph is not necessarily connected. We are not aware of any results on the behavior of push in the giant component of a sparse random graph.

The entropy (see (3.1)) of a binomial random graph is

$$-\binom{n}{2} (p \log(p) + (1-p) \log(1-p)).$$

(3.2)

It is not hard to see that (3.2) evaluates to $\Theta(n \log n)$ for sparse binomial random graphs. Every compression scheme therefore needs at least $\Omega(n \log n)$ bits in expectation.

3.2.2 Random Graphs with a Given Degree Sequence

The degree sequence of uniform random graphs and binomial random graphs follows a binomial distribution. A more general random graph model is obtained by fixing a degree sequence (or a probability distribution from which a degree sequence is sampled), and then choosing a graph uniformly at random from all graphs with that specific degree sequence. It is not very hard to see that uniform random graphs and binomial random graphs are special cases of these random graphs with a fixed degree sequence.

After fixing a degree sequence, the random graph can be sampled with the following procedure. First, all the vertices are created, and each vertex is equipped with as many half edges (an edge that is not connected to any other vertex) as specified by the degree sequence. Second, pairs of half edges are connected by a perfect matching, which is chosen uniformly at random from all perfect matchings. This approach is known as the configuration model, and it was invented by Bollobás [Bol80], and Bender and Canfield [BC78] independently. The disadvantage of this model is that it can create multiple edges (more than one edge between the same pair of vertices) and loops (an edge that contains the same vertex twice), and therefore constructs multigraphs. Its advantage is that it provides a simple process that creates every graph (and multigraph) with the same probability. Moreover, there are results (see [Jan09]), which state that the created multigraphs are with constant probability simple, given that the degree sequence satisfies certain smoothness prop-
In this scenario, every statement that holds w.h.p for random multigraphs also holds w.h.p for random simple graphs.

Early work of Bender and Canfield \cite{BC78}, and Bollobás \cite{Bol80} focused on random regular graphs, in particular on the number of $r$ regular graphs on $n$ vertices, for given integers $n$ and $r$. Bollobás and de la Vega \cite{BFdlV82} proved that the diameter of a random regular graph is $O(\log n)$, and Robinson and Wormald \cite{RW94} later showed that random $r$-regular graphs are w.h.p Hamiltonian if $r \geq 3$.

For general degree sequences, Molly and Reed \cite{MR98} described a criteria for the degree sequence that is necessary and sufficient for the existence of a giant component. Fernholz and Ramachandran studied the size of the $k$-core (the maximum subgraph of minimum degree $k$) of a random graph with power-law degree sequence \cite{FR04}. In a later paper, they used this result to prove that the diameter of a sparse random graph with a given degree sequence is w.h.p $c \cdot \ln n$, and they gave tight bounds on $c$ for many degree sequences \cite{FR07}. In particular, their result generalizes to binomial random graphs, and thereby provides the first precise bound on the diameter of sparse binomial random graphs.

It is not very hard to see that the clustering coefficient of a random graph with a given (sparse) degree sequence tends to zero for large $n$. Moreover, a similar entropy argument as the one presented for binomial random graphs shows that at least $\Omega(n \log n)$ bits are needed to store such random graphs.

The first to study the running time of push on random graphs with a given degree sequence where Fountoulakis and Panagiotou. In \cite{FP12}, they proved that the time needed to spread a rumor on $d$-regular random graphs and multigraphs is w.h.p $(1 + o(1))c_d \ln n$, and they determined the constant $c_d$ precisely. A random power-law graph contains w.h.p many vertices of constant degree that are only connected to vertices of very high degree (i.e. of degree $n^{\Theta(1)}$). It follows from a coupon collector argument (see \cite{MU05}) that the time needed to inform all those vertices is therefore at least $n^{\Theta(1)}$. For the push-pull protocol, Fountoulakis, Panagiotou and Sauerwald \cite{FPS12} showed that the time needed to spread the information to at least an $(1 - \varepsilon)$-fraction of the vertices of a random power-law graph with exponent $2 < \beta < 3$ is w.h.p of order $O(\log \log n)$, while if $\beta > 3$ it is w.h.p $\Omega(\log n)$. They used the Chung-Lu model \cite{CL04}, which has a slightly different distribution than choosing uniformly at random from the graphs with a given power-law degree sequence.
3.2.3 Random Graph Models for Complex Systems

Perhaps the first step towards a random graph model for real-world networks was made by Watts and Strogatz \cite{WS98} in 1998. They designed a model that reasonably explains the occurrence of clustering and the small-world phenomenon. However, the degree distribution of the generated graphs follows a Poisson distribution. Barabási and Albert proposed in \cite{BA99} that the cause for power-law degree distributions is preferential attachment: the networks evolve continuously by the appearance of new vertices, and each new vertex chooses its neighbors with a probability that is proportional to their current degree. Their preferential attachment model was shown by Bollobás et al. \cite{BRST01} to produce power-law degree distributions. Moreover, it was proven that push-pull manages to spread a rumor in $O(\log n)$ \cite{DFF11} on preferential attachment random graphs. On the other hand, it is also known that such graphs typically have a vanishing clustering coefficient \cite{BR02}. Nevertheless, the preferential attachment model was the beginning of a vast series of proposed models that suggested mechanisms according to which a network can evolve (see e.g. \cite{ACL01,BCDR07,BO04,CKL13,CF03,LS09,LKF05} for a non-exhaustive but representative list).

All those models have their strengths and their weaknesses. Some are able to produce a power-law distributed degree sequence, but fail in producing a constant clustering coefficient \cite{ACL01,BRST01,BO04}. Others behave exactly in the opposite direction \cite{WS98}. Unfortunately, there is a lack of rigorous results for most of those models. The preferential attachment model for example has been intensively studied, but there are still no tight bounds on its degree sequence up to the maximum degree. A further weakness which all the prominent random graph models for the web (the preferential attachment model, the ACL model, the copying model, Kronecker’s product model and Kleiberg’s model) have in common, is that they have a large entropy. Graphs generated by theses models can therefore not be stored using less than an average of $\Omega(\log n)$ bits per edge \cite{CKL13}. The authors of \cite{CKL13} suggested a model, which creates graphs that can be stored using $O(n)$ bits, and nevertheless have a power-law degree sequence and a large clustering coefficient. However, this model seems artificial as all these properties follow from its design. Thus, the question arises whether there are natural random graph models that are compressible.
3.2.4 Random Hyperbolic Graphs

An alternative and fruitful approach towards understanding the structure and the dynamics of real-world networks is to attempt to describe the similarities or dissimilarities between vertices in a well-defined and formal sense. One possibility in this direction is based on the idea of assigning virtual coordinates to the vertices, i.e., the network is embedded in some metric space such that the mutual distances abstract the resemblance among the vertices.

One natural choice for the underlying metric space is the Euclidean space. In this context, Ng et al. [NZ02] proposed to embed the Internet graph into such spaces. Their original aim was to predict distances in the network by simply comparing coordinates. The authors obtained a reasonable mapping in 5 and 7 dimensions, but not without distortion and errors. Shavitt and Tankel [ST04] later observed that this embedding becomes dramatically better when the Euclidean geometry is replaced by the negatively curved hyperbolic space.

The above considerations lead us immediately to the model of random geometric graphs. Such a graph is generated by placing independently and uniformly at random \( n \) vertices in, say, \([0, 1]^2\), and creating edges whenever the (Euclidean) distance of two vertices is at most some \( r = r(n) \). These graphs have been studied intensively by many authors because of connections to percolation, statistical physics, hypothesis testing, and cluster analysis [Pen03]. Unfortunately, these results provide strong evidence that Euclidean geometry is not the adequate choice if one wants to describe large real-world networks, as the qualitative characteristics of the resulting random networks (like the average path length or the degree sequence) are very far from the ones observed in practice. In other words, the geometry that captures the main structural characteristics of real-world networks is not Euclidean, and the important question is whether there exists an appropriate choice of a geometry that gives rise to the observed features.

A preliminary answer to this question was given by Papadopoulos, Krioukov, Boguñá and Vahdat [PKBnV10]. The authors demonstrated impressively that power-law distributed degree sequences and high clustering coefficients emerge naturally from hyperbolic metric spaces. Their model, which we will denote by random hyperbolic graph, consists in its simplest variant of the uniform distribution of \( n \) vertices on the hyperbolic disk of radius \( R = R(n) \), where two vertices are connected if their
hyperbolic distance is at most $R$. The authors show via simulations and some preliminary theoretical analysis that the generated graphs exhibit a power-law degree distribution, whose exponent can be tweaked via the models parameters. Further, the authors indicate that with a slightly more complex model, they can also control the clustering of the generated graphs to bring it in line with real-world networks.

To make their case, Boguñá, Papadopoulos and Krioukov computed in [BnPK10] an embedding of the Internet graph into the hyperbolic plane by finding the maximum likelihood match to random hyperbolic graphs. They demonstrated impressively that this embedding has many desirable properties. For example, the authors examined the performance of greedy routing using the hyperbolic coordinates, i.e. the scheme in which each node forwards an incoming message to a neighbor that is closest to the destination (see also the works of Kleinberg [Kle07] and Papadimitriou et al. [PR05]). In their embedding, this simple greedy forwarding strategy exhibits a remarkably strong performance and connects 97% of all vertex pairs. The average stretch factor between chosen and optimal path is around 1.1, suggesting that greedy paths are very close to optimal. They also showed that this performance remains strong even if a fraction of the nodes is allowed to fail.

Since the model is relatively new, there are not many rigorous results. In [GPP12], three of the authors of this paper analyzed a special case of random hyperbolic graphs, the so-called threshold model. Independently to the work presented in this paper, Fountoulakis presented an analysis of the degree sequence in [Fou12] but only for constant degrees. Bode, Fountoulakis, and Müller studied the component structure of random hyperbolic graphs in [BFM13]. Surprisingly, it turns out that the existence of a giant component solely depends on the parameter that controls the power-exponent $\beta$ of the degree sequence. If the parameter is such that $\beta > 3$ then w.h.p every component is of sub linear size. If on the other hand $\beta < 3$ then there exists w.h.p a linear sized component. Recently, Candellero and Fountoulakis obtained precise bounds on a special version of the the clustering coefficient [CF13].

Independently of the above mentioned papers, the study of algorithms on negatively curved spaces has recently become of interest in many domains of computer science. Various computational tasks are difficult if we assume general underlying metric spaces; thus, natural hypotheses on the considered spaces leading to efficient algorithms are sought. A classical assumption is that the data points lie in $\mathbb{R}^d$, where distances
are computed according to the Euclidean measure. This is the setting of computational geometry. On the other hand, such restrictions are not always justified. Especially in the context of networking, several studies deal with the geometry underlying large real-world networks. Shavitt and Tankel [ST04] show empirically that the Internet embeds with smaller average distortion into a hyperbolic space than into an Euclidean space of similar dimension. Begelfor and Werman [BW05] obtained similar results for various data sets including an actors database and genomic data. Finally, Krauthgamer and Lee [KL06] studied various computational problems, for example the Traveling Salesperson Problem, when the underlying edge costs are given by the hyperbolic distance of the respective points.
Recall that in the binomial random graph model $G(n, p)$ on $n$ vertices with edge probability $p$, each of the $\binom{n}{2}$ pairs of vertices is connected independently at random with probability $p$. In this chapter we study the containment of large subgraphs in $G(n, p)$. In Section 4.1 we state and discuss our main results. Then, in Section 4.2 and Section 4.3, we derive two graph partitioning lemmas, which we apply in Section 4.4 and Section 4.5 to prove our main results. The content of this chapter is joint work with Asaf Ferber and Rajko Nenadov and appeared in [FNP13].
4.1 Results

Let $\mathcal{H}(n, \Delta, d)$ be the family of all graphs on $n$ vertices with maximum degree at most $\Delta$ and with density at most $d$, where the density of a graph $G$ (denoted by $d(G)$) is defined as

$$d(G) = \max \left\{ \frac{2|E(H)|}{|V(H)|} : H \subseteq G \right\}.$$ 

Dellamonica, Kohayakawa, Rödl and Rucińcki proved in [DJKRR12] that for maximum degree of $\Delta \geq 3$ and an edge probability $p = \omega \left( n^{-1/\Delta} \log^{1/\Delta} n \right)$, a typical member of $\mathcal{G}(n, p)$ is $\mathcal{H}(n, \Delta, \Delta)$-universal. Recently, Kim and Lee [KL13] obtained similar bounds for $\Delta = 2$. In the following theorem we show that if $d < \Delta/2$, then the bound in [DJKRR12] can be further improved.

**Theorem 4.1.** Let $n$ be a positive integer, and let $\Delta = \Delta(n) > 1$ and $d = d(n) \geq 2$ be integers. Then for $p = \omega(\Delta^{12}n^{-1/\min\{2d, \Delta\}} \log^{3} n)$, a graph $G \sim \mathcal{G}(n, p)$ is w.h.p $\mathcal{H}(n, \Delta, d)$-universal.

Next, let $\mathcal{H}(n, \Delta, d, g) \subseteq \mathcal{H}(n, \Delta, d, g)$ denote the family of graphs which additionally have girth at least $g$ (the girth of a graph is the length of its shortest cycle). Our second main result considers the family of graphs with girth at least 7.

**Theorem 4.2.** Let $n$ be a positive integer, and let $d = d(n)$ and $\Delta = \Delta(n) > 1$ be integers. Then for $p = \omega(\Delta^{12}n^{-1/d} \log^{3} n)$, a graph $G \sim \mathcal{G}(n, p)$ is w.h.p $\mathcal{H}(n, \Delta, d, 7)$-universal.

Recall that $\mathcal{T}(n, \Delta)$ denotes the family of all forests on $n$ vertices with maximum degree at most $\Delta$. The following straightforward corollary of Theorem 4.2 is an improvement of the upper bound on the $\mathcal{T}(n, \Delta)$-universality threshold.

**Corollary 4.3.** Let $n$ be a positive integer, and let $\Delta = \Delta(n) > 1$ be an integer. Then for $p = \omega(\Delta^{12}n^{-1/2} \log^{3} n)$, a graph $G \sim \mathcal{G}(n, p)$ is w.h.p $\mathcal{T}(n, \Delta)$-universal.

We remark that all our proofs can easily be improved in terms of $\log n$ and $\Delta$ factors. Since we believe that our bounds are far from the optimal, we did no effort in optimizing those factors.
The proofs of Theorem 4.1 and Theorem 4.2 use a simple partitioning lemma for graphs of girth at least 7 (Lemma 4.10) and an embedding technique based on matchings, developed by Alon and Füredi in [AF92] and by Ruciński in [Ruc92]. Using these techniques, we also managed to prove a result on more general graph classes.

The second part of this chapter is devoted to the proof of the following statement: For a constant $\Delta$, $d \leq \Delta$ and $H \in \mathcal{H}(n, \Delta, d)$, if we randomly color the edges of a typical member of $G(n, p)$ with $(1 + o(1))|E(H)|$ colors, then one can find a rainbow copy of $H$ (that is, a copy of $H$ with all the edges colored in distinct colors). Let $G \sim G(n, p)$ and assume that each edge of $G$ is colored uniformly at random with one of the colors from the set $[c] := \{1, \ldots, c\}$. This model is referred to as $G_c(n, p)$. For a given graph $H$ we say that a typical member of $G \sim G_c(n, p)$ contains a rainbow copy of $H$, if $G$ contains as a subgraph a copy of $H$ with all the edges colored in distinct colors. In [FL14], Frieze and Loh showed that for every $p \geq (1+\varepsilon) \log n/n$, if $c = n + o(n)$, then a typical member of $G_c(n, p)$ contains a rainbow Hamilton cycle. Note that their result is asymptotically optimal in both $p$ and the number of colors $c$. In this thesis we consider a much wider family than Hamilton cycles and provide the following general (but not optimal) result.

**Theorem 4.4.** Let $\alpha > 0$, let $\Delta$ and $d$ be integers, let $n$ be a sufficiently large integer, and let $H \in \mathcal{H}(n, \Delta, d)$. Then $G \sim G_c(n, p)$ w.h.p. contains a rainbow copy of $H$, provided that $p \geq n^{-1/d} \log^{5/d} n$ and $c = (1 + \alpha)|E(H)|$.

### 4.2 Graph-Theoretic Facts

In this section we mention a few facts about graphs which are used extensively throughout the chapter.

Recall that $D_{\geq k}$, $D_{\leq k}$ and $D_k$ are the sets that contain all the vertices of degree at least $k$, at most $k$ and exactly $k$ in a graph. For any positive integer $k$ and every vertex $V$ we denote the following set as $k$-neighborhood of $v$:

$$\{v \in V| \text{the distance between } u \text{ and } v \text{ is at most } k\}$$

Moreover, we say that a set $S \subseteq V$ is $k$-independent if and only if (in $G$) the distance between any two vertices of $S$ is at least $k + 1$. The first two lemmas consider the existence of $k$-independent sets in a graph.
Lemma 4.5. Let $G$ be a graph on $n$ vertices with maximum degree $\Delta \geq 2$ and let $S \subseteq V(G)$ be such that the maximum degree of all vertices in $S$ is at most $d$ (where $d \geq 1$). Then, $S$ contains a set $U \subseteq S$ of size at least $\frac{|S|}{d\Delta^k}$ which is $k$-independent in $G$.

Proof. Build $U$ greedily as follows: start with $L := S$ and $U := \emptyset$. In each step add an arbitrary vertex $v \in L$ to $U$ and delete the $k$-neighborhood of $v$ (including $v$ itself) from $L$. Since after each addition of a vertex to $U$ we delete at most $1 + d + d(\Delta - 1) + \ldots + d(\Delta - 1)^{k-1} \leq d\Delta^k$ vertices from $L$, we obtain the required.

Lemma 4.6. Let $G$ be a graph on $n$ vertices with maximum degree $\Delta \geq 2$ and let $d$ be an integer such that $dn \geq 2|E(G)|$. Then, for any integer $k$, $G$ contains a $k$-independent set $U \subseteq G$ for which:

(i) $U \subseteq D_{\leq d}$, and

(ii) $|U| \geq \frac{n}{(d+1)d\Delta^k}$.

Proof. First, we claim that $|D_{\leq d}(G)| \geq \frac{n}{d+1}$. Indeed, let $G$ be a graph which satisfies the conditions of the lemma for some $\Delta$. Using the fact that $|D_{>d}| = n - |D_{\leq d}|$, we obtain that

$$dn \geq \sum_{v \in V(G)} \deg_G(v) \geq 0 \cdot |D_{\leq d}| + (d + 1) \cdot (n - |D_{\leq d}|).$$

Therefore, we conclude that $|D_{\leq d}| \geq \frac{n}{d+1}$.

Applying Lemma 4.5 we conclude that there exists a $k$-independent set $U \subseteq D_{\leq d}$ in $G$ of size at least

$$|U| \geq \frac{|D_{\leq d}|}{d\Delta^k} \geq \frac{n}{(d+1)d\Delta^k},$$

as required.

A graph $G$ is called $d$-degenerate if every induced subgraph $G' \subseteq G$ contains a vertex of degree at most $d$. A moment’s thought reveals that every graph $H \in \mathcal{H}(n, \Delta, d)$ is $d$-degenerate (but not vice versa). The following observation follows directly from the definition of $d$-degenerate graphs.
Observation 4.7. Let $n, \Delta$ and $d$ be positive integers and let $H$ be a $d$-degenerate graph. Then there exists an ordering $(v_1, \ldots, v_t)$ of the vertices of $H$ such that

$$|N(v_i) \cap \{v_1, \ldots, v_{i-1}\}| \leq d$$

for every $2 \leq i \leq t$.

4.3 Partitioning Lemmas

In this section we prove some lemmas about partitioning graphs from $\mathcal{H}(n, \Delta, d)$ and $\mathcal{T}(n, \Delta)$. Before that, we define a class of graphs which can be partitioned in a “nice” way, and then we show that $\mathcal{H}(n, \Delta, d)$ and $\mathcal{T}(n, \Delta)$ belong to this class for suitably chosen parameters.

Definition 4.8. Let $n, d$ and $t$ be positive integers and let $\varepsilon$ be a positive number. The family of graphs $\mathcal{F}(n, t, \varepsilon, d)$ consists of all graphs $H$ on $n$ vertices for which the following holds. There exists a partition $V(H) = W_0 \cup \ldots \cup W_t$ such that:

(i) $|W_t| = \lfloor \varepsilon n \rfloor$,

(ii) $W_0 = N(W_t)$,

(iii) $W_t$ is 3-independent,

(iv) $W_i$ is 2-independent for every $1 \leq i \leq t - 1$, and

(v) for every $1 \leq i \leq t$ and for every $w \in W_i$, $w$ has at most $d$ neighbors in $W_0 \cup \ldots \cup W_{i-1}$.

Now, we show that $\mathcal{H}(n, \Delta, d) \subseteq \mathcal{F}(n, 4\Delta^6 \log n + 1, \varepsilon, 2d)$.

Lemma 4.9. Let $n$ be a positive integer, let $\Delta = \Delta(n) \geq 2$ and $d = d(n) \geq 2$ be integers and let $\varepsilon_0 = 1/(4\Delta^6)$. Then for every $\varepsilon \leq \varepsilon_0$ we have

$$\mathcal{H}(n, \Delta, d) \subseteq \mathcal{F}(n, 4\Delta^6 \log n + 1, \varepsilon, 2d).$$

Proof. Let $H \in \mathcal{H}(n, \Delta, d)$ and $t = 4\Delta^6 \log n + 1$. We show that $H \in \mathcal{F}(n, t, \varepsilon, 2d)$, for every $\varepsilon \leq \varepsilon_0$. 
Using Lemma 4.6 one can find a 4-independent set \( U \subseteq D_{\leq d}(H) \) of size
\[
|U| \geq \frac{n}{(d + 1)d\Delta^4} \geq \varepsilon_0 n.
\]

Let \( W_t \subseteq U \) be an arbitrary subset of size \(|\varepsilon n|\), and set \( W_0 = N_H(W_t) \) and \( H_{t-1} := H \setminus (W_0 \cup W_t) \). We further partition \( H_i \), for \( i = t-1, \ldots, 1 \), as follows:

- If \( V(H_i) = \emptyset \) then set \( W_i := \emptyset \) and \( V(H_{i-1}) := \emptyset \).
- Otherwise, \( H_i \in \mathcal{H}(|H_i|, \Delta, d) \) and thus by Lemma 4.6 there exists a 2-independent set \( U \subseteq D_{\leq d}(H_i) \) of size
\[
|U| \geq \frac{|H_i|}{(d+1)d\Delta^2} \geq \frac{|H_i|}{2\Delta^2} \geq \varepsilon_0 |H_i|.
\]

Set \( W_i := U \) and \( H_{i-1} := H_i \setminus W_i \).

Using the fact that \( \log(1 - x) \leq -x \) for every \( 0 < x < 1 \), we have that
\[
t = 4\Delta^6 \log n + 1 = \log n/\varepsilon_0 + 1 \geq -\log n/\log(1 - \varepsilon_0) + 1 = -\log_{1-\varepsilon_0} n + 1.
\]

Since for each \( i \) we have that \(|V(H_i)| \leq (1 - \varepsilon_0)|V(H_{i+1})|\), and since \( t \geq -\log_{1-\varepsilon_0} n + 1 \), it follows that \(|V(H_1)| \leq 1\).

Now, let \( V(H) = W_0 \cup \ldots \cup W_t \) be the obtained partition and note that each vertex \( w \in W_i \) has at most \( d \) neighbors in \( W_1 \cup \ldots \cup W_{i-1} \) for \( 2 \leq i < t \) (it follows immediately from the construction). Since all the properties \((i) - (iv)\) of Definition 4.8 follow easily from the construction, it thus remains to show that Property \((v)\) holds. That is, we need to show that every vertex in \( w \in W_1 \cup \ldots \cup W_{t-1} \) has at most \( d \) neighbors in \( W_0 \), and then we conclude that every vertex in \( W_1 \cup \ldots \cup W_t \) sends at most \( 2d \) “back-edges”. For this aim, note first that every vertex in \( W_t \) has at most \( d \) neighbors in \( W_0 \), and that \( W_0 = N_H(W_t) \). Therefore, if there exists a vertex \( w \in W_1 \cup \ldots \cup W_{t-1} \) with at least \( d + 1 \) neighbors in \( W_0 \), then there must exist at least two vertices \( x, y \in W_t \) such that \( N_H(x) \cap N_H(w) \neq \emptyset \) and \( N_H(y) \cap N_H(w) \neq \emptyset \). Therefore, one can find a path of length four between \( x \) and \( y \), which clearly contradicts the assumption that \( W_t \) is 4-independent. This completes the proof. \( \square \)

Next, we show that \( \mathcal{H}(n, \Delta, d, 7) \subseteq \mathcal{F}(n, 16d^2\Delta^2 \log n + 1, \varepsilon, d) \).

**Lemma 4.10.** Let \( n \) be a positive integer, let \( \Delta = \Delta(n) \) and \( d = d(n) \geq 2 \) be integers, and let \( \varepsilon_0 = 1/(2d^2\Delta^6) \). Then for every \( \varepsilon \leq \varepsilon_0 \) we have
\[
\mathcal{H}(n, \Delta, d, 7) \subseteq \mathcal{F}(n, 16d^2\Delta^2 \log n + 1, \varepsilon, d).
\]
4.3. Partitioning Lemmas

Proof. Let \( \gamma = \frac{1}{8(d+1)(d-1)\Delta^2} \leq \frac{1}{d^2} \) and observe that for \( d \geq 2 \)

\[
1 - \frac{(d+1)(d-1)\Delta^2 \gamma}{(d+1)(d-1)\Delta^2} > \gamma \quad \text{and} \quad \frac{d^2(1 - (d+1)(d-1)\Delta^2 \gamma)}{d+1} > 1.
\]

(4.1)

Let \( H \in \mathcal{H}(n, \Delta, d, 7) \) and \( t = 16d^2\Delta^2 \log n + 1 \). We show that \( H \in \mathcal{F}(n, t, \varepsilon, d) \), for every \( \varepsilon \leq \varepsilon_0 \).

Using Lemma 4.6, we find a \( 6 \)-independent set \( U \subseteq D_{\leq d}(H) \) of size

\[
|U| \geq \frac{n}{d(d+1)\Delta^6} \geq \varepsilon_0 n.
\]

For a fixed \( \varepsilon \leq \varepsilon_0 \), let \( W_t \subseteq U \) be an arbitrary subset of size \( \lfloor \varepsilon n \rfloor \), and set \( W_0 = N_H(W_t) \), \( X = N_H(W_0) \), and \( H_{t-1} := H \setminus (W_0 \cup W_t) \).

For \( i = t - 1, \ldots, 1 \), we iteratively find subsets of vertices \( W_i \subseteq V(H_i) \) (and set \( H_{i-1} := H_i \setminus W_i \)), in such a way that at the end of the process the obtained partition \( V(H) = W_0 \cup \ldots \cup W_t \) satisfies Properties (i)–(v) of Definition 4.8.

If \( V(H_i) = \emptyset \) then set \( W_i := \emptyset \) and \( V(H_{i-1}) := \emptyset \). Otherwise, construct \( W_i \) as follows:

1. If there exists a \( 2 \)-independent set \( U \subseteq D_{\leq d-1}(H_i) \) of size \( U \geq \gamma |V(H_i)| \), then set \( W_i := U \).

2. Otherwise, pick a \( 2 \)-independent set \( W_i \subseteq D_{\leq d}(H_i) \setminus X \) of size \( |W_i| \geq \gamma |V(H_i)| \).

We claim that whenever (1) fails, there exists a \( 2 \)-independent set \( U \subseteq D_{\leq d}(H_i) \) of size \( |U| \geq \gamma n_i \) (where \( n_i = |V(H_i)| \)) such that \( U \cap X = \emptyset \) as required in (2). We remark that we always consider the graph \( H_i \) when we write \( D_d \), \( D_{\leq d} \) or \( D_{\geq d} \) in the following calculations.

To prove our claim, suppose that there is no \( 2 \)-independent set \( U \subseteq D_{\leq d-1}(H_i) \) of size at least \( \gamma n_i \). First, note that by Lemma 4.5 we have

\[
|D_{\leq d-1}| \leq (d-1)\Delta^2 \gamma n_i.
\]

Second, since \( H_i \in \mathcal{H}(n_i, \Delta, d, 7) \), it follows that

\[
d n_i \geq \sum_{v \in V(H_i)} \deg_{H_i}(v) \geq 0 \cdot |D_{\leq d-1}| + (|D_{\leq d}| - |D_{\leq d-1}|) \cdot d + (n_i - |D_{\leq d}|) \cdot (d+1),
\]
and therefore
\[ |D_{\leq d}| \geq n_i - |D_{\leq d-1}| \cdot d \geq n_i \cdot (1 - d(d-1)\Delta^2\gamma). \]
Using the bound on \( |D_{\leq d-1}| \), we get that
\[ |D_d| = |D_{\leq d}| - |D_{\leq d-1}| \geq n_i \cdot (1 - (d + 1)(d - 1)\Delta^2\gamma). \]

(4.2)

Next, note that if \( |X \cap D_d| \leq d|D_d|/(d + 1) \), then by Lemma 4.5 there exists a 2-independent set \( W_i \subseteq D_d \setminus X \) of size at least
\[ |W_i| \geq \frac{|D_d \setminus X|}{d\Delta^2} \geq \frac{1 - (d + 1)(d - 1)\Delta^2\gamma}{d(d + 1)\Delta^2} n_i \geq \gamma n_i, \]
as required. Therefore, assume that \( |X \cap D_d| > d|D_d|/(d + 1) \). Observe that \( X \) is a 2-independent set in \( H_i \), as every vertex in \( X \) is a neighbor of a vertex in \( W_0 = N_H(W_t) \), \( W_t \) is 6-independent and there are no cycles of length at most 6 in \( H \). It thus follows that \( N_{H_i}(X) \cap X = \emptyset \) and every vertex in \( N_{H_i}(X) \) has exactly one neighbor in \( X \). Therefore,
\[ |N_{H_i}(X)| \geq d|X \cap D_d| > d^2|D_d|/(d + 1), \]
and it follows from (4.2) that
\[ |N_{H_i}(X)| > \frac{d^2}{d + 1} \cdot (1 - (d + 1)(d - 1)\Delta^2\gamma)n_i \geq \gamma n_i, \]
which is not possible. Hence, one can always find a 2-independent set \( W_i \subseteq V(H_i) \) of size at least \( \gamma n_i \) as required.

