Approximability of Combinatorial Optimization Problems on Power Law Networks

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Mikael Gast

aus

Köln

Bonn, Juni 2013

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Marek Karpinski
2. Gutachter: Prof. Dr. Heiko Röglin

Tag der Promotion: 4. September 2013
Erscheinungsjahr: 2013

ABSTRACT

One of the central parts in the study of combinatorial optimization is to classify the NP-hard optimization problems in terms of their approximability. In this thesis we study the Minimum Vertex Cover (Min-VC) problem and the Minimum Dominating Set (Min-DS) problem in the context of so called power law graphs. This family of graphs is motivated by recent findings on the degree distribution of existing real-world networks such as the Internet, the World-Wide Web, biological networks and social networks. In a power law graph the number of nodes y_i of a given degree i is proportional to $i^{-\beta}$, that is, the distribution of node degrees follows a power law. The parameter $\beta > 0$ is the so called power law exponent.

With the aim of classifying the above combinatorial optimization problems, we are pursuing two basic approaches in this thesis. One is concerned with complexity theory and the other with the theory of algorithms. As a result, our main contributions to the classification of the problems Min-VC and Min-DS in the context of power law graphs are twofold:

- Firstly, we give substantial improvements on the previously known approximation lower bounds for Min-VC and Min-DS in combinatorial power law graphs. More precisely, we are going to show the APX-hardness of Min-VC and Min-DS in connected power law graphs and give constant factor lower bounds for Min-VC and the first logarithmic lower bounds for Min-DS in this setting. The results are based on new approximation-preserving embedding reductions that embed certain instances of Min-VC and Min-DS into connected power law graphs.

- Secondly, we design a new approximation algorithm for the Min-VC problem in random power law graphs with an expected approximation ratio strictly less than 2. The main tool is a deterministic rounding procedure that acts on a half-integral solution for Min-VC and produces a good approximation on the subset of low degree vertices. Moreover, for the case of Min-DS, we improve on the previously best upper bounds that rely on a greedy algorithm. The improvements are based on our new techniques for determining upper and lower bounds on the size and the volume of node intervals in power law graphs.

ACKNOWLEDGEMENTS

This thesis presents the major part of my research as a PhD student at the B-IT Research School and the Computer Science Department at the University of Bonn.

First and foremost, I would like to thank Professor Marek Karpinski, my *Doktorvater*, not only for his untiring guidance and supply of fruitful ideas, but also for his support of my application for a scholarship at the B-IT Research School. He further contributed in many ways to a stimulating, professional, pleasant and sort of *fun* working atmosphere at the Computer Science Department of the University of Bonn.

For the financial support I am greatly indebted to the directors of the B-IT Research School Professor Mathias Jarke and Professor Armin B. Cremers, to my advisor Marek Karpinski and to the State NRW. I would like to sincerely thank the School of Natural Sciences at the Institute for Advanced Study in Princeton for accommodating me during the PiTP program on "Computation and Biology" in July 2012.

Thanks in advance to Professor Heiko Röglin, Professor Reinhard Klein, Professor Martin Weitz and Marek Karpinski for their willingness to serve as members of my promotion committee.

Thanks and best wishes go to my fellow PhD students at the B-IT Research School and the Department of Computer Science. Special thanks go to all the members of the Algorithms and Distributed Systems group, in particular to Christiane Andrade, Mathias Hauptmann, Leszek Paszkiet, Richard Schmied, Ignatios Souvatzis and Claus Viehmann. All of the new results in this thesis have been obtained in collaboration with Mathias Hauptmann. I am grateful for the countless hours we spend on stimulating discussions and fixing bugs, as well as for the everlasting debate on the intrinsic values of coffee and beer. I have benefited and learnt a lot from his broad interest in many areas of mathematics and computer science.

SPECIAL ACKNOWLEDGEMENT: Meiner Familie und meinen Freunden danke ich für die grenzenlose Unterstützung und das stete Interesse an meiner Arbeit. Insbesondere aber auch für die großartige Ablenkung und die kleinen und großen Wunder, die ich mit ihnen speziell in dieser Zeit erleben durfte.

Diese Arbeit ist meiner Lebensgefährtin Vera und meiner Tochter Elena in Liebe und Dankbarkeit gewidmet.

CONTENTS

1 INTRODUCTION

During the last two decades the advent of computerized high-throughput methods and powerful tools for data processing and visualization has led to a rapid growth and better availability of real-world data sets and representations in the form of networks. Moreover, the recent developments in the analysis of large-scale real-world networks have a great impact on the fields of mathematics, statistics, physics, sociology, biology and computer science and have been a main driver for numerous new developments in the fields.

In fact, this particular line of investigations have also sparked a whole new area of research—called *network science* or *network theory*—that focuses solely on the properties, the analysis and the modeling of large-scale networks associated to complex real-world systems. Another specialized branch in the field of biology is the area of *systems biology* at the intersection of biology, statistics and bioinformatics, where research is conducted on the structure and function of biological networks[1]. The main aim of all these relatively new branches of research is to get a better understanding of the emergence, significance, structure and function of complex real-world systems.

In order to assess the structure and organization of the corresponding networks (or *graphs*), typical statistical parameters (or *graph invariants*) such as the *diameter*, the *clustering coefficient* and the *degree distribution* have been measured and compared to the expected values in certain random graph models[2]. Topological network analyses have been applied to a variety of real-world graphs such as the *World-Wide Web*, the *Internet*, *collaboration* and *online social networks*, *protein-protein interaction networks* and other large-scale graphs of systems in nature and in technology.

It turned out that many large-scale real-world graphs are significantly different from instances generated by the classical random models with respect to the above graph invariants. As opposed to the classical random models, the diameter in real-world networks is often very small (or even ultra small), the clustering coefficient is large and—most strikingly—a *power law* distribution of node degrees is observed. These observations raised the need for new graph models that are suited to cope with these structural properties.

1 Biological networks include e. g. *metabolic networks*, *signaling pathways*, *protein-protein interaction networks*, *neuronal networks*, *food webs* and *gene regulatory networks*.

2 Among the most prominent and most widely used model of random graphs, that serves to mimic the topological structure of "typical" networks, is the classical $G(n, p)$ model due to Gilbert [Gil59] and Erdős and Rényi [ER60].

From an algorithmic point of view, the challenges in the analysis of real-world networks are twofold. On the one hand, some of the graph invariants mentioned above are relatively easy to compute, but the sheer size of the graph may rise the need for sub-linear time algorithms that output good approximate solutions very fast. On the other hand, there exist also more sophisticated graph properties that are (up to now) computationally intractable[3]. The latter includes, for example, questions concerning the existence of tightly knotted node clusters, the identification and placement of key-nodes and a number of covering problems. Technically speaking, the main hurdles are the large graph instances of real-world systems and the computational hardness or inherent intractability of the combinatorial problems which are defined by the analysis task. These challenging circumstances gave rise to a whole new approach in the analysis of graphs in the context of complex real-world systems [AB02; New03; DM03; DM04; NW06; Boc+06; PV07; Lov09].

In the following we give more explicit examples of large-scale real-world networks and introduce some topological features that distinguish them from regular graphs or uniform random graphs. Moreover, we will introduce mathematical models that were designed to capture these topological properties qualitatively and quantitatively.

1.1 MODELING REAL-WORLD NETWORKS

Many systems in nature, society and technology consist of a large number of dynamical units—or system elements—and of more or less complex modes of interactions or interconnections between the elements of the system. A natural first approach to capture the global properties of such a system is to model it as a graph consisting of *nodes* (or *vertices*) and interconnecting *links* (or *edges*). In the *adjacency matrix representation* of a graph the rows and columns correspond to the vertices and an edge exists between two vertices if the corresponding matrix entry is non-zero. This provides a very simple and at the same time informative representation of a complex system, even though the sometimes complex interactions—dependent on time, space or other constraints—are reduced to be represented as a single number in the adjacency matrix of the graph.

3 By computational intractability, we mean that the defined problem cannot be solved deterministic algorithm in running time that is bounded by a polynomial in the size of the input.

Let us give some prominent and notorious examples of existing large-scale graphs or networks. The *World-Wide-Web* is a graph with web pages representing the nodes and hyper-links between pages representing the edges (Size comparison: The Google index records ~ 45 billion individual web pages as of June 22, 2013). In the backbone of the *Internet* routers are connected through physical data wires (Size comparison: The IPv4-Internet consist of ~ 1.3 billion active IP-addresses by the end of 2012). In the context of social networks, a *communication graph* is formed via the exchange of emails, phone calls, instant messages, et cetera, within a group of people (Size comparison: Monthly active facebook users were ~ 1.11 billion as of March 31, 2013). Regarding biological networks, prominent examples are *protein-protein interaction networks* where proteins are represented as nodes and interaction or co-expression of two proteins is denoted by an edge (Size comparison: The protein interaction network of the fruit fly *drosophila melanogaster* consists of ~ 5000 proteins with ~ 200 000 known interactions).

Next, we are going to introduce and discuss some of the observed topological features that distinguish complex real-world networks from regular graphs as well as from uniform random graphs.

SMALL-WORLDS AND THE CLUSTERING COEFFICIENT. The study of dynamical processes across large-scale real-world networks suggested the existence of bridging edges that connect distant areas of the networks via a small number of 'hops' and thus allow for fast traversal of information inside the network.

In the 1960's Travers and Milgram [Mil67; TM69] conducted a series of experiments in the context of social networks where they asked randomly selected people to send letters to a distant target person, identified only by his or her name and rough location. The letters could only be send to individuals which the current holder of the letter knows by first name and which, by any chance, were closer to the target person. Travers and Milgram kept track of the letters and the number of steps for the letters to reach their final destination. The general presumption was that the letters would take several hundreds of steps. But for those letters which finally arrived, the surprising result was that the average number of links needed to reach the target person was only six. The finding was then described as the phenomenon of the *"six degrees of separation"* of a network which also directly induced that the corresponding graphs have a small diameter. More generally, networks that displayed such a characteristic

were said to satisfy the so called *"small world"* property or were simply termed *"small worlds"*.

This property was observed later for a number of other real-world networks—also biological and technological systems—such as the neuronal network of the roundworm *C.elegans*, protein interaction networks, gene networks, power grids and networks of co-authorship in scientific publications [WS98; New01b; New01a; Wat03; Bor+04; NSH04].

As we will see from a theorem by Bollobás and Fernandez de la Vega [BF82] (cf. Theorem 2.4 in Section 2.4.2), classical random graphs also share the property of having a relatively small diameter, i. e. a small average shortest path length in relation to the graph size. An additional special characteristic of many real-world networks is that they are also heavily clustered, which is expressed by high values of the so called *clustering coefficient* of the underlying graph. This is a feature that is not captured by graph instances generated by the classical uniform random models such as the $G(n, p)$ model or the $G(n, M)$ model due to Gilbert [Gil59] and Erdős and Rényi [ER60].

In order to cope with this lack of modeling capabilities, Watts and Strogatz raised the idea and a possibility of constructing random graphs that mimic this important feature of real-world networks. In their seminal paper from 1998 on the *"Collective Dynamics of 'Small-World' Networks"* Watts and Strogatz [WS98] initiated the important field of modeling large-scale real-world networks by random graphs, which itself are defined by simple constructional rules. Instead of the usual diameter $\operatorname{diam}(G)$ of a graph G, Watts and Strogatz considered the average distance between all pairs of nodes $L(G) = \sum_{\{u,v\}\subset V, u\neq v} d(u, v)/\binom{n}{2}$ and the clustering coefficient $C(G)$—which describes how well connected the neighborhood of a node or a set of nodes is (cf. Section 2.3.2).

The proposed random model for small-world networks generates graphs that have simultaneously a small value of $L(G)$ like random graphs, and a high clustering coefficient $C(G)$ like regular lattices. Furthermore, small-world networks can be seen as an intermediate between regular graphs and random graphs. Figure 1.1 shows the transition of a *regular ring graph* into a *small-world graph* and further into a *random graph*, only by increasing the probability that an edge changes one or two of its endpoints.

The introduction of small-world network models presented a major step towards a more sophisticated modeling of large-scale real-world networks. However, further in-

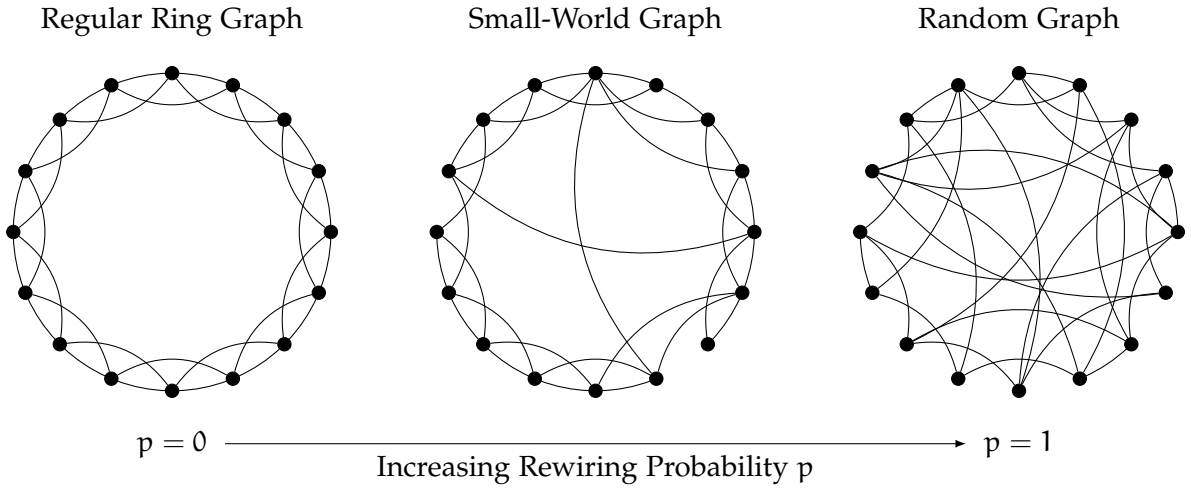


Figure 1.1: From left to right, the rewiring probability of an edge is increased, that is, the probability of an edge to change one or two of its endpoints. In this way, the initial regular ring graph with large diameter is transformed into a small-world graph by introducing short-cuts. A rewiring probability close to 1 finally transforms the initial graph into a random graph.

vestigations elucidated another special characteristic of many real-world that regards the distribution of node degrees. The crucial observation was that the degree distribution is often well approximated by a *power law distribution*.

POWER-LAW DEGREE DISTRIBUTIONS. The usual perception in the study of networked systems is that of *homogeneous networks*, meaning that all nodes of the network are topologically similar, like in regular lattices or uniform random graphs. In particular this means that the distribution of node degrees is *binomial* or *Poisson*[4] in the limit of large graph sizes. The new and remarkable observation was that the distribution of node degrees in real-world networks is often well approximated by a *power law distribution*—that is, the number of nodes y_i of a given degree i is proportional to $i^{-\beta}$, where $\beta > 0$. The parameter β is the so called *power law exponent* and a graph which has this property is called *power law graph*.

Empirical studies verified a power law distribution of node degrees for a large number of existing real-world networks such as the Internet, the World-Wide Web, protein-

4 For the definition and depiction of binomial and Poisson distributions see Section 2.2, Equation 2.1 and Equation 2.2 together with Figure 2.1 and Figure 2.2, respectively.

protein interaction networks, gene regulatory networks, peer-to-peer networks, mobile call networks, et cetera [FFF99; Kle+99; Kum+00; Bro+00; KL01; JAB01; Gue+02; Sig+03; Eub+04b; Ses+08]. Typical values of the power law exponent lie within the range $2 < \beta < 3$ (e. g. $\beta = 2.38$ for the WWW [Bro+00], $\beta = 2.4$ for protein-protein interaction networks [Jeo+01]). But there also exist examples of real-world networks with a power law exponent $\beta \leqslant 2$ or $\beta \geqslant 3$, e. g. for distributional food webs ($\beta = 1.05$, [MS00]), statistical investigations of book sales in the US ($\beta = 3.51$, [Hac67; NW06]) and networks of human sexual contacts ($\beta = 3.4$, [Lil+01]).

A simple empirical indicator, for whether the node degrees of a system follow a power law, is the observation of a straight line on a log-log plot of the degree distribution. Figure 1.2 shows the protein-protein interaction network of the flowering plant *arabidopsis thaliana* along with the corresponding node degree distribution that follows a power law with exponent $\beta = 1.8$.

Historically, power law distributions had also been observed considerably earlier in various contexts ranging from linguistics to economics and further to social sciences. For example, power law behavior in different data sets was shown for the distribution of income across a population by Pareto [Par96], for the distribution of city sizes in terms of the number of inhabitants by Auerbach [Aue13], for word frequencies in books by Estoup [Est16] and for citations amongst academic literature by Lotka [Lot26]. The idea of associating power law distributions with real-life systems—and especially its popularization—is generally attributed to the American linguist Zipf [Zip35; Zip49]. A system subject to power law behavior is therefore also often said to follow *Zipf's law* and an associated *Zipfian distribution* is observed on the data set.

Let us now introduce the background and the definitions of some modeling approaches for power law graphs that emerged in the fields of physics, computer science and mathematics.

MATHEMATICAL MODELS FOR POWER LAW GRAPHS. At the latest since the massive computer-aided acquisition of real-world data and evaluation of the associated topological information by computerized methods—accompanied by the publication of the works of Barabási and Albert [BA99]; Faloutsos, Faloutsos, and Faloutsos [FFF99]; Kleinberg et al. [Kle+99] in 1999—the idea of the ubiquity of massive power law graphs with small diameter and large clustering coefficient was pervasive in the fields

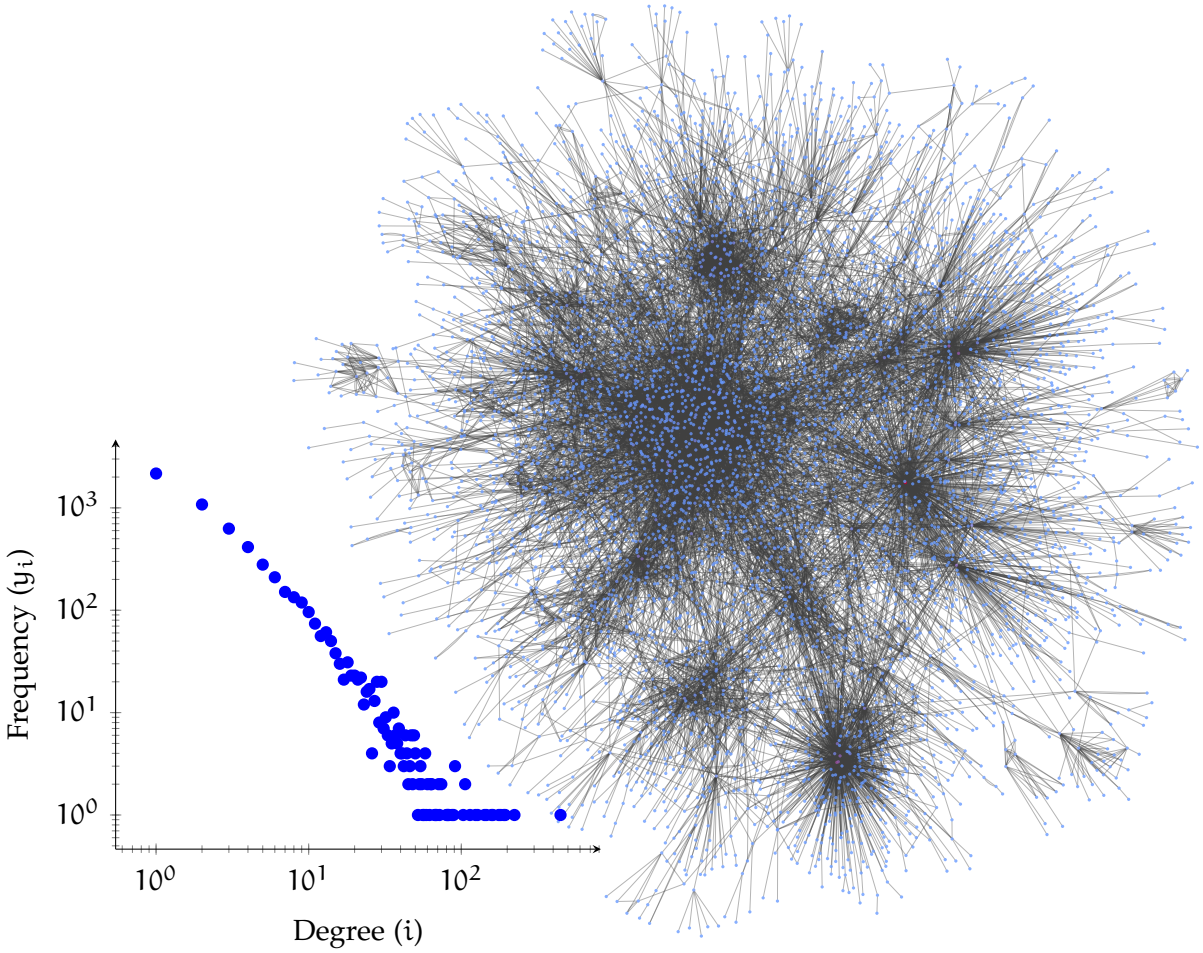


Figure 1.2: Protein-protein interaction network of the flowering plant *arabidopsis thaliana* along with the according degree distribution. The corresponding power law exponent is $\beta = 1.85$.

of computer science, mathematics and physics. In subsequent studies the aim was to describe these properties of real-world networks mathematically and to propose new and suitable models to generate graphs that display these properties quantitatively and qualitatively.

The model of *"preferential attachment"* or the concept of *"the rich get richer"* is most often referred to as a potential mechanism underlying the emergence of graphs with a power law distribution of node degrees—so called *scale-free networks* or *power law graphs*. The modern revival of these ideas is attributed to the work of the physicists Barabási and Albert [BA99], who also coined the terms *"preferential attachment"* and

"scale-free network". Indeed, very similar ideas had been mentioned and described considerably earlier by Yule [Yul25], Simon [Sim55] and de Solla Price [Pri65; Pri76]. In particular, de Solla Price was the first to proposed a mechanism of preferential attachment under the name *"cumulative advantage"*. The recently popular description of preferential attachment is due to Barabási and Albert [BA99], but the model was later more rigorously and mathematically defined by Bollobás and Riordan [BR05].

The mechanism features the role of evolutionary growth and rewiring processes in the emergence of power law graphs. In the formal definition, a new vertex is introduced one at a time and is connected to the existing graph via a prescribed number of edges. The probability of an edge to an already existing vertex is not uniform but dependent on the current degree of the target vertex. Based on this idea of preferential attachment, other models were proposed that—for example—take into account an *initial attractiveness* of nodes in the growth process [DMS00; DEM01; BO04] or that introduce new nodes by copying and rewiring existing nodes [Kum+00]. But preferential attachment is only one of several mechanisms that are capable of generating graphs with power law degree distributions.

Besides the above *evolving models*, there also exist so called *static models* for generating power law graphs. In this approach a power law degree sequence is given as an input and a graph instance representing this distribution is generated in a random fashion. Among the most widely known models of this kind is the *ACL-model* due to Aiello, Chung, and Lu [ACL01]. In this model the given degree sequence is of the form $y_1, y_2, \ldots, y_m$, where y_i is the number of vertices of degree i. This number is roughly given by $y_i \approx e^\alpha / i^\beta$, where e^α is a normalization constant which determines the size of the graph. While this model is potentially less accurate than the detailed description of a growth process of an evolving graph it has the advantage of being robust and general, that is, structural properties that are true in this model will be true for the majority of graphs with the given degree sequence.

All of the above models are well motivated and there exists a large body of literature on mathematical foundations and applications [BA99; ACL00; BR05; Eub+04b; MPS06]. In this thesis, however, we will focus on the ACL-model in order to characterize the computational complexity of some combinatorial optimization problems on power law graphs. The motivation for this theoretical research issue is presented in the upcoming section, along with some practical applications.

1.2 COMBINATORIAL OPTIMIZATION & APPLICATIONS

From the analysis of real-world networks, there exists practical evidence which entails that combinatorial optimization in power law graphs is easier than in general graphs. For example, Park and Lee [PL01] showed for the efficient placement of filters in *route-based distributed packet filtering* on power-law Internet topologies, that the greedy heuristic for the vertex cover problem generally outperforms the constant-factor approximation algorithm. Moreover, in connection with the optimal placement of sensors for disease detection inside social networks, Eubank et al. [Eub+04b] studied near-optimal dominating set problems and found that for a class of bipartite random power law graphs the problem admits a $(1 + o(1))$-approximation.

A natural question now arises whether these observations and results can be translated into provable guarantees for a more general class of power law graphs. If this is the case, one has to identify structural properties of power law graphs that allow for the design of efficient algorithms or better approximation algorithms. For the case that this is not apparent, one would like to prove fundamental results on the computational hardness of the problem. An especially interesting case would be if this classification depends on the graph parameters (e. g. the power law exponent) of instances in the graph class under study. This would imply the existence of a phase transition in the hardness of approximation as the power law exponent varies over a range of values.

One of the central parts in the study of combinatorial optimization is to classify the NP-hard optimization problems in terms of their approximability. In this thesis we will focus on two combinatorial covering problems in graphs, namely the Minimum Vertex Cover (Min-VC) problem and the Minimum Dominating Set (Min-DS) problem. In order to define the problems, let $G = (V, E)$ be a graph with vertex set V and a set of edges E.

The Min-VC problem is one of the most well-studied problems in combinatorial optimization. A *vertex cover* of a graph $G = (V, E)$ is a set of vertices $C \subseteq V$ such that each edge $e \in E$ of G has at least one endpoint in C. The Min-VC problem is the problem of finding a vertex cover of minimum cardinality in a graph. The problem is known to be NP-hard due to Karp's original proof presented in [Kar72] and APX-complete due to Papadimitriou and Yannakakis [PY91]. Moreover, Dinur and Safra

[DS05] showed that MIN-VC cannot be approximated within a factor of 1.3606, unless $P = NP$, and Khot and Regev [KR08] showed the inapproximability within $2 - \varepsilon$ for any $\varepsilon > 0$ as long as the Unique Games Conjecture (UGC) holds true.

Another well known optimization problem is MIN-DS. On a graph $G = (V, E)$ a *dominating set* is a subset of vertices $D \subseteq V$ such that every vertex in $V \setminus D$ is connected to D by at least one edge $e \in E$. The MIN-DS problem asks for a dominating set of minimum cardinality $|D|$. This problem is known to be NP-hard by a reduction from the SET COVER problem. Moreover, the result of Raz and Safra [RS97] rules out the existence of an approximation algorithm for general graphs with an approximation factor better than $c \cdot \log |V|$ for some $c > 0$, unless $P = NP$.

In the following, we list some applications of the two problems in the context of analyzing real-world graphs or networks:

- ***Identifying Key Players in Social Networks***: In the investigation of epidemic spreading of diseases across networks of travel routes or the spreading of information inside online social networks, a natural question arises about how to efficiently place key nodes at key positions inside a network such as to reach and to effect all or most of the remaining nodes[PV01; Eub+04a; Bor06; Kit+10; Wan+11]. Here, the feasibility of a solution also heavily depends on the number of key nodes needed in order to cover the whole network and thus this number is often tried to be minimized. Questions like these quickly resemble or are equivalent to classical NP-hard optimization problems, i. e. minimum covering and domination problems such as MIN-VC and MIN-DS.

- ***Clustering in Complex Networks***: In the analysis of *ad hoc* and *multi-hop wireless networks*, solutions to MIN-DS are sought in order to cluster the graph into well distinguished parts. Here, the elements of a dominating set serve as cluster-heads and the neighborhood of each cluster-head comprises a cluster in the graph [CLL04; BB06]. Under the general assumption that functionally similar or related vertices are also highly connected, the clustering provides valuable insights on the organization of the network. This may reveal a hidden modular organization of the nodes and may also help to classify nodes of unknown function [GN02; For10; LF12].

- ***Conflict Resolving in Multiple Sequence Alignments***: In the context of multiple sequence alignments (MSA) a *conflict graph* consists of vertices representing *gap blocks* in the alignment and edges that are drawn between conflicting blocks. A solution to MIN-VC in the conflict graph now corresponds to a minimum number of gap blocks such that the MSA ignoring these blocks is conflict free [RSG99].

- ***Broadcasting and Routing in Wireless Networks***: For the broadcasting of information inside a wireless network with nodes of limited transmission radius, a small dominating set with low maintenance cost is used to significantly reduce or to eliminate communication overhead for retransmissions. At the same time the covering property of the dominating set allows for reliable broadcasting [SSŽ02]. Furthermore, connected dominating sets may serve as *virtual backbones* of a wireless network to find and to update efficient routes for package transmission inside the network [DB97; WL99; WGS01].

- ***Complexes and Domination in Protein-Protein Interaction Networks***: In the large-scale identification of protein complexes in yeast protein-protein interaction graphs and hypergraphs, solutions to MIN-VC are used to choose a set of candidate *bait proteins*—that is, proteins that are directly associated to a *protein complex* [RTP04]. The number and distribution of protein complexes, which is the set of proteins expressed by the genome at a given time under defined conditions, is used to characterize the proteome of the model organism. Furthermore, in [Mil+11] connected dominating sets are sought in protein-protein interaction networks to identify proteins with a central topological and biological role. This may lead to insights on the communication structure and signaling pathways inside the network.

- ***Robustness and Vulnerability of Complex Networks***: The *robustness* of a network describes how resilient the connectivity structure is when the network is a target of node and edge deletions [Coh+00; AJB00]. The MIN-VC problem can be formulated as the problem of finding a minimum-cardinality set of vertices whose deletion makes the network edge-free. Hence, the minimum number of vertices needed to completely disconnect the network can be translated into a

robustness-measure. It was shown that Peer-to-Peer networks and co-authorship networks are highly robust against MIN-VC-*attacks* while online social networks and e-mail communication networks are rather vulnerable [Li+12].

- ***Clique Extraction in Biological Networks***: In the analysis of genetic microarray data the corresponding *gene co-expression networks* consist of nodes representing genes and weighted edges which express the amount of correlation between genes. A crucial task is reduce the high dimensionality of the data (>12 000 nodes) by grouping and extracting similarly expressed genes. This can be interpreted as a problem of extracting *cliques* from the network and can be attacked by solving the complementary dual problem MIN-VC [Bal+05; Che+05].

We conclude this introductory chapter with an outlook on the organization and main results of this thesis.

1.3 ORGANIZATION OF THE THESIS AND MAIN RESULTS

First of all, in Figure 1.3, we present a chapter dependence diagram for the relations between the various parts of this thesis. The thesis is organized as follows:

- In Chapter 2, we fix the notions and notations used throughout this thesis. This includes basic definitions from probability theory, graph theory and random graph theory. Moreover, we will introduce the concept of random graphs and power law graphs (PLG) and state basic results regarding structural properties of those families of graphs. Especially, we present the definition of the $G(\alpha, \beta)$ model for generating graphs with a given power law degree distributions—so called (α, β)-PLG.

- In Chapter 3, we introduce complexity classes of decision and optimization problems that are relevant for the classification results of this thesis. Furthermore, we define reductions between decision and optimization problems and show how to relate the hardness of approximation of decision problems with the hardness of approximation of optimization problem.