Using the fact that \( \log(1 - x) \leq -x \) for every \( 0 < x < 1 \), we have that
\[ t + 1 = 16d^2\Delta^2 \log n + 1 \geq \log n/\gamma + 1 \geq -\frac{\log n}{\log(1 - \gamma)} + 1 = -\log_{1-\gamma} n + 1. \]

Since for each \( i \) we have that \( |V(H_i)| \leq (1 - \gamma)|V(H_{i+1})| \), and since \( t \geq -\log_{1-\gamma} n + 1 \), it follows that \( |V(H_1)| \leq 1 \). Finally, let \( V(H) = W_0 \cup \ldots \cup W_t \) be the obtained partition. It follows immediately from the construction that Properties (i) – (v) of Definition 4.8 hold. This completes the proof. \( \square \)

### 4.4 Proof of Theorem 4.1 and Theorem 4.2

In this section we prove Theorem 4.1 and Theorem 4.2. These theorems follow easily from the following theorem and Lemma 4.10 and 4.9.
Theorem 4.11. Let $n$ and $t$ be positive integers, let $d = d(n) \geq 2$ be an integer, and let $\varepsilon < \frac{1}{2d}$. Then, a graph $G \sim G(n, p)$ is w.h.p $\mathcal{F}(n, t, \varepsilon, d)$-universal, provided that $p = \omega \left( \varepsilon^{-1} t n^{-1/d} \log^2 n \right)$.

In order to prove Theorem 4.11, we use a similar embedding algorithm as the one presented in [KL13] (and previously in [DJKRR12]). Let $d$ be a positive integer and $\varepsilon$ be a positive constant, and let $\mathcal{F} = \mathcal{F}(n, t, \varepsilon, d)$.

Our goal is to show that, whenever a graph $G$ is “good” with respect to some properties, then $G$ is $\mathcal{F}$-universal.

Before we state formally what a “good” graph is, we define the following auxiliary bipartite graph. For a graph $G$, an integer $k$, a subset $U \subseteq V(G)$ and a collection $\mathcal{L}$ of pairwise disjoint $k$-subsets of $V(G) \setminus U$, define the bipartite graph $B(\mathcal{L}, U)$ as follows: the parts are $\mathcal{L}$ and $U$, and two elements $L \in \mathcal{L}$ and $u \in U$ are adjacent if and only if $L \subseteq N_G(u)$.

We can define the notion of an $\mathcal{F}$-good graph $G$.

Definition 4.12. A graph $G$ on $n$ vertices is called $\mathcal{F}$-good if there exists a partition $V(G) = V_0 \cup V_1 \cup \cdots \cup V_t$ with

$$|V_i| = \frac{\varepsilon n}{16t} \quad \text{for } 1 \leq i \leq t \quad \text{and} \quad |V_0| = (1 - \frac{\varepsilon}{16})n,$$

such that for $p \geq \varepsilon^{-1} t n^{-1/d} \log^2 n$ the following properties hold.

(P1) There exists a set $K \subseteq V$ of $\varepsilon n$ vertex-disjoint $d$-cliques such that for all $U \subseteq V(G) \setminus V(K)$ with $|U| \leq (p/2)^{-d}/2$, we have

$$\{|K_d \in K | V(K_d) \subset N_G(u) \text{ for some } u \in U| \geq \frac{1}{2d+2} p^{d} |K||U|.$$ (4.3)

(P2) Let $1 \leq k \leq d$, and $\mathcal{L}$ be a collection of pairwise disjoint $k$-subsets of $V(G)$. If $|\mathcal{L}| \leq (p/2)^{-k}/2$, then for each $i = 1, \ldots, t$ with $V_i \cap (\cup_{L \in \mathcal{L}} L) = \emptyset$, we have that

$$|N_{B(\mathcal{L}, V_i)}(\mathcal{L})| \geq (p/2)^{k} |\mathcal{L}||V_i|/2.$$ (4.3)

If $|\mathcal{L}| \geq (p/2)^{-k} \log^{2(d-1)} n$, then for all $U$ with $|U| \geq (p/2)^{-k} \log^{2(d-1)} n$ and $U \cap (\cup_{L \in \mathcal{L}} L) = \emptyset$, the graph $B(\mathcal{L}, U)$ has at least one edge.

We first show that a random graph is typically good.
Lemma 4.13. Let $\varepsilon < \frac{1}{2d}$ and let $n$ be a positive integer. Then, a graph $G \sim \mathcal{G}(n, p)$ is w.h.p $\mathcal{F}$-good, provided that $p = \omega \left( \varepsilon^{-1}tn^{-1/d} \log^2 n \right)$.

Proof. Let $\varepsilon \leq \frac{1}{2d}$, let $p = \omega \left( \varepsilon^{-1}tn^{-1/d} \log^2 n \right)$ and let $G \sim \mathcal{G}(n, p)$. Furthermore, let $q \geq p/2$ be such that $1 - p = (1 - q)^2$, and note that one can expose $G \sim \mathcal{G}(n, p)$ as $G = G_1 \cup G_2$, where $G_1$ and $G_2$ are two graphs sampled from $\mathcal{G}(n, q)$ independently (for more details we refer the reader to [JLR11]). We use $G_1$ to find a family of vertex-disjoint $d$-cliques, and then $G_2$ to ensure the properties (P1) and (P2). For a simpler presentation, we assume from now on that $q$ is exactly $p/2$.

First, expose the edges of $G_1$. Since $q = \omega \left( n - \frac{2}{d} (\log n)^{1/(2d)} \right)$, it follows from [JKV08] that $G_1$ contains w.h.p $\lfloor n/d \rfloor$ disjoint $d$-cliques.

Let $K$ be a family of $\varepsilon n$ vertex-disjoint $d$-cliques. Next, fix an arbitrary partition $V(G) = V_0 \cup \ldots \cup V_t$ as in Definition 4.12, such that $V(K) \subset V_0$.

Finally, expose $G_2$. We now show that w.h.p this partition satisfies Properties (P1) and (P2).

For $U \subseteq V(G) \setminus V(K)$ with $|U| \leq (p/2)^{-d}/2$, let

$$X(U) := |\{K_d \in K|K_d \subseteq N_{G_2}(u) \text{ for some } u \in U\}|.$$

Note that $X(U)$ is the sum of i.i.d. indicator random variables $X_L$ ($L \in K$), such that $X_L = 1$ iff $L \subseteq N_{G_2}(u)$ for some $u \in U$. Since $|U| \leq (p/2)^{-d}/2$, we have that for each $L \in K$,

$$\Pr[X_L = 0] = (1 - (p/2)^d)^{|U|} \leq 1 - \frac{|U|p^d}{2^d} + \frac{|U|^2p^{2d}}{2^{2d}} \leq 1 - \frac{|U|p^d}{2 \cdot 2^d}.$$

(For the first inequality we use the fact that $(1 - a)^b \leq 1 - ab + (ab)^2$ for any positive integer $b$ and $0 < a < 1$).

Therefore, we have that $\Pr[X_L = 1] \geq 2^{-d-1}|U|p^d$, which implies that

$$\mathbb{E}[X(U)] \geq 2^{-d-1}|U|p^d|K| \geq \log^{2d} n \cdot \frac{\log^{2d} n}{2^{d+1}|U|}.$$ 

Using Chernoff’s bound we obtain that

$$\Pr \left[ X(U) < \frac{p^d}{2^{d+2}|K||U|} \right] \leq 2e^{-\frac{\log^{2d} n}{8^{d+1}}|U|} \leq \frac{2}{n^{3|U|}}.$$
(The last inequality holds since \(d > 1\)).

We can therefore upper bound the probability that there exists a set \(U\) that violates \((P1)\) by the following union bound

\[
\sum_{\ell=1}^{n} \binom{n}{\ell} \frac{2}{n^{3\ell}} = o(1).
\]

For property \((P2)\) we first assume that \(|L| \leq \left(\frac{p}{2}\right)^{-k}/2\). Note that \(X(L, V_i) := |N_{B(L, V_i)}(L)|\) is the sum of i.i.d. indicator random variables \(X_v\) (for \(v \in V_i\)), where \(X_v = 1\) iff \(L \subset N_{G_2}(v)\) for some \(L \in \mathcal{L}\). Since \((p/2)^k|\mathcal{L}| \leq 1/2\), using the fact that \((1 - a)^b \leq 1 - ab + (ab)^2/2\) holds for every integer \(b\) and a positive constant \(a\) for which \(ab < 1\) (follows from the binomial formula), we observe that

\[
E[X(L, V_i)] = |V_i| \left(1 - (1 - (p/2)^k)|\mathcal{L}|\right) \geq (1 - 1/4)(p/2)^k|\mathcal{L}||V_i|.
\]

Using Chernoff’s bound we obtain that

\[
Pr[X(L, v_i) < (p/2)^k|\mathcal{L}||V_i|/2] \leq \exp \left[-E[X(L, v_i)]/36\right] \leq \frac{1}{n^{3d|\mathcal{L}|}},
\]

where the last inequality follows as

\[
(p/2)^k|V_i| \geq (p/2)^d \cdot \frac{\varepsilon n}{16t} \geq \frac{n \log^{2d} n}{2^{d+4n}} = \omega(d \log n).
\]

Thus, the probability for having sets \(L\) and \(V_i\) such that \(|N_{B(L, V_i)}(L)| < (p/2)^k|\mathcal{L}||V_i|/2\) can be bounded by

\[
t \sum_{\ell=1}^{n} \binom{n}{k} \ell \frac{1}{n^{3d\ell}} = o(1).
\]

Next, assume that \(|\mathcal{L}| \geq (p/2)^{-k} \log^{2(d-1)} n\). Observe that each edge in \(B(L, U)\) is present with probability \((p/2)^k\), hence the probability that there are no edges is bounded by

\[
(1 - (p/2)^k)|\mathcal{L}||U| \leq \exp \left[-(p/2)^k \cdot |\mathcal{L}||U|\right].
\]

Furthermore, for \(r, \ell \geq (p/2)^{-k} \log^{2(d-1)} n\), the number of collections of \(k\)-subsets \(L\) with \(|\mathcal{L}| = \ell\) is at most \(n^{k\ell}\), and the number of subsets \(U\)
with \( |U| = r \) is at most \( n^r \). We thus have that
\[
\Pr[\exists \mathcal{L}, U \text{ with } |\mathcal{L}| = \ell, |U| = r \text{ and } e(B(\mathcal{L}, U)) = 0] \leq \exp \left[ (k\ell + r) \log n - (p/2)^{k\ell r} \right].
\]

Note that
\[
(k\ell + r) \log n \leq k \cdot (\ell \log n + r \log n) \leq 2k \cdot \frac{r\ell (p/2)^k}{\log^{2d-3} n} \leq (p/2)^{k\ell r/2}
\]
for \( n \) large enough, and hence,
\[
\exp \left[ (k\ell + r) \log n - (p/2)^{k\ell r} \right] \leq \exp \left[ -(p/2)^{k\ell r/2} \right] \leq \exp \left[ -(p/2)^{-k} \log n/2 \right] = o(1).
\]

We therefore conclude that the probability for the existence of such sets \( \mathcal{L} \) and \( U \) without an edge is \( o(1) \). \( \Box \)

Now we want to show that any \( \mathcal{F} \)-good graph is \( \mathcal{F} \)-universal. Let \( G \) be a \( \mathcal{F} \)-good graph with a partition \( V(G) = V_0 \cup \cdots \cup V_t \) and a clique-set \( \mathcal{K} \). We construct an embedding \( f : V(H) \to V(G) \) for a given graph \( H \in \mathcal{F} \) as follows.

Let \( H = W_0 \cup \cdots \cup W_t \) be the partition of \( H \) that satisfies the conditions (i) \( - \) (v) of Definition 4.12. For every \( v \in W_t \) let \( L(v) := N_G(v) \cap W_0 \) denote the neighborhood of \( v \) in \( W_0 \). In a first step we choose an arbitrary injective mapping \( f_0 : W_0 \to V(\mathcal{K}) \) such that for every \( w \in W_t \) the vertices in \( L(w) \) all map to vertices of the same clique in \( \mathcal{K} \). Note that such a mapping exists as \( \mathcal{K} \) consists of \( \lceil \varepsilon n \rceil \) \( d \)-cliques and there are exactly that many sets \( L(w) \), each of which contains at most \( d \) vertices. Moreover, such a mapping is valid as there can not be edges between \( L(u) \) and \( L(w) \) for \( u \neq w \) (because \( W_t \) is \( 3 \)-independent).

For \( i = 1, \ldots, t \), we iteratively construct \( f_i : (W_0 \cup \cdots \cup W_i) \to (V_0 \cup \cdots \cup V_i) \) from \( f_{i-1} \) as follows. Let \( V_i^* := (V_0 \cup \cdots \cup V_i) \setminus \text{Img}(f_{i-1}) \). We want to embed \( W_i \) to \( V_i^* \). For \( w \in W_i \) let \( L_i(w) := f_{i-1}(N_H(w) \cap (\cup_{j=0}^{i-1} W_j)) \) and let \( \mathcal{L}_i := \{ L_i(w) \mid w \in W_i \} \). Here it is crucial that \( W_i \) is \( 2 \)-independent and therefore \( L_i(w) \cap L_i(w') = \emptyset \) for \( w \neq w' \in W_i \). Since a vertex \( w \in W_i \) can be mapped only to the vertices in
\[
\{ v \in V_i^* \mid L_i(w) \subseteq N_G(v) \},
\]
we can extend \( f_{i-1} \) by a \( \mathcal{L}_i \) matching in \( B_i := B(\mathcal{L}_i, V_i^*) \) (recall that in \( B_i \) the set \( L_i(w) \in \mathcal{L}_i \) is connected to a vertex \( v \in V_i^* \) if and only if
$L_i(w) \subseteq N_G(v)$. More precisely, for a matching $\mathcal{M}$ which saturates $\mathcal{L}_i$ (an $\mathcal{L}_i$-matching), we define $f_i$ as follows: For $w \in W_0 \cup \cdots \cup W_{i-1}$ let $f_i(w) := f_{i-1}(w)$, and for $w \in W_i$ let $f_i(w) := v$, where $v \in V_i^*$ is the unique vertex such that $\{L_i(w), v\} \in \mathcal{M}$.

As long as we find an $\mathcal{L}_i$-matching for $1 \leq i \leq t$ we clearly construct a valid embedding of $H$ into $G$. It remains to show that we can find the required matchings.

We first show that for every $1 \leq i \leq t-1$, the auxiliary graph $B_i$ contains an $\mathcal{L}_i$-matching.

**Claim 4.14.** For every $1 \leq i \leq t-1$, there exists an $\mathcal{L}_i$-matching in $B_i$.

**Proof.** We show that Hall’s condition the existence of an $\mathcal{L}_i$-saturating matching is satisfied. First, we show that $|L_i| = |W_i| < |V_i^*| - \frac{\varepsilon n}{16}$ for $1 \leq i \leq t-1$. We have

$$|V_i^*| = |V_0 \cup \cdots \cup V_i| - |W_0 \cup \cdots \cup W_{i-1}| = |W_i \cup \cdots \cup W_t| - |V_{i+1} \cap \cdots \cup V_t|$$

and therefore

$$|V_i^*| - |W_i| = |W_{i+1} \cup \cdots \cup W_t| - |V_{i+1} \cap \cdots \cup V_t| \geq |W_t| - \frac{t-i}{16t} \varepsilon n > \frac{15\varepsilon n}{16}.$$

Thus, we have that $|L_i| = |W_i| \leq |V_i^*| - \frac{15\varepsilon n}{16} < |V_i^*| - \frac{\varepsilon n}{16}$ and the claim therefore follows by Claim 4.15 below.

**Claim 4.15.** For all $U \subseteq \mathcal{L}_i$ that satisfy $|U| \leq |V_i^*| - \frac{\varepsilon n}{16}$, we have

$$|N_{B_i}(U)| \geq |U|.$$

**Proof.** Let $U = U_0 \cup \cdots \cup U_d$, where

$$U_j := \{L \in U ||L| = j\}.$$

If $U_0 \neq \emptyset$, then $N_{B_i}(U) = V_i^*$. Therefore, we may assume that $U_0 = \emptyset$. Pick $k$ such that $|U_k| \geq |U|/d$, and distinguish between the following three cases:

**Case 1:** $|U_k| \leq (p/2)^{-k}/2$. It follows by property (P2) that

$$|N_{B_i}(U)| \geq |N_{B_i}(U_k)| \geq (p/2)^k |U_k| |V_i|/2 \geq \frac{\log^{2d} n}{2^{k+1} \cdot 16d} |U| \geq |U|.$$
Case 2: $(p/2)^{-k}/2 \leq |U_k| \leq (p/2)^{-k} \log^{2(d-1)} n$. We fix an arbitrary subset $U'_k \subset U_k$ of size $|U'_k| = (p/2)^{-k}/2$, and by the same argument as in Case 1 we get that

$$|N_{B_i}(U)| \geq |N_{B_i}(U'_k)| \geq (p/2)^k |U'_k| |V_i|/2 \geq \frac{\log^{2d} n}{2^{k+1} \cdot 16} |U'_k| \geq \frac{\log n}{2^{k+1} \cdot 16d} |U| \geq |U|. $$

Case 3: $|U_k| \geq (p/2)^{-k} \log^{2(d-1)} n$. In this case note that the induced subgraph $B_i[U_k, V_i \setminus N_{B_i}(U_k)]$ has no edges. By property $(P2)$ this yields that

$$|V_i \setminus N_{B_i}(U_k)| < (p/2)^{-k} \log^{2(d-1)} n = o(\varepsilon n),$$

which implies that $|N_{B_i}(U_k)| \geq |V_i| - o(\varepsilon n) \geq |V_i| - \frac{\varepsilon n}{16} \geq |U|$. $\square$

In the last lemma of this section we show that $B_t$ contains a perfect matching, thus we can complete the embedding of $H$.

**Lemma 4.16.** There exists a perfect matching in $B_t$.

**Proof.** We check Hall’s condition for every subset $U \subseteq \mathcal{L}_t$. For sets of cardinality $|U| \leq |V_i^*| - \frac{\varepsilon n}{16}$, Hall’s condition follows by Claim 4.15. Therefore, consider only subsets $U$ of cardinality $|U| \geq |V_i^*| - \frac{\varepsilon n}{16}$. Let $U \subseteq \mathcal{L}_t$ be such a subset. Note that by the definition of the partial embedding $f_0$, every set in $U$ is contained in one of the cliques in $K$. Suppose first that $|V_i^* \setminus N_{B_t}(U)| \geq (p/2)^{-d}/2$. We fix a subset $Y \subset V_i^* \setminus N_{B_t}(U)$ of size exactly $(p/2)^{-d}/2$. It follows by property $(P1)$ that at least $2^{-d-2} \cdot p^d |Y| \varepsilon n$ of the cliques in $K$ are completely connected to some vertices in $Y$. We conclude that

$$|N_{B_t}(V_i^* \setminus N_{B_t}(U))| \geq |N_{B_t}(Y)| \geq 2^{-3} (p/2)^d (p/2)^{-d} \cdot \varepsilon n > \varepsilon n/16,$$

which is not possible since $|U| + |N_{B_t}(V_i^* \setminus N_{B_t}(U))| \leq |V_i^*|$. Therefore, we conclude that $|V_i^* \setminus N_{B_t}(U)| \leq (p/2)^{-d}/2$. Now, using Property $(P1)$ similarly as above we obtain that

$$|N_{B_t}(V_i^* \setminus N_{B_t}(U))| \geq 2^{-d-3} (p/2)^d \cdot \varepsilon n |V_i^* \setminus N_{B_t}(U)| > |V_i^* \setminus N_{B_t}(U)|.$$

Finally, since

$$|N_{B_t}(U)| + |V_i^* \setminus N_{B_t}(U)| = |V_i^*| = |W_t| \geq |U| + |N_{B_t}(V_i^* \setminus N_{B_t}(U))|$$

$$> |U| + |V_i^* \setminus N_{B_t}(U)|$$
4.5 Proof of Theorem 4.4

In this section we prove Theorem 4.4. Before starting the proof, it will be convenient to introduce the following notation. For any bipartite graph $G = (A \cup B, E)$ with $|A| = |B| = n$ and minimum degree $\delta(G) \geq k$, let $B^t_{k-out}(G)$ denote the following set of bipartite graphs: each $D \in B^t_{k-out}(G)$ has vertex set $V(D) = V(G)$ and edge set $E(D) \subseteq E$ such that each vertex in $A$ has degree exactly $k$. Note that we can sample an element from $B^t_{k-out}(G)$ uniformly at random by choosing for each $v \in A$ uniformly at random $k$ edges from $E_G(v, B)$.

One of the main ingredients in the proof of Theorem 4.4 is the following simple lemma on the existence of perfect matchings in typical graphs from $B^t_{k-out}(G)$.

**Lemma 4.17.** Let $\varepsilon > 0$, let $n$ be a sufficiently large integer and let $k = \omega(\log n)$. Then for any bipartite graph $G = (A \cup B, E)$ with $|A| = |B| = n$ and $\delta(G) \geq \frac{n}{2} + \varepsilon n$, a graph $D$ chosen uniformly at random from $B^t_{k-out}(G)$ w.h.p contains a perfect matching.

**Proof.** Let $D$ be a graph chosen uniformly at random from $B^t_{k-out}(G)$. We show that w.h.p all subsets $S \subset A$ and all subsets $S \subset B$ with $|S| \leq n/2$ satisfy $|S| \leq |N_D(S)|$. It then follows from Hall’s theorem (see [Wes01] for more details) that $D$ has a perfect matching.

We first assume that $S \subset A$. Note that $|S| > |N_D(S)|$ implies that there exists a subset $S' \subset B$ of size $|S'| = |S| - 1$ such that $|E_D(S, B \setminus S')| = 0$. Note that in $G$, since $|S'| \leq n/2$, every vertex $v \in S$ has at least $\varepsilon n$ neighbors in $B \setminus S'$. Therefore, when choosing the $i$-th of the $k$ edges incident to $v$ and conditioning on the event that no edge in $E_G(v, B \setminus S')$ has been selected so far, the probability to miss $B \setminus S'$ is at most

$$\frac{\deg_G(v) - \varepsilon n - i + 1}{\deg_G(v) - i + 1} \leq 1 - \varepsilon.$$ 

Thus,

$$\Pr[|S| > |N_D(S)|] \leq \Pr[\exists S' \subset B||E_D(S, B \setminus S')| = 0] \leq \left(\frac{n}{|S| - 1}\right)^{|S|} (1 - \varepsilon)^{|S|} \leq e^{-\varepsilon |S| \omega(\log n)},$$
and the probability that such a bad set exists is at most
\[ \sum_{s=1}^{n/2} \binom{n}{s} e^{-\varepsilon s \omega(\log n)} \leq \sum_{s=1}^{n/2} e^{-s \omega(\log n)} = o(1). \]

Next, assume that \( S \subset B \) and observe that in order to have \( |S| > |N_D(S)| \), there must exist a set \( S' \subset A \) of size \( |S| - 1 \) such that \( |E_D(A \setminus S', S)| = 0 \). Note that \( |E_G(A, S)| \geq |S| \cdot \left( \frac{n}{2} + \varepsilon n \right) \), \( |E_G(S', S)| \leq |S| \cdot \frac{|S'|}{2} \) and therefore \( |E_G(A \setminus S', S)| \geq |S| \varepsilon n \). Since every edge of \( G \) appears in \( D \) with probability at least \( \frac{k}{n} \) (but not independently) and since this probability can only decrease if we know that another edge does not appear in \( D \), it follows that
\[
\Pr[|S| > |N_D(S)|] \leq \Pr[\exists S' \subset A |E_D(A \setminus S', S)| = 0] \\
\leq \left( \frac{n}{|S| - 1} \right) \left( 1 - \frac{\omega(\log n)}{n} \right)^{|S| \varepsilon n} \leq e^{-\varepsilon |S| \omega(\log n)}
\]
as in the previous case. \( \square \)

Now we are ready to prove Theorem 4.4.

**Proof.** Our proof is motivated by ideas of Cooper and Frieze [CF02]. Note that containing a rainbow copy of some fixed graph \( H \) is a monotone increasing property and we can therefore fix the probability \( p \) to be exactly \( n^{-1/d} \log^{5/d} n \).

Let \( \Delta \) and \( d \) be positive integers, let \( n \) be a sufficiently large integer and let \( H \in \mathcal{H}(n, \Delta, d) \). Moreover, let \( \bar{d} = \frac{2|E(H)|}{n} \) denote the average degree of \( H \) (note that \( \bar{d} \leq d \) and in fact can be much smaller than \( d \)) and let \( \alpha > 0 \) be some arbitrarily small positive constant. First, we show how to partition \( H \) in such a way that will later help us to find a rainbow copy of it in a typical member of \( G_c(n, p) \), where \( c = (1 + \alpha)|E(H)| \). For this aim we act as follows. If \( H \) contains a set \( W \) of \( \left\lfloor \frac{an}{5 \log^2 n} \right\rfloor \) isolated vertices (that is, vertices of degree 0 in \( H \)), then partition \( V(H) = \{w_1\} \cup \ldots \cup \{w_t\} \cup W \) in such a way that for each \( i \), the vertex \( w_i \) has at most \( d \) neighbors in \( \{w_1, \ldots, w_{i-1}\} \). Indeed, such a partition exists since \( H' := H - W \in \mathcal{H}(n - |W|, \Delta, d) \), and therefore is \( d \)-degenerate, so one can apply Observation 4.7. Otherwise, let \( x \) denote the number of vertices of degree larger than 0 and at most \( \bar{d} \) in \( H \). Since \( H \) contains at most \( \frac{an}{5 \log^2 n} \) isolated vertices, the following inequality holds:
\[ \tilde{d}n = 2|E(H)| \geq x + (\tilde{d} + 1) \left( n - \frac{\alpha n}{5 \log^2 n} - x \right). \]

Hence, using the fact that \( n \) is sufficiently large, we conclude that \( x \geq n/(2\tilde{d}) \). Now, let \( S \) be the set that consists of all these vertices. By applying Lemma 4.5 to \( H \) and \( S \) it follows that there exists a subset \( T \subseteq S \), such that \( T \) is 2-independent and

\[ |T| \geq \frac{|S|}{d\Delta^2} \geq \frac{n}{2d^2\Delta^2} \geq \left\lceil \frac{\alpha n}{5 \log^2 n} \right\rceil \]

for sufficiently large \( n \). Next, let \( W \subseteq T \) be an arbitrary subset of size \( \left\lceil \frac{\alpha n}{5 \log^2 n} \right\rceil \), and partition \( V(H) = \{w_1\} \cup \ldots \cup \{w_t\} \cup W \) in such a way that for each \( i \), \( w_i \) has at most \( d \) neighbors in \( \{w_1,\ldots,w_{i-1}\} \).

All in all, we have a partition \( V(H) = \{w_1\} \cup \ldots \cup \{w_t\} \cup W \) such that \( |W| = \left\lceil \frac{\alpha n}{5 \log^2 n} \right\rceil \) and one of the following holds:

1. all the vertices of \( W \) are isolated in \( H \), or
2. \( W \) is 2-independent and consists of non-isolated vertices of degree at most \( \tilde{d} \).

Note that if (2) holds then

\[ |E(W, V \setminus W)| \leq \tilde{d}|W| = \frac{2|E(H)|}{n} \cdot \left\lceil \frac{\alpha n}{5 \log^2 n} \right\rceil < \alpha|E(H)|/(2\lceil \log^2 n \rceil), \]

for \( n \) large enough.

Now we start to describe the procedure of finding a rainbow copy of \( H \). Let \( q \geq p/2 \) be such that \( 1 - p = (1 - q)^2 \) and present \( G \sim G(n, p) \) as \( G = G_1 \cup G_2 \), where \( G_1 \) and \( G_2 \) are two graphs sampled independently from \( G(n, q) \). We sample a member of \( G_c(n, p) \) by sampling a member of \( G(n, p) \) and randomly coloring exposed edges using \( c \) colors.

We find a rainbow embedding of \( H \) in \( G \sim G_c(n, p) \) in two phases. In Phase I, we find a rainbow embedding \( f \) of \( H[\{w_1 \cup \ldots \cup w_t\}] \) with edges which are taken from \( G_1 \). If \( W \) is as in (1) (that is, all the vertices in \( W \) are isolated in \( H \)), then we are done. Otherwise, in Phase II we show that one can extend \( f \) to a rainbow embedding of \( H \) in \( G \), using edges of \( G_2 \).
In what follows, we present the exact strategies of Phases I and II and prove that w.h.p everything works out well.

**Phase I:** Throughout this phase we maintain a partial rainbow embedding \( f \) of \( H \) to \( G_1 \), a set of available colors \( \mathcal{C} \) and a set of available vertices \( V' \). Initially, set \( f = \emptyset \), \( \mathcal{C} := [c] \) and \( V' := V(G) \). Additionally, we maintain for each vertex \( v \in V(G) \) a set \( U_v \subseteq V(G) \) such that \( U(v) \cap V' \) contains only unexposed potential neighbors of \( v \) in \( G_1 \). Initially, \( U_v = V(G) \setminus \{v\} \) for each \( v \in V(G) \).

We inductively build the desired partial embedding \( f \) as follows. In the first step, let \( f(v_1) := v \) for an arbitrary vertex \( v \in V' \), and set \( V' := V' \setminus \{v\} \). Assume that we have already embedded \( \{w_1, \ldots, w_{i-1}\} \) for some \( 2 \leq i \leq t \) and we wish to embed \( w := w_i \). Let \( L(w) = f(N_H(w) \cap \{w_1, \ldots, w_{i-1}\}) \) be the set of images of neighbors of \( w \) which have already been embedded (recall that \( |L(w)| \leq d \)). Let \( A_w = V' \cap (\cap_{v \in L(w)} U_v) \) be the set of all available vertices which are still unexposed neighbors of all vertices in \( L(w) \), and choose an arbitrary subset \( S_w \subset A_w \) of size \( s := \lceil \alpha n/(4\Delta \log n)^2 \rceil \) (Claim 4.18 shows that throughout Phase I this is indeed possible; that is, \( A_w \) is of size at least \( s \)). Expose all edges between \( L(w) \) and \( S_w \), and assign uniformly at random colors to all the obtained edges. Let \( x \in S_w \) be a vertex which is connected to all the vertices in \( L(w) \) and such that all the colors assigned to edges \( \{vx \mid v \in L(w)\} \) are distinct and belong to \( \mathcal{C} \). The existence of such a vertex follows from Claim 4.19 below. We extend \( f \) by defining \( f(w_i) := x \), update \( U_v := U_v \setminus S_w \) for all \( v \in L(w_i) \), \( V' := V' \setminus \{x\} \) and

\[ \mathcal{C} := \mathcal{C} \setminus \{col \in \mathcal{C} \mid \exists v \in L(w_i) \text{ such that } vx \text{ is colored in } col\}. \]

The following two claims show that w.h.p we manage to find the desired embedding in Phase I.

**Claim 4.18.** Throughout Phase I we have that \( |A_w| \geq \lceil \alpha n/(4\Delta \log n)^2 \rceil \) for every vertex \( w \in V(H) \) which has not been embedded.

**Proof.** The proof of the claim is obtained from the following four observations. First, note that at the beginning of Phase I we have that \( U_v = V(G) \setminus \{v\} \) for each \( v \in V(G) \). Second, we update \( U_v \) only after embedding a vertex \( w \) for which \( v \in L(w) \) (and then we delete the set \( S_w \) which is of size \( s = \lceil \alpha n/(4\Delta \log n)^2 \rceil \) from \( U_v \)). Third, every vertex \( v \) is a member of at most \( \Delta \) sets \( L(w) \) (recall that \( \Delta(H) \leq \Delta \)). Fourth,
note that \(|V'| \geq \lceil \alpha n / (5 \log^2 n) \rceil\) throughout Phase I (recall that we do not embed \(W\) in this phase).

Therefore, it follows that at any point during Phase I we have
\[
|U_v \cap V'| \geq |V'| - 1 - \Delta \cdot \left\lceil \frac{\alpha n}{(4\Delta \log n)^2} \right\rceil,
\]
for each vertex \(v \in V(G)\). Since \(|L(w)| \leq \Delta\), we conclude that
\[
|A_w| = |V' \cap (\cap_{v \in L(w)} U_v)| \geq |V'| - \Delta - \Delta^2 \left\lceil \frac{\alpha n}{(4\Delta \log n)^2} \right\rceil \geq s,
\]
for \(n\) large enough.

The next claim states that whenever we wish to embed a vertex \(w\), it has at least one candidate in \(V'\).

**Claim 4.19.** Let \(w \in V(H) \setminus W\). At the moment we try to embed \(w\) there exists with probability \(1 - o(1/n)\) a vertex \(x \in S_w\) for which the following holds:

(i) \(x\) is connected to all the vertices in \(L(w)\), and

(ii) all the colors assigned to the edges \(\{v, x\} : v \in L(w)\) are distinct and belong to \(C\).

**Proof.** Let
\[
X := |\{v \in S_w \mid L(w) \subseteq N_{G_1}(v)\}|.
\]
Note that \(X\) is the sum of i.i.d. indicator random variables \(X_v\) (for all \(v \in S_w\)) for which \(X_v = 1\) iff \(L(w) \subseteq N_{G_1}(v)\). Clearly, we have that (recall that \(|L(w)| \leq d\))
\[
\mathbb{E}[X] \geq sq^d \geq \frac{\alpha n}{(4\Delta \log n)^2} \cdot \Omega \left( \frac{\log^5 n}{n} \right) = \Omega(\log^3 n).
\]
Applying Chernoff’s bound we obtain that
\[
\Pr[X \leq \mathbb{E}[X]/2] = e^{-\Omega(\log^3 n)} = o(1/n).
\]
Now, note that \(|C| \geq \alpha |E(H)|\) during Phase I. Thus, the probability that for a vertex \(x \in S_w\) with \(L(w) \subseteq N_{G_1}(x)\), all the edges to \(L(w)\) have different colors from \(C\) is at least
\[
\frac{(C^\ell)}{((1 + \alpha)|E(H)|)^\ell} \geq \left( \frac{\alpha |E(H)|}{(1 + \alpha)|E(H)|^\ell} \right)^\ell \geq \left( \frac{\alpha}{(1 + \alpha)d} \right)^d =: \gamma > 0,
\]
where $|L(w)| = \ell$. Therefore, if $X \geq \mathbb{E}[X]/2$ then the probability that there is no such $x$ is at most
\[
(1 - \gamma)^{|X|} \leq e^{-\gamma|X|} = e^{-\Omega(\log^3 n)} = o(1/n).
\]
\[\square\]

Note that since we embed at most $n$ vertices, applying the union bound we obtain that for every vertex $w_i$ there exists a “good” vertex $x \in S_w$. Now, if $W$ is as in (1) (that is, all the vertices in $W$ are isolated in $H$), then we are done. Otherwise, we continue to Phase II.

**Phase II:** Let $V^* := V(G) \setminus f(V(H) \setminus W)$. Our goal is to extend $f$ with a valid embedding of $W$ into $V^*$, using edges of $G_2$, in such a way that the resulting embedding is rainbow.

For $w \in W$ let $L(w) := f(N_H(w))$ and let $L = \{L(w) | w \in W\}$. Recall that $W$ is 2-independent and thus all the $L(w)$’s are disjoint. Let $F = (\mathcal{L} \cup V^*, E_F)$ with edge set
\[
E_F := \{Lv \mid L \in \mathcal{L}, v \in V^* \text{ and } \forall u \in L \ uv \notin E(G_1)\}
\]
be the ground graph to build a bipartite auxiliary graph $\mathcal{B}(\mathcal{L}, V^*)$. Edges that appeared in $G_1$ are excluded since we can not color them again. Note that $|\mathcal{L}| = |W| = |V^*|$ and that by the following very rough estimate $F$ satisfies w.h.p the conditions of Lemma 4.17.

**Claim 4.20.** It holds with high probability that $\delta(F) \geq \frac{3}{4}|V^*|$. 

**Proof.** For every $L \in \mathcal{L}$ and $v \in V^*$ the edge $Lv \notin E_F$ if and only if there exists $u \in L$ for which $uv \in E(G_1)$. Since $G_1 \sim G(n,q)$, by applying Chernoff’s bound it follows that w.h.p $\Delta(G_1) \leq 2nq$. Moreover, since for every $L \in \mathcal{L}$ we have that $|L| \leq \bar{d}$, it follows that $\deg_F(L) \geq |V^*| - \bar{d}2nq > 3|V^*|/4$. A similar argument shows that we have $\deg_F(v) \geq 3|V^*|/4$ for every $v \in V^*$.

\[\square\]

In the following we describe a random process that tries to create a bipartite graph $\mathcal{B}(\mathcal{L}, W) \in \mathcal{B}_{[\log^2 n] \text{-out}}(F)$ by exposing edges from $G_2 \setminus G_1$ and randomly color them. First, let
\[
\mathcal{C} := \{col \in [c] \mid \exists \{u, v\} \in E(H \setminus W) \text{ s.t. } \{f(u), f(v)\} \text{ has color } col\}
\]
and note that $|\mathcal{C}| \geq \alpha|E(H)|$. Choose an arbitrary ordering $L_1, \ldots, L_{|\mathcal{L}|}$ of the elements in $\mathcal{L}$. Then, in step $1 \leq i \leq |\mathcal{L}|$, set $N_i := N_F(L_i)$ and
create $\lceil \log^2 n \rceil$ edges from $L_i$ to vertices in $N_i$ as follows: as long as $|N(B(L,V^*)(L_i))| < \lceil \log^2 n \rceil$, iteratively pick a vertex $v \in N_i$ uniformly at random, set $N_i := N_i \setminus \{v\}$ and expose all edges from $v$ to vertices in $L_i$ and color them uniformly at random with colors from $[c]$ (note that here the process can fail if at some point $N_i = \emptyset$ while $|N_B(L,V^*)(L_i)| < \lceil \log^2 n \rceil$). If all the edges are contained in $G_2$ and if they have distinct colors that are all from the set of available colors $C$, add $L_i v$ to $B(L,V^*)$.

At the end of step $i$ remove all the colors used at edges incident to $L_i$, $C := C \setminus \{col \in [c] | \exists u \in L_i, \exists v \in N_B(L,V^*)(L_i) \text{ s.t. } uv \text{ has color } col\}$.

If the process succeeds then every matching $M$ in $B(L,V^*)$ is clearly rainbow in the sense that all edges in

$$\{uv | \exists L v \in M \text{ s.t. } u \in L\}$$

have distinct colors that have not been used in the embedding in Phase I. It follows from Claim 4.21 below and Lemma 4.17 that the process succeeds w.h.p and that the constructed $B(L,V^*)$ contains a perfect matching. Finally, such a perfect matching in $B(L,V^*)$ extends $f$ into a rainbow embedding of $H$ in $G$. The following claim therefore completes the proof of Theorem 4.4.

**Claim 4.21.** The random process that creates $B(L,V^*)$ in Phase II succeeds w.h.p and it samples uniformly at random from $B^{\lceil \log^2 n \rceil - \text{out}}(F)$.

**Proof.** Note first that the process can only fail if in some round $1 \leq i \leq |L|$ we have that $N_i = \emptyset$ and $|N_B(L,V^*)(L_i)| < \lceil \log^2 n \rceil$. It therefore suffices to show that in a fixed step $1 \leq i \leq |L|$ the process creates with probability $1 - o(1/n)$ the $\lceil \log^2 n \rceil$ required edges. Let

$$X_i := \{v \in N_F(L_i) | L_i \subseteq N_{G_2}(v)\}.$$ 

Note that $|X_i|$ is the sum of i.i.d. indicator random variables $X_{i,v}$ (for all $v \in N_F(L_i)$) for which $X_{i,v} = 1$ iff $L_i \subseteq N_{G_2}(v)$. Clearly, we have that (recall that $|L_i| \leq \bar{d} \leq d$)

$$\mathbb{E}[|X_i|] \geq |N_F(L_i)| q^d \geq \delta(F) \cdot \Omega(\frac{\log^5 n}{n}) \geq \frac{3}{4} \cdot \frac{\alpha n}{5 \log^2 n} \cdot \Omega(\frac{\log^5 n}{n})$$

$$= \Omega(\log^3 n).$$
Applying Chernoff’s bound we obtain that
\[ \Pr[|X_i| \leq \mathbb{E}[|X_i|]/2] = e^{-\Omega(\log^3 n)} = o(1/n). \] (4.5)

Next, let
\[ Y_i := \{ v \in X_i \mid \text{all edges in } E(L_i, v) \text{ have distinct colors from } C \}. \]

Note that \(|Y_i|\) is the sum of i.i.d. indicator variables \(Y_{i,v}\) (for all \(v \in X_i\)) for which \(Y_{i,v} = 1\) iff all edges in \(E(L_i, v)\) have distinct colors from \(C\). Since we have by (4.4) that
\[ |E(W, V \setminus W)| \leq \alpha |E(H)|/(2[\log^2 n]) \]
and we remove for each edge in \(E(W, V \setminus W)\) at most \([\log^2 n]\) colors from \(C\), the number of available colors in \(C\) is always at least \(\alpha |E(H)|/2\). Thus, the probability that for a vertex \(v \in X_i\) all the edges to \(L_i\) have different colors from \(C\) is at least
\[ p_i = \frac{\binom{C}{\ell}}{(1 + \alpha)|E(H)|^\ell} \geq \left( \frac{\alpha |E(H)|/2}{(1 + \alpha)|E(H)|^{\ell}} \right)^\ell \geq \left( \frac{\alpha}{(1 + \alpha)2\bar{d}} \right)^{\bar{d}} =: \gamma > 0, \]

where \(|L_i| = \ell \leq \bar{d}\), and this lower bound for \(p_i\) holds independently of all other color assignments in previous steps. Therefore, if \(|X_i| \geq \mathbb{E}[|X_i|]/2\), then the expectation of \(|Y_i|\) is at least
\[ \mathbb{E}[|Y_i|] \geq |X_i| \cdot \gamma = \Omega(\mathbb{E}[|X_i|]) = \Omega(\log^3 n) \]
and it follows from Chernoff’s bound that
\[ \Pr \left[ |Y_i| < \frac{\mathbb{E}[|Y_i|]}{2} \mid |X_i| \geq \frac{\mathbb{E}[|X_i|]}{2} \right] = e^{-\Omega(\log^3 n)} = o(1/n). \] (4.6)

Combining (4.5) and (4.6) we conclude that the probability that our process fails is at most
\[ \sum_{i=1}^{\mathcal{L}} \Pr \left[ |Y_i| \leq [\log^2 n] \right] \leq |\mathcal{L}| \cdot o(1/n) = o(1). \]

Finally, since we choose a random ordering of the neighbors of \(L_i\), every \([\log^2 n]\)-tuple of neighbors of \(L_i\) has the same probability to be part of \(B(\mathcal{L}, V^*)\) and the process therefore samples an element of \(B_{[\log^2 n]}^{\ell_{out}}(F)\) uniformly at random. \(\Box\)
In this chapter we study the behavior of the push process on sparse random graphs with a given degree sequence. We focus on degree sequences that are heavy tailed, in particular on degree sequences that follow a power-law distribution, but our results hold for sparse degree sequences in general. The content of this chapter is joint work with Florian Meier and it appeared in [MP14].

5.1 Results

Recall that in the push process (or protocol/model) on a graph $G$ with $n$ vertices, initially (in round 0) only one arbitrary vertex knows the
Chapter 5. Rumor Spreading in Sparse Random Graphs

In each round \( i \geq 1 \), every vertex that knows the rumor informs uniformly at random one of its neighbors. The push process has a counterpart called pull process, in which the following procedure is executed in every round \( i \geq 1 \): every uninformed vertex asks a randomly chosen neighbor for the rumor.

For a graph or multigraph \( M = (V, E) \), let \( T(M) \) denote the time until all vertices are informed by the push protocol, and for \( 0 \leq \varepsilon \leq 1 \) let \( T_\varepsilon(M) \) denote the time until all but at most \( \varepsilon |V| \) vertices are informed. Before we state the theorem, we quickly introduce the random graph model.

A degree sequence on \( n \) vertices is a sequence of non-negative integers \( D_n = d_{1,n}, d_{2,n}, \ldots, d_{n,n} \), where \( d_{i,n} \) is the degree of the \( i \)-th vertex for \( 1 \leq i \leq n \). We consider infinite sequences \( \mathcal{D} = D_1, D_2, \ldots \), such that for each \( i \geq 1 \) \( D_i \) is a degree sequence of length \( i \). We say that such a sequence \( \mathcal{D} \) is an asymptotic degree sequence if there exist a sequence of real numbers \( \lambda_1, \lambda_2, \ldots \) such that

\[
\lim_{n \to \infty} \frac{|\{ j \mid d_{j,n} = i \}|}{n} = \lambda_i \quad \text{for every } i \geq 1 \text{ and } \lim_{n \to \infty} \sum_{j=1}^{n} d_{j,n} = n \sum_{j=0}^{\infty} j \lambda_j.
\]

Moreover, we say that \( \mathcal{D} \) is 2-smooth if

\[
\sum_{j=1}^{\infty} j^2 \lambda_j < \infty.
\]

A sufficiently large (multi)graph created from a 2-smooth asymptotic degree sequence is always sparse in the sense that the number of edges is linear in the number of vertices, since

\[
\lim_{n \to \infty} \frac{\sum_{j=1}^{n} d_{j,n}}{n} = \sum_{j=1}^{\infty} j \lambda_j \leq \sum_{j=1}^{\infty} j^2 \lambda_j < \infty. \tag{5.1}
\]

Conversely, not every sparse degree sequence is 2-smooth (we demonstrate this later on the example of a power-law with \( \beta < 3 \)).

We denote by

\[
\Delta := \Delta(\mathcal{D}, n) = \max \{d_{i,n} \mid 1 \leq i \leq n\}
\]

the maximum degree of an asymptotic degree sequence \( \mathcal{D} \). Moreover, let \( \delta(\mathcal{D}) := d - 1 \) if there exist an integer \( d \) such that \( \lambda_d = 1 \). Otherwise,
5.1. Results

Let $\delta$ be the minimum degree. By this definition of $\delta$ we assure that there are linearly many vertices of degree larger than $\delta$.

Let $M(D)$ be the sequence of random multigraphs such that the $i$-th element $M_i(D)$ is chosen uniformly at random from all multigraphs with degree sequence $D_i$. We say that $M(D)$ has a given property w.h.p if the probability that $M_n(D)$ has this property goes to 1 as $n$ tends to infinity. We use the same notation for simple graphs but we write $G(D)$ and $G_i(D)$ instead of $M(D)$ and $M_i(D)$. Similar definitions of fixed degree sequences for random graphs have been used in [MR98], [FR04] and [FR07].

**Theorem 5.1 (Main Result).** Let $D$ be a sparse asymptotic degree sequence with $\delta \geq 3$, maximum degree $\Delta$ and let $c_D := \frac{1}{\ln(2(1-\frac{1}{\delta}))}$.

If $\Delta = o(\sqrt{n})$, then for every $n$, every $\varepsilon > 0$ and every $c > c_D$ we have w.h.p

$$T_\varepsilon(M(D)) \leq c \ln n.$$  

If $\Delta$ does not depend on $n$, then we have for every $n$ with probability $1 - o(1/n)$

$$T(M(D)) = O(\ln n).$$

Let us first remark to what extent our theorem extends to simple graphs. It is well known (see Section 5.2) that for a 2-smooth degree sequence $D_n$ a non negligible fraction of all multigraphs is simple, and therefore that every graph property that holds w.h.p in $M(D_n)$ also holds w.h.p in $G(D_n)$. We restate the first statement of our main theorem for simple graphs in the following corollary. The second statement is already known for simple graphs. Since random bounded degree graphs have w.h.p a diameter of order $O(\ln n)$ (see [FR07]), it follows from a result by Feige et. al. [FPRU90] which states that $T(G) = O(\ln n)$ for every bounded degree simple graph $G$ with diameter at most $O(\ln n)$.

**Corollary 5.2.** Let $D$ be a 2-smooth asymptotic degree sequence with $\delta \geq 3$, maximum degree $\Delta$ and let $c_D := \frac{1}{\ln(2(1-\frac{1}{\delta}))}$.

If $\Delta = o(\sqrt{n})$, then for every $n$, every $\varepsilon > 0$ and every $c > c_D$ we have w.h.p

$$T_\varepsilon(G(D)) \leq c \ln n.$$  

Random regular graphs, random bounded degree graphs and random power-law graphs with $\beta > 3$ all have 2-smooth degree sequences. For
a power-law with $\beta < 3$, we have

$$\sum_{i=1}^{\infty} i^2 \lambda_i = \sum_{i=1}^{\infty} \Theta(i^{2-\beta}) = \omega(1),$$

and the degree sequence is therefore not 2-smooth. Unfortunately, this means that Corollary 5.2 does not include power-law degree sequences with $2 < \beta < 3$. But, since on power-law random graphs with $2 < \beta < 3$ push-pull is (at least) exponentially faster than push, the simple push protocol is more interesting for $\beta > 3$.

Requiring $\delta \geq 3$ is not a strong limitation, since $\lambda_2 = 0$ and minimum degree two is necessary to ensure that the random graph is w.h.p connected (see [Luc89] for more details). The only sparse degree sequences that generate w.h.p connected graphs, but are not covered by our theorem, are 3-regular and almost 3-regular graphs (where only $o(n)$ vertices do not have degree 3). For $d$-regular random graphs, Fountoulakis and Panagiotou proved in [FP12] that w.h.p $T(G(D)) = (1 + o(1))c_d \ln n$, where

$$c_d := \frac{1}{\ln(2(1 - \frac{1}{d}))} - \frac{1}{d \ln(1 - \frac{1}{d})}.$$  

Note that $c_d > c_D$ holds for $d$-regular degree sequences $D$ and $d \geq 4$. Therefore, even for very small $\varepsilon > 0$, the time needed to inform the last $\varepsilon n$ vertices is not negligible. This might come as a surprise first, but it is rather simple to explain. In a $d$-regular graph, the probability that a vertex does not receive the rumor, even if all its neighbors are informed, is $\left(\frac{d-1}{d}\right)^d$ in every round. Thus, even for $t \approx -\left(\frac{1}{d \ln(1-1/d)}\right) \ln n$ rounds, the probability that a vertex does not get the information in $t$ rounds, is $\Omega(1/n)$. For a linear fraction of non-informed vertices, we expect that constantly many remain non-informed after $t$ rounds. For large $d$ the difference between $c_d$ and $c_D$ tends to exactly this $-\frac{1}{d \ln(1-1/d)}$. For small $d$ there is a gap between $c_d - \frac{1}{d \ln(1-1/d)}$ and $c_D$, but by treating random regular graphs separately, our proof could be optimized to yield the correct constant.

Having answered these questions about push, it is natural to ask for the running time of its symmetric counterpart pull. Let $\bar{T}(G)$ denote the time pull needs to inform all vertices of a multigraph $M$ and similarly, let $\bar{T}_\varepsilon(M)$ denote the time it needs to inform all but $\varepsilon n$ vertices for a fixed $\varepsilon > 0$. It is well known (see for example [Gia11]) that for every multigraph $M$, every upper bound that holds with probability $1 - f(n)$
for $T(M)$, holds with probability $1 - O(n \cdot f(n))$ for $\bar{T}(M)$ and vice versa. For bounded degree sequences, our result therefore implies the second statement of the Corollary 5.3 below. Note that this does not follow from [FP12], since the result there holds only with probability $1 - o(1)$.

For power-law degree sequences, we observe that push and pull behave very differently. While push fails in informing the last few vertices efficiently, depending on the choice of the initial vertex, pull might need polynomially many rounds to inform the second vertex. More precisely, in a power-law random graph, there is w.h.p a vertex $v$ of degree $O(1)$ that has only neighbors of degree at least $n^{\Omega(1)}$. If $v$ is the initial vertex, then the expected time until the second vertex gets informed is $n^{\Omega(1)}$. This observation implies that for a power-law random multigraph $M$, w.h.p $\bar{T}_\epsilon(M) = n^{\Omega(1)}$ for some initial vertices. However, Theorem 5.1 implies that $\bar{T}_\epsilon(M) = O(\log n)$ holds for almost all initial vertices of a sparse random multigraph $G$.

**Corollary 5.3.** Let $D$ and $c_D$ be as defined in Theorem 5.1. If $\Delta = o(\sqrt{n})$, then w.h.p for every $n$, every $\epsilon > 0$ and every $c > c_D$ we have w.h.p

$$\bar{T}_\epsilon(M(D)) \leq c \ln n$$

for all except $o(n)$ choices of the initial vertex. If $\Delta$ does not depend on $n$, then for every $n$ we have w.h.p

$$\bar{T}(M(D)) = O(\ln n).$$

An interesting open question is whether $\bar{T}(M(D)) = O(\log n)$ holds w.h.p for almost all initial vertices of a random multigraph $M(D)$ with a sparse degree sequence $D$. We conjecture that this is indeed the case.

### 5.2 Degree Sequence and Configuration Model

In the previous section, we introduced the concept of a sparse asymptotic degree sequence. Let us here comment on two properties of such a degree sequence. Recall that we require the existence of an infinite sequence of real numbers $\lambda_1, \lambda_2, \ldots$ such that

$$\lim_{n \to \infty} \frac{|\{j | d_{j,n} = i\}|}{n} = \lambda_i.$$  \hspace{1cm} (5.2)
Thus, it follows that $\lim_{n \to \infty} \sum_{i=0}^{n} \lambda_i = 1$. Intuitively, $\lambda_i$ denotes the fraction of vertices of degree $i$ and we have $(\lambda_i \pm o(1))n$ vertices of degree $i$ in $M(D_n)$. Moreover, note that sparseness implies that

$$\lim_{i \to \infty} \lambda_i i = 0.$$  \hspace{1cm} (5.3)

This observation allows us to handle sparse graphs almost like bounded degree graphs.

A common method for generating random multigraphs with a fixed degree sequence $D_n$ is to create for every $1 \leq i \leq n$ a vertex $v_i$ with $d_{i,n}$ half edges attached to it, and connect the half edges by choosing a random configuration, i.e. a random matching of the half edges. This random graph model is well known as the configuration model, and it was introduced by Bollobás [Bol80] and Bender and Canfield [BC78] independently.

Note that instead of choosing a configuration uniformly at random, we can iteratively choose an arbitrary unmatched stub and match it to a stub that we choose u.a.r. from all unmatched half edges. This technique is called the principle of deferred decisions (see [MU05]), i.e. we delay every random choice until it is most convenient for our analysis. For the remainder of this thesis, we always construct the random matching using deferred decisions. Whenever we match a stub, we denote by $U$ the set of all unmatched half edges and by $e \in_r U$ a stub $e$ that we choose u.a.r. from $U$.

It is a weakness of the configuration model that it creates multigraphs, i.e. the created graph may contain multiple edges and loops. In [Jan09] Janson proved that for a degree sequence $D_n = d_{1,n}, \ldots, d_{n,n}$, the probability that a random multigraph $M(D_n)$ is simple is

$$\Pr[M(D_n) \text{ is simple}] = e^{-\frac{1}{16|E|^2} \left( \sum_i d_i^2 \right)^2 + \frac{1}{2}} + o(1).$$

This implies that for a 2-smooth degree sequence $D_n$ with maximum degree $o(\sqrt{n})$ a non negligible fraction of all configurations is simple, and therefore that every graph property that holds w.h.p in $M(D_n)$ also holds w.h.p in $G(D_n)$. For a sparse degree sequence $D_n$ that is not 2-smooth, an event that holds with probability at least $1 - o(\Pr[M(D_n) \text{ is simple}])$ in $M(D_n)$ holds w.h.p in $G(D_n)$. Thus, for a power-law degree sequence $D_n$ with $2 < \beta < 3$, we would have to prove that our statement holds with probability at least $1 - o(e^{-n^{3/\beta-1}})$ for $M(D_n)$, to prove with a union bound that it holds w.h.p for $G(D_n)$. 
5.3 Proof of the Main Results

We use the configuration model to prove our main theorem for \(M(D)\). Recall that we can match the stubs in the configuration model in an arbitrary order. For our purpose, it is very convenient to match a stub only when it is selected by the push process. Thereby, we combine the process that spreads the rumor with the process that matches the stubs. An important element of our proof is that we sometimes \textit{delay} a push. This means that for an unmatched stub \(e\) that is selected by the push protocol, we postpone matching \(e\) and thus also the push. A delayed stub may be matched and pushed in a later round or it may as well be omitted till the end. Note that delaying pushes can not make the spreading faster and therefore, since we are only interested in upper bounds on the rumor spreading, it is legitimate to delay pushes. We denote the process that combines the delayed push process with the matching of the stubs as the \textit{delayed random graph push process} or \textit{DRP process}. For the proof of Theorem 5.1, we split the execution of the DRP process in three phases.

Before we explain the three phases and state the main lemmas, we need to introduce some additional notation and constants. Let \(0 < \gamma \leq 1/6\) be a small enough constant, such that

\[
\frac{1 - \lambda_\delta}{64\delta^2} > \frac{2\gamma}{1 - \gamma} \quad \text{and} \quad \gamma < \frac{1}{3\delta}, \quad (5.4)
\]

which is possible since \(\lambda_\delta < 1\). Moreover, let \(M\) be the smallest integer such that

\[
\sum_{j=\delta}^{M} \lambda_{j,j} \geq 1 - \frac{\gamma}{4}\]

and let

\[
\alpha := \min \left\{ \frac{\gamma \sum_{j=\delta}^{\infty} \lambda_{j,j}}{2 \cdot M}, \frac{1 - \lambda_\delta}{1 + \lambda_\delta}, \frac{(\delta - 1)}{4\delta(1 - \frac{1}{\delta})} \right\} \Theta(1). \quad (5.5)
\]

For \(i \geq 1\) let \(I_i\) denote the set of \textit{informed} vertices after round \(i\) and let \(N_i\) denote the set of \textit{newly informed} vertices in round \(i\). Note that this definition of \(\alpha, M\) and \(\gamma\) assures that for \(|I_i| < \alpha n\) the probability to match to an uninformed vertex of degree at most \(M\), when choosing u.a.r. from the unmatched stubs \(U\) in round \(i\), is at least

\[
(1 - o(1)) \frac{n \sum_{j=\delta}^{M} \lambda_{j,j} - |I_i| M}{n \sum_{j=\delta}^{\infty} \lambda_{j,j}} \geq (1 - o(1))(1 - \frac{3\gamma}{4}) > 1 - \gamma. \quad (5.6)
\]
We neglect vertices of degree larger than $M$ whenever possible. Let therefore $\bar{N}_i$ denote the set of non sleeping vertices of degree at most $M$ that have been newly informed in round $i$.

With this definitions at hand, we are ready to explain the three phases. For the sake of simplicity, we do not repeat the whole framework in the statements of the following lemmas, but we always make the assumption of Theorem 5.1.

In the beginning, it is very unlikely that a vertex is informed more than once. Therefore, the rumor spreading builds a tree with branching factor at least $\delta - 1$. This guarantees exponential growth as $\delta - 1 \geq 2$.