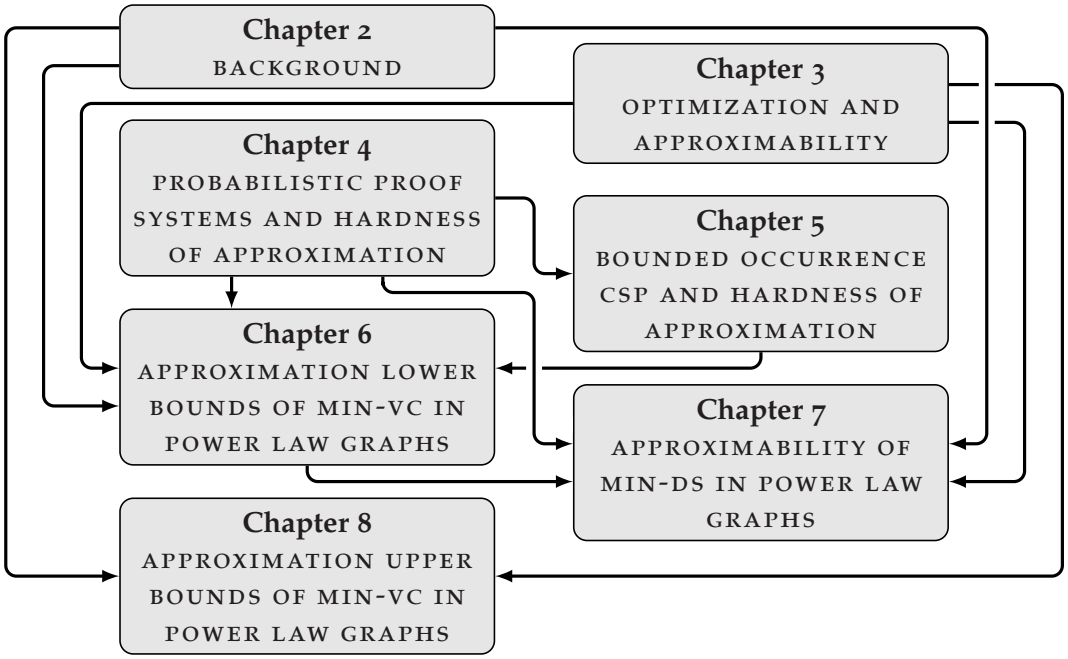


Figure 1.3: A chapter dependence diagram for the various parts of this thesis. In Figure 1.4 on page 19, we also give a detailed dependence diagram.

- In Chapter 4, we provide the necessary instruments to analyze hardness of approximation results that rely on the properties and existence of *multi-prover proof systems* (e. g. utilized by Feige [Fei98] to prove approximation lower bounds for the Set Cover and Min-DS problem) and *probabilistically checkable proof systems* (e. g. utilized by Austrin, Khot, and Safra [AKS09] to prove approximation lower bounds for bounded degree Min-VC) for certain problems in NP.

- In Chapter 5, we introduce the Hybrid problem defined by Berman and Karpinski [BK99] which is used in reductions to prove hardness of approximation results for bounded occurrence constraint satisfaction problems. Moreover, Chapter 5 presents techniques proposed by Berman and Karpinski [BK99] in order to obtain approximation hardness results for bounded occurrence constraint satisfaction problems and graph problems on instances with bounded degree. In particular, this reduction technique provides the current best explicit lower bounds on the approximability of Min-VC on degree d-bounded graphs for $d = \{3, 4, 5\}$,

which will be used in Chapter 6 to obtain new lower bounds for MIN-VC in power law graphs.

- In Chapter 6, we prove the APX-hardness of MINIMUM VERTEX COVER (MIN-VC) in *connected* power law multigraphs in the $G(\alpha, \beta)$ model for $0 < \beta < \beta_{max} \approx 2.48$ (for $\beta > \beta_{max}$, (α, β)-PLGs are not connected anymore). This partially resolves an open question posed by Ferrante, Pandurangan, and Park [FPP08]. Furthermore, we give explicit approximation lower bounds for this problem. Our results are based on reductions from bounded degree instances and the corresponding explicit lower bounds in the general case. Our reductions consist of multigraph embeddings of bounded degree graphs into (α, β)-PLG, based on appropriate wheel constructions. We also extend the $G(\alpha, \beta)$ model for power law graphs and consider degree distributions where β is of the form $\beta_f = 1 \pm 1/f(n)$, for a sufficiently fast growing function $f(n)$. We show that these distributions converge to those of (α, β)-PLG for $\beta = 1$ and can be seen as a combinatorial variant of the evolving preferential attachment model for power law graphs.

- In Chapter 7, we study the approximation complexity of MINIMUM DOMINATING SET (MIN-DS) in (α, β)-PLG. Our contributions to characterization and classification of this problem are twofold: On the one hand, we give the first logarithmic lower bounds for the approximability of this problem for the parameter range $0 < \beta \leqslant 2$, based on a reduction from the SET COVER problem combined with the logarithmic lower bound for SET COVER given by Feige [Fei98]. This improves over the previously known constant factor approximation lower bounds given by Shen et al. [She+12], which are based on reductions from the bounded degree MIN-DS problem. On the other hand it was also shown in [She+12] that, for $\beta > 2$, the MIN-DS problem in (α, β)-PLG admits a constant factor approximation upper bound, that is, for this range of the model parameter the problem is in the class APX. We improve on this result by giving new upper bounds on the approximation ratio of an algorithm based on the greedy strategy for MIN-DS. In [She+12], membership of MIN-DS in (α, β)-PLG in APX was shown by constructing a lower bound for the optimum and an upper bound for the greedy solution separately. We obtain our new results by relating the cost and structure of an optimum solution to those of a greedy-based solution. This sophisticated

analysis yields improved upper bounds for almost the whole range $\beta > 2$. Finally, we take a very close look at the phase transition at $\beta = 2$. Similar as in Section 6.8 and Section 6.9 we extend the power law model and consider the case when $\beta_f = 2 + \frac{1}{f(n)}$ is a function of the graph size n which converges to 2 from above. We obtain the following surprising result: For every function $f(n)$ with $f(n) = \omega(\log(n))$ (i.e. when β_f converges fast enough), MIN-DS in (α, β_f)-PLG still provides a logarithmic approximation lower bound and, for every function $f(n)$ with $f(n) = o(\log(n))$, the problem is in APX.

- Chapter 8 deals with the approximability of MIN-VC on random power law graphs in the $G(\alpha, \beta)$ model. More precisely, we show that the MIN-VC problem can be approximated with an expected approximation ratio strictly less than 2. This possibility may be ruled out for the general case due to the conjecture by Khot and Raman [KR02], and the result of Khot and Regev [KR08] which states that there exists no $2 - \varepsilon$ approximation for MIN-VC under this conjecture. We construct an approximation algorithm for MIN-VC with an expected approximation ratio of $2 - f(\beta)$ which, in the limit of large graph sizes, improves on the currently best upper bound of $2 - \Theta(1/\sqrt{\log n})$ for the general case due to Karakostas [Kar09]. The result is obtained by combining the linear programming approach for MIN-VC due to Nemhauser and Trotter [NT75] with a new deterministic rounding procedure which achieves an approximation ratio of $3/2$ on a subset of low degree vertices. For this subset, we show that the expected contribution to the cost of the associated linear program is sufficiently large.

Figure 1.4 provides a more detailed view on the organization and conceptual associations between the various chapters of the thesis.

The next chapter will provide the necessary background as well as the basic notions and notations that facilitate the understanding of the topics presented in this thesis.

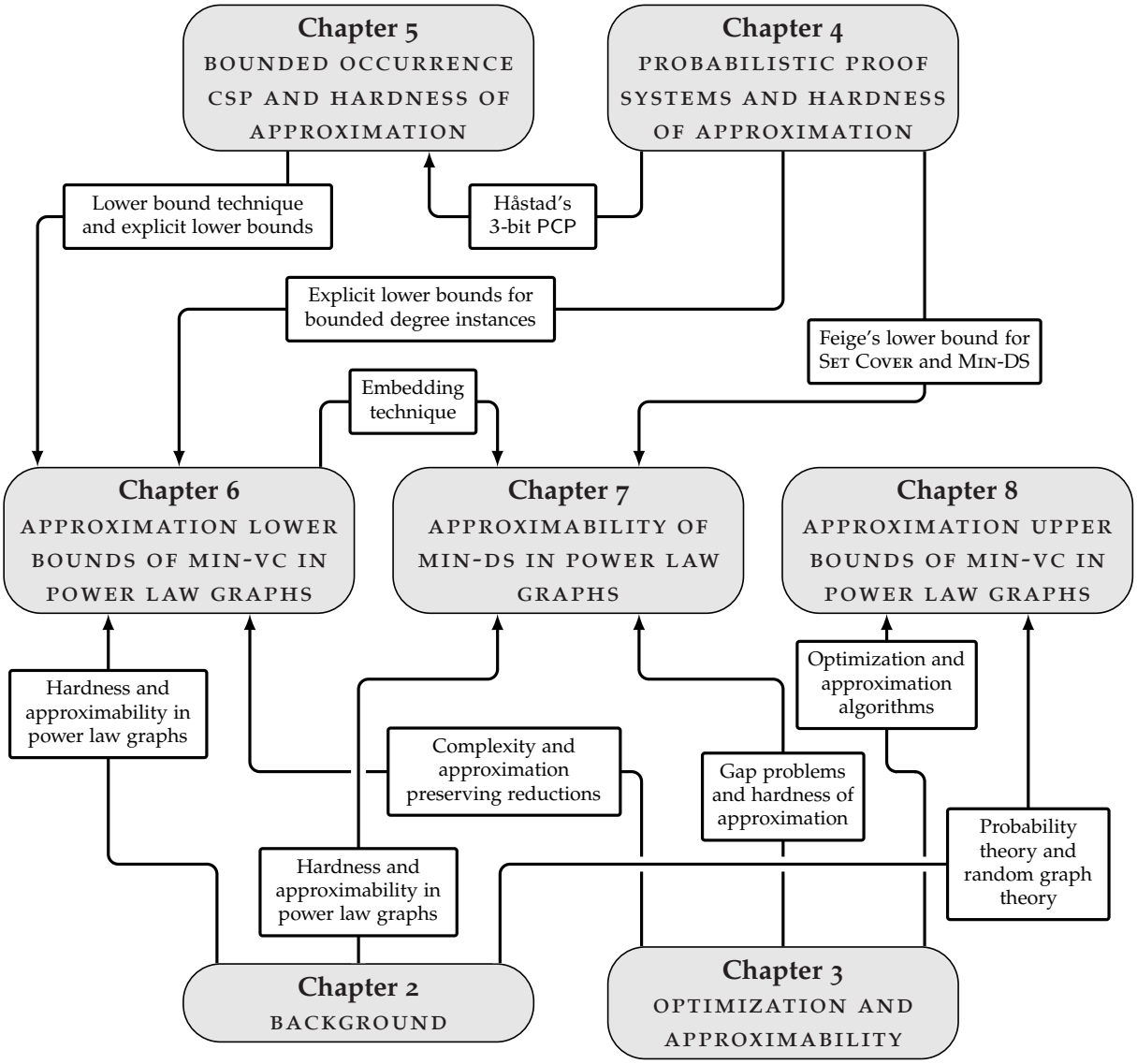


Figure 1.4: A detailed dependence diagram for the various parts of this thesis. The middle layer represents the central part of this thesis including the chapters with the main results (Chapters 6, 7 & 8), whereas the lower layer provides formal foundations (Chapters 2 & 3) and the chapters in the upper layer summarize techniques and previous results that are used to obtain our main results (Chapters 4 & 5).

2 BACKGROUND

We start with the definition of the most basic notions and notations that are used in the sequel.

2.1 BASIC NOTIONS AND NOTATIONS

The sets of integers, positive integers and natural numbers will be denoted by $\mathbb{Z} = \{0, 1, -1, 2, -2, \dots\}$, $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ and $\mathbb{N} = \mathbb{N}_0 \backslash \{0\}$, respectively. The sets of real numbers and positive real numbers will be denoted by $\mathbb{R}$ and $\mathbb{R}_+$.The sets of rational numbers and positive rational numbers will be denoted by $\mathbb{Q}$ and $\mathbb{Q}_+$. For a natural number $n \in \mathbb{N}$, we define the sets $[n] = \{1, 2, \dots, n\}$, $[n]_0 = \{0\} \cup [n]$ and $[0] = \emptyset$ in short notation. We denote the *power set* of a set A, i. e. the set of all possible subsets of A including the empty set $\emptyset$ and the set A itself, by $\mathbb{P}(A)$. By $\mathbb{P}_k(A)$ we denote the set of all subsets of cardinality k. A *multiset* is defined as a 2-tuple (A, m) where A is a set and $m \colon A \to \mathbb{N}$. The set A is called the underlying set of elements and the function m defines the number of occurrences $m(a)$ for each element $a \in A$. For a real number $x \in \mathbb{R}$, we denote by $\lfloor x \rfloor$ the largest integer not greater than x and by $\lceil x \rceil$ the smallest integer not less than x. For a real number $x \in \mathbb{R}_+$, we denote by $\log(x)$ the logarithm of x to the base 2 and by $\ln(x)$ the logarithm of x to the base e.

ASYMPTOTIC NOTATION. The *asymptotic notation*, or *Bachmann-Landau notation*, is a family of notations that was introduced by Bachmann [Bac94] and Landau [Lan09] in order to describe the limiting behavior of mathematical functions. In the realm of computational complexity theory, this notation is used to classify the behavior of algorithms in response to changes of the input size. Typically, the performance or efficiency of an algorithm is defined in terms of its *time* and *space complexity* and is measured as a function of the input length, i. e. a function $t : \mathbb{N} \to \mathbb{N}$ with $t(n)$ being the maximum number of basic operations performed or register space used by the algorithm on inputs of length n.

In the remainder of this thesis, we will make use of the following notations for the classification of functions. Let $f(n)$ and $g(n)$ be two functions $f, g : \mathbb{N} \to \mathbb{N}$. We say that

- $f(n) \in \mathcal{O}(g(n))$ if there exist a constant $c > 0$ and $n_0 \in \mathbb{N}$ such that $f(n) \leqslant c \cdot g(n)$ holds for all $n \geqslant n_0$,
- $f(n) \in \Omega(g(n))$ if there exist a constant $c > 0$ and $n_0 \in \mathbb{N}$ such that $g(n) \cdot c \leqslant f(n)$ holds for all $n \geqslant n_0$,
- $f(n) \in \omega(g(n))$ if there exist a constant $c > 0$ and $n_0 \in \mathbb{N}$ such that $g(n) \cdot c < f(n)$ holds for all $n \geqslant n_0$,
- $f(n) \in \Theta(g(n))$ if there exist a constants $c_1 > 0$, $c_2 > 0$ and $n_0 \in \mathbb{N}$ such that $c_1 \cdot g(n) \leqslant f(n) \leqslant c_2 \cdot g(n)$ holds for all $n \geqslant n_0$,
- $f(n) \in o(g(n))$ if for all $\varepsilon > 0$, there exist $n_\varepsilon \in \mathbb{N}$ such that $|f(n)| \leqslant |\varepsilon \cdot g(n)|$ holds for all $n \geqslant n_\varepsilon$, and
- $f(n) \sim g(n)$ if for all $\varepsilon > 0$, there exist $n_\varepsilon \in \mathbb{N}$ such that $\left|\frac{f(n)}{g(n)} - 1\right| < \varepsilon$ holds for all $n \geqslant n_\varepsilon$.

By a slight abuse of notation, in the following we write $f(n) = \mathcal{O}(g(n))$ instead of $f(n) \in \mathcal{O}(g(n))$ and likewise for all the asymptotic notations defined above.

2.2 PROBABILITY THEORY

The aim of this section is provide the basic definitions of probability and to introduce different kinds of probability distributions that may occur when one measures the degree distribution of real-world networks or the distribution of data points in other real-world systems. We start with the definition of a probability space.

PROBABILITY SPACES. A *probability space* is a triple $(\Omega, \mathcal{F}, \Pr)$ consisting of *elementary events* or *states* ω which are collected in the non-empty set Ω, a σ-algebra $\mathcal{F}$, which is the system of *observable* subsets or *events* $E \subseteq \Omega$ and a *measure* $\Pr$, which assigns a *probability* $\Pr(E) \in [0, 1]$ to all $E \in \mathcal{F}$. Furthermore, a system $\mathcal{F}$ of subsets $E \in \Omega$ is called σ-*algebra* on Ω if

1. $\emptyset, \Omega \in \mathcal{F}$,
2. $E \in \mathcal{F}$ implies that $E^c := \Omega \setminus E \in \mathcal{F}$,

3. $E_1, E_2, \cdots \in \mathcal{F}$ implies that $\bigcup_{i=1}^{\infty} E_i \in \mathcal{F}$.

The pair $(E, \mathcal{F})$ is a field of sets, called a *measurable space*. A map $\Pr : \mathcal{F} \to [0,1]$ is called *probability measure* if $\Pr(\Omega) = 1$ and for all $E_1, E_2, \cdots \in \mathcal{F}$ with $E_i \cap E_j = \emptyset$ for $i \neq j$ we have

$$\Pr\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \Pr(E_i) \ .$$

RANDOM VARIABLES AND EXPECTATION. Let $(\Omega, \mathcal{F}, \Pr)$ be a probability space and $(E, \mathcal{E})$ a measurable space. An $(E, \mathcal{E})$*-valued random variable* is a function $X : \Omega \to E$ which is $(\mathcal{F}, \mathcal{E})$-measurable, i. e. for every subset $B \in \mathcal{E}$, its preimage $X^{-1}(B) \in \mathcal{F}$. When the space E is the real line $\mathbb{R}$, most commonly the σ-algebra $\mathcal{E}$ is chosen to be the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ and X is called a *real-valued random variable* or simply a *random variable*.

The *probability distribution* of a random variable X is captured by its *cumulative distribution function* (CDF) $F_X(x) = \Pr(X \leqslant x)$. The *complementary cumulative distribution function* (CCDF) is defined as $\bar{F}_X(x) = \Pr(X > x) = 1 - F_X(x)$. A function $f_X : \mathbb{R} \to [0, \infty)$ is called the *probability density function* (PDF) of X if $\Pr(a \leqslant x \leqslant b) = \int_a^b f_X(x)\,dx$.

In case of a discrete random variable X, a *discrete probability distribution* is characterized by its *probability mass function* (PMF) $f_X(x) = \Pr(X = x) = \Pr(\{\, \omega \in \Omega \mid X(\omega) = x \,\})$, where for the total probability $\sum_{x \in \mathbb{R}} f_X(x) = 1$ holds and where $f_X(x) = 0$ for all $x \notin X(\Omega)$.

In general, if X is a random variable defined over a probability space $(\Omega, \mathcal{F}, \Pr)$, the *expected value* of X is defined as

$$\mathbb{E}[X] = \int_\Omega X \,\mathrm{d}\Pr = \int_\Omega X(\omega)\,\Pr(\mathrm{d}\omega) \ .$$

PROBABILITY DISTRIBUTIONS. A discrete random variable X is said to have a *binomial distribution* with parameters $n \in \mathbb{N}$ and $p \in [0,1]$ if

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{and} \quad \Pr(X \leqslant x) = \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i} \ , \quad (2.1)$$

for $x = 0, 1, \ldots, n$ (cf. Figure 2.1). The expected value is $\mathbb{E}(X) = n \cdot p$.

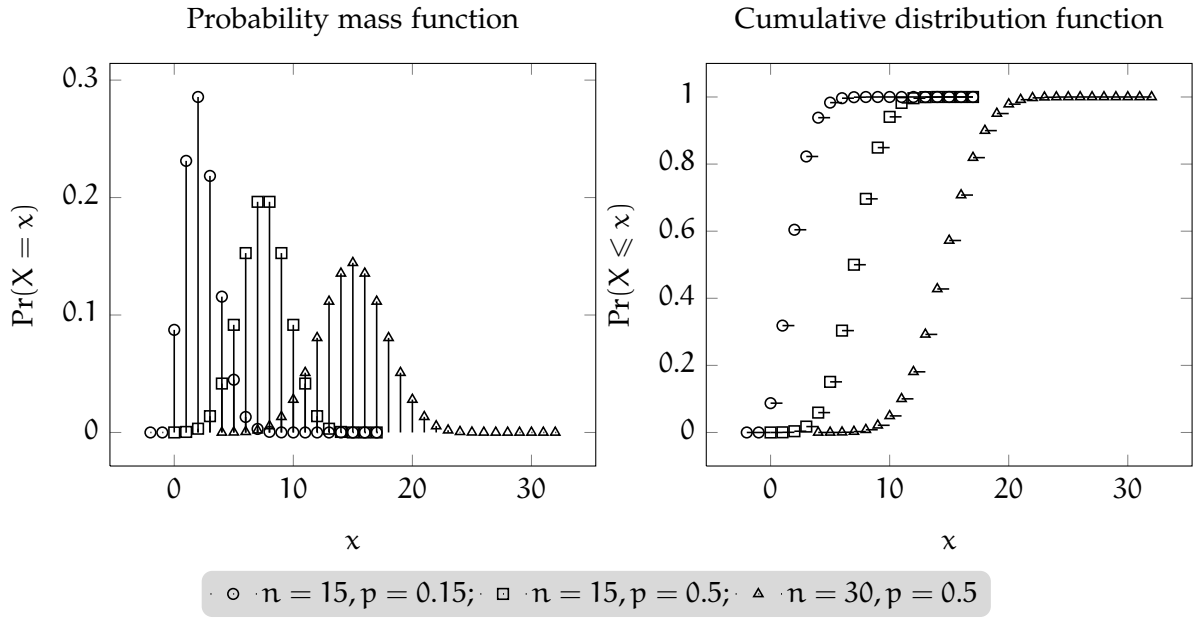


Figure 2.1: Plot of the probability mass and the cumulative distribution function of a binomial distribution as functions of the parameters n and p according to Equation 2.1.

A discrete random variable X is said to have a *Poisson distribution* with parameter $\lambda > 0$ if

$$\Pr(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{and} \quad \Pr(X \leqslant x) = e^{-\lambda} \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!} \ , \tag{2.2}$$

for $x = 0, 1, \ldots$ (cf. Figure 2.2). The expected value is $\mathbb{E}(X) = \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} = \lambda$.

A (continuous and real-valued) random variable X is said to have a *normal* (or *Gaussian*) *distribution* with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ if

$$\Pr(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{and} \quad \Pr(X \leqslant x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \, dt \ , \tag{2.3}$$

for $x \in \mathbb{R}$ (cf. Figure 2.3).

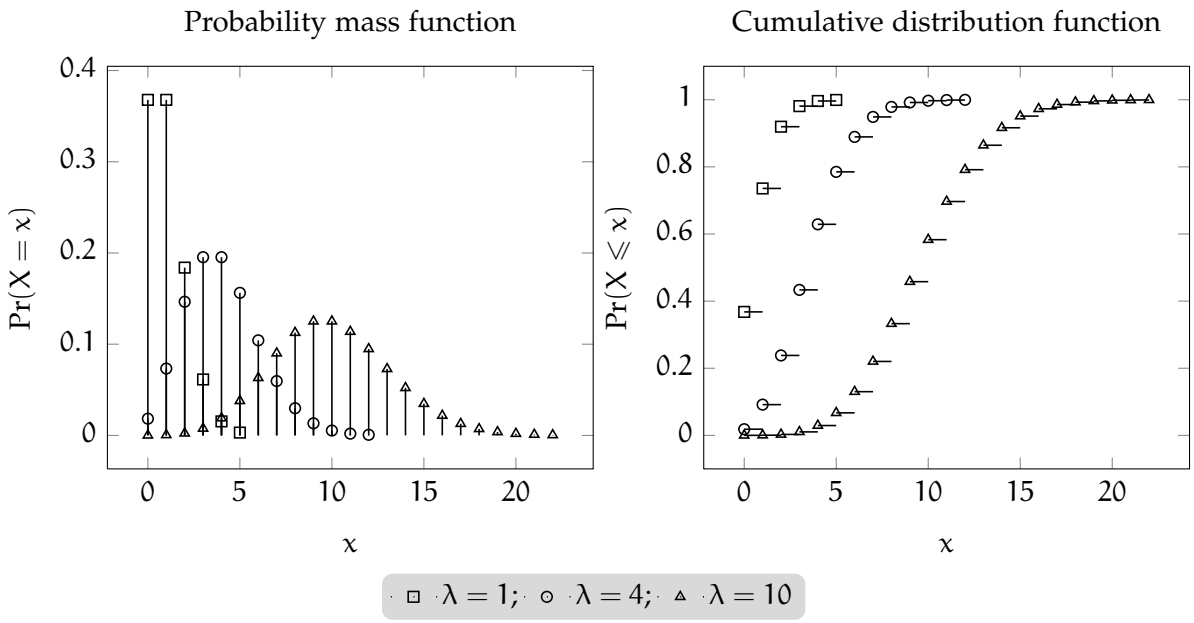


Figure 2.2: The probability mass and the cumulative distribution function of a Poisson distribution as functions of the parameter λ according to Equation 2.2.

A random variable X is said to have a *log-normal distribution* with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ if

$$\Pr(X = x) = \frac{1}{x\sqrt{2\pi\sigma^2}}\, e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} \quad \text{and} \quad \Pr(X \leqslant x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^x \frac{1}{t}\, e^{-\frac{\ln(t-\mu)^2}{2\sigma^2}}\, dt \ , \tag{2.4}$$

for $x \in [0, \infty)$ (cf. Figure 2.4).

A random variable X is said to have a *exponential distribution* with parameter $\lambda > 0$ if

$$\Pr(X = x) = \begin{cases} \lambda\, e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases} \quad \text{and} \quad \Pr(X \leqslant x) = \begin{cases} 1 - e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}, \tag{2.5}$$

for $x \in [0, \infty)$ (cf. Figure 2.5).

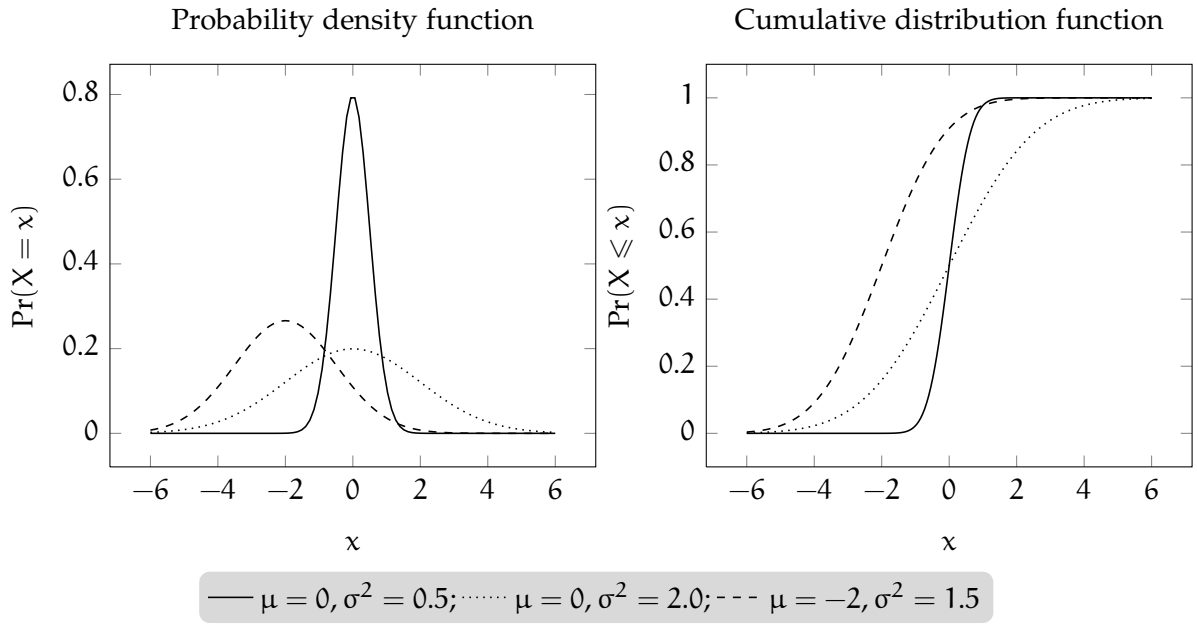


Figure 2.3: The probability density and the cumulative distribution function of a normal distribution as functions of the parameters μ and σ^2 according to Equation 2.3.

A random variable X is said to have a *Pareto distribution* with parameters $\alpha > 0$ and $x_{\min} > 0$ if

$$\Pr(X = x) = \begin{cases} \alpha \frac{x_{\min}^{\alpha}}{x^{\alpha+1}}, & x \geqslant x_{\min} \\ 0, & x < x_{\min} \end{cases} \quad \text{and} \quad \Pr(X \leqslant x) = \begin{cases} 1 - \left(\frac{x}{x_{\min}}\right)^{-\alpha}, & x \geqslant x_{\min} \\ 0, & x < x_{\min} \end{cases}, \tag{2.6}$$

for $x \in [x_{\min}, \infty)$ (cf. Figure 2.6).

A random variable X is said to have a *power law distribution* with parameters $\alpha > 0$ and $x_{\min} > 0$ if

$$\Pr(X = x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}}\right)^{-\alpha} \quad \text{and} \quad \Pr(X \leqslant x) = 1 - \left(\frac{x}{x_{\min}}\right)^{-\alpha+1}, \tag{2.7}$$

for $x \in [x_{\min}, \infty)$ (cf. Figure 2.7).

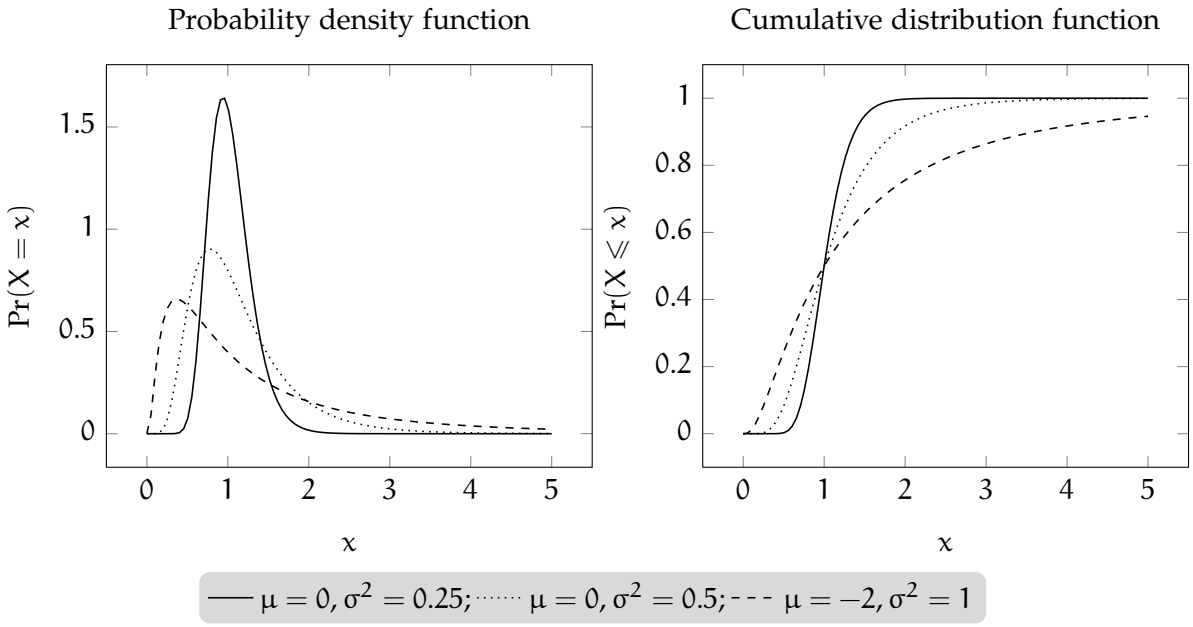


Figure 2.4: The probability density and the cumulative distribution function of a log-normal distribution as functions of the parameters μ and σ^2 according to Equation 2.4.