**Lemma 5.4.** Let $t_1$ be the smallest integer such that $|\bar{N}_{t_1}| \geq \log^5 n$. It holds w.h.p that we can delay the DRP process so that $t_1 = O((\log \log n)^2)$ and $|I_{t_1}| = O(\log^5 n)$. Moreover, if the the maximum degree $\Delta$ is constant, we can delay the DRP process so that $t_1 = O(\log n)$ holds with probability $1 - o(1/n)$.

The second phase is longer and its running time dominates the total running time of the rumor spreading. In this phase, we want to inform a linear fraction of the vertices, and we therefore have to deal with vertices that are informed more than once. But since we have many vertices of degree larger than $\delta$, we can compensate for this and show that there is still exponential growth. The proof of the following lemma contains the main ideas of our analysis.

**Lemma 5.5.** Let $t_2$ be the smallest integer such that $|I_{t_2}| \geq \alpha n$. It holds with probability $1 - o(1/n)$ that we can delay the DRP process so that $t_2 - t_1 \leq c \ln n$.

In the third phase, we handle the spreading to the remaining vertices for random graphs of bounded and unbounded maximum degree.

**Lemma 5.6.** For $\varepsilon > 0$, let $t_3$ be the first round such that $|I_{t_3}| > (1-\varepsilon)n$ and let $t_4$ be the first round such that $|I_{t_4}| = n$. It holds w.h.p that we can delay the DRP process so that $t_3 - t_2 = o(\log n)$. Moreover, if $\Delta$ does not depend on $n$, then it holds with probability $(1 - o(1/n))$ that we can delay the DRP process so that $t_4 - t_2 = O(\log n)$.

The three lemmas together clearly imply Theorem 5.1. We prove Lemma 5.4 in Section 5.3.2, Lemma 5.5 in Section 5.3.2, and Lemma 5.6 in Section 5.3.3. We believe that the presented proofs give a good intuition
on the dynamics of the rumor spreading process, and that the techniques we use could be helpful for the analysis of other processes on random multigraphs.

### 5.3.1 Phase 1

This section is devoted to the proof of Lemma 5.4.

**Proof of Lemma 5.4** We consider the process until for the first time at least $\log^5 n$ vertices become newly informed. We control the process (by delaying pushes) so that the informed vertices build a 3-regular tree. Let $T_1 := \{v\}$, where $v$ is the initial vertex and for $j > 1$, let $T_j$ denote the set of vertices on the $j$-th level of the tree. On level $j \geq 1$, we delay all pushes for $t_j$ (to be defined later) rounds. For every vertex $v \in T_j$, let $E_v$ denote the set of unmatched stubs that have been selected by the push process during those $t_j$ rounds. We select from every set $E_v$ at most two (three on the root) stubs and match them u.a.r. to stubs in $U$. Finally, we inform the first $\lceil (4/3)^{j-1} \rceil$ (or all if there are less) vertices that have been matched to vertices in $T_j$, and denote the set of newly informed vertices by $T_{j+1}$. We iterate this process as long as $j \leq h$ for

$$h := \frac{5 \log \log n}{\log(4/3)} + 2.$$
and the probability that a branch dies because the corresponding vertex did not select two stubs during the $t_j$ rounds is at most $(2/3)^{t_j}$.

Altogether, the probability that a branch at level $j$ dies is at most

$$
\chi_j := \sum_{i=1}^{j-1} |T_i| \Delta + 2|T_j| \left( n - \sum_{i=1}^{j-1} |T_i| \right) / 3 + \left( \frac{2}{3} \right)^{t_j}.
$$

Let

$$
p_j := \Pr \left[ |T_j| < \left( \frac{4}{3} \right)^{j-1} \left| T_j \right. \geq \left( \frac{4}{3} \right)^{j-2} \right],
$$

and note that under the assumption of $|T_{j-1}| \geq \left( \frac{4}{3} \right)^{j-2}$, at least $\frac{2}{3} \left( \frac{4}{3} \right)^{j-2}$ of the $2 \left( \frac{4}{3} \right)^{j-2}$ branches on level $j$ have to die, to satisfy the event $|T_j| < \left( \frac{4}{3} \right)^{j-1}$. The probability for this is at most

$$
p_j \leq \left( \frac{2}{3} \right)^{\left( \frac{4}{3} \right)^{j-2}} \cdot \chi_j^{\left( \frac{4}{3} \right)^{j-2}} \leq 2^j \chi_j^{\left( \frac{4}{3} \right)^{j-2}} = \left( 4 \chi_j^{\frac{3}{2}} \right)^{\left( \frac{4}{3} \right)^{j-2}}.
$$

Moreover, it is not very hard to see that $|T_3| < (4/3)^2$ holds only if at least three branches on the first two levels die. The probability for this is at most

$$
\Pr[|T_3| < (4/3)^2] = O(\chi_3^3).
$$

Combining (5.7) and (5.8), we conclude that the probability for $|T_h| < \left( \frac{4}{3} \right)^{h-1} = \frac{4}{3} \log^3 n$ is at most

$$
\Pr[|T_3| < (4/3)^2] + \sum_{j=4}^{h} p_j \leq O(\chi_3^3) + \sum_{j=4}^{h} \left( 4 \chi_j^{\frac{3}{2}} \right)^{\left( \frac{4}{3} \right)^{j-2}}.
$$

With this at hand, we can prove the first statement of the lemma. Let

$$
t_j = \left( \frac{3 \log \log n}{2 \log(3/2)} \right) \text{ for all } 1 \leq j \leq h.
$$

Clearly, the number of rounds we need to build $T_h$ is

$$
\sum_{j=1}^{h} t_j = O((\log \log n)^2),
$$

and, since

$$
\chi_j = O \left( \frac{2^i}{\sqrt{n}} + \frac{1}{\log^{3/2} n} \right) = O \left( \frac{1}{\log^{3/2} n} \right),
$$
5.3. Proof of the Main Results

the probability that there are less than \( \frac{4}{3} \log^5 n \) newly informed vertices in \( T_h \) is therefore by (5.9) at most \( O(\log \log n / \log^{4/3} n) = o(1) \). It remains to show that at least \( \log^5 n \) of these vertices have degree at most \( M \). It follows from (5.6) that every newly added vertex is with probability at least \( 1 - \gamma \) of degree at most \( M \). Moreover, this probability bound holds independently of the degrees of all the other vertices in the tree. Thus, it follows from Theorem 2.1 that with probability \( 1 - n^{-\omega(1)} \) at least \( \log^5 n \) of the newly informed vertices are of degree at most \( M \).

For bounded degree graphs, we set \( t_j := 36 \log n / (j^2 \log(3/2)) \) for \( 1 \leq j \leq h \), and the time needed to build the tree is therefore at most

\[
\frac{36 \log n}{\log(3/2)} \sum_{j=1}^{h} \frac{1}{j^2} = O(\log n).
\]

Moreover, since for \( 1 \leq j \leq h \) we have that

\[
\chi_j = O \left( \frac{2^j}{\sqrt{n}} \right) + O \left( n^{-\frac{36}{j^2}} \right),
\]

it is not hard to see that the probability that there are less than \( \frac{4}{3} \log^5 n \) newly informed vertices in \( T_h \) is by (5.9) at most

\[
O \left( \chi_3^3 \right) + \sum_{j=4}^{h} \left( 4 \chi_j^{2/3} \right)^{(4/3)^j - 2} = o(1/n).
\]

Finally, since we only allow \( \lceil (\frac{4}{3})^{j-1} \rceil \) vertices on level \( j \) for \( 1 \leq j \leq h \), the number of informed vertices is at most \( O(\log^5 n) \).

5.3.2 Phase 2

Lemma 5.5 states that the exponential growth, exploited in the first phase, continues until a linear fraction of the vertices is informed. Unfortunately, when \( |I_i| \) is large, we can not neglect the probability that a selected stub is matched to an already informed vertex. We call this a back match. On the other hand, since we have many vertices of degree strictly larger than \( \delta \geq 3 \), the branching factor is often larger than \( \delta - 1 \geq 2 \).

First, we introduce a \( (\delta - 1) \)-ary random tree process. It is not very hard to show that this tree process has exponential growth. Then, we
couple the DRP process to the tree process. Our coupling guarantees that the number of newly informed vertices of degree at most $M$ is always exactly the number of new born vertices in the tree process, and we thereby prove an exponential lower bound on the growth of the DRP process.

The Tree Process

The tree process with branching factor $\delta - 1$ is defined as follows. Every vertex has $\delta$ stubs attached to it. A stub is either free, if it has never been selected, or used, if it is already matched. We start in round $t_1 + 1$ with $\lceil \log^5 n \rceil$ vertices, of which each one has exactly one stub that is used and the other $\delta - 1$ stubs are free. Note that the time shift of $t_1$ is convenient to couple the tree process with phase 2 of the push process. In round $i > t_1$, we select at every vertex one stub u.a.r from all its stubs. We denote the set of free stubs that have been selected in round $i$ as $S^T_i$. Algorithm 1 then iterates through all stubs $e \in S^T_i$ and creates a new vertex $u$ at $e$.

**Algorithm 1** Tree process for $S^T_i$

```plaintext
for $e \in S^T_i$ do
    Create a new vertex $u$ and match a stub of $u$ to $e$.
end for
```

For $i \geq t_1 + 1$, let $N^T_i$ be the set of newly created vertices in round $i$, and let $P^T_i$ be the set of free stubs after round $i$. The following lemma quantifies the exponential growth.

**Lemma 5.7.** Let $|N^T_{t_1}| = \lceil \log^5 n \rceil$ and let $\delta \geq 3$. It holds with probability $1 - o(1/n)$ for all $t_1 + 1 \leq i = O(\log n)$ that

$$|P^T_i| = (1 + o(1))|P^T_{t_1}| \left(2 \left(1 - \frac{1}{\delta}\right)\right)^{i-t_1} \quad (5.10)$$

and

$$|N^T_i| = (1 + o(1)) \frac{1}{\delta} |P^T_{i-1}|. \quad (5.11)$$

**Proof.** In every round $i$, a free stub $e \in P^T_{i-1}$ creates $\delta - 1$ new stubs with probability $1/\delta$, or remains free with probability $(\delta - 1)/\delta$. This
observation already implies that the expectation of $|P^T_i|$ conditioned on $|P^T_{i-1}|$ is
\[
E\left[ |P^T_i||P^T_{i-1}| \right] = |P^T_{i-1}| \left( \frac{2(\delta - 1)}{\delta} \right)
\]
and that the expected number of new vertices in round $i$ given the number of free stubs in round $i - 1$ is
\[
E\left[ |N^T_i||P^T_{i-1}| \right] = |P^T_{i-1}| \frac{1}{\delta}.
\]
We start with $|P^T_{t_1}| \geq (\delta - 1)|N^T_{t_1}|$ free stubs and it therefore remains to show that $|P^T_i|$ and $|N^T_i|$ are for all $t_1 < i = O(\log n)$ sufficiently concentrated around its expectation. Since exactly one stub at a vertex is selected in every round, the events are far from independent. To gain independence, we distribute the free stubs in $P^T_{i-1}$ in $\delta - 1$ sets $P_{i-1,1}, \ldots, P_{i-1,\delta-1}$ such that every set has either cardinality $|P^T_{i-1}|/(\delta - 1)$ or $|P^T_{i-1}|/(\delta - 1) + 1$, and such that no two stubs of the same vertex are in the same set. Then, we apply the Chernoff bound (see Theorem 2.1) to show that with probability $1 - O(\log^2 n)$ (where we can choose the $c$ arbitrarily large) for all those sets the number of stubs that create a new vertex and the number of stubs that remain free, are within a factor $(1 \pm O(1/\log^2 n))$ of their expectation. The lemma follows by a union bound, since we have $3(\delta - 1)$ random variables, $O(\log n)$ rounds, and since
\[
(1 \pm O(1/\log^2 n))^{O(\log n)} = (1 \pm o(1)).
\]
\[\square\]

We remark that Lemma 5.7 implies that
\[
|N^T_i| \geq (1 \pm o(1)) \frac{1}{\delta} |P^T_{i-1}| = (1 \pm o(1)) \frac{1}{\delta} \left( 2 \left( 1 - \frac{1}{\delta} \right) \right)^{i-1-t_1} |P^T_{t_1}|
\]
\[
\geq (1 \pm o(1)) \frac{(\delta - 1)}{\delta} \left( 2 \left( 1 - \frac{1}{\delta} \right) \right)^{i-1-t_1} \log^5 n
\]
holds with probability $1 - o(1/n)$ for all $t_1 < i \leq O(\log n)$. Recall that $c_D = \frac{1}{\ln(2(1-\frac{1}{\delta}))}$, and note that the inequality
\[
t_1 + c_D(\ln n - 5 \ln \log n) \leq t_1 + c \ln n
\]
(5.12)
holds for $n$ large enough. Therefore, for $i \geq t_1 + c \ln n$, we have with probability $1 - o(1/n)$

$$|N_i^T| \geq (1 - o(1)) \frac{\delta - 1}{\delta} \left(2 \left(1 - \frac{1}{\delta}\right)^{t_1 + cD(\ln n - 5 \ln \log n) - t_1} \log^5 n\right) $$

$$\geq (1 - o(1)) \frac{(\delta - 1)}{2\delta(1 - \frac{1}{\delta})} n > \frac{(\delta - 1)}{4\delta(1 - \frac{1}{\delta})} n \geq \alpha n. \quad (5.13)$$

**Coupling the two Processes**

In order to apply the bound derived in (5.12) and (5.13) to $t_2$, we couple the DRP process to the tree process, in a way that $|\bar{N}_i| = |N_i^T|$ holds after every round $i \geq t_1$. We first explain the coupling and show then that we can run it long enough to inform at least $\alpha n$ vertices.

Recall that we consider the random graph process $M(D)$ with degree sequence $D_n$ and vertex set $V$, where every vertex $v \in V$ has $d_v$ stubs attached. In round $i \geq t_1 + 1$, the DRP process selects at every informed vertex $v \in I_i$ one of the $d_v$ stubs uniformly at random. We use the same definition of free and used stubs as in the tree process. However, the fundamental difference to the tree processes is that a free stub $e'$ can become the target of a back match. In this case, $e'$ is matched without being able to increase the number of informed vertices. But since the degree of most of the vertices is larger than $\delta$, we have in every round more stubs that are selected for the first time than in the tree model. Therefore, we can assign to every target of a back match $e'$ another selected (and delayed stub) twin($e'$) that can spread the rumor instead of $e'$, as soon as $e'$ is selected in for the first time. Moreover, we ignore all vertices of degree larger than $M$, as informing such vertices would increase the probability of back matches in later rounds. Whenever we match a free selected stub $e$ to a stub $e' \in_r U$, we say that $e'$ is good if it is attached to an uninformed vertex of degree at most $M$.

Recall that $\bar{N}_i$ denotes the set of non-sleeping vertices of degree at most $M$ that have been newly informed in round $i$. We start the coupling with $\lceil \log^5 n \rceil$ vertices in $\bar{N}_{t_1}$ at time $t_1 + 1$. Note that we can do so by delaying the stubs at all but $\lceil \log^5 n \rceil$ vertices in $\bar{N}_{t_1}$ forever. Clearly, for $i = t_1$ it holds that $|\bar{N}_i| = |N_i^T|$, and our coupling will maintain this invariant. At time $i > t_1$ we select at every vertex $v \in \cup_{j=t_1}^{i-1} \bar{N}_j$ one stub u.a.r. and define $\bar{S}_i$ to be the set of all free stubs (or twins of free stubs) that have been selected. For an exact simulation of the DRP process,
we have to match or delay all stubs in $\tilde{S}_i$. We couple this decision with
the random tree process, by first running round $i$ of the tree process
and then simulating the DRP process, so that we inform exactly $|N^T_i|$ new vertices.

Let $\tilde{S}_i$ and $S^T_i$ be the sets of free stubs that are selected in round $i$ of
the DRP process and the tree process. We first run Algorithm 1 on $S^T_i$
and then Algorithm 2 on $\tilde{S}_i$. Algorithm 2 iteratively selects a stub $e$
from $\tilde{S}_i$ and matches it to a randomly chosen stub $e' \in R \cup U$. If $e'$ is
good, it informs the vertex at $e'$, otherwise it assigns a stub from $\tilde{S}_i$ as
twin to $e'$. The algorithm terminates as soon as the number of newly
informed good vertices is exactly the number of new vertices in the tree
model.

**Algorithm 2** Coupled DRP

```plaintext
while $|\tilde{N}_i| < |N^T_i|$ do
    remove an arbitrary stub $e$ from $\tilde{S}_i$.
    choose $e' \in R \cup U$ and match $e$ to $e'$.
    if $e'$ is good then
        inform the vertex at $e'$.
    else
        remove an arbitrary stub $e_t$ from $\tilde{S}_i$ and set $e_t$ sleeping.
        $\text{twin}(e') := e_t$.
    end if
end while
```

Clearly, Algorithm 2 assures that, as long as $\tilde{S}_i$ is large enough, $|\tilde{N}_i|$ is
always equal to $|N^T_i|$, and we can therefore apply the bounds given by
Lemma 5.7 to $|\tilde{N}_i|$. The following lemma states that as long as we are
in Phase 2, we always have enough free stubs in $\tilde{S}_i$.

**Lemma 5.8.** The condition $\tilde{S}_i \neq \emptyset$ is with probability $1 - o(1/n)$ never violated when running Algorithm 2 with $|I_i| < \alpha n$.

Before we prove Lemma 5.8, we show how it implies this section’s main lemma.

**Proof of Lemma 5.5.** It is not very hard to see that Algorithm 2 correctly simulates the DRP process. The lemma therefore follows from
Lemma 5.8 since $\alpha n \leq |N^T_{t_1 + c \cdot \ln n}|$, as we already mentioned in (5.12),
holds with probability $1 - o(1/n)$ and since $|N^T_i| = |\bar{N}_i| \leq |I_i|$ holds for every $i \geq t_1 + 1$ for which $|I_i| \leq \alpha n$.

It remains to prove Lemma 5.8.

**Proof of Lemma 5.8.** We assume during the whole proof that (5.10) and (5.11) hold for all considered integers $i$. We fix a round $t_1 < i \leq O(\log n)$ and assume that the statement was true for all previous rounds. Based on this assumption, we prove that the statement of the lemma holds with probability $1 - o(1/n^2)$ for round $i$. The lemma then follows by induction (the statement is by definition true for $i = t_1$) and union bound.

Let $\bar{N}_{i,k} := \{v \in \bar{N}_i | \deg(v) = k\}$ and let $m_{i,k} := |\bar{N}_{i,k}|/|\bar{N}_i|$. We claim that the expected number of selected stubs in round $i$ is

$$\sum_{t=t_1}^{i-1} |N_t| \left( \frac{\delta - 1}{\delta} \right)^{i-t}$$

in the tree model, and

$$\sum_{t=t_1}^{i-1} \sum_{k=\delta}^M |\bar{N}_t| \cdot m_{t,k} \left( \frac{k-1}{k} \right)^{i-t}$$

in the push process. Indeed, let $v$ be a vertex of degree $k$ that has been informed in round $t$. It is not very hard to see that the probability that one of the stubs at $v$ is selected for the first time in round $t + j$ is exactly $\left( \frac{k-1}{k} \right)^j$. Hence, the expectation of $|S_i| := |\bar{S}_i| - |S^T_i|$ is at least

$$\sum_{t=t_1}^{i-1} \sum_{k=\delta}^M |\bar{N}_t| \cdot m_{t,k} \left[ \left( \frac{k-1}{k} \right)^{i-t} - \left( \frac{\delta - 1}{\delta} \right)^{i-t} \right]$$

$$\geq \sum_{k=\delta}^M |\bar{N}_{i-1}| \cdot m_{i-1,k} \left[ \frac{k-1}{k} - \frac{\delta - 1}{\delta} \right]$$

$$\geq \sum_{k=\delta+1}^M |\bar{N}_{i-1}| \cdot m_{i-1,k} \left[ \frac{\delta}{\delta + 1} - \frac{\delta - 1}{\delta} \right] \geq \frac{1}{\delta^2 + \delta} |\bar{N}_{i-1}| (1 - m_{i-1,\delta}).$$

If we consider the random variable $|S_i|$ in the setting of Theorem 2.2 then the coordinates are the outcome of the random experiments that select a stub at every vertex (in both processes) in every round. The effect of one coordinate $d_i$ is therefore at most one, and the number of
coordinates that can contribute is at most
\[
2 \left| \bigcup_{j=t_i}^{i-1} N_j^T \right| \cdot i = O \left( |N_{i-1}^T| \log n \right),
\]
where the equality follows by Lemma 5.7 and the induction hypothesis. Hence, by Theorem 2.2 and since $|N_{i-1}^T| \geq \log^5 n$, it follows that
\[
\Pr \left[ |S_i| \leq \frac{1}{2(\delta^2 + \delta)} |N_{i-1}^T|(1 - m_{i-1, \delta}) \right] = n^{-\Omega(\log^3 n)} = o(1/n^2)
\]
as long as $1 - m_{i-1, \delta} = \Omega(1)$, which we will prove later. For now, we assume that $|S_i| \geq \frac{1}{2(\delta^2 + \delta)} |N_{i-1}^T|(1 - m_{i-1, \delta})$.

Note that $|S_{i+1}|$ is at least $|N_i^T|$ and that therefore
\[
|S_i| \geq |N_i^T| + |S_i| \geq |N_i^T| + \frac{1}{2(\delta^2 + \delta)} |N_{i-1}^T|(1 - m_{i-1, \delta}).
\]

Whenever we match a new stub $e$ of $S_i$ to a stub $e' \in R U$, the probability that $e'$ is good is, by (5.6), at least $1 - \gamma$. Suppose that we match $|N_i^T| + |S_i|/2$ times a stub of $S_i$, and let $X$ count the number of times we match to a good stub. We will show that $X$ is with probability $1 - o(1/n^2)$ at least $|N_i^T|$, and thus $S_i$ becomes never empty, since we match at most $|N_i^T| + |S_i|/2 - X \leq |S_i|/2$ many times to stubs that are not good, so that the remaining $|S_i|/2$ stubs suffice as twins.

It follows from Lemma 5.7 that
\[
|N_i^T| + |S_i|/2 \geq |N_i^T| + \frac{|N_{i-1}^T|(1 - m_{i-1, \delta})}{4(\delta^2 + \delta)} \geq |N_i^T| \left( 1 + \frac{1 - m_{i-1, \delta}}{8(\delta + 1)(2\delta - 2)} \right) \geq |N_i^T| \left( 1 + \frac{1 - m_{i-1, \delta}}{16\delta^2} \right),
\]
and the expected number of matched stubs that are good is therefore at least
\[
\mathbb{E}[X] \geq (1 - \gamma)|N_i^T| \left( 1 + \frac{1 - m_{i-1, \delta}}{16\delta^2} \right).
\]

Note that each stub is matched to a good stub with probability at least $(1 - \gamma)$ (independently of the other matches), and that $|N_i^T| = \Omega(\log^5 n)$. In order to prove (by Theorem 2.1) that $X \geq |N_i^T|$ holds
with probability $1 - n^{-\Omega(\log^4 n)}$, it therefore suffices to show that

$$(1 - \gamma) \left( 1 + (1 - m_{i,\delta}) \frac{1}{16\delta^2} \right) \geq 1 + \gamma. \quad (5.16)$$

Recall that $m_{i-1,\delta} = \frac{|\tilde{N}_{i-1,\delta}|}{|\tilde{N}_{i-1}|}$ and observe that $|\tilde{N}_{i-1,\delta}|$ is the sum of $|\tilde{N}_{i-1}|$ indicator random variables, which are one with probability at most

$$D_{\delta} \delta \left( \frac{n - |I_{i-1}| - D_{\leq \delta}}{n - |I_{i-1}|} \right) \leq 1 - \frac{n - |I_{i-1}| - D_{\leq \delta}}{n - |I_{i-1}|} \leq 1 - \frac{1 - \alpha - (1 + o(1))\lambda_{\delta}}{1 - \alpha},$$

where $D_{\delta} := |\{j \mid d_{j,n} = \delta\}|$ and $D_{\leq \delta} := |\{j \mid d_{j,n} \leq \delta\}|$. As this bound holds for every i.r.v. deterministically, we conclude by Theorem 2.1 that

$$\Pr \left[ |\tilde{N}_{i-1,\delta}| \geq \left( 1 - \frac{1 - \alpha - \lambda_{\delta}}{2(1 - \alpha)} \right) |\tilde{N}_{i-1}| \right] \leq n^{-\Omega(\log^4 n)}.$$

It follows that with probability $1 - o(1/n^2)$

$$m_{i-1,\delta} \leq 1 - \frac{1 - \alpha - \lambda_{\delta}}{2(1 - \alpha)} \Omega(1), \quad (5.17)$$

which settles (5.14). Recalling $\alpha \leq (1 - \lambda_{\delta})/(1 + \lambda_{\delta})$ from (5.5) we finish the proof by showing (5.16) as follows

$$(1 - \gamma) \left( 1 + (1 - m_{i-1,\delta}) \frac{1}{16\delta^2} \right) \geq (1 - \gamma) \left( 1 + \left( \frac{1}{2} - \frac{\lambda_{\delta}}{2(1 - \alpha)} \right) \frac{1}{16\delta^2} \right) \geq (1 - \gamma) \left( 1 + \frac{2\gamma}{(1 - \gamma)} \right) = 1 + \gamma.$$

\[\square\]

### 5.3.3 Phase 3

The following lemma states that for a fixed set of informed vertices $I_{t_1} \subset V$ of size $\lceil \alpha n \rceil$, there exists w.h.p for almost every vertex in $V \setminus I_{t_1}$ a short path that contains only vertices of small degree and ends at a vertex in $I_{t_1}$. 
Lemma 5.9. Let $t_2$ such that $|I_{t_2}| = \lceil \alpha n \rceil$ and let $\varepsilon > 0$. Then, there exist w.h.p for all but at most $o(n)$ vertices $v \in V \setminus I_{t_2}$ a path $v, v_1, \ldots, v_k$ such that

1. $v_k \in I_{t_2}$
2. $k \in O(\log \log n)$
3. For all $1 \leq i \leq k$ it holds that $\deg(v_i) \leq \log \log n$.

If the degree sequence is bounded with constant maximum degree $\Delta$, then there exists with probability $1 - o(1/n)$ for every vertex $v \in V \setminus I_{t_2}$ a path $v, v_1, \ldots, v_k$ with properties 1. and 2.

Proof. First, we claim that whenever $|I_i| = \alpha n$ there are at least $\alpha n$ unmatched stubs at informed vertices. We observe that there are in total at least $N := \alpha n \delta$ stubs at informed vertices. If $F$ of those $N$ stubs are unmatched, then there are $(N - F)/2$ edges between vertices in $I_i$ and therefore $(N - F)/2 - \alpha n + 1$ of those edges where back matches. The probability of a back match is, by (5.6), at most $\gamma$ and it is a straightforward application of Theorem 2.1 to show that for $F < \alpha n$ the probability of this event is $e^{-\Omega(n)}$. We therefore assume that we have at least $\alpha n$ unmatched stubs at informed vertices.

Let $v \in V \setminus I_{t_2}$ be an arbitrary vertex. We show that $v$ is w.h.p the starting vertex of a path that satisfies the properties of the lemma. Similarly as in the proof of Lemma 5.4 we describe a process that builds a 3-regular tree $T$ with root $v$, by matching first three stubs at $v$ and then recursively two new stubs at every neighbour. Note that we can always assume that all the vertices in the tree are uninformed, since otherwise the desired path is already established. Whenever we match to a vertex that is already in the tree, we let the corresponding branch of the tree die.

Let $|T_i|$ count the number of vertices on the $i$-th level of the tree. It follows from the sparseness of the degree sequence, in particular from

$$\sum_{j=0}^{\infty} \lambda_j j = O(1),$$

that there exist a function $f(n) = o(1)$ such that

$$\sum_{j=\log \log n}^{n} \lambda_j j = O(f(n)).$$
and therefore that the number of stubs at vertices of degree at least \( \log \log n \) is
\[
\sum_{j=\log \log n}^{n} d_{j,n} = O(n \cdot f(n)).
\]

As long as we do not hit a stub at an informed vertex, we have at least \( |U| = \Omega(n) \) unmatched stubs and the probability that we match to a vertex of degree at least \( \log \log n \) or a vertex that is already in the tree when we choose a stub u.a.r. from \( U \) is therefore at most
\[
\chi_i = O\left( \frac{n \cdot f(n) + \sum_{j=1}^{i} |T_j| \log \log n}{n} \right),
\]
which is \( O(f(n) + \log^2 n/n) = o(1) \) for \( \sum_{j=1}^{i} |T_j| = O(\log n) \). Let
\[
p_i := \Pr \left[ |T_i| < \left( \frac{6}{5} \right)^{i-1} \left| T_{i-1} \right| \geq \left( \frac{6}{5} \right)^{i-2} \right] \quad \text{and let} \quad h := \frac{\log \log n}{\log(6/5)} + 1.
\]

If we assume that \( |T_{i-1}| \geq \left( \frac{6}{5} \right)^{i-2} \), then at least \( 4/5 \left( \frac{6}{5} \right)^{i-2} \) of the \( 2 \left( \frac{6}{5} \right)^{i-2} \) stubs at level \( i-1 \) have to die in order to have \( |T_i| < \left( \frac{6}{5} \right)^{i-2} \). This happens with probability at most
\[
p_i \leq \left( \frac{2}{5} \right)^{i-2} \left( \frac{4}{5} \right)^{i-2} \chi_i^{\frac{4}{5}} \left( \frac{6}{5} \right)^{i-2} \leq 2^2 \left( \frac{6}{5} \right)^{i-2} \chi_i \left( \frac{6}{5} \right)^{i-2} \leq \left( 4\chi_i^{\frac{4}{5}} \right)^{\left( \frac{6}{5} \right)^{i-2}}.
\]

Moreover, it is not hard to see that
\[
\Pr \left[ |T_2| < \left( \frac{6}{5} \right) \right] = \Pr[|T_2| < 2] = O(\chi_2^2),
\]
and the probability that \( |T_h| < \log n \) is therefore at most
\[
Pr \left[ |T_2| < \frac{6}{5} \right] + \sum_{i=3}^{h} p_i \leq O(\chi_2^2) + \sum_{i=3}^{15} \left( 4\chi_{15}^{4/5} \right)^{(\frac{6}{5})^{i-2}} + \sum_{i=1}^{\infty} \left( 4\chi_h^{4/5} \right)^i
\]
\[
= O(\chi_{15}^{4/5}) + \frac{4\chi_h^{4/5}}{1 - 4\chi_h^{4/5}} = o(1).
\]

Thus, if we stop the tree process on level \( h \) we have w.h.p at least \( \log n \) free stubs. On the other hand, the vertices in \( I_{t_2} \) have at least \( \alpha n = \Omega(n) \)
free stubs to which we can match. Hence, whenever we match one of the stubs at level $h$, we have a constant probability $\bar{p}$ to match to a stub of a vertex in $I_{t_2}$, and the probability that we do not match to at least one of those stubs is at most

$$\bar{p}^{\Omega(\log n)} = n^{-\Omega(1)}.$$  

We conclude that, as long as $|T|$ has enough vertices on level $h$ and we match at least one of them to $I_{t_1}$, vertex $v$ is adjacent to a path that satisfies the conditions of the lemma. Since this happens with probability at least $1 - o(1)$, the expected number of vertices for which no such path exist is at most $o(n)$. It therefore follows from Markov’s inequality that w.h.p for all but $o(n)$ vertices such a path exists.