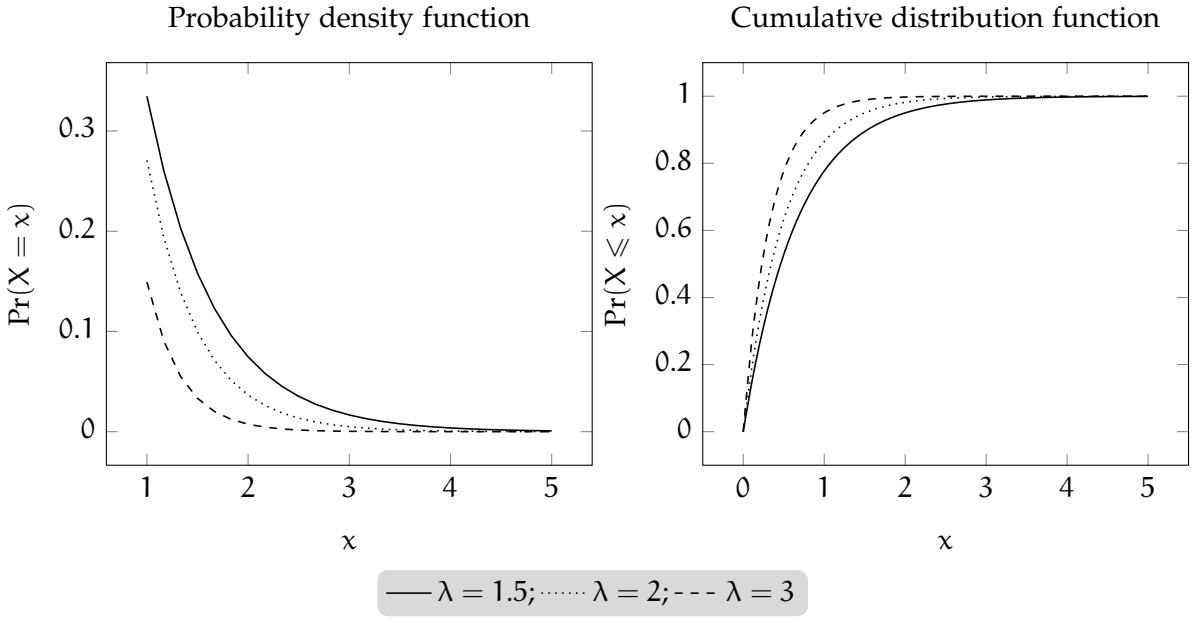


Figure 2.5: The probability density and the cumulative distribution function of a exponential distribution as functions of the parameter λ according to Equation 2.5.

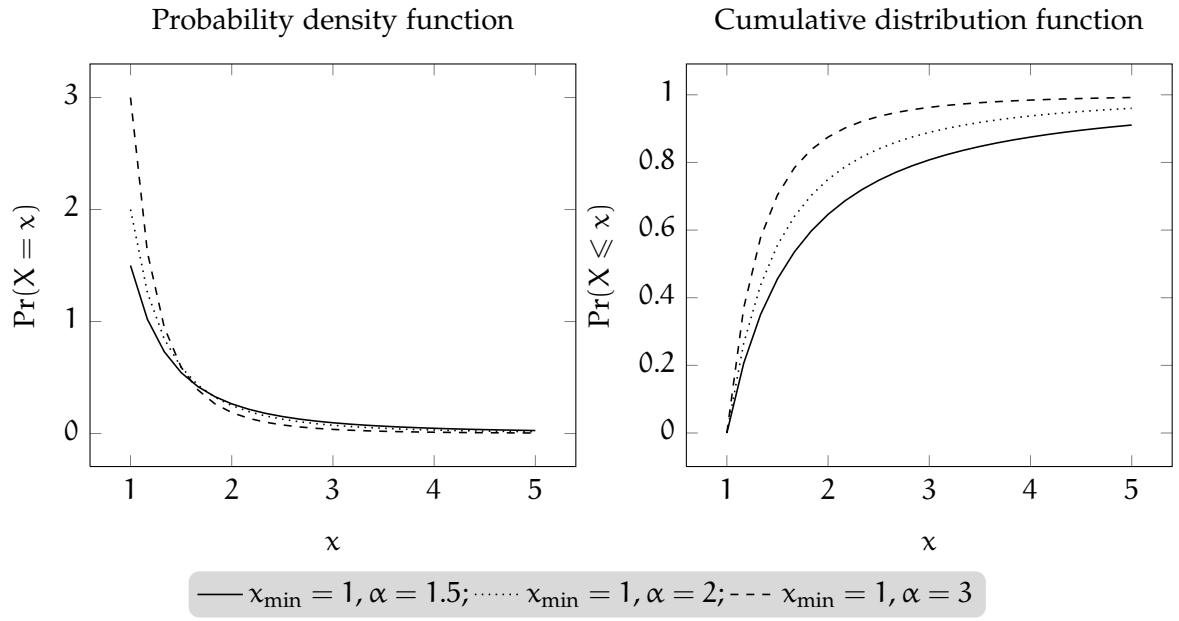


Figure 2.6: The probability density and the cumulative distribution function of a Pareto distribution as functions of the parameters α and x_{min} according to Equation 2.6.

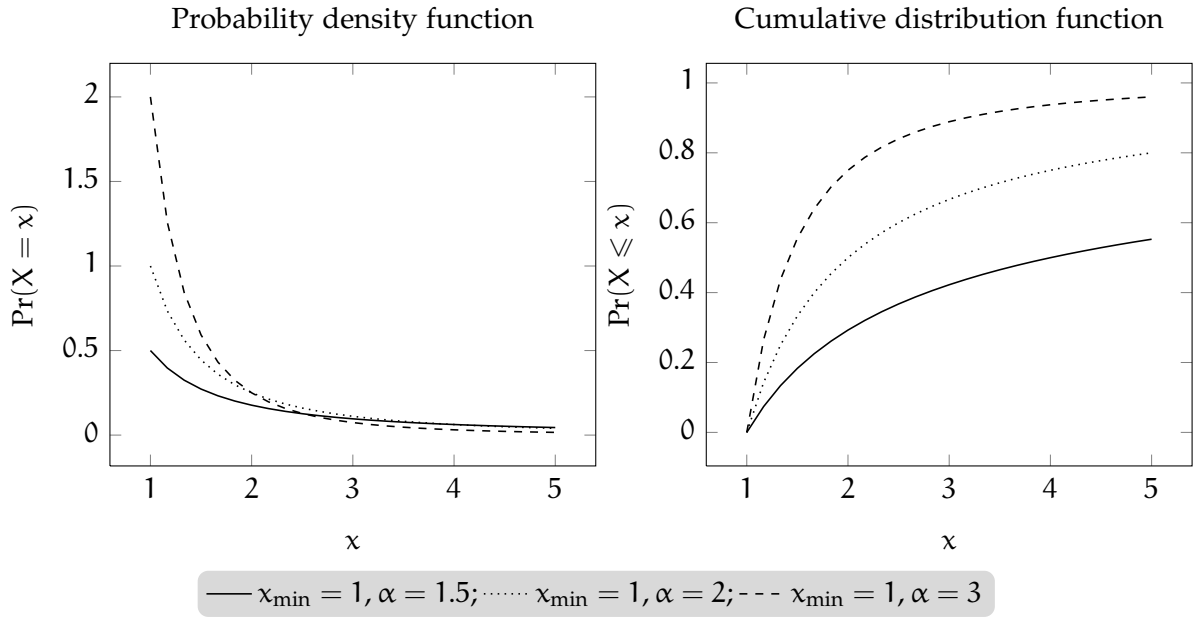


Figure 2.7: The probability density and the cumulative distribution function of a power law distribution as functions of the parameters α and x_{min} according to Equation 2.7.

2.3 GRAPH THEORY

This section gives an introduction to the graph theory terminology and notation used throughout this thesis. We start by giving a short outline of the historic beginnings of graph theory.

2.3.1 *Introduction*

Historically, the *Seven Bridges of Königsberg* is one of the first problems in graph theory and was posed in natural language by the middle of the 18th century. The formal problem description goes back to the famous mathematician Leonard Euler, who was confronted with the question if there exists a round trip through the city of *Königsberg* which passes every bridge of the city exactly once (see [Eul41]). Euler assigned points to the different districts of the city and lines to the bridges connecting those districts, and thus obtained a clearly structured schematic representation of the problems initial situation in form of an undirected graph (cf. Figure 2.8).

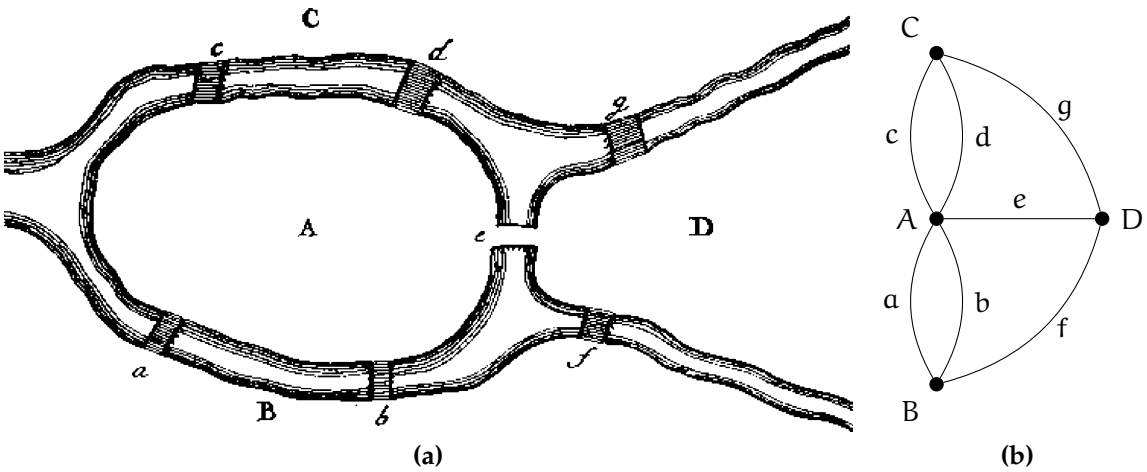


Figure 2.8: Find a tour trough the city of Königsberg, using every bridge that crosses the river Pregel exactly once. The setting of the *Seven Bridges of Königsberg* problem in a first drawing by Euler [Eul41] (Figure 2.8a) and its schematic representation as a *undirected graph* (Figure 2.8b).

This is one of the first examples for the application of graphs for the formal description and as a method of solution of real-world problems. Further milestones in the history of graph theory are listed below:

- In 1847 Kirchhoff [Kir47] introduced the concept of graphs for the investigation of *electronic circuits*.
- In 1857 Cayley [Cay57] considered a special class of graphs, so called *trees*, in connection with the enumeration of *isomers of saturated hydrocarbons*.
- Also in 1857 Hamilton invented the *Icosian Game* and an improved variant called the *Traveller's Dodecahedron*[1] and thereby introduced the well known HAMILTONIAN CYCLE problem on graphs (cf. [BLW86; Bol98]).
- In the middle of the 19th century Francis Guthrie [Gut54] formulated the *four-color conjecture* while coloring the map of counties in England. The conjecture, published by his brother Frederick, claims that four colors would suffice to color every planar map such that any two adjacent fields have different colors [Gut80]. As it turned out, the four-color conjecture could not be proven after several attempts until the computer-aided proof by Appel and Haken [AH76] in 1976.

In the 20th century graphs are ubiquitous representations of biological, technological, social and information systems and graph analysis—along with the tools of graph theory—is applied in each of these fields to understand the structure, function and organization of the networks under study. Let us now turn to the basic definitions in the realm of graph theory.

2.3.2 *Basic Definitions*

In what follows, we introduce the basic notations and invariants in connection with the notion of graphs that are used throughout this thesis. For a more detailed and complete introduction to graph theory see e. g. the textbooks by Bollobás [Bol98] and Diestel [Die10].

1 A feasible solution for this game is a round-trip along the edges of the dodecahedron that visits every node exactly one and ends at a node adjacent to the starting node.

GRAPHS AND MULTIGRAPHS. An *undirected graph* is a 2-tuple $G = (V, E)$ of finite sets V, E with $V \cap E = \emptyset$ and such that $E \subseteq \mathbb{P}_2(V)$. The elements of $V(G)$ are regarded as the *vertices* or *nodes* of the graph G, the elements of $E(G)$ are the *undirected edges* or *lines* of G. An undirected edge $e \in E$ is usually denoted as an unordered pair $e = \{u, v\}$ and we say that e is *incident* to the nodes u and v. Two edges are called *adjacent* if they share a common node as an endpoint, whereas two nodes are adjacent when there is an edge connecting the two nodes. If all vertices of a graph are pairwise adjacent the graph is *complete* and a complete graph on n vertices is denoted as $\mathbb{K}^n$. A graph $G = (V, E)$ is called *simple* if E does not contain multiple edges with the same incident nodes or self-loops $e = \{u, u\}$. By extending the notion of the set of edges to *multisets*, the corresponding *multigraph* is allowed to contain multiple *parallel edges* and *self-loops* $e = \{u, u\}$.

The vertex set of a graph G is referred to as $V(G)$ and its edge set $E(G)$, independent of the names of the two sets as denoted in the tuple. Throughout this thesis, we set the cardinality of the vertex set $|V(G)| := n$, which is also referred to as the *order* of G, denoted as $|G|$. Graphs of finite or infinite order are also called *finite* or *infinite* graphs, respectively. The cardinality of the edge set is $|E(G)| := m$, also denoted as $\|G\|$. The set of adjacent nodes to some node v is called the *neighborhood* of v and is denoted by $N(v) := \{ u \in V \mid \{u, v\} \in E \}$. For a subset $V' \subseteq V$, we set $N(V') := \bigcup_{v \in V'} N(v) \setminus V'$. We define the *degree* $d(v)$ as the number of edges incident to v. The number $\delta(G) := \min\{ d(v) \mid v \in V \}$ is the *minimum degree* of G, the number $\Delta(G) := \max\{ d(v) \mid v \in V \}$ is the *maximum degree* of G. If all of the vertices $v \in V(G)$ are of the same degree k the graph G is called k-*regular*. The number $d(G) := \frac{1}{|V|} \sum_{v \in V} d(v)$ is the *average degree* of G, and clearly $\delta(G) \leqslant d(G) \leqslant \Delta(G)$.

ISOMORPHIC GRAPHS AND SUBGRAPHS. Let $G = (V, E)$ and $G' = (V', E')$ be two graphs. We call G and G' *isomorphic*, $G \simeq G'$, if there exists a bijection $\varphi : V \to V'$ with $\{u, v\} \in E \Leftrightarrow \{\varphi(u), \varphi(v)\} \in E'$ for all $u, v \in V$. The map φ is then called an *isomorphism*.

Let $G = (V, E)$ be a finite (undirected) graph. A graph $H = (V', E')$ with $V' \subseteq V$ and $E' \subseteq E$ is called *subgraph* of G, $H \subseteq G$. Moreover, we define an *edge-induced* subgraph $G[E'] := (V[E'], E')$ with $V[E'] := \bigcup_{\{u,v\} \in E'} \{u, v\}$, i. e. the subgraph that contains exactly the edges $e \in E'$ and all incident vertices. Analogously, we define a *node-induced*

subgraph $G[V'] := (V', E[V'])$ with $E[V'] := \{\{u, v\} \in E \mid \{u, v\} \subseteq V\}$, i. e. the subgraph that contains exactly the nodes V and all edges of G that connect nodes of V. A subgraph $H \subseteq G$ is called *spanning* subgraph if V' spans all of G, i. e. $V' = V$.

PATHS AND TREES. A *path* of length l in a graph $G = (V, E)$ is a sequence $P = v_1, v_2, \ldots, v_l, v_{l+1}$ of $l + 1$ vertices $v_i \in V$ such that every two subsequent vertices in P are adjacent in G. A path is called *simple* if the vertices of the sequence are pairwise different. The sequence $C = v_1, v_2, \ldots, v_l, v_1$ is called a *cycle* of length l in G. The *distance* $d(u, v)$ of two vertices u, v in G is the length of a shortest path from u to v. If no such path exists, we set $d(u, v) := \infty$. The maximum distance between any two vertices of G is called the *diameter* of G, denoted as $\operatorname{diam}(G)$.

A non-empty graph G is called *connected* if the there exists a path between any two vertices of G. For any subset $U \subseteq V$ where $G[U]$ is connected, we may also call the subset U connected. A maximal subset U, such that $G[U]$ is connected, is called a *component* of G. A graph without cycles, i. e. an *acyclic* graph, is called *forest*. A connected forest is called *tree* and the vertices of degree 1 are called its *leaves*.

SPECIAL SUBSETS. Pairwise non-adjacent vertices or edges are called *independent* and—more generally—a subset of vertices or edges is called *independent set* of vertices or edges, if no two of its elements are adjacent. A *vertex cover* of G is a subset of vertices $C \subseteq V$ such that each edge has at least one endpoint in C. A *dominating set* in G is a subset of vertices $D \subseteq V$ such that for each $v \in V$ either $v \in D$ or $D \cap N(v) \neq \emptyset$. A *matching* of a graph G is a subset $M \subseteq E$ of edges which are pairwise not adjacent in G. A *clique* of a graph G is a subset $L \subseteq V$ of vertices such that every pair of vertices from the subset are adjacent.

DIRECTED GRAPHS. A *directed graph* is a 2-tuple $G = (V, A)$ of finite sets with $A \subseteq V \times V$ denoting the set of *directed edges* or *arcs* of G. A directed edge $e \in A$ is denoted as an ordered pair $e = (u, v)$, an we say that e is directed from its *source node* u to its *target node* v. The two arcs (u, v) and (v, u) are called *reversals* of each other. We define the *in-degree* $d^+(v)$ and the *out-degree* $d^-(v)$ as the number of ingoing and outgoing arcs of v, respectively. All the definitions here hold, in slight variations, for both directed and undirected graphs.

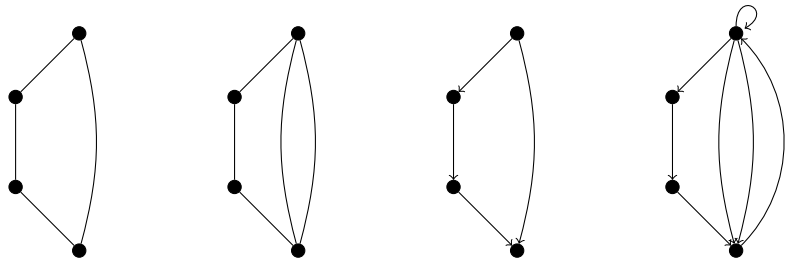


Figure 2.9: Examples for an undirected simple graph (Figure 2.9a), an undirected multigraph (Figure 2.9b), a directed simple graph (Figure 2.9c) and a directed multigraph with a self-loop (Figure 2.9d).

HYPERGRAPHS. Another generalization of the graphs introduced above are the so called *hypergraphs* $H = (V, E)$, where $E \subseteq \mathbb{P}(V) \setminus \emptyset$ is the set of *hyperedges*—that is, edges connecting an arbitrary number of vertices. In a *directed hypergraph* a directed hyperedge $e \in E$ is denoted as an ordered pair (X, Y), where $X \subseteq V, X \neq \emptyset$ and $Y \subseteq V \setminus X$. The set X is called the *source* of the directed hyperedge and Y is called the *target*.

CLUSTERING COEFFICIENTS. Besides the graph invariants mentioned above, one of the defining properties of many real-world graphs is the *clustering coefficient* $C(G)$, which describes how well connected the neighborhood of a node or a set of nodes is. Following Watts and Strogatz [WS98], given a simple undirected graph $G = (V, E)$ and a vertex v, the *local clustering coefficient* $C_v(G)$ is defined by

$$C_v(G) = 2 \frac{|\{\{u, w\} \mid u, w \in N_v, \{u, w\} \in E\}|}{d(v)(d(v) - 1)} \ .$$

Since $d(v)$ denotes the number of edges incident to v, the denominator of the equation above is the maximum possible number of edges between neighbors of v. Note that $0 \leqslant C_v(G) \leqslant 1$ and that the value of $C_v(G)$ quantifies how close the neighbors of v are to being a clique (see Figure 2.10).

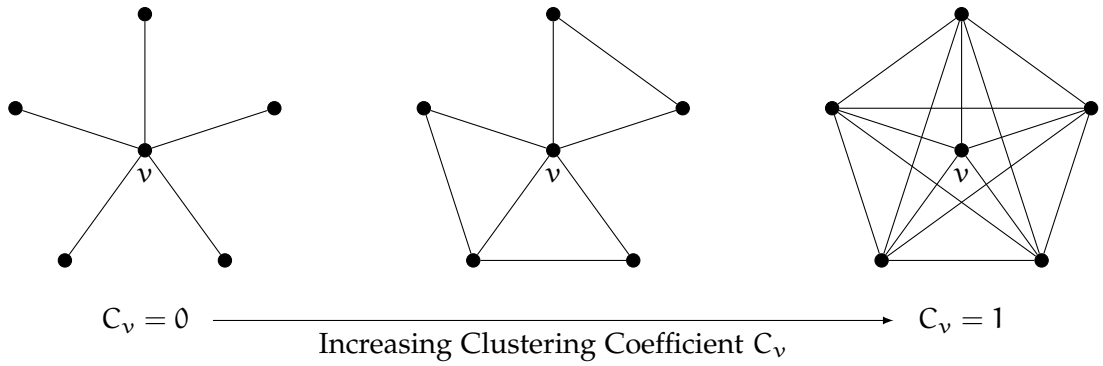


Figure 2.10: Graphs of increasing local clustering coefficient C_v at a vertex v.

For the *global clustering coefficient* we have two possible definitions, namely

$$C(G) = C_1(G) = \frac{1}{n} \sum_{v \in V} C_v(G)$$

and

$$C(G) = C_2(G) = \frac{\sum_{v \in V} (d(v)(d(v)-1))\, C_v(G)}{\sum_{v \in V} (d(v)(d(v)-1))} \ .$$

The first definition follows Watts and Strogatz [WS98] and represents some kind of an average value over averages, which is often not very informative. The second definition, which is also sometimes termed the *transitivity ratio* of a graph, is a related concept that places more weight on the nodes of higher degree and has the following natural equivalent form that also applies to multigraphs.

$$C_2(G) = \frac{\text{\# of pairs of adjacent edges } \{v,u\},\{v,w\} \text{ where } \{u,w\} \text{ is an edge}}{\text{\# of pairs of adjacent edges } \{v,u\},\{v,w\}} \ ,$$

or equivalently,

$$C_2(G) = \frac{3 \times \text{\# of triangles}}{\text{\# of pairs of adjacent edges}} \ .$$

In order to illustrate the difference between the two definitions of the clustering coefficient $C_1(G)$ and $C_2(G)$ consider the graph G on n vertices shown in Figure 2.11.

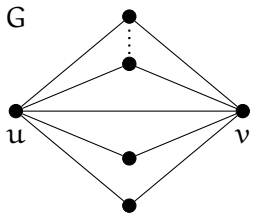


Figure 2.11: An example of a graph that has considerably different values for the two definition of the global clustering coefficient $C_1(G)$ and $C_2(G)$. It is easy to see that the local clustering coefficient is $C_u(G) = C_v(G) = \frac{2}{n-1}$ for the vertices u and v, and $C_x(G) = 1$ for all other vertices. For the global clustering coefficient it follows that $C_1(G) = 1 - o(1)$ and $C_2(G) \sim 2/n$.

DEGREE DISTRIBUTIONS. Another characteristic property of graphs is its underlying *degree distribution* or *degree sequence*. The (non-cumulative) degree distribution $f_d(k)$ is defined to be the fraction of nodes with degree k, that is, for a graph over n vertices with y_k vertices of degree k, $f_d(k) = \frac{y_k}{n}$. More formally, let $G = (V, E)$ be a graph and $k = 1, 2, \ldots$, then, if we pick a vertex $v \in V$ uniformly at random, $\Pr(d(v) = k) = f_d(k)$.

The degree sequence of a graph is defined in two different ways, namely the d-degree sequence and the y-degree sequence. First, let $G = (V, E)$ be a graph with maximum degree Δ. Then, the y-degree sequence is defined as $y(G) = (y_1, y_2, \ldots, y_\Delta)$, where y_i is the number of vertices of degree i. Second, let n be the number of vertices in G. Then, the d-degree sequence is defined as $d(G) = (d(v_1), d(v_2), \ldots, d(v_n))$, where the vertex degrees $d(v_i)$ are ordered such that the sequence is monotonic non increasing.

In the upcoming section we are going to introduce the concept of random graph models that aim to describe how "typically observed" graphs emerge and how they look like.

2.4 RANDOM GRAPH THEORY

We are going to introduce the standard models for uniform random graphs. Moreover, we will summarize some previous results that characterize the structure of uniform random graphs. This small extract from the huge field of random graphs theory will be suited to point out the differences and similarities between the classical uniform random graphs and random power law graphs.

2.4.1 *Introduction and Basic Definitions*

The research on the *theory of random graphs* was initiated by Erdős and Rényi—and independently by Gilbert—in the late 1950's when the first formal definitions of a random graph were presented in the seminal papers [ER59] and [Gil59]. The overall aim in this branch of research was to investigate the properties of a 'typical' graph with a fixed number of labeled vertices, and to devise mathematical models that generate such graphs in a random fashion. In fact, random graph theory is not the study of graphs, but the study of a *family* or a *probability space* of graphs. Such a graph class consists of many different graphs, each of which is associated with a probability denoting the likelihood of occurrence.

The $G(n,p)$ model due to Gilbert [Gil59] describes a random graph family on n vertices where en edge between any two vertices occurs with probability p. In the Erdős-Rényi model an equal probability is assigned to all graphs over a vertex set $[n] = \{1,2,\ldots,n\}$ with exactly M edges. The Erdős-Rényi model is therefore often also referred to as the $G(n,M)$ model for random graphs.

In other words, Erdős and Rényi [ER59] considered the probability space $\mathcal{G}_{n,M}$ of all $\binom{N}{M}$ graphs with vertex set $[n]$ and M edges, where $N = \binom{n}{2}$ is the number possible edges between vertices of the vertex set $[n]$. By assigning an equal probability to each element, the set $\mathcal{G}_{n,M}$ is made into a *probability space* and the graph $G_{n,M} \in \mathcal{G}_{n,M}$ will denote a random element of this probability space.

In order to have a similar notation for the $G(n,p)$ model, a family of graphs or probability space $\mathcal{G}_{n,p}$ is defined as follows: Let $\{X_{ij} \mid 1 \leqslant i < j \leqslant n\}$ be an array of independent and identically distributed (i.i.d.) Bernoulli random variables, with $\Pr(X_{ij} = 1) = p$ and $\Pr(X_{ij} = 0) = 1-p$, and let $G_{n,p} \in \mathcal{G}_{n,p}$ be the random graph on vertex set $[n]$ in which an edge connects vertices i and j if $X_{ij} = 1$. This means, in order to construct a random graph $G_{n,p} \in \mathcal{G}_{n,p}$, we draw edges with probability p, independent of each other. Erdős and Rényi also noted in [ER60] that in the $\mathcal{G}_{n,p}$ model introduced by Gilbert, the number of edges is a random variable with expectation $p\binom{n}{2}$. Thus, for $M \sim p\binom{n}{2} = pN$, the two models $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,M}$ are almost interchangeable.

Now let $G_{n,M} \in \mathcal{G}_{n,M}$ or $G_{n,p} \in \mathcal{G}_{n,p}$ be a random graph in the $G(n,M)$ model or $G(n,p)$ model, respectively. We say that $G_{n,\bullet}$ has a certain property $\mathcal{P}$ with high probability (w.h.p.) if

$$\Pr(G_{n,\bullet} \text{ has property } \mathcal{P}) \to 1 \ ,$$

as $n \to \infty$. The interesting case in the $\mathcal{G}_{n,M}$ model is when the number of edges is a function of n, i. e. $M = M(n)$, or—for the $\mathcal{G}_{n,p}$ model—if p is a function of n.

Another way to generate a random graph on n vertices is to employ a stochastic process that starts on a graph with n vertices and no edges, and that, with each step the process, adds an edge from the set of possible edges. Formally, a *graph process* $\widetilde{G}_n = (G_{n,t})_{t=0}^{N}$ on a vertex set $[n]$ is a nested sequence of graphs, $G_{n,0} \subset G_{n,1} \subset \cdots \subset G_{n,N}$ such that $G_{n,t}$ has precisely t edges. The space $\widetilde{\mathcal{G}}_n$ of *random graph processes* consists of all $N!$ graph processes on the vertex set $[n]$. We note here, that the distribution of $G_{n,t}$ as a random graph process stopped at time t is precisely the distribution of an element $G_{n,t} \in \mathcal{G}_{n,t}$. This means, elements $G_{n,M} \in \mathcal{G}_{n,M}$ can be viewed as a state or snapshot of a random graph process $\widetilde{G}_n$ at time $t = M$. Furthermore, a random graph process has a natural interpretation as a *dynamic Markov process* that acts on a sequence of graphs $G_{n,0}, \ldots, G_{n,N}$. Given a sequence of graphs or steps $G_{n,0}, \ldots, G_{n,t}$ at time t, the next step $G_{n,t+1}$ is obtained by adding one of the $N - t$ remaining possible edges to $G_{n,t}$ uniformly at random.

2.4.2 *Results on Classical Random Graphs*

The following theorem presents a fundamental result of Erdős and Rényi—stated in a slightly different form in [ER59]—about the property of *connectedness* of certain random graphs $G_{n,M} \in \mathcal{G}_{n,M}$.

Theorem 2.1 (Erdős and Rényi [ER59])**.**
Let $M_\omega = n/2 \cdot (\log n + \omega)$*, where* $\omega = \omega(n)$ *is a function of* n*. If* $\omega \to -\infty$ *then a typical* G_{n,M_ω} *is disconnected, while if* $\omega \to \infty$*, a typical* G_{n,M_ω} *is connected.*

In the study of the evolution of random graphs, Erdős and Rényi showed for a number of fundamental graph properties that there exist sharp thresholds on the

corresponding parameters, such that graphs with a parameter size slightly less that this threshold are very unlikely to have the property, while graphs of size slightly above this threshold almost surely (a.s.) have the property. This is known as a *phase transition* of a property within a family of random graphs or within a random graph process. Erdős and Rényi [ER60] proved that the component structure of a random graph process undergoes such a phase transition if $t \sim n/2$ in a random graph process $\widetilde{G}_n$ or if $p = 1/n$ for a graph $G_{n,p} \in \mathcal{G}_{n,p}$. We state the following theorem.

Theorem 2.2 (Erdős and Rényi [ER60])**.**
Let $G_{n,p} \in \mathcal{G}_{n,p}$, $c > 0$ *be a constant and set* $p = c/n$.

(i) If $c < 1$, *then* w.h.p. *every component of* $G_{n,p}$ *is of order* $\mathcal{O}(\log n)$.

(ii) If $c > 1$, *then* w.h.p. $G_{n,p}$ *has a component with* $(\gamma(c) + o(1))n$ *vertices, where* $\gamma(c) > 0$, *and all of the remaining components are of order* $\mathcal{O}(\log n)$.

If a component of order $\Theta(n)$ as described above in part (ii) of Theorem 2.2 exists, this component is called the *giant component* of the random graph $G_{n,p}$. For the case $c \gg 1$ the giant component has a large highly connected subgraph, which is sometimes termed as the *core* of the graph.

Let $y_1, \ldots, y_\Delta$ be the *degree sequence* of a graph, where y_i is the number of vertices of degree i and Δ is the maximum degree. In the $\mathcal{G}_{n,p}$ model, for p being a constant, the degree sequence of $G_{n,p}$ is well approximated by a sequence of n i.i.d. Binomial random variables with probability p and mean $n \cdot p$ (cf. Section 2.2 and Equation 2.1 on page 24). For $p = c/n$, with c being a constant, the degree sequence is close to a sequence of n i.i.d. Poisson random variables with mean c (cf. Equation 2.2 on page 25). As noted by Bollobás and Riordan [BR05], the following theorem holds regarding the degree distribution of a graph $G_{n,p} \in \mathcal{G}_{n,p}$.