In order to prove the second statement of the lemma, we need to prove that the above calculations hold with probability $1 - o(1/n^2)$ for bounded degree sequences. Let

$$h := \frac{\log \log n^3 - \log \log (\frac{\alpha}{\Delta})}{\log (6/5)} + 1$$

be the height of the tree that we build. For a constant maximum degree $\Delta$, the probability $\chi_i$ is at most

$$\chi_i = O\left(\frac{2^i \Delta}{n^\delta}\right) = O\left(\frac{2^i}{n}\right)$$

for $i \leq h$ and $|V \setminus I_{t_1}| = \Omega(n)$. It is therefore not very hard to see that

$$\Pr \left[ |T_7| < \left(\frac{6}{5}\right)^6 \right] = \Pr[|T_7| < 3] = O(1/n^3)$$

and it follows by (5.18) that for $8 \leq i \leq h$

$$p_i = O\left(\frac{\log n}{n}\right)^{\frac{4}{5}(\frac{6}{5})^6} = o(n^{-2.3}).$$

Thus, the probability that we have at least $(\frac{6}{5})^{h-1} = 3 \log n / \log (\alpha/\Delta)$ free stubs on level $h$ is at least

$$1 - \left( \Pr \left[ |T_7| < \left(\frac{6}{5}\right)^6 \right] + \sum_{i=8}^{h} p_i \right) = 1 - o(1/n^2).$$
Finally, since the probability that we match to an informed vertex when we match the stubs at \( T_h \) is at least \( \bar{p} := \frac{\alpha n}{\Delta n} = \frac{\alpha}{\Delta} \) for every stub, the probability that we match none of the stubs to an informed vertex is at most \( \bar{p}^3 \log n / \log(\alpha / \Delta) = O(1/n^3) \), and the second statement of the lemma therefore follows by a first moment argument.

With Lemma 5.9 at hand, we can prove Lemma 5.6 by exploiting only properties of the random graph.

**Lemma 5.6.** Note that Lemma 5.9 is stated for \( |I_{t_2}| = \lceil \alpha n \rceil \). This is not a problem since we can switch from phase 2 to phase 3 as soon as \( \lceil \alpha n \rceil \) vertices are informed.

We first handle unbounded degree sequences. By Lemma 5.9, we have w.h.p for all but \( o(n) \) vertices a good path. We claim that the rumor spreads over such a path with probability at least \( 1 - O(1/\log^2 n) \) in \( o(\log n) \) rounds. Let \( t := 3(\ln \log n) \cdot (\log \log n) \). Then for all \( 1 \leq j \leq k \), the probability that it takes more than \( t \) rounds until \( v_j \) pushes the rumor to \( v_{j-1} \) is at most

\[
\left( 1 - \frac{1}{\log \log n} \right)^t \leq e^{-\frac{t}{\log \log n}} = \frac{1}{\log^3 n},
\]

and the probability that there is at least one vertex on the path that needs more than \( t \) rounds to push the rumor is at most \( O(1/\log^2 n) \). Therefore, the expected number of vertices for which such a path exist but it does not transfer the rumor in time \( k \cdot t = O((\log \log n)^3) \) is \( o(n) \) and it follows by Markov’s inequality that for every \( \varepsilon > 0 \) with high probability at least \( (1 - \varepsilon)n \) vertices are informed at time \( t_2 + k \cdot t = t_2 + O((\log \log n)^3) \).

For bounded degree sequences, we have that with probability \( 1 - o(1/n) \) every vertex \( v \) has path as described in Lemma 5.9. We say that the rumor is delayed if it does not travel one step further on the path during a round of the protocol. Suppose that the rumor needs more than \( t' \) steps to travel from \( v_k \) to \( v \). This means that the rumor was delayed at least \( t' - k \) times. Let \( t' := 6 \Delta^2 \ln n \). Since the expected number of delays in \( t' \) rounds is at most \( t'(\Delta - 1)/\Delta \), it follows by the Chernoff bound that the rumor reaches \( v_k \) with probability \( 1 - o(1/n^2) \) after \( t' \) rounds and we can finish the proof with a union bound.
In this chapter we study the model of random hyperbolic graphs. The model is more sophisticated than the two models considered in the previous chapters. We therefore begin with a short introduction into hyperbolic geometry, before we introduce the model and state our results in Section 6.1. We develop a framework that might be of use for further work on hyperbolic random graphs. We state and prove our tools in Section 6.2, and we describe the framework in Section 6.3. Finally, in Section 6.4-6.8, we prove all our results. The content of this chapter is joint work with Karl Bringmann, Luca Gugelmann, Konstantinos Panagiotou, and Ralph Keusch, and it appeared in [GPP12] and [BPP].
6.1 Models and Results

Let us begin this section with a few facts about the geometry of hyperbolic planes. We will restrict ourselves to the most basic notions, and refer the reader to e.g. [And05] and many references therein for an extensive introduction.

First of all, there are many equivalent representations of the hyperbolic plane, each one highlighting different aspects of the underlying geometry. We will consider here the so-called native representation, which was described by Papadopoulos et al. in [PKBnV10], as it is most convenient for defining the model of random hyperbolic graphs.

One basic feature of the hyperbolic plane is that it is isotropic, meaning that the geometry is the same regardless of direction. In other words, we can distinguish an arbitrary point, which we call the center or the origin. In the native representation of the hyperbolic plane we will use polar coordinates \((r_v, \theta_v)\) to specify the position of any vertex \(v\), where the radial coordinate \(r_v\) equals the hyperbolic distance of \(v\) from the origin. Given this notation, the distance \(d\) of two vertices \(u\) and \(v\) with coordinates \((r_u, \theta_u)\) and \((r_v, \theta_v)\) can be computed by solving the equation

\[
\cosh(d) = \cosh(r_u) \cosh(r_v) - \sinh(r_u) \sinh(r_v) \cos(\theta_u - \theta_v),
\]

where \(\cosh(x) = (e^x + e^{-x})/2\) and \(\sinh(x) = (e^x - e^{-x})/2\). For our purposes we will denote from now on by \(d(r_u, r_v, \theta_u - \theta_v)\) the solution of \(6.1\) for \(d\).

The crucial difference between the Euclidean and the hyperbolic plane is that the latter contains in a well-defined sense more “space”. More specifically, a circle with radius \(r\) has in the Euclidean plane a length of \(2\pi r\), while its length in the hyperbolic plane is \(2\pi \sinh(r) = \Omega(e^r)\). In other words, a circle in the hyperbolic plane has a length that is exponential in its radius as opposed to linear.

Based on the above facts, the authors of [PKBnV10] defined the model of random hyperbolic graphs. In this model, each of the \(n\) vertices is equipped with polar coordinates, which are drawn uniformly at random from the hyperbolic disk of radius \(R = R(n)\). Note that the total area of a circle of radius \(r\) equals

\[
2\pi \int_0^r \sinh(t) dt = 2\pi (\cosh(r) - 1).
\]
To choose the polar coordinates uniformly at random in the hyperbolic disk of radius \( R \), it suffices to choose each \((r, \theta)\) such that \( \theta \) is chosen uniformly at random in the interval, say, \((-\pi, \pi]\) and its radial coordinate \( r \) is drawn according to the distribution with density function \( \frac{\sinh(r)}{(\cosh(R) - 1)} \), where \( 0 \leq r \leq R \). To add flexibility to the model, the authors of \([PKBnV10]\) use a slightly different density function for the radial coordinate: \( \alpha \sinh(\alpha r)/(\cosh(\alpha R) - 1) \), where \( \alpha > 1/2 \). For \( \alpha < 1 \) this favors points closer to the center, while for \( \alpha > 1 \) points with radius closer to \( R \) are favored. For \( \alpha = 1 \) this corresponds to the uniform distribution.

The generation of the edge set depends on the coordinates of the vertices. In the most general model, any two vertices are connected by an edge with a probability that depends on their (hyperbolic) distance. More precisely, based on considerations from statistical mechanics, the authors of \([PKBnV10]\) propose to connect two vertices \( u \) and \( v \) with probability

\[
\chi_{r_u,r_v}^{\theta_u-\theta_v} := \frac{1}{1 + e^{\frac{1}{2\pi}(d(r_u,r_v,\theta_u-\theta_v)-R)}},
\]

where \( d \) denotes their hyperbolic distance and \( T \) is an additional parameter of the model. This probability distribution is the Fermi-Dirac-distribution from statistical mechanics. Note, that the probability is large if the vertices have distance \( \leq R \), and becomes quickly smaller when the distance is larger than \( R \). In the limit \( T \to 0 \) it is equivalent to the step function and we get the threshold model where two points on this disk are connected by an edge if and only if they are at hyperbolic distance at most \( R \) from each other. On the other hand, for \( T = \infty \) the model corresponds to the classical Erdős-Rényi random graph \( G_{n,1/2} \), and the geometrical structure is not present anymore.

In \([PKBnV10], [Fou12], \) and \([BFM13]\) the model was defined with an additional parameter \( \zeta \) to control the curvature of the hyperbolic space. Since all our results do only depend on the ratio \( \zeta/\alpha \), we might as well assume that \( \zeta = 1 \) and use only \( \alpha \) to control the model.

Let us now proceed to a formal definition of the two models. With all the above notation at hand, the random hyperbolic graph \( G_{\alpha,C,T}(n) \) with \( n \) vertices and parameters \( \alpha, C \) and \( T \) is defined as follows.

**Definition 6.1** (Random Hyperbolic Graph \( G_{\alpha,C,T}(n) \)). Let \( \alpha > 0 \), \( C \in \mathbb{R}, n \in \mathbb{N}, T > 0 \), and set \( R = 2 \log(n) + C \). The random hyperbolic graph \( G_{\alpha,C,T}(n) \) is a graph \((V,E)\) with the following properties.
Chapter 6. Random Hyperbolic Graphs

- The vertex set $V$ of $G_{\alpha,C,T}(n)$ is $V = \{1, ..., n\}$.

- Every $v \in V$ is equipped independently with random polar coordinates $(r_v, \theta_v)$, where $r_v \in [0, R]$ has density $f(r) := \frac{\alpha \sinh(\alpha r)}{\cosh(\alpha R) - 1}$ and $\theta_v$ is drawn uniformly from $(-\pi, \pi]$.

- The edge set $E$ is built by $\binom{n}{2}$ independent random experiments where the probability that $\{u, v\} \in E$ is $\chi_{r_u, r_v}^{\theta_u - \theta_v}$.

An interesting special case of the model is if $T$ tends to $0$. Then, two vertices are connected if and only if their hyperbolic distance is at most $R$. We refer to this model as the threshold model $G_{\alpha,C}(n)$, which is defined as follows.

**Definition 6.2 (Threshold Model $G_{\alpha,C}(n)$).** Let $\alpha > 0$, $C \in \mathbb{R}$, $n \in \mathbb{N}$, and set $R = 2\log n + C$. The random hyperbolic graph $G_{\alpha,C}(n)$ is a graph $(V,E)$ with the following properties.

- The set of vertices with attached polar coordinates is built in the same way as in Definition 6.1.

- The edge set of $G_{\alpha,C}(n)$ is given by $\{\{u, v\} \in \binom{V}{2} | d(r_u, r_v, \theta_u - \theta_v) \leq R\}$.

In this thesis we only consider the general model $G_{\alpha,C,T}(n)$. Note that all our result also hold for $G_{\alpha,C}(n)$ (for more details see the remark after Theorem 6.6). Let us first comment on the various model parameters. A result of Fountoulakis [Fou12] states that for our choice of $R$ and the additional restrictions $T < 1$ and $\alpha > 1/2$, the resulting graph has a bounded average degree (depending on $\alpha, C$ and $T$ only). However, for $T = 1$ the average degree is of order $\log n$ while for $T > 1$ it is polynomially in $n$. Similarly, if $\alpha \leq 1/2$, then the degree sequence is so heavy tailed that a bounded average degree is impossible. A simple calculation shows that the expected number of vertices with radius at most $R/2$ is $n \int_0^{R/2} f(r) dr = \Theta(n^{1-\alpha})$. Since the hyperbolic metric satisfies the triangle inequality, the distance between such vertices is at most $R$. It therefore follows that the expected number of edges among these vertices is already $\Theta(n^{2-2\alpha})$, as two vertices of distance at most $R$ are with constant probability connected. More careful calculations show that already for $\alpha = 1/2$ the average degree is of order $\Omega(\log n)$. In this thesis we study only sparse random hyperbolic graphs, which
means that we consider only the parameter space where the average degree is bounded, i.e. $|E| = \Theta(|V|)$. We therefore assume that $a > 1/2$ and $T < 1$.

For the remaining part of this section we present and discuss our main results for $G_{\alpha,C,T}(n)$. First of all, we study the degree sequence, and provide sharp bounds for the number of vertices of degree $k$. For a graph $G$, let $D_k(G)$ denote the number of vertices of degree $k$ and let $D_{\geq k}(G)$ denote the number of vertices of degree at least $k$. Our first theorem demonstrates that the degree sequence of $G_{\alpha,C,T}(n)$ follows a power-law with exponent $2\alpha + 1 > 2$.

**Theorem 6.3.** Let $\alpha > 1/2$, $0 < T < 1$, $C \in \mathbb{R}$ and $G \sim G_{\alpha,C,T}(n)$. Set

$$
\delta = \min \left\{ \frac{(2\alpha-1)}{4(2\alpha+1)\alpha}, \frac{2(2\alpha-1)}{5(2\alpha+1)} \right\}
$$

and let $\delta' < \delta$. Then, with high probability for all $0 \leq k \leq n^{\delta'}$

$$
\frac{D_k(G)}{n} = (1 + o(1)) \frac{2\alpha \xi^{2\alpha}}{k!} \left( \Gamma(k - 2\alpha) - \int_0^\xi t^{k-2\alpha-1} e^{-t} dt \right), \quad (6.3)
$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ denotes the Gamma function and

$$
\xi := \frac{2\alpha}{(\alpha - 1/2)} e^{-C/2} \left( \lim_{t \to T} \frac{t}{\sin(\pi t)} \right). \quad (6.4)
$$

If $n^{\delta'} \leq k \leq \frac{n^{1/2\alpha}}{\log n}$ then with high probability

$$
\frac{D_{\geq k}(G)}{n} = (1 + o(1)) \xi^{2\alpha} k^{-2\alpha}. \quad (6.5)
$$

Note that for sufficiently large $k$ we have that $\Gamma(k - 2\alpha)/k! = \Theta(k^{-2\alpha-1})$. It is therefore not very hard to show that (6.3) and (6.5) imply that the number of vertices of degree $k$ in $G_{\alpha,C,T}(n)$ is $(1 + o(1)) c_{\alpha,C,T} k^{-2\alpha-1} n$, for an appropriate $c_{\alpha,C,T} > 0$.

Our second theorem states tight bounds for the number of edges and for the expected degree of a vertex in $G_{\alpha,C,T}(n)$.

**Theorem 6.4.** Let $\alpha > 1/2$, $0 < T < 1$, $C \in \mathbb{R}$, $G \sim G_{\alpha,C,T}(n)$ and let $v \in V(G)$ be a vertex of $G$. Then, the expected degree of $v$ is

$$
(1 + o(1)) \frac{\alpha \xi}{(\alpha - 1/2)}.
$$
Furthermore, it holds with high probability that 

$$|E(G)| = (1 + o(1)) \frac{\alpha \xi n}{(\alpha - 1/2)}.$$ 

Next, we give sharp bounds for the maximum degree in $G_{\alpha,C,T}(n)$.

**Theorem 6.5.** Let $\alpha > 1/2$, $0 < T < 1$, $C \in \mathbb{R}$ and $G \sim G_{\alpha,C,T}(n)$. Then, the maximum vertex degree of $G$ is with high probability $n^{1/2} \alpha + o(1)$.

We also study the clustering of hyperbolic random graphs. Recall that the local clustering coefficient of a vertex $v$ is defined by

$$\bar{c}_v = \begin{cases} 0 & \text{if } \deg(v) < 2 \\ \frac{|\{\{u_1,u_2\} \in E | u_1,u_2 \in N(v)\}|}{\binom{\deg(v)}{2}} & \text{else}, \end{cases} \quad (6.6)$$

where $N(v) := \{u \mid \{u,v\} \in E\}$ is the neighborhood of $v$. The global clustering coefficient of a graph $G = (V,E)$ is the average over all local clustering coefficients

$$\bar{c}(G) := \frac{1}{n} \sum_{v \in V} \bar{c}_v. \quad (6.7)$$

The following theorem gives a constant lower bound on the global clustering coefficient, which holds with high probability.

**Theorem 6.6.** Let $\alpha > 1/2$, $0 < T < 1$, $C \in \mathbb{R}$, $G \sim G_{\alpha,C,T}(n)$ and $\bar{c} := \bar{c}(G)$. Then, we have with high probability that $\bar{c} = \Theta(1)$.

Note that in [CF13], Candelleron and Fountoulakis derived tight bounds for a different formulation clustering coefficient. For the version of the clustering coefficient used in this paper, it seems to be more challenging to derive tight bounds.

The theorems stated so far are generalizations of the results in [GPP12]. In particular, we there proved the corresponding statements to Theorem 6.3-6.6 for the threshold model. These results can also be obtained by setting $T = 0$ in Theorem 6.3-6.6 (note that $\lim_{t \to 0} \frac{t}{\sin(\pi t)} = \frac{1}{\pi}$).

Our last result considers the compressibility of random hyperbolic graphs, in particular the performance of the so called $BV$-compression scheme, which is one of the building blocks of the compression algorithms in [BV04]. Its main idea can be described in two steps. First, a suitable
ordering of the vertices is fixed. Second, edges are stored in adjacency lists such that for each edge \{u, v\} only the distance between u and v in the fixed ordering is encoded. The following theorem states that a random hyperbolic graph on n edges can w.h.p be stored in \(O(n)\) bits using the BV-compression scheme on the angular ordering of the vertices.

**Theorem 6.7.** Let \(\alpha > 1/2\), \(C \in \mathbb{R}\), \(0 < T < 1\) and let \(G = (V, E) \sim G_{\alpha, C, T}(n)\), such that the vertices in \(V\) are ordered by their angle in the hyperbolic plane. Then, the BV-compression scheme needs with high probability at most \(O(n)\) bits to store \(G\).

With Theorem 6.7, we prove that a compression scheme, which performs well on the web graph, succeeds in storing random hyperbolic graphs in \(O(n)\) bits. This is in contrasts to the classical models for the web graph, which need at least \(O(n \log n)\) bits [CKL+13]. Our result therefore suggests that random hyperbolic graphs are able to capture global properties of the web graph better that these models. In our proof, we derive a more technical bound on the encoding length, which also depends on the parameters used to control the number of edges in the random graph. In Section 6.8.5, we discuss our bound and give an argument that indicates that our bound is optimal in at least one of the parameters.

### 6.2 Properties of the Model

The goal of this section is to prove some lemmas on connection probabilities, which are heavily used in the proofs of the theorems. In particular we state and prove a lemma that gives strong bounds on the probability that a vertex at a fixed radius is connected to a random vertex. This lemma is the core of most of our proofs.

Recall that the hyperbolic distance between two points is the solution to (6.1). This definition, in particular that it is stated implicitly and that it contains trigonometric and hyperbolic functions, is inconvenient for our calculations. We therefore need to approximate the solution of (6.1). By looking only at the extremal of the angles, we make a first simple observation.

**Observation 6.8.** Let \(r, r' \in \mathbb{R}\) such that \(r' \leq r\) an let \(-2\pi \leq \theta - \theta' \leq 2\pi\). Then,

\[ r - r' \leq d(r, r', \theta) \leq r + r'. \]
Proof. The maximum and minimum value for (6.1) are clearly obtained for \( \theta - \theta' = \pi \) and \( \theta - \theta' = 0 \) where the solutions of the equation are exactly \( r + r' \) and \( r - r' \).

In order to derive more precise approximations, we need estimates for the hyperbolic and trigonometric functions. Let us first repeat the basic estimates of \( \cosh(x) \) and \( \sinh(x) \) which are usually precise enough for our calculation. For all \( x \geq 0 \)

\[
\frac{e^x}{2} \leq \cosh(x) \leq e^x \quad \text{and} \quad \sinh(x) \leq \frac{e^x}{2}
\] (6.8)

and for \( x \geq \frac{\ln 3}{2} \)

\[
\frac{e^x}{3} \leq \sinh(x).
\] (6.9)

Furthermore, recall the well known trigonometric identity

\[
\cosh(x \pm y) = \cosh(x) \cosh(y) \pm \sinh(x) \sinh(y)
\] (6.10)

and the following approximation of the cosine

\[
\frac{\theta^2}{2} - \frac{\theta^4}{4!} \leq (1 - \cos(\theta)) \leq \frac{\theta^2}{2}.
\] (6.11)

By considering the angle between the two points, we can (at least asymptotically) estimate the distance more accurately.

**Lemma 6.9.** Let \((r_1, \theta_1)\) and \((r_2, \theta_2)\) be polar coordinates such that \(r_1 \geq r_2\) and let \( \theta := |\theta_1 - \theta_2| \). Then,

\[
d(r_1, r_2, \theta) = \max \{r_1 - r_2, r_1 + r_2 + 2 \log \theta\} \pm O(1)
\]

\[
= r_1 + r_2 + 2 \log (\theta + e^{-r_1} + e^{-r_2}) \pm O(1).
\]

Proof. Using the definition of the hyperbolic distance (6.1) and the trigonometric identity

\[
\cosh(x \pm y) = \cosh(x) \cosh(y) \pm \sinh(x) \sinh(y),
\]

we derive that \( d := d(r_1, r_2, \theta) \) is the solution of

\[
\cosh(d) = \cosh(r_1 - r_2) + (1 - \cos(\theta)) \sinh(r_1) \sinh(r_2).
\]
This leads to an upper bound of

\[
e^d \leq 4 \max \{ \cosh(r_1 - r_2), (1 - \cos(\theta)) \sinh(r_1) \sinh(r_2) \}
\]

\[
\leq 4 \max \{ e^{r_1-r_2}, \theta^2 e^{r_1+r_2} \}
\]

on \( e^d \). Note that if we assume \( r_1 \) and \( r_2 \) to be at least \( \frac{\ln 3}{2} \), we can derive the lower bound

\[
(1 - \cos(\theta)) \sinh(r_1) \sinh(r_2) \geq \left( \frac{\theta^2}{2} - \frac{\theta^4}{4!} \right) \frac{1}{9} e^{r_1+r_2}
\]

\[
\geq \frac{\theta^2}{18} \left( 1 - \frac{\pi^2}{12} \right) e^{r_1+r_2} \geq \frac{1}{108} \theta^2 e^{r_1+r_2},
\]

which implies

\[
e^d \geq \max \{ \cosh(r_1 - r_2), (1 - \cos(\theta)) \sinh(r_1) \sinh(r_2) \}
\]

\[
\geq \frac{1}{108} \max \{ e^{r_1-r_2}, \theta^2 e^{r_1+r_2} \}.
\]

We can always assume \( r_1 \) and \( r_2 \) to be at least \( 1/2 \ln 3 \) since the error we might make by this assumption is always covered in the term \( \pm O(1) \). Our upper and lower bound on \( e^d \) therefore prove the first equality of the lemma.

For the second part we distinguish whether the maximum is attained at \( r_1-r_2 \) or at \( r_1+r_2+2 \log \theta \). If the maximum is \( r_1-r_2 \) then \( 2r_2 \leq -2 \log \theta \) and therefore \( e^{-r_2} \geq \theta \). Thus

\[
r_1 + r_2 + 2 \log (\theta + e^{-r_1} + e^{-r_2}) \leq r_1 + r_2 + 2 \log(3e^{-r_2}) = r_1 - r_2 + O(1)
\]

and

\[
r_1 + r_2 + 2 \log (\theta + e^{-r_1} + e^{-r_2}) \geq r_1 + r_2 + 2 \log(e^{-r_2}) = r_1 - r_2.
\]

On the other hand if the maximum is \( r_1 + r_2 + 2 \log \theta \) then \( e^{-r_2} \leq \theta \). Similarly,

\[
r_1+r_2+2 \log (\theta + e^{-r_1} + e^{-r_2}) \leq r_1+r_2+2 \log(3\theta) = r_1+r_2+2 \log \theta + O(1)
\]

and

\[
r_1 + r_2 + 2 \log (\theta + e^{-r_1} + e^{-r_2}) \geq r_1 + r_2 + 2 \log \theta.
\]

\( \square \)
In particular, Lemma 6.9 implies a bound on the probability that two vertices at radius $r_1, r_2$ with angle difference $\theta$ are connected. To that matter, note that
\[
\chi^\theta_{r_1, r_2} = \frac{1}{1 + e^{\frac{1}{2\pi} (d(r_1, r_2, \theta) - R)}} \leq e^{-\frac{1}{2\pi} (d(r_1, r_2, \theta) - R)},
\]
holds because of $1/(1 + x) \leq 1/x$. Therefore, applying Lemma 6.9, we derive that there exists an absolute constant $\bar{c}$ such that
\[
\chi^\theta_{r_1, r_2} \leq e^{\bar{c} \frac{R - r_1 - r_2}{2\pi} \theta^{-1}/T}. \tag{6.12}
\]

We devote the rest of this section to prove the following lemma. This lemma is heavily used in all the later calculations and an important ingredient of our proofs.

**Lemma 6.10.** Let $\alpha > 1/2$, $0 \leq T < 1$, $0 < \beta < 1$, $C \in \mathbb{R}$ and $G \sim G_{\alpha, C; T}(n)$. Moreover, let $-\pi \leq \theta \leq \pi$ and $\beta R \leq r \leq R$, let $v_1 \in V(G)$ be a vertex with fixed polar coordinates $(r, \theta)$, and let $0 \leq x < R$ such that $R - x = \Theta(\log n)$. Then, for any vertex $v_2 \in V(G)$ with random polar coordinates $(r_2, \theta_2)$
\[
p_{r,x} := \Pr\{v_1, v_2\} \in E(G) \wedge r_2 \geq x\} = \xi \cdot e^{C/2} \cdot e^{-r/2} (1 \pm o(1)), \tag{6.13}
\]
for
\[
\xi := \frac{2\alpha}{(\alpha - 1/2)} e^{-C/2} \left( \lim_{t \to T} \frac{t}{\sin(\pi t)} \right).
\]

Before we continue with the proof of the lemma, let us give an intuitive description of the statement. Equation (6.13) states that the probability for a vertex $v$ with radius $r$ to be connected to a random vertex is, up to constants and error terms, equal to $e^{-r/2}$. Thus, the degree of $v$ is a binomial distribution with parameters $n$ and $e^{-r/2}$. In particular, if $r$ is small, then the expected degree of $v$ is large, and on the other hand, if $r = 2 \log n \approx R$, then the expected degree of $p$ is $O(1)$. In other words, the closer a vertex is located to the border of the disc, the smaller its degree will be, and (6.13) allows us to quantify precisely the dependence. In [GPP12], we stated Lemma 3.2 that gives very tight estimates for $p_{r,x}$ in the threshold model. The above lemma is clearly more general in a sense that we can look at the limit of $T \to 0$. However, in [GPP12], we calculated tighter estimates for the error terms, which makes the lemma stronger for some applications.
In order to prove the lemma we need to introduce some geometric definitions and probability measures. Recall that, according to Definition 6.1, the probability that a vertex has polar coordinates \((r, \theta)\) is \(f(r)/(2\pi)\) for

\[
f(r) = \frac{\alpha \sinh(\alpha r)}{(\cosh(\alpha R) - 1)} = (1 + o(1))\alpha e^{-\alpha(R-r)}, \tag{6.14}
\]

and does only depend on the radial coordinate \(r\). For every Borel measurable subset of the hyperbolic disk of Radius \(R\), the probability that a vertex has coordinates in \(S\) is

\[
\mu(S) = \frac{1}{2\pi} \int_S f(y) dy. \tag{6.15}
\]

The sets that we usually consider are combinations of balls. We denote the ball of radius \(x\) around a point \((r, \theta)\) as

\[
B_{r,\theta}(x) := \{(r', \theta') | d(r, r', |\theta - \theta'|) \leq x\}.
\]

Since the probability that two fixed vertices at hyperbolic distance \(d\) are adjacent is

\[
\frac{1}{1 + e^{1/(2T)(d-R)}}, \tag{6.16}
\]

by inverting this relation, it follows that the distance of two vertices which are adjacent with probability \(p\) is exactly

\[
d(p) = 2T \log \left( \frac{1}{p} - 1 \right) + R. \tag{6.17}
\]

Let \(0 \leq r \leq R\) and \(0 \leq p \leq 1\). Then (6.17) leads to a simple characterization of the set of all points which would be connected to \(v = (r, \theta')\) with probability at least \(p\), namely

\[
\{ u = (y, \theta) \mid \Pr\{|u, v\} \in E| \geq p \} = B_{r,\theta'}(d(p)) \cap B_{0,0}(R).
\]

Note that there are two special cases where the geometric situation is considerably simpler than usual. If \(p\) is very large, in particular if

\[
p \geq \frac{1}{1 + e^{-r/(2T)}}
\]

then \(d(p) \leq R - r\) which means that \(B_{r,0}(d(p)) \subseteq B_{0,0}(R)\). In contrast to this, whenever

\[
p \leq \frac{1}{1 + e^{r/(2T)}} \tag{6.18}
\]
then $d(p) \geq R + r$ which means that a vertex of radius $r$ is connected to every other vertex on the disk with probability at least $p$.