Theorem 2.3.
Let y_i *be the number of vertices of degree* i *in* $G_{n,p} \in \mathcal{G}_{n,p}$ *where* $p = c/n$ *and* $c > 0$ *a constant. Then for* $i = 0, 1, \ldots$

$$\Pr\left((1-\varepsilon)\frac{c^i e^{-c}}{i!} \leqslant \frac{y_i}{n} \leqslant (1+\varepsilon)\frac{c^i e^{-c}}{i!}\right) \to 1$$

as $n \to \infty$.

As defined in Section 2.3 on page 33 the *diameter* of a graph is the maximum length of a shortest path between to vertices. The diameter of a random graph is a property which has been studied in great detail (cf. Bollobás [Bol81], Bollobás and Fernandez de la Vega [BF82], Chung and Lu [CL01]). In the $\mathcal{G}_{n,p}$ model, if $\frac{pn}{\log n} \to \infty$ and $\frac{\log n}{\log(pn)} \to \infty$, then the diameter of $G_{n,p} \in \mathcal{G}_{n,p}$ is asymptotic to $\frac{\log n}{\log(pn)}$ w.h.p. Furthermore, we write here a result on the diameter of random *regular* graphs, which closely resembles a result on random power law graphs that will be stated in Section 2.5.

Let $3 \leqslant r \leqslant n$ be a fixed constant and let $n \cdot r$ be even. Then $\mathcal{G}_{n,r\text{-reg}}$ denotes the family of random r-regular graphs on n vertices. We state the following simplified form of the main theorems from [BF82].

Theorem 2.4 (Bollobás and Fernandez de la Vega [BF82])**.**
Let $r \geqslant 3$ and $\varepsilon > 0$ be fixed. Then, for the diameter of $G_{n,r\text{-}reg} \in \mathcal{G}_{n,r\text{-}reg}$, the following holds as $n \to \infty$:

$$\Pr\left((1-\varepsilon)\frac{\log n}{\log(r-1)} \leqslant \operatorname{diam}(G_{n,r\text{-}reg}) \leqslant (1+\varepsilon)\frac{\log n}{\log(r-1)}\right) \to 1 \ .$$

Now, we are going to introduce another family of graphs which is also adapted to modeling typical graphs stemming from large-scale real-world systems.

2.5 POWER-LAW GRAPHS

Recently the study of *large-scale* real-world networks revealed common topological signatures and statistical features that are not easily captured by classical *uniform* random graphs—such as generated by the $G(n,p)$ and $G(n,M)$ introduced in Section 2.4.

2.5.1 *Introduction*

Within the interdisciplinary study of complex networks one of the most interesting properties of the graphs under study is the so called *scale-freeness*, a property that contains a set of features such as small (and ultra-small) diameters, large clustering coefficients and particular distributions of node degrees. In contrast to uniform random graphs that exhibit a binomial or Poisson distribution of node degrees, this family is

characterized by the property of having a *power law degree distribution*. This means that the number of nodes y_i of a given degree i is proportional to $i^{-\beta}$, where $\beta > 0$ is the so called *power law exponent*. The corresponding graphs of this class are therefore called *power law graphs* or—with a stronger focus on the set of features that define scale-freeness—*scale-free networks*. Figure 2.12 shows a comparison of a random graph in the $G(n, p)$ model and a random power law graph, both with the same numbers of vertices and edges.

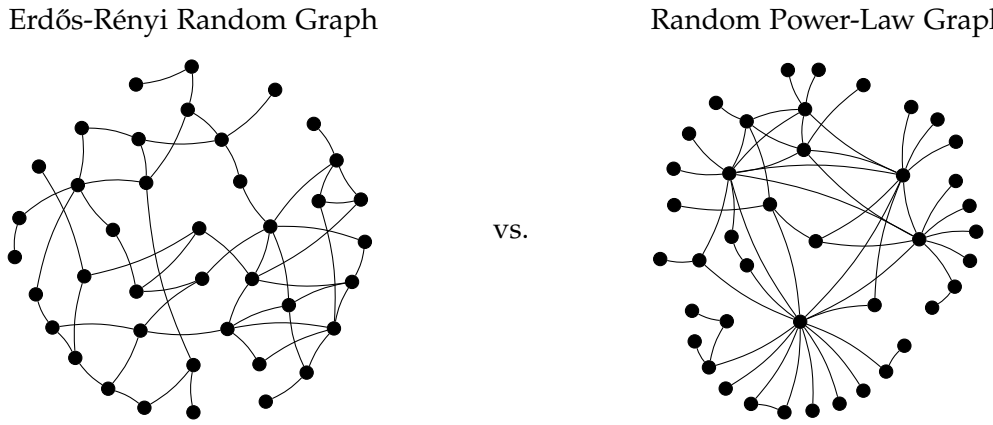


Figure 2.12: Comparison of a random graph in the $G(n, p)$ model due to Erdős and Rényi [ER60] and Gilbert [Gil59] versus a random power law graph in the generalized preferential attachment model on n vertices and m edges, with $n = 42$ and $m = 56$.

Starting in 1999 Kumar et al. [Kum+00; Bro+00], Kleinberg et al. [Kle+99; KL01] and Faloutsos, Faloutsos, and Faloutsos [FFF99; Sig+03] measured the degree sequence of the World-Wide Web and independently observed that it is well approximated by a power law distribution. In the sequel, this was verified for a large number of existing real-world networks such as *protein-protein interaction networks*, *gene regulatory networks*, *peer-to-peer networks*, *e-mail networks*, *mobile call networks*, *social networks* et cetera [JAB01; Gue+02; EMB02; Ses+08; Eub+04b].

In fact, power law distributions in real-world data sets had also been observed considerably earlier. For example for the distribution of income, city sizes, word frequencies and for citations of academic literature [Par96; Aue13; Est16; Lot26]. Besides these and other early investigations, the idea of associating power law distributions with real-life systems and also the popularization of this idea is generally attributed

to the American linguist Zipf [Zip35; Zip49]. Zipf considered the frequency of occurrence of words in a natural language and observed a power law distribution. He is also the originator of the widespread term *Zipf's Law* as a synonym for power law behavior, and associated power law distribution is therefore often referred to as a *Zipfian distribution*.

2.5.2 *Basic Definitions*

First, we will give a non-stochastic definition of a power law relationship. A finite sequence $y = (y_1, y_2, \ldots, y_n)$, with $y_i \in \mathbb{R}$ and ordered such that $y_1 \leqslant y_2 \leqslant \ldots \leqslant y_n$, follows a *power law* or exhibits a *scaling relationship* if

$$i = c \cdot y_i^{-\alpha} \ ,$$

where i is the *rank* of y_i, c is a normalizing constant and α is the *scaling index*.

In a stochastic context, assuming an underlying probability model Pr for a non-negative random variable X, let $F_X(x) = \Pr(X \leqslant x)$ denote the cumulative distribution function (CDF) of X, and let $\bar{F}_X(x) = \Pr(X > x) = 1 - F(x)$ be the complementary cumulative distribution function (CCDF) of X. A random variable X or its corresponding distribution function F_X is said to follow a *power law* if, as $x \to \infty$,

$$\Pr(X > x) = 1 - F(x) \sim c \cdot x^{-\alpha} \ ,$$

for constants $0 < c < \infty$ and a $\alpha > 0$. Here, $f(x) \sim g(x)$ means that the limit of the ratio $f(x)/g(x)$ goes to 1 as $x \to \infty$. The constant α of a distribution following a power law is also referred to as the *power law exponent*.

If a random variable X has a power law distribution, then a doubly logarithmic plot of the CCDF of X asymptotically yields a straight line of slope $-\alpha$. This fact can be used as an empirical test for whether a random variable X follows a power law on a given sample[2]. In Figure 2.13 we show a linear plot of the CCDF of a power law distribution in comparison with a log-log plot of the same CCDF.

2 This observation has already been made by Auerbach [Aue13] for the distribution of city sizes in 1913.

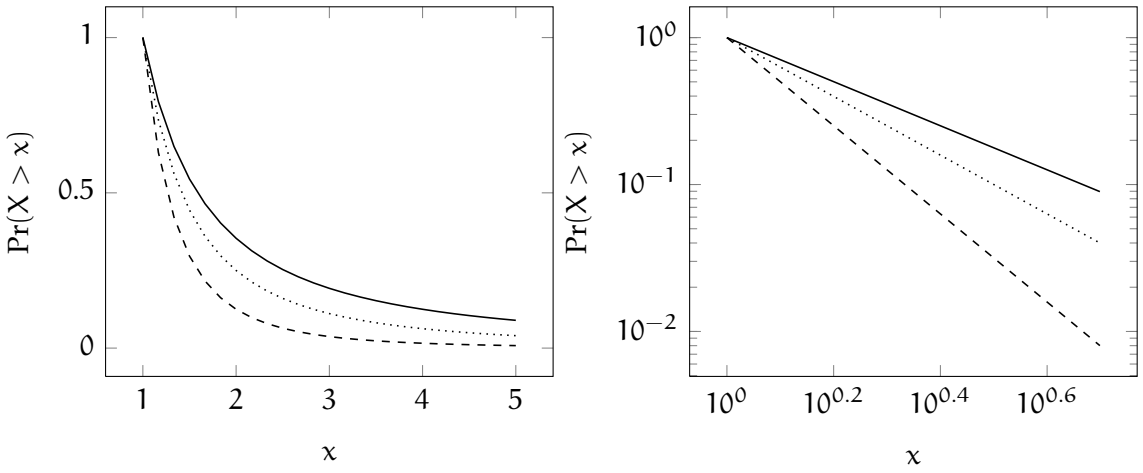


Figure 2.13: Left: A plot of the CCDF $\Pr(X > x) = x^{-\alpha}$ of a power law distribution with $\alpha = 1.5$ (—), $\alpha = 2$ (·····) and $\alpha = 3$ (- - -). Right: A doubly logarithmic plot of the same CCDF yielding a straight line of slope $-\alpha$.

One specific power law distribution is the *Pareto distribution*, which was introduced by Pareto [Par96] in 1896 as a mathematical argument to explain the distribution of income inside a population. It satisfies

$$\Pr(X > x) = \left(\frac{x}{k}\right)^{-\alpha} ,$$

for some constants $\alpha > 0$ and $k > 0$. Figure 2.14 shows a plot of the above CCDF of the Pareto distribution on a linear scale and a log-log scale.

In the following, we describe some random graph models that aim to explain how power law degree distributions in evolving graphs arise.

2.5.3 *Evolving Models*

The model of *"preferential attachment"* is most often referred to as a potential mechanism underlying the emergence of scale-free graphs and features the role of evolutionary growth or rewiring in the construction process. Also known under the name of *"cumulative advantage"*, the model was first mentioned and described by Yule [Yul25], Simon [Sim55] and de Solla Price [Pri65; Pri76]. Later this concept was reintroduced to a broader audience by Barabási and Albert [BA99] and was more rigorously and

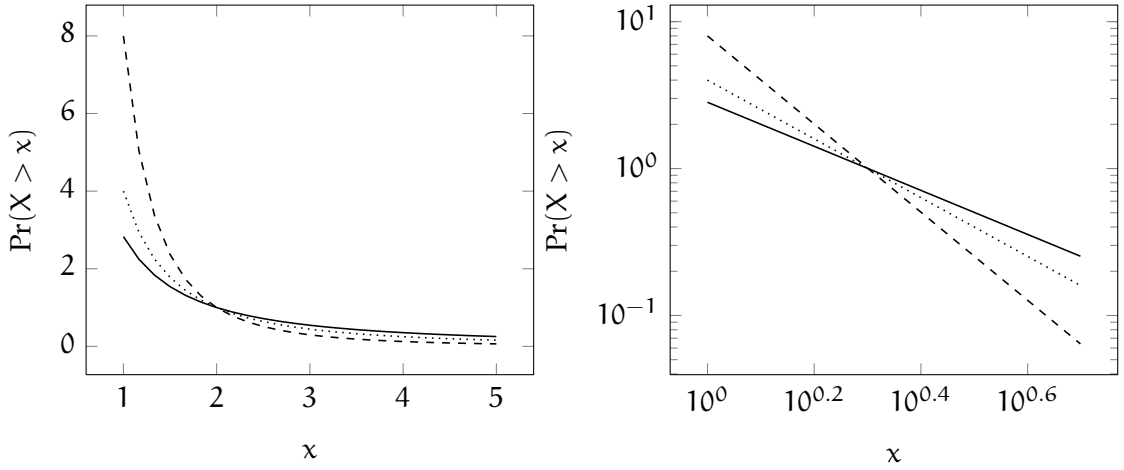


Figure 2.14: Left: A plot of the CCDF $\Pr(X > x) = (x/2)^{-\alpha}$ of a Pareto distribution with $\alpha = 1.5$ (—), $\alpha = 2$ (·····) and $\alpha = 3$ (- - -). Right: A doubly logarithmic plot of the same CCDF yielding a straight line of slope $-\alpha$.

mathematically defined by Bollobás and Riordan [BR05]. In the model, a newly introduced vertex will connect to already existing vertices with a probability depending on their current degree. This principle of network growth is therefore customary described as a phenomenon of *"the rich get richer"* or *preferential attachment*.

THE BARABÁSI-ALBERT MODEL. Let us now give a slightly more formal description of the *preferential attachment model* introduced by Barabási and Albert [BA99]. Starting on an initial *seed graph* with a small number m_0 of vertices, at each time step t, a new vertex v_t along with $m \leqslant m_0$ edges is added to the existing graph. The m edges are added randomly to m different vertices and the probability that v_t connects to a vertex v_i is proportional to the degree $d_t(v_i)$ of v_i at time t, i. e. the probability of an edge $\{v_t, v_s\}$ at time t is given by

$$\Pr(i = s) = \frac{d_t(v_s)}{\sum_{v_j \in V, v_j \neq v_s} d_t(v_j)} \ .$$

However, the model description above gives rise to some problems, as pointed out by Bollobás and Riordan [BR05, pp. 7-10]. The first problem is the initial situation of a seed graph on m_0 vertices and no edges, since the attachment probability is dependent

on the vertex degrees, which are all 0 in the initial situation. The second problem is with the preferential attachment rule for the case $m \geqslant 2$. Suppose we add $m \geqslant 2$ edges at time $t+1$. Since the above description does not allow to add the edges one by one independently, we must fully describe the probability that one specific set S out of the $\binom{t}{m}$ possible sets—consisting of vertices of the previous step—is chosen to be connected by the newly added edge. This probability is not uniquely specified by giving the marginal probabilities that $v_i \in S$ for each vertex v_i from time step t.

THE LCD MODEL. In order to be able to prove results for preferential attachment random graphs, Bollobás and Riordan [BR05] defined a precise random graph model—the *LCD model*—satisfying the description of Barabási and Albert [BA99]. In this model, multi-edges and self-loops are allowed and can be interpreted for example in the context of the WWW that two websites can have multiple links between them and also links to themselves. Bollobás and Riordan considered a fixed sequence of vertices $v_1, v_2, \ldots$ and defined inductively a random graph process $(G_1^{(t)})_{t\geqslant 1}$ so that $G_1^{(t)}$ is a graph on the vertex set $V = \{v_i \mid 1 \leqslant i \leqslant t\}$. The initial situation is the empty graph $G_1^{(0)}$ or the graph $G_1^{(1)}$ with one vertex and one self-loop. Given the graph $G_1^{(t-1)}$ at time $t-1$, the graph $G_1^{(t)}$ is formed by adding the vertex v_t and a single edge $\{v_t, v_i\}$. The target index i is chosen randomly with

$$\Pr(i = s) = \begin{cases} \frac{d(v_s)}{2t-1} & \text{for } 1 \leqslant s \leqslant t-1 \\ \frac{1}{2t-1} & \text{for } s = t \ , \end{cases}$$

where $d(v_s)$ is the degree of v_s in $G_1^{(t-1)}$. In other words, an edge $e = \{v_t, v_i\}$ is constructed with probability proportional to the degree of v_i and counting e as already contributing to the degree of v_t. For $m > 1$, m edges originating at v_t are added one by one and counting all the previous edges and the fixed end of the current edge as already contributing to the node degrees. These precise rules lead to a equivalent alternative definition, where the graph process $(G_m^{(t)})_{t\geqslant 0}$ is defined by running the process $(G_1^{(t)})_{t\geqslant 0}$ on a sequence $v_1', v_2', \ldots$, and to construct the graph $G_m^{(t)}$ from $G_1^{(mt)}$ by identifying the vertices $v_1', v_2', \ldots, v_m'$ to form the single vertex v_1, $v_{m+1}', v_{m+2}', \ldots, v_{2m}'$

to form the vertex v_2, and so on. In this way, $G_m^{(t)}$ is defined in terms of a much simpler object $G_1^{(mt)}$ and results about $G_m^{(t)}$ can be proven in terms of $G_1^{(mt)}$.

Another important property, which also gave the model its name, is the following: While $G_m^{(t)}$ is a dynamic process, the distribution of the graph $G_m^{(n)}$ obtained at a particular time $t = n$ has a simple static description, namely the *linearized chord diagram* (LCD) description defined in [BR04]. A LCD with n chords, or equivalently, an n-*pairing* is a partition of the set $\{1, 2, \ldots, 2n\}$ into pairs, which yields a number of $(2n)!/(n!2^n)$ n-pairings. A pair is formed by a chord of a LCD L, that is, a line connecting two points of the set $\{1, 2, \ldots, 2n\}$. Moreover, every chord is said to have a *left* and *right endpoint*. Figure 2.15 shows a LCD with n chords over $2n$ points.

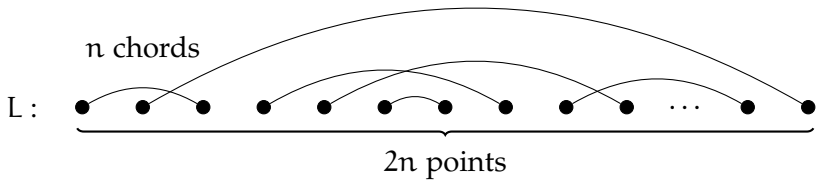


Figure 2.15: A linearized chord diagram L with n chords over $2n$ points, as defined by Bollobás and Riordan [BR04].

A graph $\phi(L)$ from a LCD L is generated as follows: starting from the left, vertex v_1 is formed by accumulating all endpoints to the right including the first right endpoint of a chord. All further endpoints up to the next right endpoint form vertex v_2, and so on. The edge set of the graph $\phi(L)$ is generated by replacing each chord by an edge joining the vertices corresponding to the two endpoints of the chord (see Figure 2.16).

Now let $\mathcal{G}_m^{(n)}$ denote the probability space of random graphs in the LCD model over the set $[n]$, where a random element $G_m^{(n)} \in \mathcal{G}_m^{(n)}$ has the distribution of the process described above. Bollobás and Riordan [BR04] and independently Cohen and Havlin [CH03] proved the following property regarding the diameter and connectedness of random graphs $G_m^{(n)} \in \mathcal{G}_m^{(n)}$.

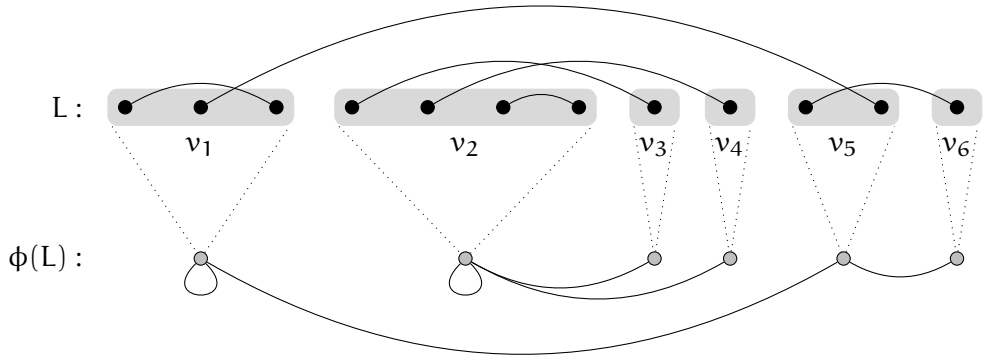


Figure 2.16: A LCD L and its corresponding graph $\phi(L)$. The edges of $\phi(L)$ are generated by joining the vertices corresponding to the two endpoints of a chord in L.

Theorem 2.5 (Bollobás and Riordan [BR04],Cohen and Havlin [CH03])**.** *For any fixed* $m \geqslant 2$ *and a real number* ε, w.h.p. $G_m^{(n)} \in \mathcal{G}_m^{(n)}$ *is connected and has a diameter satisfying*

$$(1-\varepsilon)\frac{\log n}{\log\log n} \leqslant \operatorname{diam}(G_m^{(n)}) \leqslant (1+\varepsilon)\frac{\log n}{\log\log n} \ .$$

OTHER PREFERENTIAL ATTACHMENT MODELS. Based on the vague description of the Barabási-Albert model and the formal description of the LCD model, a number of other preferential attachment models have been introduced. For example, Dorogovtsev, Mendes, and Samukhin [DMS00] and independently Drinea, Enachescu, and Mitzenmacher [DEM01] proposed a variation of the Barabási-Albert model with *initial attractiveness* which is a numerical constant that is added to the neighboring probability of a newly introduced node. This model was made precise along the lines of the LCD model by Buckley and Osthus [BO04] and hence is called the *Buckley-Osthus model*. Furthermore, Kumar et al. [Kum+00] proposed a model for the graph structure of the WWW where a new vertex is added to the graph by copying an already added vertex and changing some of its links. The links are dependent on a *copy factor* which determines if the link connects to a vertex chosen uniformly at random or if the link destination is copied from the prototype vertex. Hence, this model is also called the *"copying model"*.

All the above models are designed to explain how power law degree distributions arise. A general drawback of these evolving models is that they are intrinsically hard to analyze. For example, it is very hard to specify the probability of an edge in a generated graph instance, since every static instance of the evolving process is dependent on all previous steps. This is hindering, especially, when attempting to derive approximability results for combinatorial optimization problems for a family of power law graphs. More generally speaking, the difficulty lies in the task of deriving structural properties of the graph instances generated by these models. For this purpose, a second class of models for power law graphs was introduced where the target degree distribution is fully specified and fixed in the beginning.

2.5.4 *Static Models*

Motivated by the behavior of massive graphs derived from data in telecommunications, Aiello, Chung, and Lu [ACL00; ACL01] proposed a graph model that ensures a power law degree distribution by fixing a degree sequence via two parameters α, β and then to take the space of random graphs with this degree sequence. Thus their approach somehow complements the above evolving models in that it does not aim to explain how power laws arise, but—given that a graph has a power law degree sequence—allows to derive structural properties and statistical features which hold with asymptotically high probability (that is, a probability tending to 1 as the size of the graph goes to ∞). Furthermore, the derived results are true not only for certain instances of the random graph model, but for the majority of graphs with the given degree sequence. This model will be referred to as the $G(\alpha, \beta)$ *model* or *ACL model* for random power law graphs, and the corresponding graph class will be denoted as $\mathcal{G}_{\alpha,\beta}$. A graph or random element from this graph class is denoted by $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$. This graph model will be the underlying model for studies and the results of this thesis.

THE AIELLO-CHUNG-LU MODEL. In this section we describe in detail the random $G(\alpha, \beta)$ model proposed by Aiello, Chung, and Lu [ACL01]. The $G(\alpha, \beta)$ model considers a random graph with the following degree distribution and maximum degree

$\Delta = \left\lfloor e^{\alpha/\beta} \right\rfloor$: Depending on two given values α and β, for each $1 \leqslant i \leqslant \Delta$ there are y_i vertices of degree i with

$$y_i = \begin{cases} \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor & \text{if } i > 1 \text{ or } \sum_{i=1}^{\Delta} \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor \text{ is even,} \\ \lfloor e^\alpha \rfloor + 1 & \text{otherwise.} \end{cases}$$

Here, i and y_i satisfy $\log y_i = \alpha - \beta \log i$. The value of the parameter α corresponds to the logarithm of the size of the graph and the parameter β can be interpreted as the log-log growth rate.

Let $\mathcal{G}_{\alpha,\beta}$ be the family of all undirected graphs with multi-edges and self-loops on $n = \sum_{i=1}^{\Delta} y_i$ vertices which have y_i vertices of degree i ($1 \leqslant i \leqslant \Delta$). The number of vertices and edges of $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ can be computed as $n = \sum_{i=1}^{\Delta} \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor$ and $m = \frac{1}{2} \sum_{i=1}^{\Delta} \left\lfloor \frac{e^\alpha}{i^{\beta-1}} \right\rfloor$, respectively. As in [ACL01], in the following we will work with the real numbers $\frac{e^\alpha}{i^\beta}$, $e^{\alpha/\beta}$ instead of their integer counterparts when $\beta \geqslant 2$. In this case the error is a lower order term[3] in the size of the graph. When dealing with the parameter range $0 < \beta < 2$, we will give a detailed error term analysis whenever the above numbers are rounded down to integers.

According to [ACL00; ACL01], the parameters n, m, α and β are related roughly as follows:

$$n \approx \begin{cases} \zeta(\beta) e^\alpha & \text{if } \beta > 1 \\ \alpha e^\alpha & \text{if } \beta = 1 \\ \frac{e^{\frac{\alpha}{\beta}}}{1-\beta} & \text{if } 0 < \beta < 1 \end{cases} \quad \text{and} \quad m \approx \begin{cases} \frac{1}{2}\zeta(\beta-1) e^\alpha & \text{if } \beta > 2 \\ \frac{1}{4}\alpha e^\alpha & \text{if } \beta = 2 \\ \frac{1}{2}\frac{e^{\frac{2\alpha}{\beta}}}{2-\beta} & \text{if } 0 < \beta < 2 \end{cases}$$

In order to define the random graph model, we consider α to be large and β to be fixed and define that $G(\alpha, \beta)$ is the distribution on $\mathcal{G}_{\alpha,\beta}$ obtained in the following way, given a target d-degree sequence $d(G_{\alpha,\beta}) = (d(v_1), \ldots, d(v_n))$ (cf. [ACL01]):

1. Generate a set L of $d(v)$ distinct copies of each vertex v according to the degree sequence.

3 For example, when considering the number of vertices, the error is at most $e^{\alpha/\beta}$ which is in $o(n)$ (cf. [ACL01, p. 6]).

2. Generate a random matching on the elements of L.

3. For each pair of vertices u and v, the number of edges joining u and v in $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ is equal to the number of edges in the matching of L which join copies of u to copies of v.

Figure 2.17 shows the generation of edges, self-loops and multi-edges in the resulting graph $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ of the random model described above.

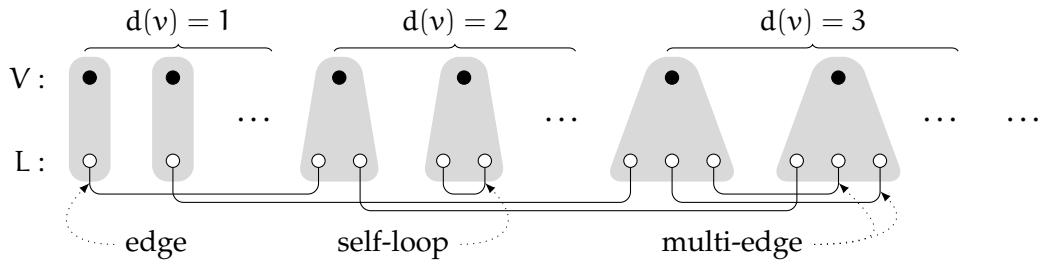


Figure 2.17: Generation of edges, self-loops and multi-edges via random matching in the random model for $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ as described by Aiello, Chung, and Lu [ACL01].

We conclude this section with a formal definition of a power law graph in the model described above.

Definition 2.1 ((α,β)-power law graph ((α,β)-PLG))**.**
A graph $G_{\alpha,\beta} = (V, E)$ is called a (α,β)-PLG where multi-edges and self-loops are allowed if the maximum degree is $\Delta = \lfloor e^{\alpha/\beta} \rfloor$ and the number of vertices of degree i is

$$y_i = \begin{cases} \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor & \text{if } i > 1 \text{ or } \sum_{i=1}^{\Delta} \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor \text{ is even} \\ \lfloor e^\alpha \rfloor + 1 & \text{otherwise.} \end{cases}$$

In order to point out the differences and similarities between the classical uniform random graphs and random power law graphs, in the following section we mention some previous results on structural properties of random power law graphs.

2.5.5 *Results on Static Power Law Graphs*

We start with some previous results on the component structure of graphs $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$, so called (α, β)-power law graphs.

CONNECTED COMPONENTS. In the context of random graphs with prescribed degree sequences, Molloy and Reed [MR95; MR98] investigated the existence of giant components for random graphs with a fixed number $(\lambda_i + o(1))n$ of vertices of degree i, where $\sum_i \lambda_i = 1$. They showed that a giant component emerges when $Q = \sum_{i \geqslant 1} i(i-2)\lambda_i > 0$ and provided that for the maximum degree Δ, $\Delta < n^{1/4-\varepsilon}$ holds. Moreover, a giant component almost surely (a.s.) does not exist for the case when $Q = \sum_{i \geqslant 1} i(i-2)\lambda_i < 0$ and $\Delta < n^{1/8-\varepsilon}$.

Aiello, Chung, and Lu [ACL01] computed the sum Q for the case of (α, β)-PLG as follows:

$$\begin{aligned} Q &= \sum_{i=1}^{e^{\alpha/\beta}} i(i-2) \left\lfloor \frac{e^{\alpha}}{i^{\beta}} \right\rfloor \\ &\approx \sum_{i=1}^{e^{\alpha/\beta}} \frac{e^{\alpha}}{i^{\beta-2}} - 2 \sum_{i=1}^{e^{\alpha/\beta}} \frac{e^{\alpha}}{i^{\beta}} \\ &\approx (\zeta(\beta-2) - \zeta(\beta-1))\, e^{\alpha} \text{ if } \beta > 3. \end{aligned}$$

As a critical point the value $\beta_0 \approx 3.478$ is considered, which is a solution to $\zeta(\beta-2) - \zeta(\beta-1) = 0$. Especially if $\beta > \beta_0$, we have for the above sum that $Q < 0$ and a.s. no giant component emerges.

Next, we summarize the results of Aiello, Chung, and Lu [ACL01] regarding the component structure of graphs $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ for different values of β.

Theorem 2.6 (Aiello, Chung, and Lu [ACL01])**.**
Let $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ be an (α, β)-PLG, then the following holds for the connectivity and the existence of a giant component of the graph:

1. *For $\beta > \beta_0 \approx 3.478$,* $G_{\alpha,\beta}$ a.s. *has no giant component. When $\beta < \beta_0 \approx 3.478$, there is* a.s. *a unique giant component.*

2. *For* $\beta = \beta_0 \approx 3.478$*, the situation corresponds to the* phase transition *of the connectivity of random graphs in the* $G(n,p)$ *model with* $p = 1/n$*, as described in Section 2.4 and Theorem 2.2.*

3. *For* $2 < \beta < \beta_0$*, the second largest components are* a.s. *of size* $\Theta(\log n)$*. For any* $2 \leqslant x < \Theta(\log n)$*, there is* a.s. *a component of size* x*.*

4. *When* $\beta = 2$*, the second largest components are* a.s. *of size* $\Theta\left(\frac{\log n}{\log\log n}\right)$*. For any* $2 \leqslant x < \Theta\left(\frac{\log n}{\log\log n}\right)$*, there is* a.s. *a component of size* x*.*

5. *For* $1 < \beta < 2$*, the second largest components are* a.s. *of size* $\Theta(1)$*. The graph is* a.s. *not connected.*

6. *For* $\beta = 1$*, there is a nontrivial probability for both cases that the graph is connected or disconnected.*

7. *For* $0 < \beta < 1$*, the graph is* a.s. *connected.*

DIAMETER. Another important graph invariant under study is the *diameter* of random power law graphs, that is, the length of a longest shortest path in the graph. Lu [Luo1] determined the diameter of random power law graphs in the $G(\alpha,\beta)$ model for all $\beta > 0$. The results are summarized in the following theorem:

Theorem 2.7 (Lu [Luo1])**.**
Let $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ *be a random power law graph in the* $G(\alpha,\beta)$ *model. For the diameter* $\operatorname{diam}(G_{\alpha,\beta})$ *the following holds:*

(i) For $0 < \beta < 2$*, almost surely* $\operatorname{diam}(G_{\alpha,\beta}) \leqslant 2\left\lfloor \frac{1}{2-\beta} \right\rfloor + 5$.

(ii) For $\beta > 2$*, almost surely* $\operatorname{diam}(G_{\alpha,\beta}) = \Theta(\log n)$.

The second result (ii) implies that there exist two constants c_1 and c_2 such that

$$c_1 \cdot \log n \leqslant \operatorname{diam}(G_{\alpha,\beta}) \leqslant c_2 \cdot \log n \ .$$

Currently there is a big gap between c_1 and c_2 and it is an open question whether the limit $\lim_{n\to\infty} \frac{\operatorname{diam}(G_{\alpha,\beta})}{\log n}$ exists and—if the answer is *yes*—what the value would be.

We finish the background chapter with some bibliographical notes and suggestions for further reading.

2.6 BIBLIOGRAPHIC NOTES

For further reading on the topics in Section 2.1—in particular regarding the asymptotic notation—we refer the reader to sections in the following textbooks: [Knu97, Section 1.2.11, pp. 107-123], [Sip97, Section 7.1, pp. 226-228] and [Cor+09, Section 3.1, pp. 41-50]. The definitions in Section 2.2 follow standard introductions to probability theory, e. g. found in the monograph by Feller [Fel71] or the book by Grinstead and Snell [GS98]. For a more detailed and complete introduction to graph theory and random graph theory (Section 2.3 and Section 2.4) see e. g. the monographs by Bollobás [Bol98] and Diestel [Die10].

In Section 2.4.2 we present some standard results in the context of random graphs. The proof of Theorem 2.1 on page 38 regarding the connectedness of random graphs in the $G(n, M)$ model can be found in [ER59]. The proof of Theorem 2.2 on page 39 regarding the phase transition of component sizes in the $G(n, p)$ model appeared in [ER60]. A reformulation of Theorem 2.3 on page 39 describing degree distributions in the $G(n, p)$ model is found in [BR05]. Regarding the diameter of regular random graphs, the proof of Theorem 2.4 on page 40 can be found in [BF82].

Section 2.5 introduces the concept of power law graphs along with the basic definitions and formal descriptions of generative model. For further reading, we recommend the survey articles by Mitzenmacher [Mit03]; Bollobás and Riordan [BR05]; Li et al. [Li+05]; Newman [New05]; Chung and Lu [CL06]; Simkin and Roychowdhury [SR11]. Moreover, we present some previous results in the context of power law graphs in Section 2.5.5. The proof of Theorem 2.5 on page 46 appeared in [BR04]; the proof of Theorem 2.6 on page 51 can be found in [ACL01] and the proof of Theorem 2.7 on page 52 can be found in [Lu01].

3 OPTIMIZATION AND APPROXIMABILITY

The main purpose of computational complexity theory is to classify computational problems according to their inherent difficulty. Especially this means to classify the problems into those that are efficiently solvable and to those that are not. In the context of computational resources, efficiently solvable means for a certain problem that there exists an algorithm which produces an exact solution and runs in deterministic polynomial time. The class P consists of all problems that are solvable in deterministic polynomial time, and while many computational problems are known to be in P, there are many other problems where one cannot decide whether they are in P or outside of P. Among the latter type of problems there are many which are known to be in the class NP, which is the class problems whose solutions can be verified in polynomial time.

In the following two sections we will introduce the notions and notations of computational decision and optimization problems along with the precise definition of the classes P and NP and certain subclasses of NP optimization problems. Then we will provide a set of techniques to prove the hardness of approximation of NP optimization problems.

3.1 DECISION PROBLEMS AND THE CLASSES P AND NP

We start with the definition of decision problems. Let Σ be a finite alphabet, we define a string s of length n over Σ as a mapping $s : [n] \to \Sigma$. A *decision problem* is a set $\Pi \subseteq \Sigma^*$ of strings over Σ. A string s is a *yes*-instance for Π if $s \in \Pi$ and a *no*-instance if $s \notin \Pi$. By $\overline{\Pi}$ we denote the set of all *no*-instances of Π. As in common notation, we denote by P the class of decision problems which are decidable in deterministic polynomial time and by NP the class of decision problems for which a non-deterministic polynomial-time bounded Turing Machine (TM) exists. Let $t(n)$ be an arbitrary function, we denote by $\mathsf{DTIME}(t(n)), \mathsf{NTIME}(t(n))$ the deterministic and non-deterministic *time complexity classes*, and by $\mathsf{DSPACE}(t(n)), \mathsf{NSPACE}(t(n))$ the deterministic and non-deterministic *space complexity classes*. We define

$$\mathsf{NP} := \bigcup_{t(n) \in n^{\mathcal{O}(1)}} \mathsf{NTIME}(t(n))$$

and

$$\mathsf{P} := \bigcup_{t(n)\in n^{\mathcal{O}(1)}} \mathsf{DTIME}(t(n)) \ .$$

Intuitively, NP is the class of all decision problems where the *yes*-instances have efficiently verifiable proofs. More precisely, these proofs are verifiable in polynomial time by a deterministic TM. The inclusion $\mathsf{P} \subseteq \mathsf{NP}$ is obvious, but if $\mathsf{NP} \subseteq \mathsf{P}$ holds is still—at the time of writing this thesis—a big open question.

In order to give a formal notion of probabilistic computation, we shortly introduce the concept of *probabilistic Turing Machines* (PTM). A probabilistic Turing Machine is a non-deterministic Turing Machine which randomly chooses between the available transitions at each computational step according to some probability distribution. As a consequence, a probabilistic Turing Machine has a stochastic output in form a random variable. Let $t(n)$ be an arbitrary function, we denote by $\mathsf{ZTIME}(t(n))$ the set of languages for which there is a PTM that always returns the correct answer and runs in expected polynomial time. We define class $\mathsf{ZPP} := \bigcup_{t(n)\in n^{\mathcal{O}(1)}} \mathsf{ZTIME}(t(n))$.

In order to define reductions among decision problems, let Σ be a finite alphabet and $\Pi, \Pi' \subseteq \Sigma^*$ be two decision problems. We say that Π is *polynomial-time reducible* to Π', $\Pi \leqslant_p \Pi'$, if and only if there exists a polynomial-time computable function $f : \Sigma^* \to \Sigma^*$ such that for all $x \in \Sigma^*$ we have that $x \in \Pi$ if and only if $f(x) \in \Pi'$. A decision problem Π is *NP-hard* if for all decision problems $\Pi' \in \mathsf{NP}$, $\Pi' \leqslant_p \Pi$. The decision problem Π is *NP-complete* if $\Pi \in \mathsf{NP}$ and, additionally, Π is NP-hard.

3.2 OPTIMIZATION PROBLEMS AND SUBCLASSES OF NPO

We give the definition of an optimization problem. An *optimization problem* is a tuple $(\mathcal{I}, \mathcal{F}, \text{cost}, \text{goal})$, where $\mathcal{I} \in \Sigma^*$ is the set of *instances*, $\mathcal{F}$ is a collection of sets $\mathcal{F}(I) \in \Sigma^* \times \Sigma^*$ of *feasible solutions* for each instance $I \in \mathcal{I}$, cost is a *cost function* $\text{cost} : \mathcal{F}(I) \to \Sigma^*$ assigning a cost $\text{cost}(F)$ to every feasible solution $F \in \mathcal{F}(I)$, and $\text{goal} \in \{\min, \max\}$ is the *optimization goal*. If $\text{goal} = \min$, we speak of a *minimization* problem, and if $\text{goal} = \max$, we speak of a *maximization* problem. For every instance

$I \in \mathcal{I}$, we call a feasible solution $F \in \mathcal{F}(I)$ *optimal* if $\mathrm{cost}(F) = \min_{F' \in \mathcal{F}(I)} \mathrm{cost}(F')$ or if $\mathrm{cost}(F) = \max_{F' \in \mathcal{F}(I)} \mathrm{cost}(F')$ for a minimization problem or a maximization problem, respectively. This optimal solution is denoted by $\mathsf{OPT}(I)$. The *optimal value* $\mathsf{opt}(I)$ of an instance $I \in \mathcal{I}$ is the cost of an optimal solution $\mathsf{OPT}(I)$ of the corresponding optimization problem. An optimization problem $\Pi = (\mathcal{I}, \mathcal{F}, \mathrm{cost}, \mathrm{goal})$ is an NP *optimization problem* if the following holds:

- The set of instances $\mathcal{I}$ is recognizable in deterministic polynomial time.
- For all $I \in \mathcal{I}$ and $F \in \mathcal{F}(I)$, there exists a polynomial p such that $|I| \leqslant p(|F|)$ and the question if $F \in \mathcal{F}(I)$ can be decided deterministically in time polynomial in $|F|$.
- For all $I \in \mathcal{I}$ and $F \in \mathcal{F}(I)$, the cost function $\mathrm{cost}(F)$ can be evaluated deterministically in time polynomial in $|F|$.

An optimization problem is a P *optimization problem* if for each instance $I \in \mathcal{I}$, an optimal solution can be computed in time polynomial in the size of the instance $|I|$. The class NPO is the set of all NP optimization problems, and the class PO is the set of all P optimization problems.

Let Π be an NP optimization problem. Given an instance $I \in \mathcal{I}$ and a feasible solution $F \in \mathcal{F}(I)$, we define the *performance ratio* of F with respect to I as

$$R(I, F) = \max \left\{ \frac{\mathrm{cost}(F)}{\mathsf{opt}(I)}, \frac{\mathsf{opt}(I)}{\mathrm{cost}(F)} \right\}.$$

We have that $R(I, F) \geqslant 1$ and $R(I, F)$ is close to 1 if F is close to the optimum solution $\mathsf{OPT}(I)$. Now let Π be an NP optimization problem and let $\mathcal{A}$ be an algorithm that, for any instance $I \in \mathcal{I}$ of Π, returns a feasible solution $F_{\mathcal{A}} \in \mathcal{F}(I)$. Given an arbitrary function $r : \mathcal{I} \to (1, \infty)$, we say that $\mathcal{A}$ is an $r(I)$-*approximate algorithm* for Π if for any instance $I \in \mathcal{I}$ a performance ratio of $R(I, F_{\mathcal{A}}) \leqslant r(I)$ is verified for every $F_{\mathcal{A}} \in \mathcal{F}(I)$. If $\mathcal{A}$ is a polynomial time $r(I)$-*approximate algorithm* for $\Pi \in \mathsf{NPO}$, then the ratio $R(I, F_{\mathcal{A}})$ is also called the *approximation ratio* of the *approximation algorithm* $\mathcal{A}$ and, equivalently, the problem $\Pi \in \mathsf{NPO}$ is said to be *approximable* within $r(I)$.

We define the following *complexity classes* for NP optimization problems $\Pi \in \mathsf{NPO}$ (i. e. subclasses of NPO):

- Π belongs to the class PO if it is optimally solvable within polynomial time; i. e. there exists an polynomial time algorithm $\mathcal{A}$ with $R(I, F_{\mathcal{A}}) = 1$ for Π.
- Π belongs to the class PTAS if it admits a *polynomial time approximation scheme*; i. e. there exists a polynomial time algorithm $\mathcal{A}$ such that for every instance I and every $n \in \mathbb{N}$, $\mathcal{A}(I, n)$ returns a feasible solution F with $R(I, F_{\mathcal{A}}) \leqslant 1 + 1/n$ and whose time complexity is bounded by a polynomial in $\mathcal{O}(|I|^{f(n)})$ for some $f : \mathbb{N} \to \mathbb{N}$.
- Π belongs to the class APX if it is approximable within c, for some constant $c \geqslant 1$; i. e. Π admits a polynomial time approximation algorithm $\mathcal{A}$ with approximation ratio c.

The inclusion PO $\subseteq$ PTAS $\subseteq$ APX $\subseteq$ NPO is obvious, and it is also clear that these inclusions are strict if P $\neq$ NP.

3.3 APPROXIMATION PRESERVING REDUCTIONS AND EMBEDDING REDUCTIONS

A number of inapproximability results for optimization problems were shown via reductions. For this purpose, several different kinds of reductions between optimization problems have been proposed (see e. g. [Cre97]). For our purposes it is sufficient to define the notion of *approximation-preserving reductions* among combinatorial optimization problems, so called AP-reductions.

Definition 3.1 (Approximation-Preserving Reduction)**.**
Let $\Pi = (\mathcal{I}, \mathcal{F}, \text{cost}, \text{goal})$ and $\Pi' = (\mathcal{I}', \mathcal{F}', \text{cost}', \text{goal}')$ be two NP optimization problems. We say that Π *AP-reduces* to Π', denoted by $\Pi \leqslant_{\text{AP}} \Pi'$, if there exist two functions $f : \{0,1\}^* \times \mathbb{Q}_+ \to \{01,\}^*$ and $g : \{0,1\}^* \times \{0,1\}^* \times \mathbb{Q}_+ \to \{01,\}^*$ and a constant $\alpha \geqslant 1$ such that for every fixed $\beta \in \mathbb{Q}_+$ the following properties hold:

1. For every instance $I \in \mathcal{I}$ of Π, then $f(I, \beta) = I' \in \mathcal{I}'$ and if $\mathcal{F}(I) \neq 0$, then $\mathcal{F}'(I') \neq 0$.
2. For all $F' \in \mathcal{F}'(I')$, we have $F = g(I, F', \beta) \in \mathcal{F}(I)$.

3. For every $I \in \mathcal{I}$ and $F' \in \mathcal{F}'(I')$, an approximation ratio $R'(I', F') \leqslant \beta$ with respect to Π' implies that $R(I, F) \leqslant 1 + \alpha \cdot (\beta - 1)$.

4. The functions g and f are computable in polynomial time.

An important property is that the classes APX and PTAS are closed under AP-reductions. That is, assuming that $\Pi \leqslant_{AP} \Pi'$, then $\Pi' \in APX$ implies $\Pi \in APX$ and $\Pi' \in PTAS$ implies $\Pi \in PTAS$. In order to define *complete* problems for some subclasses of NPO, let C be a class of NP optimization problems. An optimization problem Π is C-*hard* if for all optimization problems $\Pi' \in C$, $\Pi' \leqslant_{AP} \Pi$. The optimization problem Π is C-*complete* if $\Pi \in C$ and, additionally, Π is C-hard.

For the sake of completeness, we give also the definition of L-*reductions* which is due to Papadimitriou and Yannakakis [PY91].

Definition 3.2 (L-Reduction)**.**
Again, let $\Pi = (\mathcal{I}, \mathcal{F}, \text{cost}, \text{goal})$, $\Pi' = (\mathcal{I}', \mathcal{F}', \text{cost}', \text{goal}')$ be two NP optimization problems. We say that Π L-*reduces* to Π', denoted by $\Pi \leqslant_L \Pi'$, if there exist two functions $f : \{0,1\}^* \times \mathbb{Q}_+ \to \{01,\}^*$ and $g : \{0,1\}^* \times \{0,1\}^* \times \mathbb{Q}_+ \to \{0,1\}^*$ and constants $\alpha, \beta \geqslant 0$ such that for every instance $I \in \mathcal{I}$ of Π the following properties hold:

1. The function f produces an instance $I' = f(I) \in \mathcal{I}'$ of Π' with $\text{opt}'(I') \leqslant \alpha \cdot \text{opt}(I)$. And if $\mathcal{F}(I) \neq \emptyset$, then $\mathcal{F}'(I') \neq \emptyset$

2. For all $F' \in \mathcal{F}'(I')$, the function g produces a solution $F = g(I, F') \in \mathcal{F}(I)$ of Π with $|\text{cost}(F) - \text{opt}(I)| \leqslant \beta \cdot |\text{cost}'(F') - \text{opt}'(I')|$, where $|F'| \leqslant p(|I|)$ for some polynomial p.

3. The functions g and f are computable in polynomial time.

The main difference is that the class PTAS is closed also under L-reductions, but it is an open question whether the class APX is also closed under L-reductions. Crescenzi [Cre97] conjectured that this is not the case.

In Section 6.4.1, we will introduce a special kind of *embedded-approximation-preserving reductions* that are defined for combinatorial optimization problems on graphs that in-

herit a certain structural property. Reductions of this kind will enable us to transfer approximation hardness results for combinatorial optimization problems on general graphs to the case of power law graphs.

3.4 GAP PROBLEMS

In the main results of this thesis, we consider exclusively *optimization* problems, but, in order to characterize the computational complexity of a problem, it is more convenient to consider decision problems. The aim of this section is to show the connection of hardness of approximation of decision problems and the hardness of approximation of optimization problems. Therefore, we introduce the notion of so called *promise problems* and *gap problems*.

Promise problems are a natural generalization of decision problems, where only a subset of inputs are considered as valid instances. More precisely, in the formulation of a promise problem we are given the set of yes-instances and no-instances and do not care about inputs outside of the union of the two sets.

Let us start with the formal definition of a promise problem.

Definition 3.3 (Promise Problem)**.**
A *promise problem* Π is a pair $(\Pi_{\mathrm{yes}}, \Pi_{\mathrm{no}})$, where $\Pi_{\mathrm{yes}}, \Pi_{\mathrm{no}} \subseteq \{0,1\}^*$ and $\Pi_{\mathrm{yes}} \cap \Pi_{\mathrm{no}} = \emptyset$ such that all elements in Π_{yes} are accepted and all elements in Π_{no} are rejected. The set $\Pi_{\mathrm{yes}} \cup \Pi_{\mathrm{no}}$ is called the *promise*.

The elements of Π_{no} are called no-instances and the elements of Π_{yes} are called yes-instances. For the remaining inputs $\{0,1\}^* \setminus (\Pi_{\mathrm{yes}} \cup \Pi_{\mathrm{no}})$ there are no requirements on the output. Decision problems are a special case of promise problems, where $\Pi_{\mathrm{yes}} \cup \Pi_{\mathrm{no}} = \{0,1\}^*$.

Polynomial time reductions can easily be extended to promise problems Π, Π' by giving a polynomial-time computable mapping that maps yes-instances of Π to yes-instances Π' and no-instances of Π to no-instances of Π'. A promise problem $\Pi = (\Pi_{\mathrm{yes}}, \Pi_{\mathrm{no}})$ is *NP-hard* if all problems in NP can be polynomial time reduced to Π.

With this notion at hand, we can give an alternate characterization of the classes P and NP in terms of the decidability of membership of an input x in the sets Π_{yes} and Π_{no}. Consider the following definition of the class P:

Definition 3.4.
The promise problem $\Pi = (\Pi_{yes}, \Pi_{no})$ is in the class P if it is solvable in deterministic polynomial time, that is, there exists a polynomial time algorithm $\mathcal{A}$ such that the following conditions hold:

$$\begin{aligned} \text{Completeness:} \quad & x \in \Pi_{yes} \Rightarrow \mathcal{A}(x) = 1 \\ \text{Soundness:} \quad & x \in \Pi_{no} \Rightarrow \mathcal{A}(x) = 0 \ , \end{aligned}$$

For the characterization of the class NP the requirement of efficient verification is adapted in an analogous way. In this case the deciding algorithm acts as a *verifier* that, on input x and given a membership proof π, decides whether to accept or reject.

We give the following definition of the class NP:

Definition 3.5.
The promise problem $\Pi = (\Pi_{yes}, \Pi_{no})$ is in the class NP if there exists a polynomial p and a polynomial time algorithm V such that the following two conditions hold:

$$\begin{aligned} \text{Completeness:} \quad & x \in \Pi_{yes} \Rightarrow \exists \pi,\ V(x, \pi) = 1 \\ \text{Soundness:} \quad & x \in \Pi_{no} \Rightarrow \forall \pi,\ V(x, \pi) = 0 \ , \end{aligned}$$

where π is a proof of length at most $p(|x|)$.

In order to characterize the reducibility among promise problems, we say that a promise problem $\Pi = (\Pi_{yes}, \Pi_{no})$ is polynomial time reducible to $\Pi' = (\Pi'_{yes}, \Pi'_{no})$ if and only if there exists a polynomial time computable mapping $f\colon \{0,1\}^* \to \{0,1\}^*$ such that for every $x \in \Pi_{yes}$ ($x \in \Pi_{no}$) it holds that $f(x) \in \Pi_{yes}$ ($f(x) \in \Pi_{no}$). Inputs or queries that violate the promise $\Pi_{yes} \cup \Pi_{no}$ are not allowed.

Let us now introduce the notion of *gap problems*, which are promise decision problems intended to capture the difficulty of a corresponding (approximate) optimization problem. Gap problems are a special type of promise problems in which there exists

a *gap* between the yes-instances and no-instances. In case that the corresponding optimization problem Π is a maximization problem, there are yes-instance objects that have a relatively high optimum value and no-instance objects that have a relatively low optimum value. This gap is typically described with respect to two bounding functions $a, b\colon \{0,1\}^* \to \mathbb{R}$ that are computable in polynomial time.

Consider the following definition of $[a, b]$-gap problems:

Definition 3.6.
Let $\Pi = (\mathcal{I}, \mathcal{F}, \text{cost}, \text{goal})$ be an optimization problem and $a < b \in \mathbb{Q}_+$. The corresponding $[a, b]$-*gap problem* of Π is the following promise decision problem, denoted by $\textsc{Gap}\text{-}\Pi_{[a,b]} = (\Pi_{\text{yes}}, \Pi_{\text{no}})$:

$$\Pi_{\text{yes}} = \begin{cases} \{\, I \in \mathcal{I} \mid \text{opt}(I) \geqslant b \,\}, & \text{if goal} = \max \\ \{\, I \in \mathcal{I} \mid \text{opt}(I) < a \,\}, & \text{if goal} = \min, \end{cases}$$

and

$$\Pi_{\text{no}} = \begin{cases} \{\, I \in \mathcal{I} \mid \text{opt}(I) < a \,\}, & \text{if goal} = \max \\ \{\, I \in \mathcal{I} \mid \text{opt}(I) \geqslant b \,\}, & \text{if goal} = \min. \end{cases}$$

The following lemma shows the important connection between the hardness of approximation of decision problems and the hardness of approximation of optimization problems.

Lemma 3.1.
Let $\Pi_{[a,b]}$ be a promise decision problem corresponding to an optimization problem Π with $a < b$. If $\textsc{Gap}\text{-}\Pi_{[a,b]}$ is NP-hard, then there is no α-approximation algorithm for Π with $\alpha < b/a$, assuming $P \neq NP$.

We conclude the chapter with some bibliographic notes. In particular, we give some pointers to the corresponding sections of introductory textbooks, which deal with the topics presented in this chapter.

3.5 BIBLIOGRAPHIC NOTES

The notations and definitions of decision and optimization problems, presented in Section 3.1 and Section 3.2, can be found in a more complete context in the textbooks by Cormen et al. [Cor+09, Section 34.2, pp. 979-983] and Sipser [Sip12, Section 7.3-7.5, pp. 241-271].

Regarding the different kinds of reductions between optimization problems defined in Section 3.3, an introduction to approximation-preserving reductions can be found in [Cre97; Cre+99] and L-reductions were originally defined in [PY88; PY91]. A textbook covering these topics is by Ausiello et al. [Aus+99].

For a survey on promise problems with applications, we refer the reader to the survey of Goldreich [Gol06]. A definition and some examples for the use of gap problems are given in the monograph by Goldreich [Gol08, Section 10.1.1.2, pp. 446-449].

4 PROBABILISTIC PROOF SYSTEMS AND HARDNESS OF APPROXIMATION

To give a first intuition for the characterization of mathematical proofs and proof systems, let us quote from the textbook by Goldreich [Golo8][Chapter 9, p. 350]:

> The commonly agreed principles of reasoning are associated with a verification procedure that distinguishes proper applications of these principles from improper ones. A *line of reasoning* is considered valid with respect to such fixed principles (and is thus deemed a proof) if and only if it proceeds by proper applications of these principles. Thus, a line of reasoning is considered valid if and only if it is accepted by the corresponding verification procedure.

This concept is best illustrated by the formal notion and notation of a *proof system*. In the common notation of a proof system, the verification procedure is represented by a personification called the *verifier*. The proof is presented to the verifier in form of a string, which is provided by the so called *prover*. The verifier then examines the proof and decides if it is valid or not.

We will start with an introduction to *interactive proof systems*, where the reasoning between the verifier and the prover is interactive.

4.1 INTERACTIVE PROOFS

By *interactive* we mean that the verifier and the prover are deterministic functions that go through rounds of reasoning where the verifier asks a question to the prover and gets a response. The next question and response is computed as a function of the input based on the questions and responses of the previous rounds. For any $k \geqslant 0$, a k-*round interaction* of two deterministic functions $f, g : \{0,1\}^* \to \{0,1\}^*$ on input $x \in \{0,1\}^*$ is denoted by $\langle f, g\rangle(x)$ and is a sequence of strings defined as follows:

$$
\begin{aligned}
a_1 &= f(x) \\
a_2 &= g(x, a_1) \\
&\dots \\
a_{2i+1} &= f(x, a_1, \dots, a_{2i}) && \text{for } 2i < k \\
a_{2i+2} &= g(x, a_1, \dots, a_{2i+1}) && \text{for } 2i+1 < k\ .
\end{aligned}
$$

The output $f(x, a_1, \ldots, a_k) \in \{0,1\}$ at the end of the interaction between verifier and prover is denoted by $\mathrm{out}_f\langle f, g\rangle(x)$.

Let us first give an alternative definition of the class NP in terms of *deterministic proof systems*. This will serve as the canonical formulation of a proof system with efficient verification procedures.

Definition 4.1 (Deterministic proof systems)**.**
A language L is in NP or, equivalently, has a k-*round deterministic interactive proof system* for some $k \geqslant 0$ if and only if there exists a polynomial-time deterministic TM V (a *verifier*) that on input $x, a_1, \ldots, a_i$ can have a k-round interaction with any function P (a *prover*) such that the following two conditions hold:

$$\begin{aligned} \text{Completeness:} \quad & x \in L \Rightarrow \exists P : \{0,1\}^* \to \{0,1\}^*, \text{ s.t. } \mathrm{out}_V\langle V, P\rangle(x) = 1 \\ \text{Soundness:} \quad & x \notin L \Rightarrow \forall P : \{0,1\}^* \to \{0,1\}^*, \ \mathrm{out}_V\langle V, P\rangle(x) = 0 \ . \end{aligned}$$

Now we consider the case of a *probabilistic verifier* whose questions are computed using a probabilistic algorithm and whose conclusions are error-prone with some error probability. The probability of accepting a proof for a wrong statement can be non-zero but should be "small", whereas the rejection of a proof for a wrong statement is required to be with "good" probability, independent of the strategy used by the prover. We give the following definition of the class IP[k], that is, the set of languages that have a k-*round interactive proof system with probabilistic verifier*.

Definition 4.2 (Interactive proof systems)**.**
For any $k \geqslant 1$, a language L is in IP[k] if and only if there exists a probabilistic polynomial-time TM V that can have a k-round interaction with a function $P : \{0,1\}^* \to \{0,1\}^*$ such that the following two conditions hold:

$$\begin{aligned} \text{Completeness:} \quad & x \in L \Rightarrow \exists P : \{0,1\}^* \to \{0,1\}^*, \text{ s.t. } \mathrm{out}_V\langle V, P\rangle(x) = 2/3 \\ \text{Soundness:} \quad & x \notin L \Rightarrow \forall P : \{0,1\}^* \to \{0,1\}^*, \ \mathrm{out}_V\langle V, P\rangle(x) = 1/2 \ . \end{aligned}$$

The class IP is defined as $IP = \bigcup_{c \geqslant 1} IP[n^c]$.

Note here, that the probabilities $2/3$ and $1/2$ in the above definition can be replaced by probabilities arbitrary close to 1 and 0, respectively, by repeating the whole procedure several times in order to achieve the desired probability. The full characterization of the class IP was a long standing open question until the following result from 1992.

Theorem 4.1 (Shamir [Sha92])**.**

$$IP = PSPACE \ .$$

In other words, Theorem 4.1 states that the proofs that can be verified by an interactive probabilistic verifier in polynomial time are exactly those proofs that can be generated within *polynomial space*, so that the class of languages having interactive proof systems is PSPACE.

In order to assess the full power of interactive proof systems, Ben-Or et al. [Ben+88] defined so called *multi-prover interactive proof systems* with more that one prover. One important assumption is that the provers do not communicate once the verifier has send his first query. They may communicate with each other before the interaction with the verifier starts, especially they can agree upon a shared strategy. Furthermore, in each round of interaction, the verifier sends a query to each provers and each prover sends a response.

The set of languages with multi-prover interactive proof systems is called MIP and is defined analogously to definition 4.2. Note here that we may assume that there are exactly two provers in a multi-prover interactive proof system, since allowing a polynomial number of proves does not change the class.

Clearly, $\mathsf{IP} \subseteq \mathsf{MIP}$ due to the fact that the verier can always ignore one prover. Moreover, it turns out that the class MIP is strictly larger than IP unless $\mathsf{PSPACE} \neq \mathsf{NEXP}$. The following theorem due to Babai, Fortnow, and Lund [BFL91] states that that the class of languages having two-prover interactive proof systems is *nondeterministic exponential time*.

Theorem 4.2 (Babai, Fortnow, and Lund [BFL91])**.**

$$MIP = NEXPTIME \ .$$

This result paved the way for the celebrated PCP theorem, which can be seen as a "scaled-down" version of Theorem 4.2. In Section 7.4.3 we will analyze the one-round

multi-prover proof systems used by Feige in the proof of logarithmic lower bounds for the Set Cover problem (cf. [Fei98]) in order to proof a first logarithmic lower bound for the Minimum Dominating Set problem on power law graphs in Chapter 7.

4.2 PROBABILISTICALLY CHECKABLE PROOFS

This new notion of *probabilistically checkable proof systems* (PCP) coincides with the notion of multi-prover interactive proof systems for the case of only a one-round interaction between the verifier and provers.

Let us first give the definition of a restricted probabilistic verifier (PCP verifier) that has oracle access to a proof π over some alphabet Σ and has *randomness complexity* $r : \mathbb{N} \to \mathbb{N}$ and *query complexity* $q : \mathbb{N} \to \mathbb{N}$. That is, the verifier may use at most $r(n)$ random bits and queries at most $q(n)$ bits of the proof π.

Definition 4.3 (PCP Verifier)**.**
A $(r(n), q(n))$-*probabilistic verifier* is a probabilistic polynomial time TM V that acts on input $x \in \{0,1\}^n$, and has oracle access to a proof π over Σ.

(i) $r(n)$ denotes the maximal number of *random bits* used by V on of size n.

(ii) $q(n)$ denotes the maximal number of *queries* used by V on of size n.

We let $V^\pi(x)$ denote the random variable representing the output of V on input x and with random access to proof π.

A PCP verifier is said to be *non-adaptive* if it decides on all of its queries before it enters the query state and receives the answers of any of its queries.

Figure 4.1 shows a $(r(n), q(n))$-restricted probabilistic verifier in form of a probabilistic Turing Machine (PTM, cf. Section 3.1). The PTM may request the information of individual locations of the proof π via a $q(n)$ queries. With each query $q(i)$ it receives the corresponding bit $\pi_{q(i)}$ of the proof π. Moreover, the PTM has read-only access to a tape of length $r(n)$, which contains $r(n)$ random bits that may be used on a given input x of length n.