Recalling (6.15), we can derive $\mu(B_{r,\theta'}(d(p)) \cap B_{0,0}(R))$ by integrating $f(y)$ over all $y \in [0, R]$ and then over all $\theta$ satisfying $d(r, y, \theta) \leq d(p)$. Let us consider the range of $\theta$ where this inequality is satisfied. If there exists such a range, then by (6.1) one of its extremals is

$$\theta_{r,p}(y) := \arg \max_{0 \leq \varphi \leq \pi} \{d(r, y, \varphi) \leq d(p)\}$$

$$= \arccos \left( \frac{\cosh(r) \cosh(y) - \cosh(d(p))}{\sinh(r) \sinh(y)} \right). \quad (6.19)$$

By symmetry the other extremal is $-\theta_{r,p}(y)$. If $d(r, y, \theta) > d(p)$ holds for all $\theta$, then define $\theta_{r,p}(y) = 0$. If $d(r, y, \theta) \leq d(p)$ for all $\theta$, then define $\theta_{r,p}(y) = \pi$. Thus, by (6.15) we get

$$\mu(B_{r,\theta'}(d(p)) \cap B_{0,0}(R)) = \frac{1}{2\pi} \int_0^R \int_{\theta_{r,p}(y)}^{\theta_{r,p}(y)} f(y)d\theta dy = \frac{1}{\pi} \int_0^R \theta_{r,p}(y) f(y)dy$$

(6.20)

Since $\mu(B_{r,\theta'}(d(p)) \cap B_{0,0}(R))$ does not depend on $\theta'$ we write for the remaining part of this thesis $\mu(B_r(d(p)) \cap B_0(R))$ instead. Note that for the special case where $r = 0$, Observation 6.8 implies that

$$\mu(B_0(x)) = \int_0^x f(y)dy = \int_0^x \frac{\alpha \sinh(\alpha y)}{\cosh(\alpha R) - 1} dy$$

$$= \frac{\cosh(\alpha x) - 1}{\cosh(\alpha R) - 1} = (1 + o(1))e^{-\alpha(R - x)}. \quad (6.21)$$

The next technical lemma gives almost tight bounds on $\theta_{r,p}(y)$.

**Lemma 6.11.** Let $1 \leq r \leq R$ and $0 \leq p \leq 1$ such that $|d(p) - r| \leq y \leq d(p) + r$. Then

$$\theta_{r,p}(y) = 2e^{\frac{d(p) - r - y}{2}} (1 \pm \Theta(e^{d(p) - r - y} + e^{-d(p) - r + y} + e^{-d(p) + r - y})).$$

**Proof.** Note that the bound stated by the lemma is trivial for $|d(p) - r| \leq y \leq 1$. We therefore assume that $y > 1$. By using (6.19) and the
trigonometric identity (6.10) we infer that

$$\cos(\theta_{r,p}(y)) = \frac{\sinh(r) \sinh(y) + \cosh(r - y) - \cosh(d(p))}{\sinh(r) \sinh(y)}$$

$$= 1 + 2 \frac{e^{r-y} + e^{-r+y}}{(e^r - e^{-r})(e^y - e^{-y})} - 2 \frac{e^{d(p)} + e^{-d(p)}}{(e^r - e^{-r})(e^y - e^{-y})}$$

$$= 1 + 2 \frac{e^{-2r} + e^{-2y}}{(1 - e^{-2r})(1 - e^{-2y})} - 2 \frac{e^{d(p)-r-y} + e^{-d(p)-r-y}}{(1 - e^{-2r})(1 - e^{-2y})}.$$  

By applying the identity

$$\frac{1}{1 - e^{-2x}} = \sum_{i \geq 0} e^{-2ix} x \geq 1 + \Theta(e^{-2x}),$$

we obtain

$$\cos(\theta_{r,p}(y)) = 1 + 2 \left( e^{-2y} + e^{-2r} - e^{d(p)-r-y} - e^{-d(p)-r-y} \right)$$

$$\cdot \left( 1 + \Theta(e^{-2r}) \right) \left( 1 + \Theta(e^{-2y}) \right)$$

$$y \leq d(p) + r \leq 1 - 2e^{d(p)-r-y} + \Theta(e^{-2y} + e^{-2r}).$$

In our next estimates we will get rid of the cosine in the above expression. By $\cos(\theta) \geq 1 - \frac{\theta^2}{2}$ we derive

$$\theta_{r,p}^2(y) \geq 4e^{d(p)-r-y} - \Theta \left( e^{-2r} + e^{-2y} \right).$$

Note that whenever $|x| \leq 1$

$$\sqrt{1 + x} = 1 + \frac{x}{2} + \Theta(x^2). \quad (6.22)$$

Applied to the previous equation, this gives a lower bound of

$$\theta_{r,p}(y) \geq 2e^{\frac{d(p)-r-y}{2}} \left( 1 - \Theta \left( e^{-d(p)-r+y} + e^{-d(p)+r-y} \right) \right). \quad (6.23)$$

Next we will derive an almost matching upper bound for $\theta_r(y)$. First we exploit that $\cos(\theta) \leq 1 - \frac{\theta^2}{2} + \frac{\theta^4}{4!}$ and thereby

$$\frac{\theta_r^2(y)}{2} - \frac{\theta_r^4(y)}{4!} \leq 2e^{d(p)-r-y} - \Theta \left( e^{-2r} + e^{-2y} \right).$$
This quadratic equation can be solved exactly by using basic tools. We omit the detailed calculations, and show just the final outcome which is

\[ \theta_{r,p}^2(y) \leq 6 - 6 \sqrt{1 - \frac{4}{3} \left( e^{d(p)-r-y} - \Theta(e^{-2r} + e^{-2y}) \right)}. \]

Note that we can assume that \( \frac{4}{3} (e^{d(p)-r-y} - \Theta(e^{-2r} + e^{-2y})) < 1 \), as otherwise \( e^{d(p)-r-y} = \Theta(1) \), and the lemma therefore states trivial bounds for that regime. Thus, we can apply (6.22) to obtain

\[
\theta_{r,p}^2(y) = 4e^{d(p)-r-y} + \Theta(e^{-2r} + e^{-2y}) + \Theta \left( e^{d(p)-r-y}(e^{-2r} + e^{-2y}) \right) \\
= 4e^{d(p)-r-y} \left( 1 + \Theta \left( e^{d(p)-r-y} + e^{-d(p)-r+y} + e^{-d(p)+r-y} \right) \right).
\]

Hence by applying again (6.22) we get

\[ \theta_{r,p}(y) \leq 2e^{\frac{d(p)-r-y}{2}} \left( 1 + \Theta \left( e^{d(p)-r-y} + e^{-d(p)-r+y} + e^{-d(p)+r-y} \right) \right) \]

which together with (6.23) concludes the proof. \( \square \)

Having this estimate for \( \theta_{r,p}(y) \), we are ready to prove the main lemma.

**Proof of Lemma 6.10.** Note that the probability that \( v_1 \) and \( v_2 \) are connected, and that \( r_2 \geq r \) is

\[ p_{r,x} = \lim_{m \to \infty} \sum_{k=1}^{m} \frac{1}{m} \mu \left( B_0(R) \cap B_r \left( d \left( \frac{k}{m} \right) \right) \setminus B_r(x) \right). \]

Since the above sum is a Riemann sum, its limit can be written as the following integral

\[ p_{r,x} = \int_0^1 \mu \left( B_0(R) \cap B_r \left( d(p) \right) \setminus B_0(x) \right) dp. \quad (6.24) \]

We choose a small enough constant \( \varepsilon > 0 \) such that

\[ e^{-r+\varepsilon \log n} = o(e^{-r/2}), \]

which is possible as \( r \geq \beta R \). Next, let

\[ a := \frac{1}{1 + e^{(r-\varepsilon \log n)/(2T)}} \quad \text{and} \quad b := \frac{1}{1 + e^{(-r+\varepsilon \log n)/(2T)}}. \]
In order to calculate the integral in (6.24) we treat the three intervals \([0, a], [b, 1]\) and \([a, b]\) separately. First note that

\[
\int_0^a \mu(B_0(R) \cap B_r(d(p)) \setminus B_0(x)) \, dp \leq \int_0^a 1 \, dp = a = O(e^{-r + \varepsilon \log n}) = o(e^{-r/2}).
\] (6.25)

For large \(p\), since \(\mu(B_0(R) \cap B_r(d(p)) \setminus B_0(x))\) is decreasing in \(p\), we can bound the integral over \([b, 1]\) by

\[
\int_b^1 \mu(B_0(R) \cap B_r(d(p)) \setminus B_0(x)) \, dp \leq \int_b^1 1 \, dp \leq e^{-r + \varepsilon \log n} = o(e^{-r/2}).
\] (6.26)

For the third interval, we claim for \(p \in [a, b]\) that

\[
\mu(B_0(R) \cap B_r(d(p)) \setminus B_0(x)) = (1 \pm o(1))2\alpha \frac{e^{d(p) - R - r}}{\pi(\alpha - 1/2)^2} + O(e^{-\alpha R}).
\] (6.27)

Note that

\[
\frac{(1 \pm o(1))2\alpha}{\pi(\alpha - 1/2)^2} e^{d(p) - R - r} = \frac{2\alpha(1/p - 1)^T}{\pi(\alpha - 1/2)} e^{-r/2}(1 \pm o(1)).
\] (6.17)

Before we prove (6.27), we show that it implies the statement of the lemma. Note that

\[
\int_0^a \frac{2\alpha(1/p - 1)^T e^{-r/2}}{\pi(\alpha - 1/2)} \, dp = \frac{2\alpha e^{-r/2}}{\pi(\alpha - 1/2)} \int_0^a \left(\frac{1}{p} - 1\right)^T \, dp
\]

\[
\leq \frac{2\alpha e^{-r/2}}{\pi(\alpha - 1/2)} \int_0^a \frac{1}{p} \, dp = \frac{2\alpha e^{-r/2}}{\pi(\alpha - 1/2)} \cdot \frac{a^{1-T}}{1-T} = o(e^{-r/2}),
\] (6.28)

and

\[
\int_b^1 \frac{2\alpha(1/p - 1)^T e^{-r/2}}{\pi(\alpha - 1/2)} \, dp \leq \frac{2\alpha e^{-r/2}}{\pi(\alpha - 1/2)} (1 - b) \left(\frac{1}{b} - 1\right)^T = o(e^{-r/2}).
\] (6.29)
Figure 6.1: The two cases for which \(|d(p) - r| > x\) for a vertex \(v\) with radius \(r\). In the situation of the left picture have \(d(p) > r\) and \(d(p) - r > x\), while in the right picture \(d(p) < r\) and \(r - d(p) > x\).

Thus, by (6.24)-(6.29)

\[
pr,x = (1 + o(1)) \int_0^1 \frac{2\alpha e^{-r/2}}{\pi(\alpha - 1/2)} \left(\frac{1}{p} - 1\right)^T dp = (1 + o(1)) \frac{2\alpha e^{-r/2}}{\pi(\alpha - 1/2)} \int_0^1 \left(\frac{1}{p} - 1\right)^T dp.
\]

Since \(\int_0^1 \left(\frac{1}{p} - 1\right)^T dp = \lim_{t \to T} \int_0^1 \left(\frac{1}{p} - 1\right)^t dp = \frac{\pi t}{\sin \pi t}\) for \(0 \leq T < 1\), it only remains to prove (6.27).

Recall (6.17) and note that \(a \leq p \leq b\) implies

\[
R - r + \varepsilon \log n \leq d(p) \leq R + r - \varepsilon \log n.
\] (6.30)

Moreover, we observed in (6.20) that we can calculate the measure for a set of points by integrating over their radius and angle. Hence,

\[
\mu(B_0(R) \cap B_r(d(p)) \setminus B_0(x)) = \frac{1}{\pi} \int_x^R \theta_{r,p}(y)f(y)dy.
\]

Unfortunately, we can not directly apply Lemma 6.11 since \(|d(p) - r| \leq y\) is not guaranteed. We therefore have to deal with the two cases (Figure 6.1) where \(|d(p) - r| > y\).

Suppose \(d(p) > r\) and \(x \leq d(p) - r\). It follows from the triangle inequality
that

\[ \mu(B_0(R) \cap B_r(d(p)) \setminus B_0(x)) = \mu(B_0(d(p) - r) \setminus B_0(x)) + \frac{1}{\pi} \int_{d(p) - r}^{\min\{R,d(p) + r\}} \theta_{r,p}(y) f(y) dy \]

\[ \text{[6.21, 6.30]} O \left( e^{-\alpha(R - (d(p) - r))} \right) + \frac{1}{\pi} \int_{d(p) - r}^{R} \theta_{r,p}(y) f(y) dy. \]

In the second case, where \( d(p) \leq r \) and \( x \leq r - d(p) \), it holds again by the triangle inequality and [6.30] that

\[ \mu(B_0(R) \cap B_r(d(p)) \setminus B_0(x)) = \frac{1}{\pi} \int_{r - d(p)}^{R} \theta_{r,p}(y) f(y) dy. \]

We conclude that in any case we have

\[ \mu(B_0(R) \cap B_r(d(p)) \setminus B_0(x)) = \frac{1}{\pi} \int_{u}^{R} \theta_{r,p}(y) f(y) dy + o \left( e^{\frac{d(p) - R - r}{2}} \right), \]

where \( u = \max\{|d(p) - r|, x\} \), and therefore, by [6.30], \( R - u = \Theta(\log n) \).

It remains to show

\[ \frac{1}{\pi} \int_{u}^{R} \theta_{r,p}(y) f(y) dy = \frac{2\alpha}{\pi(\alpha - 1/2)} e^{\frac{d(p) - R - r}{2}} (1 \pm o(1)). \]

Using Lemma [6.11] we obtain

\[ \frac{1}{\pi} \int_{u}^{R} \theta_{r,p}(y) f(y) dy = \frac{1}{\pi} \int_{u}^{R} 2e^{\frac{d(p) - r - y}{2}} (1 \pm h(p, r, y)) \frac{\alpha \sinh(\alpha y)}{\cosh(\alpha R) - 1} dy, \]

with error function

\[ h(p, r, y) = O \left( e^{d(p) - r - y} + e^y - r - d(p) + e^{r - d(p) - y} \right). \]

We split [6.31] into the leading term and the error term and integrate

\[ \int e^{-y/2} \sinh(\alpha y) dy = \frac{2}{4\alpha^2 - 1} e^{-y/2} (2\alpha \cosh(\alpha y) + \sinh(\alpha y)), \]
which leads to

\[
\frac{1}{\pi} \int_{u}^{R} 2e^{\frac{d(p)-r-y}{2}} \frac{\alpha \sinh(\alpha y)}{\cosh(\alpha R) - 1} dy = \\
= \frac{4\alpha e^{\frac{d(p)-R-r}{2}} (2\alpha \cosh(\alpha R) + \sinh(\alpha R) - e^{\frac{R-u}{2}} (2\alpha \cosh(\alpha u) + \sinh(\alpha u)))}{\pi (4\alpha^2 - 1) (\cosh(\alpha R) - 1)}.
\]

(6.33)

Using \(1/(\cosh(\alpha R) - 1) = 2 \exp(-\alpha R)(1 + \Theta(e^{-\alpha R}))\) and expanding the trigonometric terms gives

\[
(6.33) = \frac{8\alpha}{\pi (4\alpha^2 - 1)} e^{\frac{d(p)-R-r}{2}} (1 + \Theta(e^{-\alpha R})) \\
\left[ \left( \alpha + \frac{1}{2} \right) \left( 1 - e^{-(\alpha-1/2)(R-u)} \right) + \left( \alpha - \frac{1}{2} \right) \left( e^{-2\alpha R} - e^{(R-u)/2 - \alpha(R+u)} \right) \right] \\
= \frac{2\alpha e^{\frac{d(p)-R-r}{2}}}{\pi (\alpha - 1/2)} (1 + o(1)).
\]

(6.34)

For the integral over the error term in (6.31) we treat the three terms of the error function \(h(p, r, y)\) separately. For the term \(e^{d(p)-r-y}\) we observe that

\[
\int_{u}^{R} O \left( e^{(3/2)(d(p)-r-y)+\alpha(y-R)} \right) dy \\
= O \left( e^{(3/2)(d(p)-R-r)} + e^{(3/2)(d(p)-r-u)+\alpha(u-R)} \right) \\
= O \left( e^{(3/2)(d(p)-R-r)} + e^{(1/2)(d(p)-r-R)+(\alpha-1/2)(u-R)+(d(p)-r-u)} \right) \\
= o \left( e^{\frac{d(p)-R-r}{2}} \right),
\]

since \(u > d(p) - r\) and by (6.30) \(r + R - d(p) = \Theta(\log n)\). For the second error term \(e^{y-r-d(p)}\) it holds similarly that

\[
\int_{u}^{R} O \left( e^{(1/2)(y-d(p)-3r)+\alpha(y-R)} \right) dy \\
= O \left( e^{(1/2)(R-d(p)-3r)} + e^{(1/2)(u-d(p)-3r)+\alpha(u-R)} \right) \\
= O \left( e^{(1/2)(d(p)-r-R)+(R-d(p)-r)} + e^{(1/2)(d(p)-r-R)+(\alpha-1/2)(u-R)+(u-d(p)-r)} \right) \\
= o \left( e^{\frac{d(p)-r-R}{2}} \right),
\]
6.3. General Proof Strategies

because \( d(p) + r > R > u \) and \( d(p) + r - R = \Theta(\log n) \). Finally, for the third term \( e^{r-d(p)-y} \) of \( h \) we have

\[
\int_u^R O \left( e^{(1/2)(r-d(p)-3y)+\alpha(y-R)} \right) dy
\]

\[
= O \left( e^{(1/2)(r-d(p)-3R)} + e^{(1/2)(r-d(p)-3u)+\alpha(u-R)} \right)
\]

\[
= O \left( e^{(1/2)(d(p)-r-R)+(r-d(p)-R)} + e^{(1/2)(d(p)-r-R)+(\alpha-1/2)(u-R)+(r-d(p)-u)} \right)
\]

\[
= o \left( e^{\frac{d(p)-r-R}{2}} \right),
\]

where we used that \( R > u \geq |d(p) - r| \geq r - d(p) \) and \( R + d(p) - r = \Theta(\log n) \), as \( d(p) = \Theta(\log n) \). The analysis of the error terms together with (6.34) completes the proof.

\[ \square \]

6.3 General Proof Strategies

Before we prove our theorems, let us briefly describe the technique which we use to analyze random hyperbolic graphs. To study a specific property, we will define a suitable random variable \( X \) that characterizes the property. Then, we usually calculate the expectation of \( X \). Next, we wish to apply an Azuma-Hoeffding-type large deviation inequality (see Theorem 2.2) to show that \( X \) is concentrated around its expectation.

In a typical setting, such concentration inequalities require some kind of Lipschitz condition that is satisfied by the random variable under consideration. In particular, we have to bound the effect that every random experiment in \( G_{\alpha,C,T}(n) \) can have on \( X \). This is usually infeasible since a vertex that is in the center of the disk can have a strong effect on almost every meaningful random variable.

We will overcome this obstacle as follows. Instead of considering all the vertices in the graph, we will consider only vertices that lie far away from the center of the disc, i.e., which have radial coordinate larger than \( \beta R \), for some appropriate \( \beta > 0 \). Moreover, we will consider only vertices such that all their neighbors have a large radial coordinate as well. This restriction will allow us to bound the maximum effect on the target function, as with high probability all these vertices do not have too large degree.

More formally, we proceed as follows. We partition the vertex set of
A instance of \( G_{\alpha,C,T}(n) \) is created by a randomized process which operates in two steps. First the vertices get random coordinates. Then, for every pair of vertices, a random experiment decides whether it is connected or not. The probability that two vertices are connected depends on the distance between the vertices, hence on the coordinates of the two vertices. Unfortunately, the above Theorem can be applied only if all random variables are independent. We have to reformulate our random variables such that the two random processes fit into the setting of Theorem 2.2.

First, we use the random variables \( X_1, \ldots, X_n \), where \( X_i \) denotes the coordinates of the \( i \)-th vertex, such that they completely define the vertex set of \( G_{\alpha,C,T}(n) \). Next, we introduce a second set of random variables: For every \( i \in \{2, \ldots, n\} \), we use a random variable \( Y_i \) which takes \( i - 1 \) values independently and uniformly at random out of the interval \( [0, 1] \). The value \( Y_i(j) \) for \( j \in \{1, \ldots, i - 1\} \) then corresponds to the edge \( \{v_i, v_j\} \): The edge \( \{v_i, v_j\} \) is in \( E \) if and only if \( p(d(v_i, v_j)) > Y_i(j) \), where \( d(v_i, v_j) \) is the hyperbolic distance between the two vertices. In this framework all variables \( X_i \) and \( Y_i \) are independent and define \( G_{\alpha,C,T}(n) \) fully.

Let \( 1 > \beta > 0 \). As our set of random variables completely determines the graph, our random variable \( X \) is a function of those random variables. As long as we can assume that the degree of some vertex \( v \) is small, we can usually bound the effect of \( X_v \) and \( Y_v \) on \( X \) sufficiently well. The next lemma states that with high probability all vertices in \( O(\beta) \) have a small degree.

**Lemma 6.12.** Let \( \alpha > 1/2 \), \( 0 < T < 1 \), \( C \in \mathbb{R} \) and \( 0 < \beta < 1 \). There is a constant \( c > 0 \) such that the probability for the bad event

\[
B := \{ \text{there is a vertex in } O(\beta) \text{ with degree at least } cn^{1-\beta} \}
\]

in \( G_{\alpha,C,T}(n) \) is at most \( \Pr[B] = e^{-\Omega(n^{1-\beta})} \).

**Proof.** It follows from Lemma 6.10 that the expected degree in \( O \) of a vertex with radius \( r \geq \beta R \) is at most

\[
n \cdot p_{r,0} = O \left( n \cdot e^{-\frac{\beta R}{2}} \right) = O(n^{1-\beta}).
\]
6.4 Vertices of Small Degree

Hence the expected degree of a vertex of radius at least $\beta R$ is at most $c'n^{1-\beta}$ for some constant $c'$. For $c := 2ec'$ it suffices to apply a Chernoff bound (note that the degree of a vertex is the sum of $n-1$ Bernoulli-distributed random variables) to show that for a vertex $v$ of radius at least $\beta R$

$$\Pr[\deg_{O(\beta)}(v) > cn^{1-\beta}] \leq 2^{-cn^{1-\beta}}.$$  

The statement of the lemma follows by union bound over all vertices. \(\square\)

Suppose now that our random variable $X$ has the following very natural property: Whenever we modify the graph, the effect on $X$ is bounded by the number of edges that we add or remove in $O$. Unless $B$ holds, every vertex in $O := O(\beta)$ is connected to at most $cn^{1-\beta}$ other vertices in $O$. Changing the coordinates of the $i$-th vertex can influence its own degree, the degree of at most $cn^{1-\beta}$ former neighbors and the degree of at most $cn^{1-\beta}$ new neighbours. Therefore, any change in $X_i$ can add or remove at most $2cn^{1-\beta} + 1$ edges. If we still assume that $B$ does not hold, then the same argument is true for the variables $Y_i$, as any change of some $Y_i$ affects only some edges incident to the vertex $v_i$. With that we know that the considered random variable $X$ is Lipschitz with coefficients $c_i, \tilde{c}_i \leq 2cn^{1-\beta} + 1$ ($c_i$ for the random variables $X_i$, $\tilde{c}_i$ for the random variables $Y_i$) and bad event $B$.

6.4 Vertices of Small Degree

In this section we prove the first part of Theorem 6.3. Before we give all technical details, let us briefly describe the main proof idea. Given the estimates in the previous sections, in particular Lemma 6.10, it is conceptually not very difficult to compute the expected number of vertices of degree $k$ in $G_{\alpha,C}(n)$. In order to prove the claimed concentration bounds, we will use the ideas described in the previous section.

Recall our partitioning of the vertex set of $G_{\alpha,C}(n)$ into the inner set $I = I(\beta)$ and the outer set $O = O(\beta)$. Moreover, let $e(I,O)$ count the number of edges with one endpoint in $I$ and the other in $O$. The next two lemmas show that $e(I,O)$ and $|I|$ are small. This indicates that most of the vertices of degree $k$, for not too large $k$, will lie in $O$ and have all their neighbours in $O$. Let $\bar{D}_k$ denote the number of vertices in $O$ which have degree $k$ in $O$, i.e., set

$$\bar{D}_k(\beta) = \left| \{ v \in O(\beta) \mid |N(v) \cap O(\beta)| = k \} \right|.$$
In Lemma 6.15 we derive the expectation of $\bar{D}_k(\beta)$ and finally, we combine everything to show that for small $k$, $D_k$ is tightly concentrated around the expectation of $\bar{D}_k(\beta)$.

**Lemma 6.13.** Let $\alpha > 1/2$, $0 < T < 1$, $C \in \mathbb{R}$ and $0 < \beta < 1$. Then, in $G_{\alpha,C,T}(n)$, with probability at least $1 - e^{-n\Omega(1)}$

$$|I(\beta)| \leq \max\{n^{1/2}, 4en^{1-2\alpha(1-\beta)}e^{-\alpha C(1-\beta)}\}.$$ 

**Proof.** The number of vertices in $I$ follows a $\text{Bin}(n, \mu(B_0(\beta R)))$ distribution. The expected number of vertices in $I$ is therefore bounded by

$$\mathbb{E}[|I(\beta)|] \leq n\mu(B_0(\beta R)) = (1+o(1))ne^{-\alpha(R-\beta R)} \leq 2n^{1-2\alpha(1-\beta)}e^{-\alpha C(1-\beta)}$$

and it follows by the Chernoff bound that

$$\Pr[|I(\beta)| > t] \leq 2^{-t} = e^{-n\Omega(1)}$$

for every number $t = \max\{n^{1/2}, 4en^{1-2\alpha(1-\beta)}e^{-\alpha C(1-\beta)}\}$. \qed

**Lemma 6.14.** Let $\alpha > 1/2$, $0 < T < 1$, $C \in \mathbb{R}$ and $0 < \beta < 1$. Then, in $G_{\alpha,C,T}(n)$, it holds with high probability that

$$e(I(\beta), O(\beta)) = O\left(n^{1-2\alpha(1-\beta)\log n}\right).$$

**Proof.** Define $r_0 := (1 - \frac{1}{2\alpha})\log n + C$. The expected number of vertices of radius at most $r_0$ is at most

$$n \cdot \mu(B_0(r_0)) = (1 + o(1))n e^{-\alpha(R-r_0)} = n e^{-(\alpha+1/2)\log n} = o(1).$$

Thus, there is w.h.p no vertex of radius at most $r_0$. If $r_0 \geq \beta R$, we are done. For $r_0 < \beta R$, suppose that $r_0 \leq r \leq \beta R$ and note that the expected degree in $O$ of a vertex at radius $r$ is by Lemma 6.10 (where the $\beta$ for the lemma is $r_0/R$) at most

$$O(np_{r,\beta R}) = O\left(ne^{-r/2}\right). \quad (6.35)$$

We integrate over $\bar{6.35}$ in $I$ to bound the expected number of edges between $I$ and $O$. As there are w.h.p no vertices of radius at most $r_0$, it
6.4. Vertices of Small Degree

suffices to integrate from \( r_0 \) to \( \beta R \). Then the expected value is at most

\[
O \left( n^2 \int_{r_0}^{\beta R} e^{-r/2} f(r) dr \right) = O \left( n^2 e^{-\alpha R} \int_0^{\beta R} e^{(\alpha-1/2)r} dr \right) = O \left( n^{2-2\alpha+2\beta(\alpha-1/2)} \right) = O \left( n^{1-(2\alpha-1)(1-\beta)} \right).
\]

By Markov’s inequality it follows that

\[
Pr \left[ e(I(\beta), O(\beta)) = \omega \left( n^{1-(2\alpha-1)(1-\beta) \log n} \right) \right] < O \left( \frac{1}{\log n} \right).
\]

In the next lemma we obtain tight bounds on the expected degree sequence of the subgraph spanned by \( O \).

**Lemma 6.15.** Let \( \alpha > 1/2, C \in \mathbb{R}, \) and \( 3/5 < \beta < 1 \). Then, for all \( k := k(n) \) with \( 0 \leq k = o(n^{1/2}) \) we have that

\[
E [ \bar{D}_k(\beta) ] = (1 + o(1)) \frac{2n\alpha}{k!} \xi^{2\alpha} \left( \Gamma(k - 2\alpha) - \int_0^{\xi} t^{k-2\alpha-1} e^{-t} dt \right),
\]

where \( \xi = \frac{2\alpha}{(\alpha-1/2)e^{C/2}} \left( \lim_{t \to T} \frac{t}{\sin(\pi t)} \right) \).

**Proof.** Let \( v \) be an arbitrary vertex with coordinates \((r, \theta)\), where \( r > \beta R \), and let \( \bar{q}_r := p_{r, \beta R} \).

The probability that \( v \) has \( k \) neighbors with radius larger than \( \beta R \) corresponds to the probability that a binomial random variable \( \text{Bin}(n - 1, \bar{q}_r) \) has value \( k \). Therefore, the expected value of \( \bar{D}_k \) can be computed by

\[
E [ \bar{D}_k ] = n \int_{\beta R}^R \binom{n-1}{k} \bar{q}_r^k (1-\bar{q}_r)^{n-1-k} f(r) dr.
\]  

(6.36)

By Lemma [6.10] and the observation that \( R - \beta R = \Theta(\log n) \), it follows that

\[
\bar{q}_r = \xi \cdot e^{C/2} \cdot e^{-r/2} \left( 1 + o(1) \right).
\]

Recall that \( R = 2 \log n + C \) and therefore \( e^{-r} \leq e^{-\beta R} = o(1/n) \). We claim that

\[
n \int_{\beta R}^R \binom{n-1}{k} (\bar{q}_r)^k (1-\bar{q}_r)^{n-1-k} f(r) dr = (1+o(1)) \frac{n}{k!} \int_{\beta R}^R (n\bar{q}_r)^k e^{-n\bar{q}_r} f(r) dr.
\]  

(6.37)
First, we consider the binomial coefficient. Since $k \ll \sqrt{n}$, we have

$$\binom{n-1}{k} = \frac{(n-1)^k}{k!} \prod_{i=0}^{k-1} \left(1 - \frac{i}{n-1}\right) = \left(1 + o(1)\right) \frac{n^k}{k!}.\]$$

Next, we have that

$$(1 - \bar{q}_r)^{-k} = e^{(n-1-k) \log(1 - \bar{q}_r)} = e^{-n\bar{q}_r + O((n-1-k)\bar{q}_r^2) + (k+1)\bar{q}_r} (1 + o(1))$$

where the last equality follows from the fact that $n\bar{q}_r^2 = o(1)$ and $k\bar{q}_r = o(1)$.

We now estimate the integral on the right hand side of Equation (6.37). To do so we perform the variable transformation $t = n\bar{q}_r$. Then, the density $f(r)$ can be expressed as

$$f(r) = \frac{\alpha \sinh(\alpha r)}{\cosh(\alpha R) - 1} = \frac{\alpha}{2(\cosh(\alpha R) - 1)} \left(e^{\alpha r} - e^{-\alpha r}\right)$$

Since $1/(\cosh(\alpha R) - 1) = 2n^{-2\alpha}e^{-\alpha C(1+o(1/n))}$, the above calculation yields that

$$f(r) = \alpha e^{-\alpha C} \xi^{2\alpha} (n\bar{q}_r)^{-2\alpha} (1 + o(1)).$$

Further, we have that

$$dt = (1 + o(1))n\xi e^{C/2}(-1/2)e^{-r/2}dr = -\frac{1}{2}t\xi dr \quad \Rightarrow \quad dr = -2t^{-1}dt,$$

and for the upper and lower bounds that

$$r_0 = \beta R \quad \Rightarrow \quad t_0 = \frac{(1 + o(1))2\alpha T}{(\alpha - 1/2)\sin(\pi T)} e^{-\beta R/2}n = \frac{(1 + o(1))2\alpha T}{(\alpha - 1/2)\sin(\pi T)} e^{-\beta C/2}n^{1-\beta}$$

and

$$r_1 = R \quad \Rightarrow \quad t_1 = \frac{(1 + o(1))2\alpha T}{(\alpha - 1/2)\sin(\pi T)} e^{-R/2}n = \frac{(1 + o(1))2\alpha T}{(\alpha - 1/2)\sin(\pi T)} e^{-C/2}.$$

Putting everything together we have that the right hand side of (6.37) can be approximated by

$$(1 + o(1)) \frac{2n\alpha}{k!} \xi^{2\alpha} \int_{t_1}^{\infty} t^{k-2\alpha-1} e^{-t} dt$$

$$= \left(1 + o(1)\right) \frac{2n\alpha}{k!} \xi^{2\alpha} \left(\Gamma(k - 2\alpha) - \int_0^{t_1} t^{k-2\alpha-1} e^{-t} dt\right).$$
Since $t_1 = (1-o(1))\xi$, this completes the proof.

We now combine all lemmas in this section to prove Theorem 6.3.

**Proof of the first part of Theorem 6.3.** Set $\beta := \max\{\frac{3}{5}, 1 - \frac{1}{4\alpha}\} + \varepsilon$, where $\varepsilon > 0$ is a sufficiently small constant, such that $\beta < 1$. The total number of vertices of degree $k$, for any $k$ in the considered range, is clearly at most

\[ \bar{D}_k(\beta) + |I(\beta)| + e(I(\beta), O(\beta)) \quad (6.38) \]

and at least

\[ \bar{D}_k(\beta) - e(I(\beta), O(\beta)), \quad (6.39) \]

since every vertex in $I$ could possibly have degree $k$, and since each edge counted in $e(I, O)$ may affect the degree of one vertex in $O$. We will argue in the sequel that the contribution of $|I|$ and $e(I, O)$ is with high probability negligible in the above equations.

First, since $\beta > 1 - \frac{1}{4\alpha}$, by applying Lemma 6.13 we obtain that with probability at least $1 - e^{-n^{\Omega(1)}}$,

\[ |I(\beta)| = O\left(n^{1-2\alpha(1-\beta)}\right). \quad (6.40) \]

Moreover, Lemma 6.14 yields that w.h.p

\[ e(I(\beta), O(\beta)) = O\left(n^{1-(2\alpha-1)(1-\beta)} \log n\right). \quad (6.41) \]

It remains to determine the value of $\bar{D}_k$. Our assumptions on $k$ guarantee that $k \leq n^{\delta'}$, where $\delta' < \frac{(2\alpha-1)(1-\beta)}{2\alpha+1} < 1/2$, provided that $\varepsilon > 0$. With all these facts at hand we can apply Lemma 6.15, which implies that

\[ \mathbb{E}[\bar{D}_k(\beta)] = \Theta(nk^{-(2\alpha+1)}) = \Omega(n^{1-(2\alpha+1)\delta'}) \]

\[ (\delta' < \frac{(2\alpha-1)(1-\beta)}{2\alpha+1}) \omega\left(n^{1-(2\alpha-1)(1-\beta)} \log n\right). \quad (6.42) \]

Recall the definition of the coordinate space $X_1, Y_1, \ldots, X_n, Y_n$ and the critical event $B$ (Section 6.3). As long as $\bar{B}$ holds, it is guaranteed that the effect of one coordinate on $\bar{D}_k(\beta)$ is at most $2cn^{1-\beta} + 1$ (the plus one is for the vertex itself). We therefore apply Theorem 2.2 with
Chapter 6. Random Hyperbolic Graphs

f := \bar{D}_k(\beta), t := n^{3/2-\beta+2\alpha\varepsilon} and ‘bad’ event B as stated in Lemma \ref{lem:6.12}.

Note that M = n and therefore M \cdot \Pr[B] = o(1). It follows that

\Pr[|\bar{D}_k(\beta) - \mathbb{E}[\bar{D}_k(\beta)]| \geq t + o(1)] \leq 2e^{-\Omega(t^2/n^{1+2(1-\beta)})} + e^{-n^{\Omega(1)} = e^{-n^{\Omega(1)}}}.

However, \beta \geq 1 - \frac{1}{4\alpha} + \varepsilon implies for all 0 \leq k \leq n^{\delta'} that

\mathbb{E}[\bar{D}_k(\beta)] = \omega\left(n^{1-(2\alpha-1)(1-\beta)}\right) = \omega\left(n^{3/2-\beta+2\alpha\varepsilon}\right) = \omega(t).

This shows that \bar{D}_k(\beta) = (1 + o(1))\exp[\bar{D}_k(\beta)] holds with probability 1 - e^{-n^{\Omega(1)}} for every k in the considered range. The theorem therefore follows from (6.38)–(6.42).

6.5 Vertices of Large Degree

In the previous section we derived the degree sequence for vertices of degree \( k \leq n^{\delta} \), for a small constant \( \delta \). For \( k > n^{\delta} \) we will observe that the radius of almost all vertices of degree \( k \) is concentrated around a specific \( r_k \). We effectively show that it suffices to bound the number of vertices of radius at most \( r_k \) to get a tight bound on the number of vertices of degree at least \( k \).

Proof of the second part of Theorem \ref{thm:6.3}. We fix a \( k \geq n^{\delta'} \) and define

\[ r_k := 2 \left( \log(n-1) - \log k + \log \xi + \frac{1}{2} \log C \right). \]

Thus, the expected degree of a vertex at radius \( r_k \) is by Lemma \ref{lem:6.10}

\( (n-1)p_{r_k,0} = (n-1)\xi \cdot e^{C/2} \cdot e^{-r_k/2} (1 + o(1)) = k (1 + o(1)). \)

Note that Lemma \ref{lem:6.10} guarantees the existence of a function \( f(n) = o(1) \), such that for every fixed \( 0 \leq r' \leq R \) and all \( r' \leq r \leq R \)

\[ p_{r,0} \leq \xi \cdot e^{C/2} \cdot e^{-r'/2} (1 + f(n)), \]

while for all \( 0 \leq r \leq r' \)

\[ p_{r,0} \geq \xi \cdot e^{C/2} \cdot e^{-r'/2} (1 - f(n)). \]
We set \( \varepsilon := \max \left\{ \frac{\log n}{\sqrt{k}}, k^{-(1-2\alpha)^2}, 4f(n) \right\} \).

Next, let \( L^<_k \) count the vertices of degree at least \( k \) and radius at most \( r_k - \varepsilon \), let \( L^>_k \) count those of degree at least \( k \) and radius at least \( r_k + \varepsilon \), and let \( L^\pm_k \) count those of degree at least \( k \) and radius in \([r_k - \varepsilon, r_k + \varepsilon]\).

Using these conditions on the radius of the points we can write \( D \geq k \) as

\[
D \geq k = L^<_k + L^>_k + L^\pm_k.
\] (6.43)

We will show that \( L^<_k \gg L^>_k \) and \( L^<_k \gg L^\pm_k \) both hold with probability \( 1 - o(1/n) \), and further that this dominating random variable is almost surely close to its expectation.

Let us first inspect \( L^>_k \). The degree distribution of a vertex at distance at least \( r_k + \varepsilon \) is clearly dominated by \( X \sim \text{Bin}(n-1, p^+_k) \) for

\[
p^+_k := \xi \cdot e^{C/2} \cdot e^{-\frac{r_k + \varepsilon}{2}} (1 + \varepsilon/4).
\]

The expectation of such a random variable is

\[
\mathbb{E}[X] = (n - 1) \cdot \xi \cdot e^{C/2} \cdot e^{-\frac{r_k + \varepsilon}{2}} (1 + \varepsilon/4) = k \cdot (1 - \varepsilon/4(1 + o(1))).
\]

Using the Chernoff bound, we bound the probability that such a vertex has degree at least \( k \) by

\[
\Pr[X \geq (1 + \varepsilon/4(1 + o(1))) \mathbb{E}[X]] \leq e^{-\frac{\mathbb{E}[X](\varepsilon/4(1+o(1)))^2}{3}} = e^{-\Omega(\log^2 n)} = n^{-\Omega(\log n)},
\]

and we thereby conclude that \( L^>_k = 0 \) holds with probability \( 1 - o(1/n) \).

For points of distance at most \( r_k - \varepsilon \) a similar argument holds. The degree distribution of all these points dominates \( Y \sim \text{Bin}(n - 1, p^-_k) \) for

\[
p^-_k = \frac{2\alpha T e^{-\frac{r_k - \varepsilon}{2}}}{(\alpha - 1/2) \sin(\pi T)} (1 - \varepsilon/4).
\]

This random variable has expectation

\[
\mathbb{E}[Y] = (n - 1) \cdot \xi \cdot e^{C/2} \cdot e^{-\frac{r_k - \varepsilon}{2}} (1 - \varepsilon/4) = k \cdot (1 + \varepsilon/4(1 + o(1))).
\]

Hence, the probability that such a vertex has degree smaller than \( k \) is at most

\[
\Pr[Y \leq (1 - \varepsilon/4(1 + o(1))) \mathbb{E}[Y]] \leq e^{-\frac{\mathbb{E}[Y](\varepsilon/4(1+o(1)))^2}{2}} = e^{-\Omega(\log^2 n)} = n^{-\Omega(\log n)},
\]
and we therefore have $L_k^\leq = X_{\leq r_k - \varepsilon}$ with probability $1 - o(1/n)$, where $X_{\leq r_k - \varepsilon}$ denotes the number of points with radius at most $r_k - \varepsilon$. The expectation of $X_{\leq r_k - \varepsilon}$ is

$$
\mathbb{E}[X_{\leq r_k - \varepsilon}] = (1 + o(1)) n e^{-\alpha(R - (r_k - \varepsilon))} = (1 + o(1)) \xi^{2\alpha} n k^{-2\alpha} .
$$

(6.44)

Note that $\mathbb{E}[X_{\leq r_k - \varepsilon}] = \Omega(\log^{2\alpha} n)$ for $k \leq \frac{1}{\log n}$, and it therefore follows from the Chernoff bound that

$$
\Pr \left[ |X_{\leq r_k - \varepsilon} - \mathbb{E}[X_{\leq r_k - \varepsilon}]| \geq \frac{1}{\log \log n} \mathbb{E}[X_{\leq r_k - \varepsilon}] \right] \leq o(n^{-1}).
$$

For the third term in (6.43) it suffices to look only at points that have their radius in $[r_k - \varepsilon, r_k + \varepsilon]$ and neglect whether their degree is at least $k$. Clearly

$$
L_k^z \leq X_{\leq r_k + \varepsilon} - X_{\leq r_k - \varepsilon}
$$

and

$$
\mathbb{E}[X_{\leq r_k + \varepsilon} - X_{\leq r_k - \varepsilon}] = n \int_{r_k - \varepsilon}^{r_k + \varepsilon} f(y) dy = o(nk^{-2\alpha}).
$$

It therefore follows again by the Chernoff bound that $L_k^z = o(X_{\leq r_k - \varepsilon})$ with probability $1 - o(1/n)$, and we conclude that with high probability

$$
D_{\geq k} = (1 + o(1)) \mathbb{E}[X_{\leq r_k - \varepsilon}] \overset{(6.44)}{=} (1 + o(1)) \xi^{2\alpha} n k^{-2\alpha}
$$

for all $k$ in the desired range.

\[\square\]

### 6.6 The Expected & Maximum Degree

In the subsequent proofs for Theorem 6.4 and Theorem 6.5 we present two simple applications of Lemma 6.10. In particular, we show how it can be used to determine the expected degree and the maximum degree in $G_{\alpha,C,T}(n)$.

**Proof of Theorem 6.4**. The expected degree of a fixed vertex at radius $r$ is $\overline{k}(r) = (n - 1)p_{r,0}$. Moreover, we can derive the expected degree of an arbitrary vertex by integrating $\overline{k}(r) \cdot f(r)$ from $0$ to $R$. To be able to apply Lemma 6.10 we split this integral into

$$
(n - 1) \int_0^R f(r) p_{r,0} dr + (n - 1) \int_{\beta R}^R f(r) p_{r,0} dr
$$

(6.45)
where $\beta := 1/2 - 1/(4\alpha)$. For the second term in (6.45) we get by Lemma 6.10
\[(1 + o(1))n \int_{\beta R}^{R} f(r)\xi \cdot e^{C/2} \cdot e^{-r/2} dr \quad (6.14) \quad (1 - o(1))\frac{\alpha \xi}{(\alpha - 1/2)}. \quad (6.46)\]

The following upper bound therefore suffices to show that the first term is negligible
\[(n - 1) \int_{0}^{\beta R} f(r) dr = O \left(n^{1-2\alpha+2\alpha\beta}\right) = O \left(n^{1/2-\alpha}\right) = o(1). \quad (6.47)\]

To derive the number of edges we set $\beta := 1 - 1/(3\alpha)$ and denote by $e(O)$ the number of edges between vertices in $O(\beta)$. In a similar way as above we obtain that the expected number of edges in $O(\beta)$ is
\[(1 - o(1))\frac{n^2}{2} \int_{\beta R}^{R} f(r)p_{r,\beta R} dr \quad (6.46) \quad (1 - o(1))\frac{\alpha \xi n}{(\alpha - 1/2)}.\]

Furthermore, it follows from Lemma 6.12 and Theorem 2.2 that $e(O) = (1 + o(1))E[e(O)]$ w.h.p. Finally, the maximum number of edges inside $I(\beta)$ is at most $\Theta(|I(O)|^2)$, and the lemma therefore follows from Lemma 6.13 and Lemma 6.14 as w.h.p
\[|E| - e(O) = o(n). \]

The proof of Theorem 6.5 uses very much the ideas we exploited when we derived the degree sequence for large degrees. In particular, we show that if we condition on the radius of the vertices, then there is a radius $r_0$ such that on one hand there is with high probability a vertex whose degree is close to $np_{r_0}$ while on the other hand $\text{Bin}(n - 1, p_{r_0,0})$ almost surely dominates the degree distribution of every vertex in the graph.

Proof of Theorem 6.5. Let $X_{\leq r}$ denote the number of vertices with radius at most $r$, and define $X_{\geq r}$ analogously. Set $r_0 := (2 - 1/\alpha)\log n$. We will consider vertices with radius $r \in [r_0 - \varphi(n), r_0 + \varphi(n)]$ for $\varphi(n) = \log \log \log n$. Note that for sufficiently large $n$
\[\mathbb{E}[X_{\leq r_0 - \varphi(n)}] = n \cdot \mu \left(B_0 \left(r_0 - \varphi(n)\right)\right) \leq 2e^{-\alpha(\varphi(n)+C)} = o(1). \quad (6.47)\]
It follows that there are w.h.p no vertices of radius at most \( r_0 - \varphi(n) \). Next, by applying (6.21) we obtain the following bounds

\[
\frac{e^{\alpha(\varphi(n) - C)}}{2n} \leq \mu(B_0(r_0 + \varphi(n))) \leq \frac{2e^{\alpha(\varphi(n) - C)}}{n}.
\]

(6.48)

Since the number of vertices with radius at most \( r_0 + \varphi(n) \) is binomially distributed, we infer that

\[
\Pr[X \leq r_0 + \varphi(n) = 0] \leq (1 - \mu(B_0(r_0 + \varphi(n))))^n \leq e^{-e^{\alpha(\varphi(n) - C)/2}} = o(1).
\]

(6.49)

That is, there is w.h.p a vertex with radius at most \( r_0 + \varphi(n) \). On the other hand, we claim that the number of such vertices is very small. To see this, note that (6.48) implies that

\[
\mathbb{E}[X \leq r_0 + \varphi(n)] \leq 2e^{\alpha(\varphi(n) - C)},
\]

and by applying Markov’s inequality we conclude that

\[
\Pr[X \leq r_0 + \varphi(n) > \log n] \leq \frac{2e^{\alpha(\varphi(n) - C)}}{\log n} = o(1).
\]

In other words, the number of vertices with radius larger than \( r_0 + \varphi(n) \) is w.h.p at least \( n - \log n \).

Note that the degree of a vertex at radius \( r \) is distributed like Bin\((n - 1, p_{r,0})\). As there are w.h.p no vertices of radius smaller than \( r_0 - \varphi(n) \), the degree distribution of every vertex is w.h.p dominated by Bin\((n, p_{r_0}^-)\), where by Lemma 6.10

\[p_{r_0}^- := 2p_{r_0 - \varphi(n),0} = (1 + o(1))2\xi e^{C/2} \cdot e^{-\frac{r_0 - \varphi(n)}{2}}.
\]

In this case the expected degree of a fixed vertex is at most

\[np_{r_0}^- \leq (1 + o(1))2\xi e^{C/2} \cdot n^{\frac{1}{2\pi}} e^{\frac{\varphi(n)}{2}}.
\]

and by the Chernoff bound and a union bound there is w.h.p no vertex of degree larger than \((1 + o(1))4\xi e^{C/2} \cdot n^{\frac{1}{2\pi}} e^{\frac{\varphi(n)}{2}}\).

The lower bound can be established by the fact (shown in (6.49)) that there is w.h.p a vertex of radius at most \( r_0 + \varphi(n) \). There are w.h.p at most \( \log n \) such vertices, which means that we can fix one and have
that its degree distribution dominates \( \text{Bin}(n - \log n, p_{r_0}^+) \), where by Lemma 6.10

\[
p_{r_0}^+ := \frac{1}{2} p_{r_0 + \varphi(n), r_0 + \varphi(n)} = (1 + o(1)) \frac{\xi e^{C/2} \cdot e^{-\frac{r_0 + \varphi(n)}{2}}}{2}.
\]

Note that the expectation of \( \text{Bin}(n - \log n, p_{r_0}^+) \) is

\[
(1 + o(1)) np_{r_0}^+ = (1 + o(1)) \frac{\xi e^{C/2} \cdot n^{\frac{1}{2n}} e^{-\frac{\varphi(n)}{2}}}{2},
\]

and the degree of the fixed vertex is therefore w.h.p (by applying the Chernoff bound) not smaller than

\[
\frac{\xi e^{C/2} \cdot n^{\frac{1}{2n}} e^{-\frac{\varphi(n)}{2}}}{4}.
\]

\[\square\]

6.7 The Clustering Coefficient

Recall the definition of the local and global clustering coefficient in (6.6) and (6.7). In this section we present the proof of Theorem 6.6.

Proof of Theorem 6.6. Let \( \beta := 2/3 \), and for a graph \( G = (V, E) \sim G_{\alpha, C, T}(n) \) let

\[
X := \sum_{\substack{v \in V \\
\deg(v) \geq 2}} \frac{|\{\{u_1, u_2\} \in E \mid u_1, u_2 \in N(v)\}|}{\binom{\deg(v)}{2}}
\]

and

\[
Y := \sum_{\substack{v \in O(\beta) \\
\deg(v) \geq 2}} \frac{|\{\{u_1, u_2\} \in E \mid u_1, u_2 \in N(v) \cap O(\beta)\}|}{\binom{\deg(v)}{2}}.
\]

Clearly, \( \bar{c} = \frac{X}{n} \) and \( X \geq Y \). It therefore suffices to derive a constant lower bound on \( \mathbb{E}[Y] \), and to show that \( Y \) is concentrated around its expectation. For a fixed vertex \( v \) with radius \( r_v = r \geq \beta R \) and angle \( \theta_v = 0 \), let

\[
Y_r := \frac{|\{\{u_1, u_2\} \in E \mid u_1, u_2 \in N(v) \cap O(\beta)\}|}{\binom{\deg(v)}{2}},
\]
and let $\mathcal{E}_r$ be the event that the vertex has degree at least 2. We observe that $\mathbb{E}[Y_r \mid \mathcal{E}_r]$ is exactly the probability that a randomly chosen pair of neighbors $\{u_1, u_2\}$ of a vertex $v$ at radius $r$ is connected and that $u_1$ and $u_2$ have radius at least $\beta R$. In order to derive this probability for a fixed vertex $v$ at radius $r$, let us suppose that $u_1$ has coordinates $(r_1, \varphi_1)$ and that $u_2$ has coordinates $(r_2, \varphi_2)$. If we would choose each of the vertices $u_1$ and $u_2$ independently and uniformly at random, the probability for the event that $u_1$ is at $(r_1, \varphi_1)$ would be

$$\text{Pr}[(r_1, \varphi_1) \mid \{u_1, v\} \in E] = \frac{f(r_1) \cdot \chi_{r_1, r}^{\varphi_1}}{2\pi p_{r,0}},$$

and similarly, the probability for the event that $u_2$ is at $(r_2, \varphi_2)$ would be

$$\text{Pr}[(r_2, \varphi_2) \mid \{u_2, v\} \in E] = \frac{f(r_2) \cdot \chi_{r_2, r}^{\varphi_2}}{2\pi p_{r,0}},$$

where $p_{r,0}$ is as defined in Lemma 6.10. Furthermore, the probability that $u_1$ and $u_2$ are connected would be $\chi_{r_1, r_2}^{\varphi_1 - \varphi_2}$. Note that choosing a pair of vertices $\{u_1, u_2\}$ uniformly at random from a set is equivalent to choosing the two vertices independently and conditioning on the event that $u_1 \neq u_2$. However, since the degree of $v$ is at least 2 this event happens with probability at least $1/2$. Putting all the above facts together implies that

$$\mathbb{E}[Y_r \mid \mathcal{E}_r] \geq \int_{\beta R}^{R} \int_{\beta R}^{R} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{f(r_1)f(r_2) \cdot \chi_{r_1, r}^{\varphi_1} \cdot \chi_{r_2, r}^{\varphi_2} \cdot \chi_{r_1, r_2}^{\varphi_1 - \varphi_2}}{8\pi^2 p_{r,0}^2} \cdot \chi_{r_1, r_2}^{\varphi_1 - \varphi_2} d\varphi_2 d\varphi_1 dr_1 dr_2. \tag{6.50}$$

We lower bound this term by integrating only over $r \leq r_1, r_2, \leq R$, $-e^{-r_1/2} \leq \varphi_1 \leq e^{-r_1/2}$ and $-e^{-r_2/2} \leq \varphi_2 \leq e^{-r_2/2}$. Thus by (6.14) and Lemma 6.10 the above term is at least

$$\frac{\alpha^2 e^r}{8\xi^2 \pi^2 e^{2\alpha R + C}} \int_{r}^{R} \int_{r}^{R} \int_{-e^{-r_1/2}}^{e^{-r_1/2}} \int_{-e^{-r_2/2}}^{e^{-r_2/2}} e^{\alpha(r_1 + r_2)} \cdot \chi_{r_1, r}^{\varphi_1} \cdot \chi_{r_2, r}^{\varphi_2} \cdot \chi_{r_1, r_2}^{\varphi_1 - \varphi_2} \cdot \chi_{r_1, r_2}^{\varphi_1 - \varphi_2} d\varphi_2 d\varphi_1 dr_2 dr_1 \tag{6.50}$$

where $\xi$ is the constant defined in (6.4). Let us now inspect the three connection probabilities. Applying Lemma 6.9, we observe that since
\[ \frac{r_1}{2} \leq r \leq R \]

\[ \chi_{|r_1, r|}^{\varphi_1} = \frac{1}{1 + e^{\frac{1}{2T} (d(r, r_1, |\varphi_1|) - R)}} = \frac{1}{1 + (|\varphi_1| + e^{-r} + e^{-r_1})^\frac{1}{T} e^{\frac{1}{2T} (r + r_1 + \Theta(1) - R)}} \tag{6.51} \]

\[ \geq \frac{1}{1 + (3e^{\frac{1}{2} (r - R + \Theta(1))})^\frac{1}{T}} \geq \frac{1}{1 + 3e^{\frac{\Theta(1)}{2T}}} \tag{6.52} \]

and similarly

\[ \chi_{|r_2, r|}^{\varphi_2} = \frac{1}{1 + 3e^{\frac{\Theta(1)}{2T}}} \text{ and } \chi_{|r_1, r_2|}^{\varphi_1 - \varphi_2} = \frac{1}{1 + 4e^{\frac{\Theta(1)}{2T}}} \tag{6.53} \]

It therefor follows from (6.50)-(6.53) that there exists a constant \( c_{\alpha, T} \) such that

\[ \mathbb{E}[Y_r | \mathcal{E}_r] \geq c_{\alpha, T} e^{r - 2\alpha R} \int_r^R \int_r^R e^{(\alpha - 1/2)(r_1 + r_2)} dr_2 dr_1 \]

\[ \geq \frac{c_{\alpha, T}}{(\alpha - 1/2)^2} \left( e^{-\frac{1}{2} (R - r)} - e^{-\alpha (R - r)} \right)^2. \tag{6.54} \]

Next, we inspect the probability for the event \( \mathcal{E}_r \). For \( n \) large enough

\[ \Pr[\mathcal{E}_r] = \Pr[\deg(v) \geq 2 \mid r_v = r] \geq \Pr[\deg(v_r) = 2 \mid r_v = R] \]

\[ = \left( \binom{n - 1}{2} \right) p_{R, 0}^2 (1 - p_{R, 0})^{n-3} \]

\[ \geq \frac{n^2}{3} \xi^2 e^{C} e^{-R} e^{-n\xi} e^{-C/2} e^{-R/2} = \frac{\xi}{3} e^{-\xi} = \Theta(1). \tag{6.55} \]

By integrating \( \mathbb{E}[Y_r | \mathcal{E}_3] \) over all \( \beta R \leq r \leq R \) and multiplying with \( \Pr[\mathcal{E}_3] \), \( f(y) \) and \( n \), we see that the expected value of \( Y \) is bounded from below by

\[ \mathbb{E}[Y] = n \int_{\beta R}^R f(y) \Pr[\mathcal{E}_r] \mathbb{E}[Y_r | \mathcal{E}_r] dr \]

\[ \geq \frac{n \Pr[\mathcal{E}_r] c_{\alpha, T} \alpha}{2(\alpha - 1/2)^2} \int_{\beta R}^R \left( e^{-\frac{1+\alpha}{2} (R - r)} - e^{-\frac{3\alpha}{2} (R - r)} \right)^2 dr \tag{6.56} \]

for \( n \) sufficiently large. Since

\[ \int_{\beta R}^R \left( e^{-\frac{1+\alpha}{2} (R - r)} - e^{-\frac{3\alpha}{2} (R - r)} \right)^2 dr = \frac{(1 - 2\alpha)^2}{3\alpha(1 + \alpha)(1 + 4\alpha)} - o(1), \]
it follows from (6.54), (6.55) and (6.56) that $\mathbb{E}[Y] = \Theta(n)$.