We are now ready to give the following definition of probabilistically checkable proofs with $(r(n), q(n))$-probabilistic verifiers.

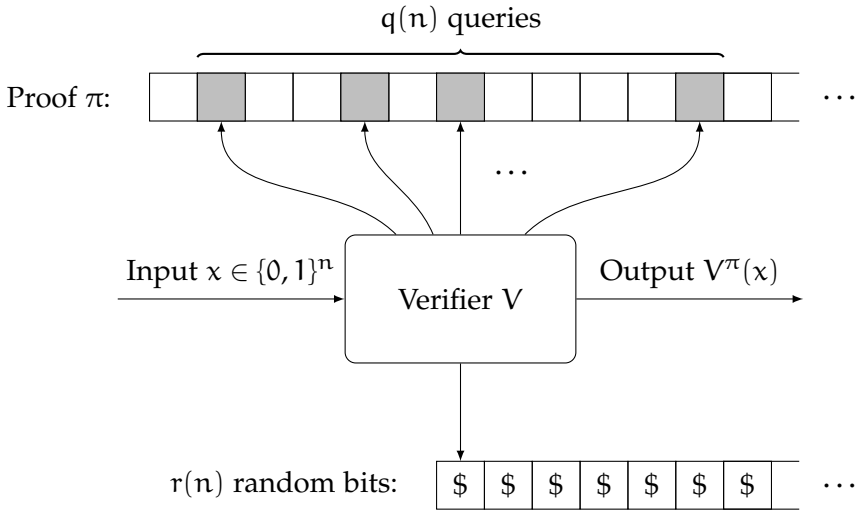


Figure 4.1: A $(r(n), q(n))$-restricted probabilistic verifier with access to $r(n)$ random bits and oracle access to a proof π via $q(n)$ queries.

Definition 4.4 (Probabilistically checkable proof systems)**.**
A language L is in the class $\mathsf{PCP}_{c(n),s(n)}(r(n), q(n))$ or, equivalently, has a *probabilistically checkable proof system* if there exists a non-adaptive $(r(n), q(n))$-probabilistic verifier that, on input x and oracle access to a proof π, satisfies the following properties:

$$\begin{aligned} \text{Completeness:} \quad & x \in L \Rightarrow \exists \pi, \text{ s.t. } \Pr[V^{\pi}(x) = 1] \geqslant c(n) \\ \text{Soundness:} \quad & x \notin L \Rightarrow \forall \pi, \Pr[V^{\pi}(x) = 1] \leqslant s(n) \ . \end{aligned}$$

The complexity class PCP is the class of languages that have efficient $(r(n), q(n))$-probabilistic verifiers and is defined as $\mathsf{PCP}_{1,1/2}(\log(n), \mathcal{O}(1))$. There exists a common shorthand notation for this class, i. e. $\mathsf{PCP}_{1,1/2}(\log(n), \mathcal{O}(1)) = \mathsf{PCP}(\log(n), \mathcal{O}(1))$. The famous PCP-theorem states that every language in NP has highly efficient $(r(n), q(n))$-probabilistic verifiers.

Theorem 4.3 (The PCP-Theorem [AS98; Aro+98])**.**

$$NP = PCP(\log(n), \mathcal{O}(1)) \ .$$

4.3 HARDNESS OF APPROXIMATION

One of the most important consequences of the PCP-theorem is that it shows approximation hardness for many NP optimization problems, i. e. that computing approximate solutions is no easier than computing exact solutions.

For constants $q \in \mathbb{N}$ and $\varepsilon > 0$, consider the gap problem Gap-Max-qSat$_{[1-\varepsilon,1]}$ which refers to instances that are a sequence of q-variable CNF formulas. The *yes*-instances are sequences that are simultaneously satisfiable, whereas the *no*-instances are sequences for which no assignment satisfies more than a $1-\varepsilon$ fraction of the clauses in the sequence. Arora et al. [Aro+98] proved the following

Theorem 4.4 (Arora et al. [Aro+98])**.**
The following three conditions are equivalent:

1. *There exists a constant* q *such that* $NP \subseteq PCP(\log(n), q)$ *(Theorem 4.3).*
2. *There exists a constant* q *such that* Gap-Max-qSat$_{[1/2,1]}$ *is NP-hard.*
3. *There exists a constant* $\varepsilon > 0$ *such that* Gap-Max-3Sat$_{[1-\varepsilon,1]}$ *is NP-hard.*

We conclude the chapter with some bibliographic notes.

4.4 BIBLIOGRAPHIC NOTES

For a comprehensive introduction to interactive proof systems, multi-prover proof systems and probabilistically checkable proof systems, we recommend the reader to the textbooks by Goldreich [Golo8, Chapter 9, pp. 349-415] and Arora and Barak [ABo9, Chapter 8, pp. 127-150; Chapter 11, pp. 205-220].

The proof of Theorem 4.1 on page 68 is due to Shamir [Sha92]. The proof of Theorem 4.2 on page 68 is due to Babai, Fortnow, and Lund [BFL91]. The original proofs of the PCP-theorem (Theorem 4.3 on page 70) as well as for Theorem 4.4 appeared in [AS98; Aro+98].

For a different and more simple proof of the PCP-theorem—using expander graphs and gap amplification techniques—we refer the interested reader to [Dino7; RSo7].

5 BOUNDED OCCURRENCE CSP AND HARDNESS OF APPROXIMATION

In this chapter we will focus on results regarding the hardness of approximation of optimization problems (or *constraint satisfaction problems* (CSP)) restricted to instances with bounded occurrences of variables or—in case of graph problems—restricted to instances of bounded degree. Especially, we will present a sequence of results that in the end proves hardness of approximation of the problems Max-IS and Min-VC.

5.1 THE MAX-E3LIN2 PROBLEM

Let us first introduce the notion of *constraint satisfaction problems* (CSP) which are a generalization of satisfiability problems, such as the 3Sat problem, that allow clauses of arbitrary form and may also depend on more than 3 variables.

Definition 5.1 (Constraint satisfaction problem (CSP))**.**
A CSP is defined as a tuple $\Pi = (X, D, C)$, where

- $X = \{x_1, \ldots, x_n\}$ is the set of variables or *domain variables*;
- $D = \{D_1, \ldots, D_n\}$ is the set of *domains*, containing a finite set of possible values for the corresponding variable;
- $C = \{C_1, \ldots, C_m\}$ is the set of *constraints*.

A state of the CSP is defined by an *assignment* $\sigma\colon X \to D$ that assigns to every variable a value from its domain. A constraint C_i which depends on a specific set of variables $x_{i_1}, \ldots, x_{i_k}$ is any subset $C_i \subseteq D_{i_1} \times \cdots \times D_{i_k}$ that specifies the combinations of values that are allowed by the constraint. The number k is called the *arity*[1] of the constraint. We say that an assignment σ satisfies constraint C_i if $C_i(\sigma) = 1$. By $\mathrm{val}(\Pi)$ we denote the maximum number of satisfied constraints over all σ, which can be defined as

$$\mathrm{val}(\Pi) = \max_{\sigma} \left(\frac{\sum_{i=1}^{m} C_i(\sigma)}{m} \right) .$$

1 If all constraints $C_i \in C$ of a CSP Π have arity at most k, we call Π a kCSP and denote it by kΠ. If all constraints have arity exactly k, we denote the problem by EkΠ.

We say that the CSP Π is *satisfiable* if $\mathrm{val}(\Pi) = 1$ and the corresponding assignment σ is called a *solution* for Π.

In order to prove the hardness of approximation of the problems MAX-IS and MIN-VC, we start with the definition of the CSP MAX-E3LIN2 which will be the starting point of the reduction.

Problem 1 (The MAX-E3LIN2 problem)**.**

INPUT: A system of linear equations $\mathcal{L}$ with n variables $\{x_1, \ldots, x_n\}$, and the equations over three variables are of the form $x_i \oplus x_j \oplus x_k = c_{ijk} \mod 2$.

OUTPUT: An assignment $\sigma : \{x_1, \ldots, x_n\} \to \{0, 1\}$.

OBJECTIVE: Maximize the number of satisfied equations in $\mathcal{L}$.

The hardness of the above problem is due to the famous result of Håstad [Hås01], which characterizes the class NP in terms of the existence of PCP verifiers with low query complexity (cf. Section 4.2). The result is summarized in the following theorem:

Theorem 5.1 (Håstad's 3-bit PCP)**.**
For every $\varepsilon, \eta > 0$ the class NP is in the class $PCP_{1-\varepsilon,1/2+\eta}(\mathcal{O}(\log n), 3)$. That is, for every language $L \in NP$ there exists a PCP verifier V that, on input x, uses $\mathcal{O}(\log n)$ random bits, queries exactly 3 bits of the proof π and evaluates a linear predicate on the 3 bits such that the following holds:

$$\begin{aligned} \textit{Completeness:} \quad & x \in L \Rightarrow \exists \pi, \textit{ s.t. } \Pr[V^\pi(x) = 1] \geqslant 1 - \varepsilon \\ \textit{Soundness:} \quad & x \notin L \Rightarrow \forall \pi, \Pr[V^\pi(x) = 1] \leqslant 1/2 + \eta \ . \end{aligned}$$

When viewing the bits of the proof π as boolean variables, then the tests of the verifier can be viewed as system of linear equations where each equation is of the form $x \oplus y \oplus z = b \mod 2$. From this characterization, we immediately obtain the hardness of approximation of the problem MAX-E3LIN2 as a corollary.

Corollary (due to Theorem 5.1)**.**
For every $\varepsilon, \eta > 0$ the problem GAP-MAX-E3LIN2$_{[1-\varepsilon,1/2+\eta]}$ *is NP-hard.*

By Lemma 3.1 we conclude that there exists no approximation algorithm for MAX-E3LIN2 with an approximation ratio less than $\frac{1-\varepsilon}{1/2+\eta} \approx 2$, unless $P = NP$. This result

is tight since a random assignment to the variables satisfies half of the equations in expectation and thus provides a trivial 2-approximation.

Now we will focus on the setting of bounded occurrence CSP and the results presented by Berman and Karpinski [BK99; BK01; BK03]. In order to achieve the results, Berman and Karpinski constructed special instances of the problem MAX-E3LIN2 with a mixed set of equations over two and three variables and, moreover, a bounded occurrence of variables in the system of linear equations. Because of the occurrence of equations of different lengths, this variant of the 3OCC-MAX-E3LIN2 problem will be called HYBRID.

Let us first describe the so called r-*regular amplifiers*, which are the basic elements in the construction of the restricted instances of the HYBRID problem.

5.2 AMPLIFIERS

In [PY91] Papadimitriou and Yannakakis introduced the concept of *amplifier* graphs, that generalize a specific class of *expander* graphs that are used in proving inapproximability results (see also [Pap94]). We present here the definition used by Berman and Karpinski [BK99] to construct the bounded occurrence instances of the HYBRID problem.

Definition 5.2 (Amplifier)**.**
Let $G = (V, E)$ be a graph and $X \subset V$ a subset of vertices.

- G is an *amplifier* for a vertex set $X \subset V$ if it contains no *bad sets* for X, that is, for all $A \subset V$, we have for the cut-size of the set A, $|\mathrm{cut}(A)| < \min\{|X \cap A|, |X - A|\}$.
- An amplifier G for $X \subset V$ is called r-*regular* if for each vertex $u \in X$, $N(u) = r - 1$, and for each $v \in V \setminus X$, $N(v) = r$.
- Let G be a r-regular amplifier for $X \subset V$, then the vertices $u \in X$ are called *contact vertices* and the vertices $v \in V \setminus X$ are called *checker vertices*.

Figure 5.1 shows amplifier constructions from the definition above for various values of $|X|$.

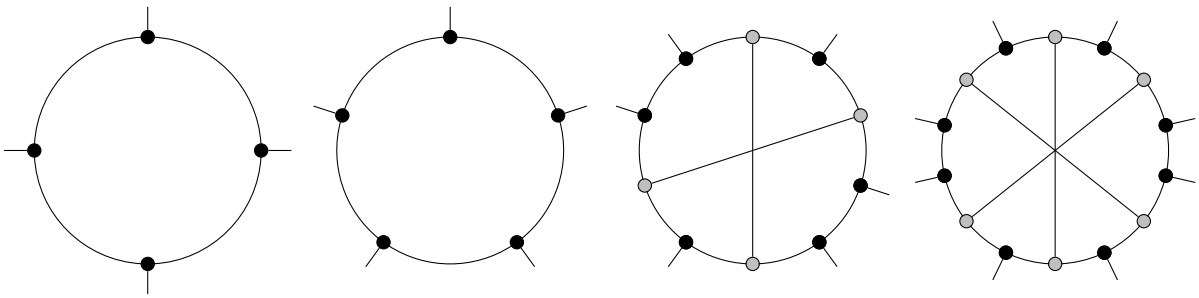


Figure 5.1: Examples of 3-regular amplifiers for $|X| = 4, 5, 6, 8$, where the black nodes (●) denote checker vertices in X and the gray nodes (○) denote contact vertices in $V \setminus X$.

The following theorem shows a result by Berman and Karpinski on the existence and constructibility of 3-regular amplifiers.

Theorem 5.2 (Berman and Karpinski [BK99])**.**
For a set of n vertices, a 3-regular amplifier with $7n$ nodes and $10n$ edges can be constructed in random linear time.

The amplifiers constructed in Theorem 5.2 have a special structure and are therefore also customary described as *wheels*. Figure 5.2 shows an example construction for $n = |X| = 8$. Note that the edges are added in such a way that they constitute a *Hamiltonian cycle* as the rim of the wheel and a *perfect matching* among all the checker vertices.

Now we introduce the Hybrid problem and state a theorem regarding the hardness of approximation of a bounded occurrence variant of this problem.

5.3 THE HYBRID PROBLEM

In [BK99] (see also [BK01; BK03]), Berman and Karpinski introduced the so called Hybrid *problem* in order to achieve new lower bounds for bounded occurrence CSP. The Hybrid problem can be seen as a variant of the CSP Max-E3Lin2 with constraints over two and three variables. Let us first give a formal definition of the Hybrid problem.

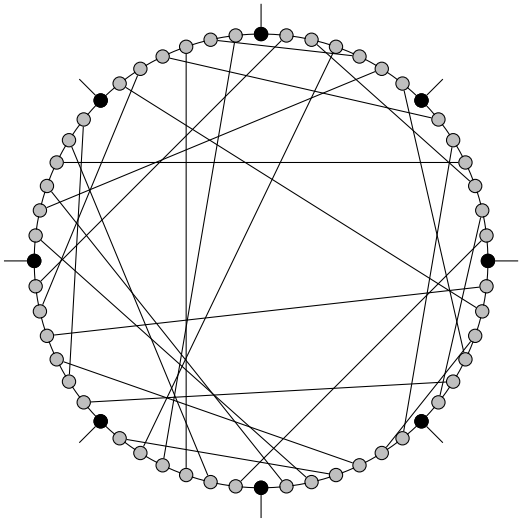


Figure 5.2: A 3-regular amplifier for $|X| = 8$ with $7n$ nodes and $10n$ edges, constructed by the means of Theorem 5.2. The black nodes (●) denote the checker vertices and the gray nodes (○) denote the contact vertices.

Problem 2 (The HYBRID problem)**.**

INPUT: A system of linear equations $\mathcal{L}$ with n variables $\{x_1, \ldots, x_n\}$, m_2 equations over two variables, and m_3 equations over three variables.

OUTPUT: An assignment $\sigma : \{x_1, \ldots, x_n\} \to \{0, 1\}$.

OBJECTIVE: Maximize the number of satisfied equations in $\mathcal{L}$.

Now, to introduce the bounded occurrences of variables, Berman and Karpinski constructed special instances of HYBRID which are similar to instances of the 3OCC-MAX-E3LIN2 problem but again with a mixed set of equations as in the definition above. For these bounded occurrence instances of HYBRID, Berman and Karpinski proved the following hardness result.

Theorem 5.3 (Berman and Karpinski [BK99])**.**
For every constant $\varepsilon > 0$, there exist bounded occurrence instances of HYBRID *with $42n$ variables, $m_2 = 60n$ equations over two variables, and $m_3 = 2n$ equations over three variables such that each variable occurs exactly 3 times and it is NP-hard to decide which of the following two conditions holds:*

(i) There exists an assignment σ *that satisfies at least* $(62-\varepsilon)n$ *equations;*

(ii) Every assignment σ *satisfies at most* $(61+\varepsilon)n$ *equations.*

The result is obtained via a polynomial time reduction from the MAX-E3LIN2 problem to the HYBRID problem. Let us give an outline of the proof of Theorem 5.3:

We start from an instance of MAX-E3LIN2 with $2n$ equations. By Theorem 5.1 we know that it is NP-hard to distinguish instances where $(2-\varepsilon)n$ equations can be satisfied from instances where at most $(1+\eta)n$ equations can be satisfied. We replace this instance MAX-E3LIN2 by an instance of the same problem where each variable occurs in at least n equations. This is done by replicating the original equations in an appropriate way. Then each variable x_i is replaced by a r-*regular amplifier* W_i (cf. Section 5.2), where r is the new number of occurrences of the variable. That is, given that a variable occurred m times, we replace the original occurrences of the variable by a graph of newly created *contact variables/vertices* and then add $6m$ so called *checker variables/vertices*. To the original $2n$ equations of length 3, we add for the remaining checker variables new equations of the form $x^i_j \oplus x^i_k = 0$. More precisely, for each amplifier W_i, the equations are formed by choosing a random bipartite matching on each chain of 6 checker vertices separating two contact vertices.

This yields a new system of equations $\mathcal{L}$ where each variable occurs exactly three times. Moreover, $\mathcal{L}$ consists of $2n$ equations of length 3 and $60n$ equations of length 2. This is the exact characterization of an instance of the HYBRID problem as described in the initial setting of Theorem 5.3. Moreover, due to Theorem 5.2, the r-regular amplifiers can be constructed efficiently and, together with the result of Theorem 5.1, the theorem follows.

The inapproximability result of Theorem 5.3 served as a starting point for many other results regarding bounded occurrence CSP and problems on bounded degree graph instances. We are especially interested in the results that regard the bounded degree variants of MAX-IS and MIN-VC, which will be presented in the following.

5.4 HARDNESS OF APPROXIMATION

A number inapproximability results were proved by Berman and Karpinski via reductions that involved the bounded occurrence HYBRID instances described above. We

will present here the results for MAX-IS and MIN-VC for degree d-bounded instances and $d \in \{3,4,5\}$. Let d-MAX-IS (d-MIN-VC) denote instances of MAX-IS (MIN-VC) on degree d-bounded graphs. The following theorem is due to Berman and Karpinski [BK03].

Theorem 5.4 (Inapproximability of d-MAX-IS ([BK03]))**.**
For every $\varepsilon > 0$, the following holds:

- 3-MAX-IS *is NP-hard to approximate within* $\frac{98}{97} - \varepsilon$;
- 4-MAX-IS *is NP-hard to approximate within* $\frac{50}{49} - \varepsilon$;
- 5-MAX-IS *is NP-hard to approximate within* $\frac{50}{49} - \varepsilon$.

More precisely, Berman and Karpinski showed that for every $\varepsilon > 0$, it is NP-hard to decide whether an instance of 3-MAX-IS on $200n$ nodes has an independent set of size at least $(98 - \varepsilon)n$ or at most $(97 + \varepsilon)n$ and whether an instance of 4-MAX-IS on $104n$ nodes has an independent set of size at least $(50 - \varepsilon)n$ or at most $(49 + \varepsilon)n$. The result is obtained via polynomial time reductions from the MAX-E3LIN2 problem to the HYBRID problem and further to the d-MAX-IS problem. Let us give an outline of the proof for the case of 4-MAX-IS:

We start from an instance of MAX-E3LIN2 and construct the corresponding HYBRID instance using r-regular amplifiers as in the proof of Theorem 5.3. Given this initial situation, we construct a 4-MAX-IS instance G based on the graph of the HYBRID instance $\mathcal{L}$. For each variable x in $\mathcal{L}$ that corresponds to an amplifier node, we create an edge $e_x = \{x_0, x_1\}$ with endpoints x_0 and x_1 corresponding to the possible values of x. For each amplifier edge $\{x, y\}$ in $\mathcal{L}$, we connect the endpoints of the edges e_x and e_y via edges $\{x_0, y_1\}$ and $\{x_1, y_0\}$. For every contact variable x, we also create an additional *direct contact* $d_x = \{\check{x}_0, \check{x}_1\}$ that is connected to the endpoints of e_x via edges $\{x_0, \check{x}_1\}$ and $\{x_1, \check{x}_0\}$.

Now let c be an equation in $\mathcal{L}$ of the form $x \oplus y \oplus z = b$. For each c we form an equation quadruple q_c of nodes, that is, a set of four nodes that correspond the four assignments of variables that yield one of the two outcomes of b. The nodes are labeled $\alpha\beta\gamma$, which corresponds to an equation $\alpha \oplus \beta \oplus \gamma = b \mod 2$. Each node is then connected to the direct contact endpoints that correspond to their negated

variable values, i. e. $\check{x}_{\neg\alpha}, \check{y}_{\neg\beta}$ and $\check{z}_{\neg\gamma}$. Additionally, the four nodes of q_c are connected via two arbitrary disjoint edges. Now, for every equation c of the form $x \oplus y \oplus z = b$ there is a corresponding subgraph $A_c = q_c \cup d_x \cup d_y \cup d_z$. Now let $J_{\alpha\beta\gamma} = \{\check{x}_\alpha, \check{y}_\beta, \check{z}_\gamma\}$ and let $I_{\alpha\beta\gamma} = J_{\alpha\beta\gamma} \cup \{\text{the node with label } \alpha\beta\gamma\}$. Berman and Karpinski showed that every maximum independent set A_c is of the form $I_{\alpha\beta\gamma}$. Moreover, an independent set I in the graph G can be normalized in such a way that every pair has exactly one element in I and an equation quadruple contains an element of I if and only if the corresponding equation is satisfied by the values given to the variables.

In this way, the HYBRID instance $\mathcal{L}$ with $6n$ equations is transformed into a graph G that has $8 \cdot 6n$ pairs, corresponding to equations of length 2, and $2n$ q-quadruples, corresponding to equations of length 3. Altogether, we transformed the original instance of MAX-E3LIN2 over $2n$ equations into an instance of 4-MAX-IS over $50n$ equations and, together with the corollary of Theorem 5.1, the theorem follows.

The MAX-IS instance G with maximum degree 4 described above can be transformed into an instance with maximum degree 5. Therefore, the same inapproximability result holds for the case of 5-MAX-IS. The proof for 3-MAX-IS is along the same lines as the outline presented above but needs a different construction of the subgraphs corresponding to the equations of length 2 and 3 in the HYBRID instance. Because of the complementary relationship between the two problems MAX-IS and MIN-VC, we immediately get the following corollary of Theorem 5.4.

Corollary (due to Theorem 5.4)**.**
For every $\varepsilon > 0$, the following holds:

- 3-MIN-VC *is NP-hard to approximate within* $\frac{103}{102} - \varepsilon$;
- 4-MIN-VC *is NP-hard to approximate within* $\frac{55}{54} - \varepsilon$;
- 5-MIN-VC *is NP-hard to approximate within* $\frac{55}{54} - \varepsilon$.

In the following Chapter 6, we will use the above results for d-MIN-VC to obtain new lower bounds for the MIN-VC problem on power law graphs. We conclude the present chapter with some bibliographic notes.

5.5 BIBLIOGRAPHIC NOTES

The material presented in this chapter is based on publications by Berman and Karpinski [BK99; BK01; BK03]. The formulation of the HYBRID problem and the full proof of the hardness of approximation result in Theorem 5.3 on page 78 is presented in [BK99]. The concept of amplifier graphs is due to Papadimitriou and Yannakakis [PY91], whereas the formal definition on page Definition 5.2 and the proof of Theorem 5.2 regarding the existence and constructibility of regular amplifiers appeared in [BK99].

Regarding the inapproximability of the MAX-IS problem and the MIN-VC problem on degree d-bounded graphs with $d \in \{3, 4, 5\}$, a series of improvements appeared in [BK99; BK01; BK03]. The explicit approximation lower bounds presented in Theorem 5.4 on page 80 are due to [BK03].

6 APPROXIMATION LOWER BOUNDS OF MINIMUM VERTEX COVER IN POWER LAW GRAPHS

6.1 INTRODUCTION

The MINIMUM VERTEX COVER (MIN-VC) problem is one of the most well-studied problems in combinatorial optimization. A *vertex cover* of a graph $G = (V, E)$ is a set of vertices $C \subseteq V$ such that each edge $e = \{u, v\}$ of G has at least one endpoint in C. The MIN-VC problem is the problem of finding a vertex cover of minimum cardinality in a graph. The MIN-VC problem is known to be NP-hard due to Karp's original proof [Kar72] and APX-complete due to Papadimitriou and Yannakakis [PY91], which implies that the problem is hard to approximate to within some constant factor.

A 2-approximation on general graphs is relatively easy to achieve by computing a *maximal matching* and taking the endpoints of each of its edges into the vertex cover. Since the maximal matching touches all vertices and each edge of the graph needs at least one of its endpoints in the vertex cover, a 2-approximative solution is obtained. Furthermore, there exits strong evidence that this approximation ratio is best possible in general graphs (see Section 6.2).

In the context of real-world networks, on the other hand, there exists practical evidence that the MIN-VC problem and other combinatorial optimization problems are easier to solve. For example, Park and Lee [PL01] showed for the efficient placement of filters in *route-based distributed packet filtering* on power-law Internet topologies, that the greedy heuristic for the MIN-VC problem generally outperforms the constant-factor approximation algorithm. A natural question now arises, whether there exist any provable guarantees that MIN-VC is indeed easier to solve on general power law graphs. In particular, one wants to know if there exists a PTAS in this case, that is, an approximation algorithm with approximation ratio arbitrarily close to 1.

In this chapter, we will somehow complement the above practical findings for the more general class of power law graphs, and show that the MIN-VC problem cannot be approximated with a ratio arbitrarily close to 1. To be more precise, we will show the APX-hardness of MIN-VC in *connected* (α, β)-PLG multigraphs for $0 < \beta < \beta_{\max} \approx 2.48$ and therefore rule out the existence of a PTAS. Furthermore, we give explicit approximation lower bounds for this problem.

Note here that in Chapter 8 we will construct an approximation algorithm for the MIN-VC problem with an expected approximation ratio of $2 - f(\beta)$ for *random* (α, β)-*PLG*, where $f(\beta)$ is a strictly positive function of the model parameter β. Here $f(\beta)$

does not depend on the size $|V|$ of the graph and thus—for large graph sizes—our approximation ratio falls below current upper bounds for MIN-VC in general graphs (which is $2 - \Theta(1/\sqrt{\log n})$ as stated by Karakostas [Kar09]).

6.2 PREVIOUS RESULTS

In this section we will summarize previous results on the approximability and hardness of approximation of MIN-VC on general graphs and power law graphs. Before we start, let us first give the formal definition of the MIN-VC problem.

Problem 3 (MINIMUM VERTEX COVER (MIN-VC))**.**

INPUT: A graph $G = (V, E)$.

OUTPUT: A *vertex cover* of G, i. e. a set $C \subseteq V$ such that each edge $\{u, v\} \in E$ has at least one endpoint in C.

OBJECTIVE: Minimize $|C|$.

APPROXIMABILITY AND HARDNESS IN GENERAL GRAPHS. The MIN-VC problem is known to be NP-complete due to Karp's original proof [Kar72] and APX-complete due to Papadimitriou and Yannakakis [PY91], which implies that the problem is hard to approximate to within some constant factor. Moreover, the problem is closely related to the MAXIMUM INDEPENDENT SET problem, which can be stated as follows:

Problem 4 (MAXIMUM INDEPENDENT SET (MAX-IS))**.**

INPUT: A graph $G = (V, E)$.

OUTPUT: An *independent set* of G, i. e. a set $I \subseteq V$ such that each edge $\{u, v\} \in E$ has at most one endpoint in I.

OBJECTIVE: Maximize $|I|$.

As we see from the above definitions, for every graph G, a solution of the MIN-VC problem is the complement of a solution to the MAX-IS problem. However, the

behavior of the two problems in terms of the approximability and approximation ratios is rather different.

For the MIN-VC problem, the value for the approximation ratio is basically at most 2. A ratio of $2 - \frac{\ln \ln n}{2 \ln n}$ has been found independently by Monien and Speckenmeyer [MS85] and Bar-Yehuda and Even [BE85] and an improvement to $2 - \Theta(1/\sqrt{\log n})$ is due to Karakostas [Kar09]. In the case of MAX-IS the value is at least $n^{1-\varepsilon}$ as shown by Håstad [Hås99], which has been further improved to $\frac{n}{2^{(\log n)^{3/4+\varepsilon}}}$ by Khot and Ponnuswami [KP06]. As mentioned above, a simple 2-approximation can be obtained by computing a *maximal matching* and taking all the endpoints of the edges this matching into the cover.

Regarding approximation lower bounds, Håstad [Hås01] showed that the general MIN-VC problem cannot be approximated within a factor of $7/6$, assuming $\mathsf{P} \neq \mathsf{NP}$. Under the same assumption, Dinur and Safra [DS05] improved the inapproximability factor to 1.3606. Moreover, Khot and Regev [KR08] proved inapproximability within $2 - \varepsilon$ for any $\varepsilon > 0$ as long as the Unique Games Conjecture (UGC) holds true. The conjecture is based on statements regarding the power of *unique 2-prover-1-round games* presented in [Kho02]. This means that in the general case the achieved approximation ratios are conjectured to be essentially best possible and hence the simple constant upper bound cannot be improved.

For this reason MIN-VC has been considered for certain restricted problem instances, along with properties that allow for improved performance guarantees for these restricted instances. Since our reductions in this chapter will map from bounded degree MINIMUM VERTEX COVER (d-MIN-VC) to MIN-VC on power law graphs, we present here some previous results regarding problem instances where the maximum degree is bounded by a constant.

APPROXIMABILITY AND HARDNESS IN BOUNDED DEGREE GRAPHS. For the case of degree d-bounded graphs, Hochbaum [Hoc83] presented a $2 - 2/d$ approximation algorithm, which was improved to $2 - \frac{\log d + \mathcal{O}(1)}{d}$ by Halldórsson and Radhakrishnan [HR94] and further improved to $2 - (1 - o(1))\frac{2 \ln \ln d}{\ln d}$ by Halperin [Hal02]. The result of Halperin also essentially matches the lower bound $2 - (2 + o_d(1))\frac{\log \log d}{\log d}$ due to Austrin, Khot, and Safra [AKS09] for large values of d. Like in the general case, this lower bound is dependent on the UGC.

When establishing new approximation lower bounds for the MIN-VC problem on power law graphs, we are also interested in explicit lower bounds for the cases of a small degree bounds, i. e. $d = 3, 4, 5$. For this reason, in Table 6.1, we list here some of the results due to Berman and Karpinski [BK99; BK03] for d-MAX-IS and d-MIN-VC. In Chapter 5 we presented the general technique that is used to obtain the above lower

Table 6.1: Inapproximability factors for d-MAX-IS and d-MIN-VC in degree d-bounded graphs for $d = 3, 4, 5$ due to Berman and Karpinski [BK99; BK03].

	d-MAX-IS lower bound	d-MIN-VC lower bound
$d = 3$	$98/97$	$103/102$
$d = 4$	$50/49$	$55/54$
$d = 5$	$50/49$	$55/54$

bounds. In Section 6.6 and Section 6.7 we will apply these explicit lower bounds in order to establish new inapproximability results for MIN-VC in power law graphs.