Set $f := Y, t := n^{6/7}$ and ‘bad’ event $\mathcal{B}$ and $c$ as stated in Lemma 6.12. Note that each coordinate change can influence $f$ by at most $c_i := 2cn^{1-\beta} + 1$ as long as $\bar{\mathcal{B}}$ holds. It therefore follows from Theorem 2.2 and Lemma 6.12 that $\Pr[X \leq \mathbb{E}[Y] - n^{6/7} - Pr[\mathcal{B}] = o(1)$, and therefore that the clustering coefficient is with high probability at least $\mathbb{E}[Y]/n = \Theta(1)$.

## 6.8 Compressibility

In this section we discuss the compressibility of random hyperbolic graphs and thereby prove Theorem 6.7. First, in Section 6.8.1 and Section 6.8.2 we introduce the considered compression scheme, and discuss the encoding of integers. Then, in Section 6.8.3 we prove an upper bound on the expected encoding length of a random hyperbolic graph, and finally, in Section 6.8.4 we show that the length of the encoding is concentrated around its expectation.

### 6.8.1 BV-Compression

Algorithm 6.8.1 describes the considered BV-type compression scheme. In [BV04], this approach has been further refined by encoding only the

---

**Algorithm 6.8.1 BV**

Input: $(G = (V, E)$ and an ordering such that $V = \{v_1, v_2, \ldots, v_{|V|}\}$

Output: Bit string $BV(G)$

For each vertex $v \in V$, let $\Gamma(v) = \{u_1, \ldots, u_k\}$ be all neighbors of $v$ such that $v \leq u_i$ (for $1 \leq i \leq v$) and let them be ordered such that $u_1 \leq \cdots \leq u_{\deg(v)}$. The algorithm creates for every $v \in V$ a list of integers $L(v) = u_1 - v, u_2 - v, \ldots, u_k - v$.

---

distances between consecutive neighbors of a vertex, and by using some more involved heuristics and compression methods.

For many applications it is very important to be able to find all the neighbors of a given vertex, without having to decode the whole graph. In the encoding scheme presented in Algorithm 6.8.1 it is only possible to read all the neighbors that are larger in the given ordering. However,
6.8. Compressibility

by encoding the graph again with the reverse ordering, we only need twice as much storage, and are able to instantly access all the neighbors of a given vertex.

6.8.2 Storing Lists of Integers

In the described BV-encoding (Algorithm 6.8.1), adjacency lists are encoded as lists of integers. Using the Elias $\gamma$-code (see for example [WMB99]), it is possible to store lists of positive integers such that each $x \in \mathbb{Z}_+$ needs at most $1 + 2 \lceil \log x \rceil \leq 3 \log(x + 1)$ bits. Since our gaps are positive, we can therefore assume that an edge of length $x$ can be stored using at most $3 \log(x + 1)$ bits.

6.8.3 Upper Bound on the Expected Code Length

In order to prove Theorem 6.7 we consider a graph $G \sim G_{\alpha,C,T}(n)$. Clearly, the length of the BV-encoding of $G$, from here on denoted as $|BV(G)|$, is a random variable that depends only on the structure of $G$.

The following lemma bounds the expected code length of a random hyperbolic graph from above.

**Lemma 6.16.** Let $\alpha > 1/2$, $C \in \mathbb{R}$, $0 < T < 1$ and let $G = (V,E) \sim G_{\alpha,C,T}(n)$, such that the vertices in $V$ are ordered by their angle in the hyperbolic plane. Then, there exists an absolute constant $\zeta$ that does not depend on $C, \alpha$ and $T$, such that

$$
\mathbb{E}[|BV(G)|] \leq \zeta \cdot \sin(\pi T) \cdot \left( \frac{2}{T} + \frac{e^{\bar{c} (T+2)}}{(1-T)^2} \right) \cdot \max\{1,-C,\frac{1}{\alpha-1/2}\} \cdot |E|
$$

$$
+ \frac{48\alpha^2}{(\alpha-1/2)^2} |V|,
$$

where $\bar{c}$ is the absolute constant given by (6.12).

**Proof.** For the sake of a simple notation, let

$$
C_0^- := \min\{0,C\} \quad \text{and} \quad C_0^+ := \max\{0,C\}.
$$

If vertex $u$ and vertex $v$ are connected, let $X_{u,v}$ count the number of vertices between $u$ and $v$ in the angular ordering, and else let $X_{i,j} = 0$. Note that the length of an edge $\{u,v\}$ that has to be encoded is exactly...
and this number can be stored (see Section 6.8.2) using at most $3 \log(X_{u,v} + 2)$ bits. The compression scheme therefore needs in total at most $X := \sum_{u,v \in V} 3 \log(X_{u,v} + 2)$ bits and a small additional overhead that is at most linear in $|E|$ and therefore considered in the constant $\zeta$. Furthermore, without fixing the coordinates of two vertices $u$ and $v$, we observe that

$$\mu := \mathbb{E}[3 \log(X_{u,v} + 2)]$$

is the same for any two $u,v \in V$ and therefore

$$\mathbb{E}[|BV(G)|] = O(\mathbb{E}[X]) = O(n^2 \mu). \quad (6.57)$$

By conditioning on the radiuses $r_1, r_2$ and an angle difference of $\theta$ between $u$ and $v$, it follows from Jensen’s inequality that

$$\mathbb{E} [\log(X_{u,v} + 2) | r_u = r_1, r_v = r_2, |\theta_u - \theta_v| = \theta] \leq \log \left( \frac{(n-2)\theta}{\pi} + 2 \right) \cdot \chi^{\theta}_{r_1,r_2} \leq \log(n\theta + 2) \cdot \chi^{\theta}_{r_1,r_2},$$

and therefore that

$$\mu \leq 3 \int_0^\pi \int_0^R \int_0^R \log(n\theta + 2) \cdot f(r_1) \cdot f(r_2) \cdot \chi^{\theta}_{r_1,r_2} dr_2 dr_1 d\theta. \quad (6.58)$$

Whenever $r_1 + r_2 \leq R$, then the distance $d(r_1, r_2, \theta)$ is at most $R$ and the probability that the two vertices are connected by an edge is close to one. Using this observation, we bound (6.58) from above by

$$3 \int_0^\pi \int_0^R \int_0^{R-r_1} \log(n\theta + 2) \cdot f(r_1) \cdot f(r_2) \cdot \chi^{\theta}_{r_1,r_2} dr_2 dr_1 d\theta + 3 \int_0^\pi \int_0^R \int_0^{R-r_1} \log(n\theta + 2) \cdot f(r_1) \cdot f(r_2) \cdot \chi^{\theta}_{r_1,r_2} dr_2 dr_1 d\theta. \quad (6.59)$$

The first term of (6.59) is by (6.14) at most

$$6\alpha^2 e^{-2\alpha R} \int_0^\pi \int_0^R \int_0^{R-r_1} \ln (n\theta + 2) e^{\alpha(r_1 + r_2)} dr_2 dr_1 d\theta \leq 6\alpha e^{-\alpha R} \int_0^\pi \int_0^R \ln (n\theta + 2) dr_1 d\theta \leq 6\alpha e^{-\alpha R} R \left[ \ln(n\theta + 2)(\theta + \frac{2}{n} - \theta) \right]^\pi_0 = o(1/n). \quad (6.60)$$

For very small angles, in particular if $\theta \leq e^{R-r_1-r_2}$ we observe that $\chi^{\theta}_{r_1,r_2}$ is also close to 1 and we therefore bound the second term of
Next, we change the order of integration of the first term of (6.61) to bound it by

\[
6.59 \quad 3 \int_0^e \int_0^R \int_0^R \ln(n\theta + 2) \cdot f(r_1) \cdot f(r_2) d\theta dr_2 dr_1
\]
\[
+ 3 \int_0^{\pi} \int_0^R \int_0^R \ln(n\theta + 2) \cdot f(r_1) \cdot f(r_2) \cdot \chi_{r_1, r_2} \ dr_2 dr_1 d\theta
\]

Next, we change the order of integration of the first term of (6.61) to bound it by

\[
6.60 \quad 6 \alpha^2 e^{-2\alpha R} \int_0^R \int_{R-r_1}^R \int_0^{e^{\frac{R-r_1-r_2}{2}}} \ln(n\theta + 2) \ e^{\alpha(r_1+r_2)} d\theta dr_2 dr_1
\]
\[
\leq 6 \alpha^2 e^{-2\alpha R} \int_0^R \int_{R-r_1}^R \int_0^{e^{\frac{R-r_1-r_2}{2}}} \left[ \ln(n\theta + 2)(\theta + \frac{2}{n} - \theta) \right] e^{\frac{R-r_1-r_2}{2}} dr_2 dr_1
\]

Since \(1/n = e^{-\frac{R-C}{2}}\), the above term is at most

\[
6.61 \quad 12 \alpha^2 e^{-2\alpha R} \int_0^R \int_{R-r_1}^R \ e^{\alpha(r_1+r_2)} \ln \left( ne^{\frac{R-r_1-r_2}{2}} + 2 \right) \left( e^{\frac{R-r_1-r_2}{2}} + e^{-\frac{R-C}{2}} \right) dr_2 dr_1.
\]

Note that \(ne^{\frac{R-r_1-r_2}{2}} = e^{\frac{2R-C-r_1-r_2}{2}}\) and that therefore \(\ln(ne^{\frac{R-r_1-r_2}{2}} + 2) \leq 2\alpha R - C - r_1 - r_2 + 2\) and \(e^{\frac{R-r_1-r_2}{2}} \geq e^{-\frac{R-C}{2}}\) for \(2R \geq r_1 + r_2 + C\), while \(\ln(ne^{\frac{R-r_1-r_2}{2}} + 2) \leq 2\) and \(e^{\frac{R-r_1-r_2}{2}} \leq e^{-\frac{R-C}{2}}\) else. This observation allows us to bound (6.63) from above by

\[
6.62 \quad 24 \alpha^2 e^{-2\alpha R} \int_0^R \int_{R-r_1}^{R-C_0^+} \ e^{\alpha(r_1+r_2)} \left( \frac{2R - C + 4 - r_1 - r_2}{2} e^{\frac{R-r_1-r_2}{2}} \right) dr_2 dr_1
\]
\[
+ 48 \alpha^2 e^{-2\alpha R} \int_0^R \int_{R-C_0^+}^R \ e^{\alpha(r_1+r_2)} e^{-\frac{R-C}{2}} dr_2 dr_1 \leq \frac{48 \alpha^2}{(\alpha - 1/2)^2 n}
\]
\[
+ 12 \alpha^2 e^{-(2\alpha-1/2)R} \int_0^R \int_{R-r_1}^{R-C_0^+} \ e^{(r_1+r_2)(\alpha-1/2)} (2R - C + 4 - r_1 - r_2) dr_2 dr_1.
\]

Next, since

\[
\int e^{r(\alpha-1/2)}(X-r) dr = e^{r(\alpha-1/2)}(\alpha-1/2)^2 (1 + (\alpha - 1/2)(X-r))
\]
we substitute first $X = 2R - C + 4 - r_1$ and then $X = R - C + C_0^+ + 4$

to bound the integral in (6.64) from above by

\[
\begin{align*}
&\frac{e^{R(\alpha-1/2)}}{(\alpha - 1/2)^2} \int_0^R e^{r_1(\alpha-1/2)} (1 + (\alpha - 1/2)(R - C + C_0^+ + 4 - r_1)) dr_1 \\
&\leq \frac{e^{R(\alpha-1/2)}}{(\alpha - 1/2)^2} \left( \frac{e^{R(\alpha-1/2)}}{(\alpha - 1/2)} + (\alpha - 1/2) \frac{e^{R(\alpha-1/2)}}{(\alpha - 1/2)^2} (1 - (C - C_0^+ - 4)(\alpha - 1/2)) \right) \\
&\leq \frac{e^{2R(\alpha-1/2)}}{(\alpha - 1/2)^3} \left( 2 - (C^- - 4)(\alpha - 1/2) \right).
\end{align*}
\]

Thus, the second term of (6.64) is at most

\[
\frac{12\alpha^2(2 - (C^- - 4)(\alpha - 1/2))}{(\alpha - 1/2)^3} \frac{e^{-C/2}}{n}. \quad (6.67)
\]

For the second term of (6.61), it follows by (6.12) and the fact that
\[
\ln(n\theta + 2) \leq \ln(n\theta) + 2 \quad \text{for} \quad \theta \geq e^{\frac{R + C_0^+ - r_1 - r_2}{2}}
\]
and \(\ln(n\theta + 2) \leq 2\) else, that it is at most

\[
\begin{align*}
&6\alpha^2 e^{\frac{c}{2\pi}} e^{-R(2\alpha - 1/2)} \left( 2 \int_0^R \int_{R-r_1}^R e^{(r_1+r_2)(\alpha - 1/2)} \int_{e^{\frac{R + C_0^+ - r_1 - r_2}{2}}}^{e^{\frac{R + C_0^+ - r_1 - r_2}{2}}} \theta^{-\frac{1}{\pi}} d\theta dr_2 dr_1 \\
&\quad + \int_0^R \int_{R-r_1}^R e^{(r_1+r_2)(\alpha - 1/2)} \int_{e^{\frac{R + C_0^+ - r_1 - r_2}{2}}}^{\pi} \theta^{-\frac{1}{\pi}} (\ln(n\theta) + 2) d\theta dr_2 dr_1 \right),
\end{align*}
\]

where the first term of (6.68) is at most

\[
\frac{12\alpha^2 Te^{\frac{c}{2\pi}}}{(\alpha - 1/2)^2(1-T)} \frac{e^{-C/2}}{n}. \quad (6.69)
\]

Moreover, since

\[
\int \theta^{-\frac{1}{\pi}} (\ln(n\theta) + 2) = \frac{T}{(T - 1)^2} \left[ \theta^{1 - \frac{1}{\pi}} ((T - 1) \ln(n\theta) - T - 2) \right],
\]
which is negative as \(T < 1\), the inner most integral of the second term in (6.68) is at most

\[
\frac{T}{(T - 1)^2} e^{(1 - \frac{1}{\pi}) \frac{R + C_0^+ - r_1 - r_2}{2}} \left( (T + 2) + (1 - T) \left( \frac{2R - C + C_0^+ - r_1 - r_2}{2} \right) \right).
\]
Thus, the second term of (6.68) is at most
\[\int_0^R \int_{R-r_1}^R e^{(\alpha-1/2)(r_1+r_2)} \left(\frac{2R - C + C_0^+ - r_1 - r_2}{2(1-T)^{-1}} + 3\right) dr_2 dr_1\]
\[\cdot \frac{6\alpha^2 e^{\frac{C}{T^2}} T e^{-2\alpha R + \frac{R}{T}}}{(1-T)^2 e^{\frac{1-T}{T} C_0^+}} \leq \frac{3\alpha^2 e^{\frac{C}{T^2}} T}{(1-T)^2} e^{-2\alpha R + \frac{R}{T}}\]
\[\cdot \int_0^R \int_{R-r_1}^R e^{(\alpha-1/2)(r_1+r_2)} (2R - C + C_0^+ - r_1 - r_2) dr_2 dr_1\]
\[+ \frac{18\alpha^2 e^{\frac{C}{T^2}} T}{(\alpha - 1/2)^2(1-T)^2} e^{-\frac{C}{T}}.\]  \hspace{1cm} (6.70)

By substituting \(X = R - C + C_0^+ - r_1\) into (6.65) the first term of (6.70) is at most
\[\frac{3\alpha^2 e^{\frac{C}{T^2}} T}{(\alpha - 1/2)^2(1-T)} e^{-\alpha R} \int_0^R e^{(\alpha-1/2)r_1} (1+(\alpha-1/2)(R-C+C_0^+ - r_1)) dr_1\]  \hspace{1cm} (6.71)
and by applying (6.65) a second time with \(X = R - C + C_0^+\) we get an upper bound of
\[\frac{3\alpha^2 e^{\frac{C}{T^2}} T}{(\alpha - 1/2)^3(1-T)} (2 - C_0^- (\alpha - 1/2)) e^{-\frac{C}{T}} \frac{n}{\alpha}.\]  \hspace{1cm} (6.72)

Summing over the bounds in (6.60), (6.64), (6.67), (6.69), (6.70) and (6.72), we conclude that
\[n^2 \mu \leq e^{-C/2} n \frac{72\alpha^2 T}{(\alpha - 1/2)^2} \left(-C_0^- + \frac{2}{\alpha - 1/2}\right) \left(\frac{2}{\frac{T}{n}} + \alpha \frac{\pi}{(1-T)^2}\right) + \frac{48\alpha^2 n}{(\alpha - 1/2)^2}.\]  \hspace{1cm} (6.73)

The lemma therefore follows by comparing (6.73) with the number of edges in \(G\) (see Theorem 6.4).

### 6.8.4 Concentration

In order to show that it is unlikely to need much more than \(\mathbb{E}[|BV(G)|]\) bits, we wish to apply Theorem 2.2. We consider again the coordinate space \(X_1, Y_1, \ldots, X_n, Y_n\) described in Section 6.3. In this setting \(|BV(G_{\alpha,C,T}(n))|\) is clearly a function of \(X_1, Y_1, \ldots, X_n, Y_n\), but there arise two problems when we try to bound the effect of one coordinate.
The obvious problem are points which are close to the center, since they can have lots of adjacent edges. The other case that might raise problems are long edges. Long edges are edges between points of distance much longer than $R$, which means that those edges might need lots of storage as they can connect two vertices that are far from each other in our cyclic arrangement. We say that an edge is bad if it is long or adjacent to a central vertex, and we denote all other edges as good. We first show that there are w.h.p only few bad edges. Then, we derive a suitable Lipschitz condition for all the other edges and show that the code length we need for the good edges is w.h.p in $O(\mathbb{E}[|\text{BV}(G)|])$.

More formally, for $0 < \beta < 1$ let $I = I(\beta)$ denote the set of all inner vertices of $G_{\alpha, C}(n)$ with radius at most $\beta R$ while the outer set $O = O(\beta)$ denotes the set of vertices of radius larger than $\beta R$. Furthermore let $e(I, O)$ be the set of edges with one endpoint in $I$ and one in $O$ and let $e(I)$ be the set of edges with both endpoints in $I$ ($O$). The following lemma state that the number of edges with an endpoint in $I$ is small.

**Lemma 6.17.** Let $0 < \beta < 1$. Then, we have with high probability

$$e(I, O) + e(I) = o\left(\frac{n}{\log n}\right).$$

**Proof.** The expected number of such edges is at most

$$n^2 \int_0^{\beta R} \int_0^R \int_0^\pi f(r_1)f(r_2)\chi_{r_1, r_2}^\theta d\theta dr_2 dr_1. \quad (6.74)$$

Since $\chi_{r_1, r_2}^\theta$ is close to one if $r_2 \leq R - r_1$ or if $\theta \leq e^{\frac{R-r_1-r_2}{2}}$, we split the integral in three parts and upper bound $\chi_{r_1, r_2}^\theta$ by one in the first two parts

$$\leq n^2 \int_0^{\beta R} \int_0^{R-r_1} \int_0^\pi f(r_1)f(r_2) d\theta dr_2 dr_1 \quad (6.74)$$

$$+ n^2 \int_0^{\beta R} \int_{R-r_1}^R \int_0^{e^{\frac{R-r_1-r_2}{2}}} f(r_1)f(r_2) d\theta dr_2 dr_1 \quad (6.75)$$

$$+ n^2 \int_0^{\beta R} \int_{R-r_1}^R \int_{e^{\frac{R-r_1-r_2}{2}}}^\pi f(r_1)f(r_2)\chi_{r_1, r_2}^\theta d\theta dr_2 dr_1.$$

The first term in (6.75) is clearly at most

$$O\left(n^2 e^{-\alpha R} R\right) = O(n^{2-2\alpha} \log n), \quad (6.76)$$
and the second is at most
\[ O \left( n^2 e^{-\alpha R} \int_0^{\beta R} e^{r_1(\alpha - \frac{1}{2})} dr_1 \right) = O \left( n^{1-(2\alpha-1)(1-\beta)} \right). \tag{6.77} \]

For the third term of \((6.75)\) we use \((6.12)\) to derive an upper bound of
\[ O \left( n^2 e^{-2\alpha R + \frac{R}{2\pi}} \int_0^{\beta R} \int_{R-r_1-2 \log \pi}^R e^{(r_1+r_2)(\alpha - \frac{1}{2})} \int_{e^{\frac{R-r_1-r_2}{2}}}^{\pi} \theta^{-\frac{1}{2}} d\theta dr_2 dr_1 \right). \tag{6.78} \]

Since \(T < 1\), by calculating the integrals in \((6.78)\), we obtain an upper bound of
\[ O \left( n^{1-(2\alpha-1)(1-\beta)} \right) \]
for the third term of \((6.75)\). All in all, we conclude that
\[ \mathbb{E}[e(I, O) + e(I)] = O \left( n^{1-(2\alpha-1)(1-\beta)} \right) \]
and the theorem follows by Markov’s inequality since \(1-(2\alpha-1)(1-\beta) < 1\).

**Lemma 6.18.** Let \(\varepsilon < 1\). Then, the number of edges for which the end vertices have distance at least \((1 + \varepsilon)R\) is w.h.p \(o \left( \frac{n}{\log n} \right)\).

**Proof.** We fix the the radiuses of two vertices to be \(r_1\) and \(r_2\). The probability that the two vertices are connected by an edge of length at least \((1 + \varepsilon)R\) is
\[ \frac{1}{\pi} \int_{\theta \leq \pi \text{ s.t. } d(r_1, r_2, \theta) \geq (1+\varepsilon)R} \chi_{r_1, r_2}^\theta d\theta. \]
Since, by Lemma [6.9] the integration bounds for \(\theta\) can be expressed as
\[ \pi \geq \theta \geq e^{\frac{1}{2}((R+\varepsilon)-r_1-r_2) \pm \Theta(1)}, \]
and since we can therefore restrict \(r_2\) to \(r_2 \geq (1+\varepsilon)R - r_1 - \Theta(1)\) as otherwise \(e^{\frac{1}{2}((R+\varepsilon)-r_1-r_2) + \Theta(1)} \geq \pi\), the expected number of long edges is at most
\[ \binom{n}{2} \frac{1}{\pi} \int_0^R f(r_1) \int_{(1+\varepsilon)R-r_1 \pm \Theta(1)}^R f(r_2) \int_{e^{\frac{1}{2}((R+\varepsilon)-r_1-r_2) \pm \Theta(1)}}^{\pi} \chi_{r_1, r_2}^\theta d\theta dr_2 dr_1. \tag{6.79} \]
Using (6.12), we upper bound this expectation by

\[ O\left(n^2 e^{-(2\alpha - \frac{1}{2}T)R} \int_0^R \int_{(1+\varepsilon)R-r_1 \pm \Theta(1)}^R e^{(r_1+r_2)(\alpha - \frac{1}{2}T)} \int_{\pi}^{\frac{\pi}{2}} (R(1+\varepsilon)-r_1-r_2) \pm \Theta(1) \theta^{-\frac{1}{2}} d\theta dr_2 dr_1 \right). \]

(6.80)

Similar calculations as in Lemma 6.16 and Lemma 6.17 show that

(6.80) = O\left(n^{1-\varepsilon(\frac{1}{T}-1)}\right),

and the lemma therefore follows by Markov’s inequality as \(1 - \varepsilon(\frac{1}{T}-1) < 1\).

With Lemma 6.16, Lemma 6.17, and Lemma 6.18 at hand, we can finally prove our main result.

Proof of Theorem 6.7: Set \(\varepsilon := 1/17\) and \(\beta := 16/17\). We showed in Lemma 6.17 and Lemma 6.18 that there are w.h.p at most \(o\left(\frac{n}{\log n}\right)\) bad edges. Note that in \(BV\), the total number of bits used to encode bad edges is therefore w.h.p at most \(o\left(\frac{n}{\log n} \cdot \log n\right) = o(n)\). It therefore suffices to show that the remaining good edges can w.h.p be encoded by using at most \(t := n^{67/68}\) bits more than in expectation. In the setting of Theorem 2.2, let \(X_i\) and \(Y_i\) (for \(1 \leq i \leq n\)) be the random variables described in Section 6.3. Let \(BV := BV(X_1, Y_1, \ldots, X_n, Y_n)\) count the number of bits we need to encode the good edges and note that \(0 \leq BV \leq n \log n\).

To prove our main theorem, it thus suffices to define a good event \(\tilde{E}\) that holds with probability \(1 - 2^{-n^{O(1)}}\) and that guarantees that the effect of \(X_i, Y_i\) on the the amount of bits needed to encode the good edges is at most \(c_i = O(n^{8/17})\). To that manner, we divide the disk into segments \(S_1, S_2, \ldots, S_k\) of angle \(e^{-R(\frac{1}{2} - \varepsilon - (1-\beta))}\). Note that whenever two vertices \(u\) and \(v\) of radius at least \(\beta R\) are in non adjacent segments then their distance is at least

\[d(\beta R, \beta R, e^{-R(\frac{1}{2} - \varepsilon - (1-\beta))}) \geq 2\beta R + 2 \log(e^{-R(\frac{1}{2} - \varepsilon - (1-\beta))}) - \Theta(1) = 2\beta R - R(1 - 2\varepsilon - 2(1 - \beta)) - \Theta(1) \geq R(1 + 2\varepsilon) - \Theta(1) > R(1 + \varepsilon)\]

(6.81)
for $n$ large enough. It is therefore not possible that we have good edges between vertices in non adjacent segments. Let $Z_i$ count the number of vertices in segment $i$ and note that for all $1 \leq i \leq k$

$$\mathbb{E}[Z_i] = n \frac{e^{-R(\frac{1}{2} - \varepsilon) + (1 - \beta)}}{2\pi} \leq \frac{e^{-C(\frac{1}{2} - \varepsilon) + (1 - \beta)}}{2\pi} n^{2(\varepsilon + (1 - \beta))} \leq \frac{e^{-13C/34}}{2\pi} n^{4/17}.$$ 

It follows from the Chernoff bound that the probability for the event that $Z_i$ is at least $e^{-13C/34} n^{4/17}$ is at most $2^{-n^{\Omega(1)}}$. Let $E$ be the event that for some $1 \leq i \leq k$ the random variable $Z_i$ is at least $e^{-13C/34} n^{4/17}$. Note that $k = \Theta(e^{R(\frac{1}{2} - \varepsilon) - (1 - \beta)}) = \Theta(n^{15/17})$ and that the probability of $E$ is therefore at most $\Theta(n^{15/17}) 2^{-n^{\Omega(1)}} = 2^{-n^{\Omega(1)}}$.

Next, we bound the effect of a single coordinate change under the assumption of $\bar{E}$. The effect of $X_i$ and $Y_i$ is at most the maximal amount of bits needed to encode the good edges adjacent to $v_i$ plus at most one bit for every good edge that goes over $v_i$. Moreover, all good neighbors of $v_i$ are either in in the same sector as $v_i$ or in one of the two adjacent sectors. Thus, under the assumption of $\bar{E}$, the number of bits needed to encode the edges to those good neighbors is at most $O(n^{8/17} \log n)$. Similarly, for every good edge $\{u, w\}$ that goes over $v_i$, both vertices $u$ and $w$ have to be in the same sectors as $v_i$ or in one of the two adjacent sectors. Hence, there are at most $O(n^{8/17})$ such edges. All in all, if $\bar{E}$ holds, the effect of one random variable $X_i$ or $Y_i$ (for $1 \leq i \leq n$) is at most $c_i = O(n^{8/17})$. It therefore follows from Theorem 2.2 that

$$\Pr[f > \mathbb{E}[BV] + t + (n \log n) \Pr[\mathcal{E}]] \leq e^{-2t^2/n^{33/17}} = o(1),$$

and the theorem follows as $\mathbb{E}[BV] + t + (n \log n) \Pr[\mathcal{E}]$ is clearly at most $(1 + o(1))\mathbb{E}[BV]$. 

6.8.5 Open Problems

In this section we obtained an upper bound of $O(n)$ on the number of bits used by the BV-scheme to compress a random hyperbolic graph with $n$ vertices. More precisely, we derive an upper bound (Lemma 6.16) that increases in the number of edges as well as with $C \to -\infty$, $T \to 1$ and $\alpha \to 1/2$. Note that any compression scheme that uses adjacency lists needs at least $\Omega(|E|)$ bits. Moreover, for every graph with an average degree of $\bar{k}$, the BV-scheme needs at least $\Omega(\log(\bar{k}) \cdot |E|)$ bits, since
the average length of an edge increases in $\bar{k}$ and since the length $x$ of an edge can be encoded using $O(\log(x+1))$ bits (see Section 6.8.2). Therefore, our dependency in $-C$ seems to be optimal as (by Theorem 6.4) the average degree increases in $e^{-C/2}$ and our bound on $|BV(G)|$ increases proportional to $-C$. Determining the correct dependency of $T$ and $\alpha$ is an interesting open question.
Bibliography


[BV05] ———, *Codes for the world wide web*, Internet Mathematics **2** (2005), no. 4, 407–429. 17

[BW05] Evgeni Begelfor and Michael Werman, *The world is not always flat or learning curved manifolds*, Tech. report, School of Engineering and Computer Science, Hebrew University of Jerusalem, 2005. 28


[Jan09] Svante Janson, The probability that a random multigraph is simple, Combinatorics, Probability & Computing 18 (2009), no. 1-2, 205–225. 23 56


Bibliography


Curriculum Vitae

Ueli Peter
born on March 23, 1982
in Zürich, Switzerland

1989 - 1995 Primarschule in Bassersdorf
1995 - 1998 Sekundarschule in Bassersdorf
1998 - 2002 Apprenticeship at Paninfo AG, Brüttisellen
2002 - 2004 Kantonale Maturitätsschule für Erwachsene in Zürich
   Degree: Swiss Federal Maturity
2004 - 2007 Studies in Computer Science at ETH Zurich
   Degree: Bachelor of Science ETH in Computer Science
2007 - 2009 Studies in Computer Science at ETH Zurich
   Degree: Master of Science ETH in Computer Science
2009 - 2010 Software Engineer at Comerge, Zurich
since 2010 Ph.D. student at ETH Zurich