APPROXIMABILITY AND HARDNESS IN POWER LAW GRAPHS. In a series of papers Ferrante, Pandurangan, and Park [FPP08] and Shen, Nguyen, and Thai [SNT10; She+12] studied the approximation hardness of classical optimization problems such as MIN-VC, MAX-IS and MINIMUM DOMINATING SET (MIN-DS) in *combinatorial power law graphs* in the ACL model and showed NP-hardness and APX-hardness for simple (α, β)-PLG and (α, β)-PLG multigraphs, respectively.

In Table 6.2 we list some of the main results of [She+12], in particular the previously best approximation lower bounds for MIN-VC and MIN-DS in (α, β)-PLG and (α, β)-PLG multigraphs for $\beta > 1$.

Let us now give an overview of our contributions and new results of the present chapter, regarding the classification of MIN-VC in power law graphs.

6.3 OVERVIEW AND OUR RESULTS

In this chapter, we show the APX-hardness of MIN-VC in *connected* (α, β)-PLG multigraphs for $0 < \beta < \beta_{\max} \approx 2.48$ and give explicit approximation lower bounds for this

Table 6.2: Previously known lower bounds due to Shen et al. [She+12] for the inapproximability of MIN-DS under condition $P \neq NP$ and for MIN-VC under UGC on disconnected (α, β)-PLG with $\beta > 1$.

Problem	(α, β)-PLG multigraphs	(α, β)-PLG
MAX-IS	$1 + \frac{1}{140(2\zeta(\beta)3^{\beta}-1)} - \varepsilon$	$1 + \frac{1}{1120\zeta(\beta)3^{\beta}} - \varepsilon$
MIN-DS	$1 + \frac{1}{390(2\zeta(\beta)3^{\beta}-1)}$	$1 + \frac{1}{3120\zeta(\beta)3^{\beta}}$
MIN-VC	$1 + \frac{2\left(1-(2+o_c(1))\frac{\log\log c}{\log c}\right)}{\left(\zeta(\beta)c^{\beta}+c^{\frac{1}{\beta}}\right)(c-1)}$	$1 + \frac{2-(2+o_c(1))\frac{\log\log c}{\log c}}{2\zeta(\beta)c^{\beta}(c+1)}$

problem. For $\beta > \beta_{max}$, (α, β)-PLG are not connected anymore. As in [She+12], our reductions consist of multigraph embeddings of bounded degree graphs into (α, β)-PLG, based on appropriate wheel constructions that embeds any graph bounded degree graph G_d into a connected power law multigraph $G_{\alpha,\beta}$. We also extend the model of (α, β)-PLG to a functional case and consider degree distributions where β is of the form $\beta_f = 1 \pm 1/f(n)$ for a sufficiently fast growing function $f(n)$. These distributions converge to those of (α, β)-PLG for $\beta = 1$ and can be seen as a combinatorial variant of the evolving models for power law graphs introduced in Section 2.5.3. We show that our new inapproximability results for the case $\beta = 1$ in the original model also hold for this functional model.

ORGANIZATION OF THE CHAPTER. The chapter is organized as follows. The embedding reduction of Shen et al. [She+12] will be presented in detail in Section 6.4. In Section 6.5 an outline of our methodology for the reduction is presented, i. e. we describe an algorithm for the wheel construction of a general embedding technique of d-bounded graphs into connected (α, β)-PLG. In Section 6.6 we give the detailed description of our reduction from MIN-VC in d-bounded degree graphs to MIN-VC in connected (α, β)-PLG for the parameter β in the interval $\beta \in (1, \beta_{max}]$. Section 6.7 deals with the case $\beta \in (0, 1]$ and gives the details of the reduction of MIN-VC in d-bounded degree graphs to MIN-VC in d-bounded degree graphs which provide a *perfect matching* and then further to MIN-VC in connected (α, β)-PLG. Furthermore, we give a thorough error term analysis for this case.

Figure 6.1 shows the global organization of the chapter, pointing to the different ranges and the phase transitions dependent on the model parameter β.

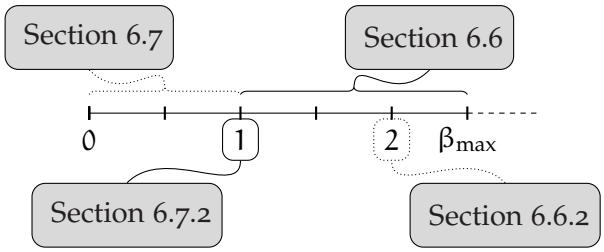


Figure 6.1: Chapter guide with respect to the phase transitions and different ranges of the model parameter β.

For all the above cases, we prove explicit lower bounds for the approximability of MIN-VC in (α, β)-PLG which only depend on the degree bound d, the parameter β and on the lower bounds ε_d for d-bounded MIN-VC. The resulting inapproximability factors are summarized in Table 6.3.

Table 6.3: Inapproximability factors of MIN-VC in connected (α, β)-PLG multigraphs for the three half-open intervals between 0 and β_{max}.

	Inapproximability factors
$\beta \in (0, 1]$	$1 + \frac{\varepsilon_d}{1+2d}$
$\beta \in (1, 2]$	$1 + \frac{\varepsilon_d}{1+d\cdot\frac{(\zeta(\beta)-1)\cdot(d+1)^\beta-1}{2}}$
$\beta \in (2, \beta_{max})$	$1 + \frac{\varepsilon_d}{1+d(d+1)^\beta\left(\frac{1}{2^{\beta+1}}+\zeta(\beta)-1-\frac{1}{2^\beta}-\frac{1}{(d+1)^\beta}\right)}$

In order to show the interdependence of the lower bounds and the parameters d, ε_d and β, Figure 6.2 presents a comparison of the resulting inapproximability factors for various d and the corresponding lower bounds ε_d for d-bounded MIN-VC on the subintervals $\beta \in (0, 1]$, $\beta \in (1, 2]$ and $\beta \in (2, \beta_{max}]$.

To illustrate the behavior of the inapproximability factor at the phase transition points, in Figure 6.3 we plot the inapproximability factors over the whole interval $(0, \beta_{max}]$ for $d = 3$ with $1 + \varepsilon_d = {}^{103}/_{102}$ and for $d = 4, 5$ with $1 + \varepsilon_d = {}^{55}/_{54}$. As mentioned in Section 6.2, the lower bounds ε_d are due to Berman and Karpinski [BK03] and the general technique for obtaining lower bounds for bounded occurrence

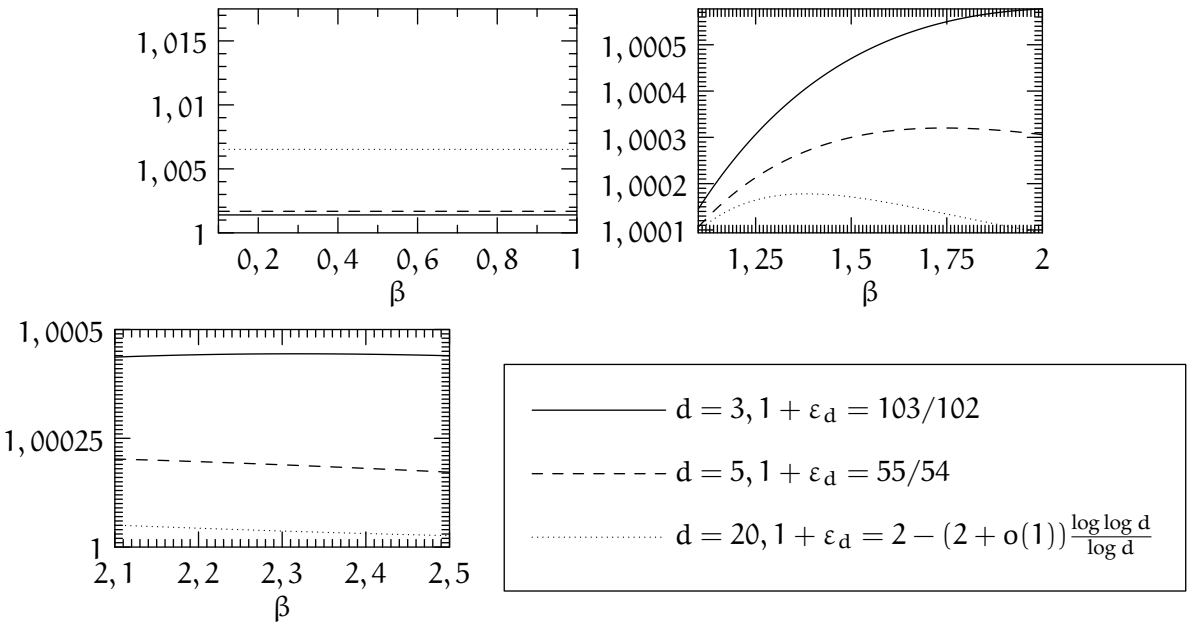


Figure 6.2: Comparison of the inapproximability factors for MIN-VC in connected (α, β)-PLG multigraphs for various d and lower bounds $1 + \varepsilon_d$ on the subintervals $\beta \in (0, 1]$, $\beta \in (1, 2]$ and $\beta \in (2, \beta_{max})$.

CSPs is presented in Chapter 5. The points of discontinuity (*jumps*) correspond to the phase transitions at $\beta = 1$ and $\beta = 2$, treated in Section 6.7.2 and Section 6.6.2.

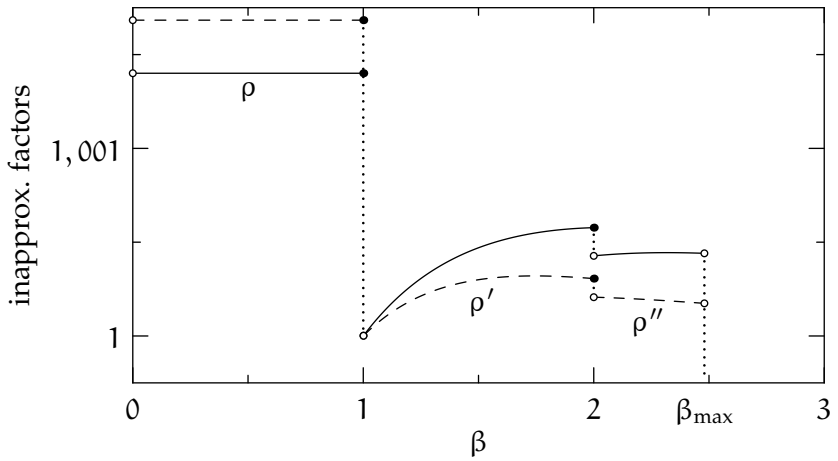


Figure 6.3: Plot of the inapproximability factor for MIN-VC in connected (α, β)-PLG multigraphs for $d = 3$ with $1 + \varepsilon_d = 103/102$ (—) and for $d = 4, 5$ with $1 + \varepsilon_d = 55/54$ (- - -) on the interval $(0, \beta_{max})$.

Now we are going to introduced the notions and methods used in this chapter and some preliminary results.

6.4 PRELIMINARIES

We start with the presentation of techniques and results that regard the hardness and inapproximability of combinatorial optimization on power law graphs.

6.4.1 *Approximability of Combinatorial Optimization Problems on Power-Law Graphs*

The NP-hardness for a certain class of combinatorial optimization problems in power law graphs was shown by Ferrante, Pandurangan, and Park [FPP08] in 2008.

NP-HARDNESS OF COMBINATORIAL OPTIMIZATION PROBLEMS WITH OPTIMAL SUBSTRUCTURE PROPERTY. Ferrante, Pandurangan, and Park [FPP08] showed that if a optimization problem is NP-hard on simple general graphs and satisfies the so called *optimal substructure property*, then it is also NP-hard on simple (α, β)-PLG. The property mentioned above is defined as follows:

Definition 6.1 (Optimal Substructure Property)**.**
Let Π be a NP-hard optimization problem that takes a graph G as input. Π has the *optimal substructure property* if and only if every optimal solution OPT of Π on G contains an optimal solution $\mathsf{OPT}[C]$ of Π on each of the graphs maximal connected components.

A number of important combinatorial optimization problems satisfy the above property, in particular the MIN-VC problem, the MIN-DS problem and the MAX-IS problem. In order to prove NP-hardness of the above optimization problems, Ferrante, Pandurangan, and Park [FPP08] showed how to embed any simple undirected graph $G = (V, E)$ into a simple undirected (α, β)-PLG $G_{\alpha,\beta} = (V_{\alpha,\beta}, E_{\alpha,\beta})$ with $\beta > 0$ such that G is a set of maximal connected components of $G_{\alpha,\beta}$ and $|V_{\alpha,\beta}| = \text{poly}(|V|)$. Altogether, they obtained the following theorem on the NP-hardness of optimization problems with optimal substructure property on simple (α, β)-PLG.

Theorem 6.1 (Ferrante, Pandurangan, and Park [FPP08])**.**
Let Π be an optimization problem on graphs with the optimal substructure property. If Π is NP-hard on simple general graphs, then for all $\beta > 0$, Π is NP-hard also on simple undirected (α, β)-PLG.

Note that the embedding construction for the above theorem only produces *disconnected* (α, β)-PLG. Therefore, an open question regarding the hardness in *connected* and simple (α, β)-PLG is posed in [FPP08]. Next, we are going to present the first inapproximability results that were shown for the case of (α, β)-PLG.

APX-HARDNESS AND EXPLICIT LOWER BOUNDS. In a series of papers [SNT10; She+12] Shen et al. showed the APX-hardness of several optimization problems with optimal substructure property on disconnected (α, β)-PLG multigraphs and presented the first explicit approximation lower bounds for the problems that are summarized in Table 6.2.

In order to achieve the lower bounds, Shen, Nguyen, and Thai [SNT10] proposed a cycle-based embedding technique to embed any d-bounded graph G_d into a PLG $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ with $\beta > 1$. The most important feature of this embedding construction is that the graph G_d will be a set of maximal components in $G_{\alpha,\beta}$ and combinatorial optimization problems with the optimal substructure property (see Definition 6.1) are polynomially solvable in the residual graph $G_{\alpha,\beta} \setminus G_d$ and that $|G_{\alpha,\beta}| = \mathcal{O}(|G_d|)$.

Let us give a brief description of the embedding construction presented in [SNT10; She+12]. With β given by the instance, the embedding is constructed by the algorithm `Cyclic_Embedding`, which embeds any d-bounded graph G_d into a power law graph $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$. Let $\tau(i) = \lfloor e^{\alpha}/i^{\beta} \rfloor - n_i$ be the number of leftover degrees i with $n_i = 0$ if $i > d$.

Considering step 4 of the algorithm, a $\vec{d}$-regular cycle ($RC_n^{\vec{d}}$) is defined as follows:

Definition 6.2 ($\vec{d}$-regular cycle ($RC_n^{\vec{d}}$))**.**
Let $\vec{d} = (d_1, \ldots, d_n)$ be a degree sequence (vector) on n vertices. A $\vec{d}$-regular cycle $RC_n^{\vec{d}}$ consists of two cycles, each of size n, where the two i-th vertices of degree d_i are adjacent via $d_i - 2$ multi-edges.

An example of a $\vec{d_1}$-regular cycle $RC_n^{\vec{d}}$ is shown in Figure 6.4.

Algorithm 6.1 : `Cyclic_Embedding`

Input : d-bounded graph $G_d = (V, E)$.
Output : Power law graph $G_{\alpha,\beta}$ with G_d embedded as a maximal component.

(1) Choose α such that $e^{\alpha} = \max_{1 \leqslant i \leqslant d}\{n_i \cdot i^{\beta}\}$ and $e^{\alpha/\beta} \geqslant d$;
forall the $v \in V, d(v) = 1$ **do**
(2) Construct $\lfloor \tau(1)/2 \rfloor$ separate cliques K_2;
forall the $v \in V, d(v) = 2$ **do**
(3) Construct a cycle of size $\tau(2)$;
forall the $v \in V, 2 \leqslant d(v) = i \leqslant \left\lfloor e^{\alpha/\beta} \right\rfloor$ **do**
(4) Construct a $\vec{d}$-regular cycle where $\vec{d}$ is a vector composed of $2\lfloor \tau(i)/2 \rfloor$ i-elements, for all i satisfying $\tau(i) > 0$;
(5) For all leftover isolated vertices L with $\tau(i) - 2\lfloor \tau(i)/2 \rfloor = 1$, construct a $\vec{k}$-branch-$\vec{d}$-cycle, where $\vec{k}$ ($\vec{d}$) is a vector composed of the vertices in L with odd degrees (even degrees), respectively;

As we see from Figure 6.4 and step 4 of algorithm 6.1, vertices of the same residual degree are paired in the construction of the $\vec{d_1}$-regular cycle and the remaining isolated vertices with $\tau(i) - 2\lfloor \tau(i)/2 \rfloor = 1$ are processed in step 5 of the algorithm. Here a $\vec{k}$-branch-$\vec{d}$-cycle is constructed where $\vec{k}$ ($\vec{d}$) is a vector composed of the vertices in L with odd degrees (even degrees), respectively. We give the following definition of a $\vec{k}$-branch-$\vec{d}$-cycle ($\vec{k}$-$BC_n^{\vec{d}}$):

Definition 6.3 ($\vec{k}$-branch-$\vec{d}$-cycle ($\vec{k}$-$BC_n^{\vec{d}}$))**.**
Let $\vec{d} = (d_1, \ldots, d_n)$, $\vec{k} = (k_1, \ldots, k_m)$ be the two degree vectors of the remaining even degrees (d_i) and odd degrees (k_i). A $\vec{k}$-branch-$\vec{d}$-cycle consists of n vertices such that each vertex on the cycle has residual degree d_i, and of $|\vec{k}|/2$ branches. Note that, in any case, the number of residual vertices $|\vec{k}|$ of odd degree is even, such that on each branch two vertices of odd degree are matched by an edge and all remaining residual degrees are even. The even residual degrees are then realized via self-loops (cf. Figure 6.5).

In this framework Shen et al. [She+12] proposed the following definition of an *embedded-approximation-preserving reduction*.

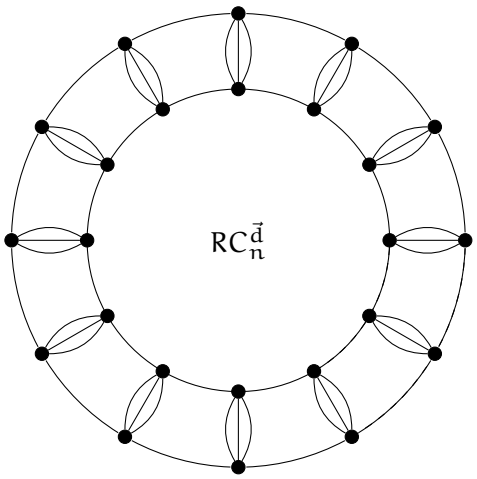


Figure 6.4: The double wheel construction $RC_n^{\vec{d}}$ for the embedding of degree d bounded graphs into combinatorial power law graph due to Shen et al. [She+12]. The construction shown in the figure is an example for $d = 5$ and $n = 12$.

Definition 6.4 (Embedded-Approximation-Preserving Reduction)**.**
Let Π be an optimization problem on graphs with the optimal substructure property. A reduction from a graph instance G of Π to an instance on a power law graph $G_{\alpha,\beta}$ is called *embedded-approximation-preserving* if it satisfies the following two properties:

(i) G is a subset of maximal connected components of $G_{\alpha,\beta}$.

(ii) For any optimal solution $\mathsf{OPT}(G_{\alpha,\beta})$ of Π on $G_{\alpha,\beta}$, we have that the cost

$$\mathsf{opt}(G_{\alpha,\beta}) \leqslant c \cdot \mathsf{opt}(G) \ ,$$

where c is a constant that corresponds to the growth of the optimal solution.

The following theorem shows how to relate the ε-inapproximability of optimization problem Π on general graphs G to an inapproximability-factor ρ of Π on (α, β)-PLG.

Theorem 6.2 (Shen et al. [She+12])**.**
Let $\Pi = \{\mathcal{I}, \mathcal{F}, \mathrm{cost}, \mathrm{goal}\}$ *be an optimization problem with optimal substructure property and* ε*-inapproximability on general graphs* G*. If there exists an embedded-approximation-*

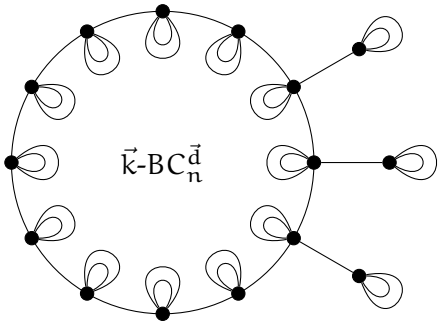


Figure 6.5: Example of the residual wheel construction $\vec{k}$-BC$_n^{\vec{d}}$ in order to realize the residual degrees of the sequence $d_1, \ldots, d_\Delta$ in the construction due to Shen et al. [She+12].

preserving reduction from G *to an* (α, β)*-PLG* $G_{\alpha,\beta}$*, then we can extract the following bounds on the inapproximability factor* ρ *of* Π *on* $G_{\alpha,\beta}$*:*

$$\rho \leqslant \begin{cases} \frac{\varepsilon \cdot c}{(c-1)\varepsilon+1} & \text{, if } \Pi \text{ is a maximization problem, i.e. goal} = \max \\ \frac{\varepsilon+c-1}{c} & \text{, if } \Pi \text{ is a minimization problem, i.e. goal} = \min \end{cases} .$$

Proof. Suppose there exists an ρ-approximation algorithm $\mathcal{A}$ for Π on $G_{\alpha,\beta}$. Let $A = \mathrm{val}(\mathcal{A}(G))$ and $B = \mathrm{val}(\mathcal{A}(G_{\alpha,\beta} \setminus G))$ be the sizes of solution produced by $\mathcal{A}$ on G and $G_{\alpha,\beta} \setminus G$, respectively, and let $A^* = \mathsf{opt}(G)$ and $B^* = \mathsf{opt}(G_{\alpha,\beta} \setminus G)$ their corresponding optimal values. Since there exists an embedded-approximation-preserving reduction from G to $G_{\alpha,\beta}$, we have that $B^* \leqslant (c-1)A^*$. With completeness it follows that $\mathsf{opt}(G) = A^* \Rightarrow \mathsf{opt}(G') = B^*$ and with the soundness property one can obtain a lower bound for ρ, dependent on goal $\in \{\min, \max\}$, as follows:

- If goal $= \max$, the following holds since $B \leqslant B^*$ and $B^* \leqslant (c-1)A^*$

$$\begin{aligned} & A^* + B^* \leqslant \rho(A + B) \\ \iff \quad & A^* \leqslant \rho \cdot A + (\rho - 1)B^* \\ \iff \quad & A^* \leqslant \rho \cdot A + (\rho - 1)(c-1)A^* \ . \end{aligned}$$

Furthermore, we assumed that Π is hard to approximate within ε on G. This yields that $A^* > \varepsilon \cdot A$ and therefore

$$A^* < \frac{\rho \cdot A^*}{\varepsilon} + (\rho - 1)(c - 1)A^* \Leftrightarrow \frac{\varepsilon \cdot c}{(c-1)\varepsilon + 1} \ .$$

- If goal = min, the following holds since $B^* \leqslant B$

$$\begin{aligned} & A + B \leqslant \rho(A^* + B^*) \\ \Longleftrightarrow \quad & A \leqslant \rho \cdot A^* + (\rho - 1)B^* \\ \Longleftrightarrow \quad & A \leqslant \rho \cdot A^* + (\rho - 1)(c - 1)A^* \ . \end{aligned}$$

Since $A > \varepsilon \cdot A^*$, this yields

$$\varepsilon < \rho + (\rho - 1)(c - 1) \Leftrightarrow \frac{\varepsilon + c - 1}{c} \ .$$

□

This relation allowed Shen et al. to transfer the known inapproximability results of the above mentioned combinatorial optimization problems for the case of d-bounded graphs to the case of power law graphs as summarized in Table 6.2.

6.4.2 *Connectivity and Expected Cut Sizes*

In the course of this chapter, we will deal with the expected cut sizes in the random PLG model $G(\alpha, \beta)$ while we establish connectivity in the embedding construction of the reduction. We are going to show that the expected number of edges between any two components of a graph $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ is linear in the size of the smaller component. This fact will be used to motivate our connected embedding construction in Section 6.5.2.

Consider the random PLG model $G(\alpha, \beta)$ introduced in Section 2.5.4. For a given degree sequence $d_1, \ldots, d_{n'}$, let the set L of vertex copies be defined as $L = \bigcup_{i=1}^{n'} L_i$

with $L_i = \{v_{i,1}, \dots, v_{i,d_i}\}$. Note that in the case of an (α, β)-PLG, the degree sequence is

$$\underbrace{1 \dots 1}_{\lfloor e^\alpha \rfloor} \dots \underbrace{j \dots j}_{\left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor} \dots \Delta = \left\lfloor e^{\alpha/\beta} \right\rfloor \ .$$

Now let $n = \sum_{j=1}^{n'} d_j$, and let $\mathcal{M}^n$ denote the set of all matchings on the set L. Furthermore, for $v, v' \in L$, let $\mathcal{M}^n_{v,v'}$ be the set of all matchings which contain the edge $\{v, v'\}$. We have

$$|\mathcal{M}^n| = \frac{n!}{2^{n/2} \cdot \left(\frac{n}{2}\right)!} \quad \text{and} \quad |\mathcal{M}^n_{v,v'}| = \frac{(n-2)!}{2^{(n-2)/2} \cdot \left(\frac{n-2}{2}\right)!} \ .$$

Hence, the probability for an edge $e = \{v, v'\}$ to be an element of a random matching M over n vertices is

$$\Pr(e = \{v_i, v_j\} \in \mathcal{M}^n) = \frac{|\mathcal{M}^n_{v,v'}|}{|\mathcal{M}^n|} = \frac{\frac{(n-2)!}{2^{(n-2)/2} \cdot \left(\frac{n-2}{2}\right)!}}{\frac{n!}{2^{n/2} \cdot \left(\frac{n}{2}\right)!}} = \frac{1}{n-1} \ .$$

Thus we obtain the following result.

Lemma 6.1.
Consider the random PLG model $G(\alpha, \beta)$ and let A, B be disjoint subsets of vertices of the resulting PLG. Then the expected number of edges between A and B is

$$\mathbb{E}(\#\textit{edges between } A \textit{ and } B) = \frac{\left(\sum_{w \in A} d(w)\right) \cdot \left(\sum_{u \in B} d(u)\right)}{\sum_{v \in V} d(v)} \ .$$

6.5 OUTLINE OF THE METHOD

Let us give an outline of the methods used in this chapter.

6.5.1 *Lower Bound Technique*

In order to obtain approximation lower bounds for the MINIMUM VERTEX COVER problem in connected (α, β)-PLG, we will construct for each parameter $\beta \in [0, \beta_{max})$ as well as for the functional cases $\beta = 1 \pm \frac{1}{f(n)}$ polynomial-time reductions R_β from MIN-VC$_d$, that is MIN-VC in d-bounded graphs, to MIN-VC$_{\alpha,\beta}$. R_β embeds any d-bounded graph G_d into a connected multigraph $R_\beta(G_d) = G_{\alpha,\beta}$, which consists of a multigraph copy $\mu(G_d)$ of G_d attached to a multigraph wheel W. Furthermore, all the degree 1 nodes of $G_{\alpha,\beta}$ are attached to wheel nodes of W. In the case $1 < \beta < \beta_{max}$, it is sufficient to choose $\mu(G_d) = G_d$ for the embedding. For $0 < \beta \leqslant 1$, we have to construct a multigraph copy of G_d in order to artificially increase the node degrees inside G_d and to be able to choose a vertex cover in $\mu(G_d)$ independent of the embedding. In any case the size of the fully composed $G_{\alpha,\beta}$ will be linear in the size of G_d.

Let us now describe how the multigraph copy of G_d is constructed and which approximation lower bounds can obtained under this construction.

MIN-VC IN BOUNDED DEGREE GRAPHS WHICH PROVIDE A PERFECT MATCHING

We will now describe the polynomial time reduction from MIN-VC$_d$ to MIN-VC$^{PM}_{d+2}$. Given a graph G_d of maximum degree d with a vertex set $V = \{v_1, \ldots, v_n\}$, we construct the graph $\widetilde{G}_{d+2} = (\widetilde{V}, \widetilde{E})$ as follows:

1. The set of vertices $\widetilde{V}$ consists of four disjoint copies of the vertex set V, namely $\widetilde{V} := V_1 \cup V_2 \cup V_3 \cup V_4$ with $V_i = \{v_{i,j}, 1 \leqslant j \leqslant n\}, i = 1, \ldots, 4$.
2. $\widetilde{E} := E_1 \cup E_2 \cup P$, where $G^1_d = (V_1, E_1)$ and $G^2_d = (V_2, E_2)$ are disjoint copies of G_d, i. e. $E_i = \{\{v_{i,j}, v_{i,l}\} \mid \{v_j, v_l\} \in E\}$.
3. $P := \bigcup_{j=1}^{n} \{\{v_{1,j}, v_{3,j}\}, \{v_{3,j}, v_{4,j}\}, \{v_{4,j}, v_{2,j}\}, \{v_{3,j}, v_{2,j}\}, \{v_{1,j}, v_{4,j}\}\}$.

The construction is shown in Figure 6.6. Let R_{PM} be this reduction, i. e. for any d-bounded degree graph G_d, $R_{PM}(G_d) = \widetilde{G}_{d+2}$. Now suppose C is a vertex cover in $\widetilde{G}_{d+2}$ with $C = C_1 \cup C_2 \cup C_3 \cup C_4$, where $C_i = C \cap V_i, 1 \leqslant i \leqslant 4$. We observe that C_1 and C_2 are vertex covers of G^1_d and G^2_d, respectively. Furthermore, for every $j \in \{1, \ldots, n\}$ the following holds for any vertex cover C on $\widetilde{G}_{d+2}$ (see also Figure 6.6):

- If $v_{1,j} \in C$ and $v_{2,j} \in C$, then C also contains one of the nodes $v_{3,j}, v_{4,j}$.

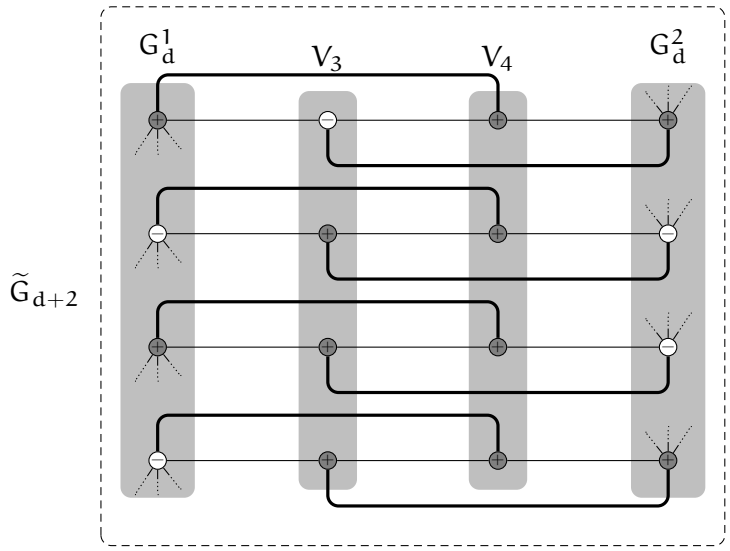


Figure 6.6: Example $\widetilde{G}_{d+2}$ after the construction step of reduction R_{PM} that converts any d-bounded graph G_d into a $(d+2)$-bounded graph which provides a perfect matching M (e. g. via the **thick** edges of the set P). The nodes $\oplus$ denote covering vertices and $\ominus$ denote non-covering vertices.

- If $v_{1,j} \notin C$ and $v_{2,j} \notin C$, then C contains both nodes $v_{3,j}, v_{4,j}$.
- If $v_{1,j} \in C$ and $v_{2,j} \notin C$—or vice versa—, then C contains both nodes $v_{3,j}, v_{4,j}$.

Hence,

$$\begin{aligned} |C| &= |C_1| + |C_2| + |C_1 \cap C_2| + 2 \cdot |V \setminus (C_1 \cup C_2)| + 2 \cdot |C_1 \triangle C_2| \\ &= 3 \cdot |C_1 \triangle C_2| + 3 \cdot |C_1 \cap C_2| + 2 \cdot |V \setminus (C_1 \cup C_2)| \\ &= 1 \cdot |C_1 \triangle C_2| + |C_1 \cap C_2| + 2 \cdot |V| \\ &= |C_1 \cup C_2| + 2 \cdot |V| \ . \end{aligned} \tag{6.1}$$

From Equation 6.1, we get that a minimal vertex cover is obtained by choosing $C_1 = C_2$ and minimizing the cardinality of this set and thereby minimizing the cardinality of $C_1 \cup C_2$. Hence, we can restrict ourselves to vertex covers C with the property $C_1 = C_2$, i. e. C_1 and C_2 being copies of the same vertex cover C_d in G_d. Thus, we obtain the following lemma for the construction described above.

Lemma 6.2.
There is a polynomial time algorithm $\mathcal{T}$ which transforms any vertex cover $C = C_1 \cup C_2 \cup$

$C_3 \cup C_4$ *of a graph* $\widetilde{G}_{d+2} = R_{PM}(G_d)$ *into a vertex cover* $C' = \mathcal{T}(G_d, C)$ *of* $\widetilde{G}_{d+2}$ *such that* $C'_1 = C'_2 = \operatorname{argmin}\{|C_1|, |C_2|\}$ *and* $|C'| = \min\{|C_1|, |C_2|\} + 2 \cdot |V| \leqslant |C|$.

Proof. Without loss of generality let $|C_1| \leqslant |C_2|$, $|C_1| = k$ and $C_1 = \{v_{1,1}, \ldots, v_{1,k}\}$. If $v_{1,j} \in C$ and $v_{2,j} \in C$, then either $v_{3,j} \in C$ or $v_{4,j} \in C$ and we are done.

If $v_{1,j} \notin C$ and $v_{2,j} \in C$, we have that $v_{3,j} \in C$ and $v_{4,j} \in C$ and we transform the cover C into a cover C', with $|C'| \leqslant |C|$, by setting $v_{1,j}, v_{2,j} \notin C'$ and $v_{3,j}, v_{4,j} \in C'$. If $v_{1,j} \in C$ and $v_{2,j} \notin C$, again we have that $v_{3,j} \in C$ and $v_{4,j} \in C$ and we set $v_{1,j}, v_{2,j} \notin C'$ and $v_{3,j}, v_{4,j} \in C'$. □

RESULTING LOWER BOUND. We will now give an estimate of the resulting lower bound of the reduction. Suppose that MIN-VC$_d$ is hard to approximate within approximation ratio $1 + \varepsilon_d$. Furthermore, suppose that algorithm $\mathcal{A}$ is a polynomial time $(1+\varepsilon)$-approximation algorithm for MIN-VC$^{PM}_{d+2}$. Then the algorithm $\mathcal{B}$, which on input G_d constructs the vertex cover $\widetilde{C} = \mathcal{T} \circ \mathcal{A} \circ R_{PM}(G_d) = \bigcup_{i=1}^{4} \widetilde{C}_i$ and then returns $\widetilde{C}_1$, is an approximation algorithm for MIN-VC$_d$. We have,

$$|\widetilde{C}| \leqslant (1+\varepsilon) \cdot \mathsf{OPT}(\widetilde{G}_{d+2}) = (1+\varepsilon) \cdot (\mathsf{OPT}(G_d) + 2n) \ . \tag{6.2}$$

Furthermore, due to the proof of the previous Lemma 6.2, $|\widetilde{C}| = |\widetilde{C}_1| + 2n$. Thus from Equation 6.2, we obtain

$$|\widetilde{C}_1| \leqslant (1+\varepsilon) \cdot \mathsf{OPT}(G_d) + \varepsilon \cdot 2n \leqslant ((1+\varepsilon) + d \cdot 2\varepsilon) \cdot \mathsf{OPT}(G_d) \ ,$$

where the second inequality holds due to the fact that $\mathsf{OPT}(G_d) \geqslant \frac{n}{d}$. Thus it must be that $\varepsilon_d \leqslant (d \cdot 2 + 1) \cdot \varepsilon$. Finally we note that if G_d has maximum degree at most d, then the maximum degree of $\widetilde{G}_{d+2} = R_{PM}(G_d)$ is at most $d+2$. Altogether we obtain the following result.

Theorem 6.3.
Suppose that MIN-VC$_d$ *is hard to approximate within approximation ratio* $1 + \varepsilon_d$. *Then* MIN-VC$^{PM}_{d+2}$ *is hard to approximate within approximation ratio* $1 + \frac{\varepsilon_d}{1+2d}$.

6.5.2 *Embedding Technique*

We let Γ denote the set of neighbors of G_d in the wheel W of the embedding construction and let W_1 denote the set of wheel nodes which are adjacent to at least one degree 1 node in $G_{\alpha,\beta}$ (see Figure 6.7).

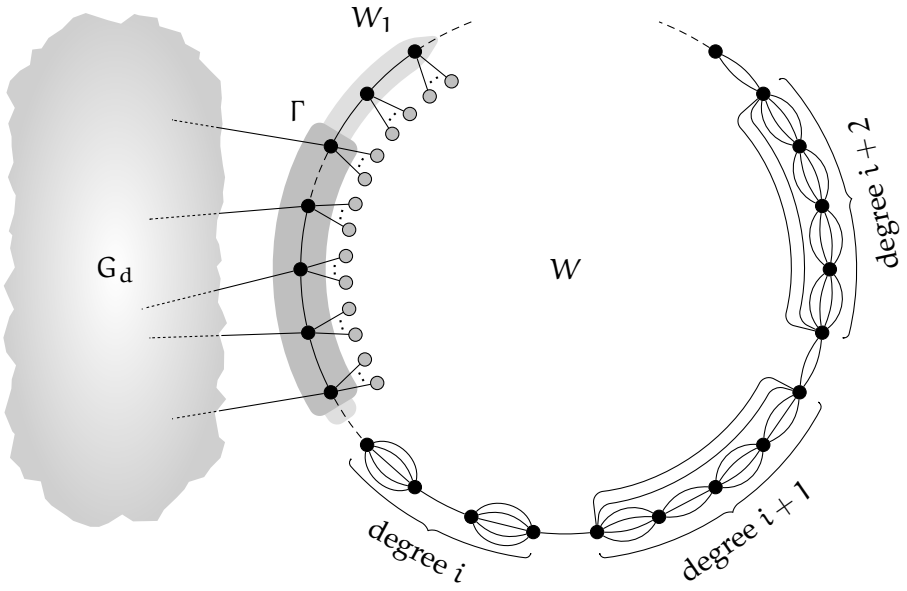


Figure 6.7: Embedding construction of the reduction R_β. Any d bounded graph G_d is attached to a multigraph wheel W (more precisely, to a subset of vertices Γ) with a number of edges linear in $|G_d|$. The residual degrees of the power law degree sequence are realized inside W in a cyclic increasing order via degree 1 vertices and multi-edges. $W_1 \in V(W)$ denotes the subset of wheel nodes which have a degree 1 vertex attached.

The rest of the wheel nodes is used to implement the power law degree sequence by systematically constructing multi-edges between nodes of similar target degree.

Due to the different behavior of the power law distributions at the phase transition points of the parameter β, we have to distinguish the two cases $1 < \beta < \beta_{max} \approx 2.48$ and $0 < \beta \leqslant 1$.

For $1 < \beta < \beta_{max}$, we construct the PLG $G_{\alpha,\beta} = \mu(G_d) \cup W$ in such a way that the set Γ of neighbors of G_d in the wheel W satisfies $\Gamma \subseteq W_1$ and $|\Gamma| = \Theta(n)$. This means every neighbor of nodes from G_d in the wheel is also adjacent to at least one node of degree 1. Neighbors of degree 1 nodes have the property that every vertex cover

either contains this node or all its degree 1 neighbors. This implies that any optimum vertex cover C^{OPT} in $G_{\alpha,\beta}$ contains the set Γ, and hence the intersection of C^{OPT} with $\mu(G_d)$ corresponds to an optimum vertex cover C_d^{OPT} in G_d.

In the case $0 < \beta \leqslant 1$, and also in the functional cases $\beta = 1 - \frac{1}{f(n)}$ and $\beta = 1 + \frac{1}{f(n)}$, the behavior of the power law distributions is rather different. In these cases, the number of degree 1 nodes in (α, β)-PLG is not sufficient to attach a degree 1 node to every neighbor of $\mu(G_d)$ in W. Hence, we cannot guarantee anymore that every optimum vertex cover in $G_{\alpha,\beta}$ contains an optimum vertex cover in G_d. Furthermore, in order to obtain that $|G_{\alpha,\beta}| = \mathcal{O}(|G_d|)$, the nodes of $\mu(G_d)$ must have high degree in $G_{\alpha,\beta}$. Since the set Γ is too small to realize this degree of nodes in G_d via parallel edges into Γ, we need to replace the edges inside G_d by sufficiently many multi-edges. In order to keep track of the node-degrees and to implement the power law distribution, we will first map G_d to a graph $\widetilde{G}_{d+2}$ which contains a perfect matching. This allows us to increase the node degrees inside $\widetilde{G}_{d+2}$ in a controlled manner, namely pairwise along the edges of a perfect matching. Then we construct $G_{\alpha,\beta} = R_\beta(G_d)$ in such a way that $|\Gamma| = o(n)$.

This means in the cases $0 < \beta \leqslant 1$ and in the functional cases $\beta = 1 - \frac{1}{f(n)}$ and $\beta = 1 + \frac{1}{f(n)}$, our reduction from MIN-VC$_d$ to MIN-VC$_{\alpha,\beta}$ is the composition of a reduction from MIN-VC$_d$ to MIN-VC$^{PM}_{d+2}$, that is MIN-VC restricted to $(d+2)$ bounded degree graphs which provide a perfect matching, and a reduction from MIN-VC$^{PM}_{d+2}$ to MIN-VC$_{\alpha,\beta}$. We will show that any approximation algorithm for MIN-VC$_{\alpha,\beta}$ also yields an approximation algorithm with almost the same approximation ratio for the problem of constructing a minimum size vertex cover for $G_{\alpha,\beta}$ which contains the set Γ. This special version of the vertex cover problem for graphs $G_{\alpha,\beta} = R_\beta(G_d)$ will be denoted as $\widehat{\text{MIN-VC}}_{\alpha,\beta}$.

In both cases, our polynomial time reduction from MIN-VC$_d$ to MIN-VC$_{\alpha,\beta}$ has the following general structure.

1. Map G_d to a multigraph $\mu(G_d)$.
   In the case $\beta > 1$, $\mu(G_d)$ *is equal to* G_d. *In the case* $0 < \beta \leqslant 1$ *and in the functional cases* $\beta = 1 - \frac{1}{f(n)}$ *and* $\beta = 1 + \frac{1}{f(n)}$, *we first apply our polynomial time reduction from* MIN-VC$_d$ *to* MIN-VC$^{PM}_{d+2}$, *i.e. the* MIN-VC *problem restricted to* $(d+2)$*-bounded degree graphs which provide a perfect matching. This yields a graph* $\widetilde{G}_{d+2}$ *of size* $4 \cdot |G_d|$

which contains a perfect matching M*. Then we replace the edges of* M *by multi-edges such as to increase the degree of nodes in* $\widetilde{G}_{d+2}$ *appropriately.*

2. Choose the parameter α as small as possible such that each node in $\mu(G_d)$ can be connected to a set $\Gamma \subseteq W$, where $|\mu(G_d)| \leqslant \mathrm{vol}(\Gamma)$ and $|\Gamma| = o(n)$.

3. Construct the set of wheel-nodes W.
Assign to every node v *in* $\mu(G_d) \cup W$ *a target node degree* $d_{\alpha,\beta}(v)$*, i. e. the desired degree in the power law degree distribution. Generate the edges (multi-edges) from* $\mu(G_d)$ *to the set* $\Gamma \subset W$*. The wheel will be constructed in such a way that wheel nodes of the same degree always form an induced connected subgraph in* W*.*

4. Connect the degree 1 nodes to the wheel W.

5. Construct edges inside W such that the resulting multigraph is an (α, β)-PLG.

In order to keep track of the node degrees and the edges being already constructed in steps 1–5 of this reduction, we make use of the notion of the *residual degrees* $d_r(v)$ of nodes v in the graph $G_{\alpha,\beta}$.

As we see from the above description of the embedding construction, the embedding algorithm partitions the wheel nodes $w \in W$ into the subsets $W_1 \subseteq W, \Gamma \subseteq W_1$ and $W \setminus W_1$. In step 3 of the embedding construction, a target degree and residual degree is assigned to every node $v \in \mu(G_d)$ and $w \in W$, respectively $d_{\alpha,\beta}(v), d_{\alpha,\beta}(w)$ and $d_r(v), d_r(w)$. Then the nodes of $\mu(G_d)$ and the degree 1 nodes are attached to the wheel W, where the residual degrees of $v \in \mu(G_d)$ are resolved immediately and the residual degrees of $w \in W$ are updated. The completion step 5 of the reduction consists of the algorithm `Fill_Wheel`, which we will describe now. This algorithm gets as an input the set of wheel nodes W with residual degrees $d_r(w)$ for all $w \in W$. It generates the missing edges degree wise in a cyclic order by placing multi-edges between nodes of similar degree (cf. Figure 6.8).

If $w_{j,1}, \ldots, w_{j,n_j}$ are the nodes of degree $d_{\alpha,\beta}(w_{j,l}) = j$ in the wheel W, then the following invariant will be maintained.

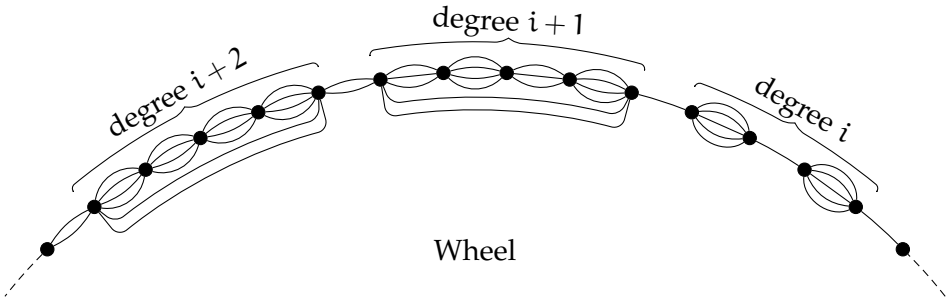


Figure 6.8: Procedure `Fill_Wheel` that realizes the residual degrees on the wheel nodes in W.

Invariant 1.
In every stage of the construction, for every $j \in \{1, \ldots, \Delta\}$, $d_r(w_{j,1}) \leqslant \ldots \leqslant d_r(w_{j,n_j})$ and $d_r(w_{j,n_j}) - d_r(w_{j,1}) \leqslant 1$.

We are now ready to give the pseudo-code description of the algorithm `Fill_Wheel` on page 105. This completes the description of our reduction and embedding technique.

Next, we are going to treat the dependency of the above reduction techniques with regard to the model parameter β. Especially, we will distinguish between the cases where $0 < \beta \leqslant 1$ and $1 < \beta < \beta_{max} \approx 2.48$.

6.6 THE CASE $\beta > 1$

We will now consider the case when the parameter β is in the range $1 < \beta < \beta_{max} \approx 2.48$. We distinguish the subcases $1 < \beta < 2$, $\beta = 2$ and $2 < \beta < \beta_{max}$ which differ by the choice of the intervals and the analysis of the construction.

We will make use of the notion of an *interval* in a PLG. Let $G_{\alpha,\beta} = (V, E)$ an (α, β)-PLG. An *interval* of nodes in $G_{\alpha,\beta}$ is a set $[a, b] = \{v \in V \mid a \leqslant d(v) \leqslant b\}$, where $1 \leqslant a \leqslant b \leqslant \Delta = \left\lfloor e^{\alpha/\beta} \right\rfloor$. Firstly, we consider the case when $1 < \beta < 2$.

6.6.1 *Subcase* $1 < \beta < 2$

Given a degree d bounded graph G_d, we construct $G_{\alpha,\beta} = R_\beta(G_d)$ as follows. Let W be the set of wheel nodes, $W_1 \subseteq W$ is the set of nodes $w \in W$ which are adjacent to at

Algorithm 6.2 : `Fill_Wheel`

Input : The set of wheel nodes $\{w_{j,l}\} \in V(W)$ with $j \in \{3, \dots, \Delta\}$, $l \in \{1, \dots, n_j\}$ and residual degrees $d_r(w_{j,l})$.

Output : A graph W with residual degrees $d_r(w_{j,l}) = 0$.

for $j = 3, \dots, \Delta$ **do**
  while $d_r(w_{j,n_j}) > 0$ **do**
    choose l min such that $d_r(w_{j,l})$ is max;
    if $l < n_j$ **then**
      generate edge $\{w_{j,l}, w_{j,l+1}\}$;
      $d_r(w_{j,l}) := d_r(w_{j,l}) - 1$;
      $d_r(w_{j,l+1}) := d_r(w_{j,l+1}) - 1$;
    else if $l = n_j, d_r(w_{j,1}) > 0$ **then**
      generate edge $\{w_{j,l}, w_{j,1}\}$;
      $d_r(w_{j,l}) := d_r(w_{j,l}) - 1$;
      $d_r(w_{j,1}) := d_r(w_{j,1}) - 1$;
    else if $l = n_j, d_r(w_{j,1}) = 0, j < \Delta$ **then**
      generate edge $\{w_{j,l}, w_{j+1,1}\}$;
      $d_r(w_{j,l}) := d_r(w_{j,l}) - 1$;
      $d_r(w_{j+1,1}) := d_r(w_{j+1,1}) - 1$;
    else
      take degree 1 node w_1 and generate edge $\{w_{j,l}, w_1\}$;
      $d_r(w_{j,l}) := d_r(w_{j,l}) - 1$;

least one node of degree 1. Furthermore, let n_j be the number of nodes of degree j in the graph G_d. The set Γ of neighbors of G_d in the wheel should satisfy $\Gamma \subseteq W_1 \subseteq W$. We let n_j be the number of nodes of degree j in the graph G_d and we want to choose $W_1 = [j_0, \Delta]$ as small as possible such as to meet the following requirements:

1. *Sufficient amount of node degree in the wheel:*

$$\lfloor e^\alpha \rfloor + n \leqslant \sum_{j=j_0}^{\Delta} (j-2) \cdot \left(\left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor - n_{j-1} \right).$$

2. *Enough degree* 1 *nodes:* $\lfloor e^\alpha \rfloor \geqslant |\Gamma|$, which holds if

$$\lfloor e^\alpha \rfloor \geqslant \sum_{j=j_0}^{\Delta} \left(\left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor - n_{j-1} \right).$$

3. *Node degrees of* G_d*:* $\left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \geqslant n_{j-1}$ for all $j = 2, \ldots, d+1$.

The first constraint ensures that a sufficient amount of node degree is available in the set W_1, such as to let all the nodes from G_d and all the degree 1 nodes be adjacent to nodes from W_1. The second constraint guarantees that every node in the neighborhood Γ of G_d can have least one degree 1 node adjacent to it. Since we may assume that G_d does not contain any node of degree 1 and since every node in G_d will have one neighbor in Γ, the third constraint ensures that the degree distribution of the embedded graph $\mu(G_d)$ fits into the power law distribution of the fully composed graph $G_{\alpha,\beta}$. Considering constraint 3, the following lemma shows how to choose the size parameter α in order to achieve the requirement.

Lemma 6.3.
If $\left\lfloor \frac{e^\alpha}{(d+1)^\beta} \right\rfloor \geqslant n$*, then the requirement 3 is satisfied.*

Proof. Recall that the function $f(n) = \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor$ has negative slope for all $\beta > 0$. Hence, we have that $\left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \geqslant \left\lfloor \frac{e^\alpha}{(j-1)^\beta} \right\rfloor$. And since $n_j \leqslant n$, for all $j = 1, \ldots, d$, the lemma follows. □

Hence we choose $e^\alpha = (d+1)^\beta n$ to fulfill the constraint. Let us now consider constraint 1. In order to minimize the size of W_1, we want to choose j_0 as large as possible. This yields the following alternate version of requirement 1, which omits the subtraction of the term n_{j-1}.

1*. $\sum_{j=j_0}^{\Delta} (j-2) \cdot \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \geqslant n + \lfloor e^\alpha \rfloor$.

The above inequality is equivalent to

$$\begin{aligned} & e^\alpha \cdot \left(\left[\frac{x}{2-\beta} \right]_{j_0}^{\Delta} - 2 \cdot \zeta(\beta) \right) \geqslant n + (d+1)^\beta \cdot n \\ \iff \quad & \Delta^{2-\beta} - j_0^{2-\beta} \geqslant (2-\beta) \cdot \left(1 + \frac{1}{(d+1)^\beta} + 2 \cdot \zeta(\beta) \right) . \end{aligned} \tag{6.3}$$

Lemma 6.4.
Inequality 6.3 holds for $j_0 = \Delta - h(n)$ *with* $h(n) = \Delta^u, 1 > u > \beta - 1$.

Proof. We give a proof for the special case $\frac{1}{2-\beta} = l \in \mathbb{Z}$. In this case the requirement is equivalent to

$$\begin{aligned} & \Delta^{2-\beta} - \text{const.} \geqslant (\Delta - h(n))^{2-\beta} \\ \iff \quad & \left(\Delta^{2-\beta} - \text{const.} \right)^{\frac{1}{2-\beta}} \geqslant \Delta - h(n) \\ \iff \quad & \Delta - q\left(\Delta^{1/l} \right) \geqslant \Delta - h(n), \end{aligned}$$

where q is a polynomial of degree $l - 1 = \frac{1}{2-\beta} - 1 = \frac{\beta-1}{2-\beta}$. □

Hence, we can choose the parameter $u = \frac{1+(\beta-1)}{2} = \frac{\beta}{2}$ and $W_1 = [\Delta - \Delta^u, \Delta]$ and obtain $|W_1| = \sum_{j=\Delta-\Delta^u}^{\Delta} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor = o\left(|G_{\alpha,\beta}|\right) = o(n)$. We are now ready to give the pseudo-code description of our reduction in the case $1 < \beta < 2$ in algorithm `Reduction`$_{\beta>1}$ on page 108.

RESULTING LOWER BOUND. We will now give an estimate of the resulting lower bound of our reduction. Suppose MIN-VC$_d$ is hard to approximate within $1 + \varepsilon_d$. Furthermore, suppose that $\mathcal{A}_\beta$ is an approximation algorithm for MIN-VC$_{\alpha,\beta}$ on (α, β)-PLG with an approximation ratio of $1 + \varepsilon_\beta$. This yields an approximation algorithm

Algorithm 6.3 : `Reduction`$_{\beta>1}$

Input : $G_d = (V, E)$ degree d bounded graph with $V = \{v_1, \ldots, v_n\}$ such that $2 \leqslant d_{G_d}(v_1) \leqslant \ldots \leqslant d_{G_d}(v_n) \leqslant d$

Output : (α, β)-PLG $G_{\alpha,\beta} = (V_{\alpha,\beta}, E_{\alpha,\beta})$ with $V_{\alpha,\beta} = V \cup W$

(1) **choose** $u = \frac{\beta}{2}$;

(2) **let** $\alpha = \min\left\{\alpha' \middle| [\Delta - \Delta^u, \Delta]| \geqslant n \text{ and } \left\lfloor \frac{e^{\alpha}}{(d+1)^{\beta}} \right\rfloor \geqslant n\right\}$;

(3) $n_j :=$ ♯nodes of degree j in G_d $(j = 2, \ldots, d)$; `/* Generate wheel nodes */`
$V_{\alpha,\beta} := V_d \cup W$ $(W = \bigcup_{j=1}^{\Delta} W_{(j)}$ with
$W_{(j)} = \left\{w_{j,l} \middle| 1 \leqslant l \leqslant \left\lfloor \frac{e^{\alpha}}{j^{\beta}} \right\rfloor - n_{j-1}\right\}, j = 1, \ldots, \Delta)$;

(4) Generate edges $\{w_{j,l}, w_{j,l+1}\}$ $(j = 2, \ldots, \Delta$ and $l = 1, \ldots, |W_{(j)}|)$; `/* Generate rim edges */`
Generate edges $\{w_{j,|W_{(j)}|}, w_{j+1,1}\}$ and one edge $\{w_{\Delta,1}, w_{2,1}\}$ $(j = 2, \ldots, \Delta)$;
let $d_r(w_{j,l}) := j - 2$ $(j = 2, \ldots, \Delta, 1 \leqslant l \leqslant |W_{(j)}|)$;
`/* Generate edges from` G_d `to` W `*/`

(5) **for** $(c = 1, j = \Delta - \Delta^u; c \leqslant n; j{+}{+})$ **do**
  for $(l = 1; l < |W_{(j)}| \wedge c \leqslant n; l{+}{+}, c{+}{+})$ **do**
    Generate an edge $\{v_c, w_{j,l}\}$ and set $d_r(w_{j,l}) := d_r(w_{j,l}) - 1$;

`/* Connect degree 1 nodes */`

(6) **for** $(j = \Delta - \Delta^u, c_1 = 0; j < \Delta \wedge c_1 < \lfloor e^{\alpha} \rfloor; j{+}{+})$ **do**
  for $(l = 1; l < |W_{(j)}| \wedge c_1 < \lfloor e^{\alpha} \rfloor; l = (l == |W_{(j)}|?1 : l + 1))$ **do**
    Generate one edge $\{w_{1,c}, w_{j,l}\}$ and set $c_1 := c_1 + 1, d_r(w_{j,l}) := d_r(w_{j,l}) - 1$;

`/* Generate remaining edges */`

(7) Apply algorithm `Fill_Wheel`;
let $E_{\alpha,\beta}$ be the union of E and the set of all edges generated in steps (4)-(7);
return $(V_{\alpha,\beta}, E_{\alpha,\beta})$

$\mathcal{A}_d$ for MIN-VC$_d$. As we have seen before, we may assume that on given input $R_\beta(G_d)$, algorithm $\mathcal{A}_\beta$ constructs a vertex cover which consists of a cover C_d in G_d, the set W_1 of nodes which are adjacent to degree 1 nodes and a cover on the set $W \setminus W_1$. Since the wheel W is constructed in such a way that an optimal cover can be computed in polynomial time, we may assume that $\mathcal{A}_\beta$ constructs a cover $C_d \cup W_1 \cup \mathsf{OPT}(W \setminus W_1)$. Furthermore, we assume that $\mathcal{A}_d$ on input G_d returns the cover C_d. By the above assumptions, we have that

$$\begin{aligned} & |C_d| + |W_1| + |\mathsf{OPT}(W \setminus W_1)| \leqslant (1+\varepsilon_\beta) \cdot (|\mathsf{OPT}_d| + |W_1| + |\mathsf{OPT}(W \setminus W_1)|) \\ \iff \quad & |C_d| \leqslant (1+\varepsilon_\beta) \cdot |\mathsf{OPT}_d| + \varepsilon_\beta \cdot (|W_1| + |\mathsf{OPT}(W \setminus W_1)|) \end{aligned} \tag{6.4}$$

We will now use the fact that any optimal cover on $W \setminus W_1$ can be transformed into an optimal cover that consists of every second wheel node. Since $|W \setminus W_1| = \zeta(\beta) \cdot (d+1)^\beta \cdot n - n - e^\alpha - o(n)$, we obtain from inequality 6.4 that

$$\begin{aligned} |C_d| &\leqslant (1+\varepsilon_\beta) \cdot |\mathsf{OPT}_d| + \varepsilon_\beta \cdot \left(o(n) + \frac{(\zeta(\beta)-1) \cdot (d+1)^\beta \cdot n - n - o(n)}{2} \right) \\ &\leqslant (1+\varepsilon_\beta) \cdot |\mathsf{OPT}_d| + \varepsilon_\beta \cdot \left(o(n) + \frac{(\zeta(\beta)-1) \cdot (d+1)^\beta - 1 - o(1)}{2} \cdot d \cdot |\mathsf{OPT}_d| \right) \\ &= |\mathsf{OPT}_d| \cdot \left(1 + \varepsilon_\beta \cdot \left(1 + o(1) + \frac{(\zeta(\beta)-1) \cdot (d+1)^\beta - 1 - o(1)}{2} \cdot d \right) \right), \end{aligned}$$

and hence the following result.

Theorem 6.4.
Suppose that MIN-VC$_d$ *is hard to approximate within approximation ratio* $1+\varepsilon_d$. *Then, for* $1 < \beta < 2$, MIN-VC$_{\alpha,\beta}$ *is hard to approximate within approximation ratio*

$$1 + \varepsilon_\beta = 1 + \frac{\varepsilon_d}{1 + d \cdot \frac{(\zeta(\beta)-1)\cdot(d+1)^\beta - 1}{2}} \quad .$$

We are now going to consider the case when $\beta = 2$, and show that one can obtain the same approximation lower bound as in Theorem 6.4.

6.6.2 *Subcase* $\beta = 2$

For this case again we choose the set W_1 to reside in the interval $[j_0, \Delta]$ and consider the following corresponding optimization problem, resembling the requirements 1, 1* and 3.

$$\begin{aligned} &\text{minimize} && j_0, \\ &\text{subject to} && \left\lfloor \frac{e^\alpha}{(d+1)^\beta} \right\rfloor \geqslant n, \qquad \text{and} \\ & && \sum_{j=j_0}^{\Delta} (j-2) \cdot \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \geqslant n + \lfloor e^\alpha \rfloor. \end{aligned} \tag{6.5}$$

Suppose we first choose α such that the inequality

$$e^\alpha \geqslant (d+1)^\beta \cdot (n+1) \tag{6.6}$$

holds. Then the first constraint holds as well. For $\beta = 2$, assuming inequality 6.6, we have the following chain of implications, starting from inequality 6.5:

$$\begin{aligned} \text{(6.5)}: \quad & \sum_{j=j_0}^{\Delta} (j-2) \cdot \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \geqslant n + \lfloor e^\alpha \rfloor \\ \Longleftarrow \quad & \sum_{j=j_0}^{\Delta} \frac{e^\alpha}{j} - (\Delta - j_0 + 1) - 2 \sum_{j=j_0}^{\Delta} \frac{e^\alpha}{j^2} \geqslant n + e^\alpha \\ \Longleftarrow \quad & \sum_{j=j_0}^{\Delta} \frac{1}{j} - \frac{\Delta - j_0 + 1}{e^\alpha} - 2 \sum_{j=j_0}^{\Delta} \frac{1}{j^2} \geqslant \frac{1}{(d+1)^\beta} - \frac{1}{e^\alpha} + 1 \\ \Longleftarrow \quad & \ln(\Delta) - \ln(j_0) - \frac{\Delta - j_0 + 1}{e^\alpha} - 2 \cdot \zeta(2) \geqslant \frac{1}{(d+1)^\beta} - \frac{1}{e^\alpha} + 1 \end{aligned}$$

Hence we choose $j_0 = e^{\frac{\alpha}{2} - c}$ with $c = 2 + \frac{1}{(d+1)^\beta} + 2 \cdot \zeta(2) = \mathcal{O}(1)$. This implies $|W_1| = o(n)$, and thus we obtain the same lower bound as for the case $1 < \beta < 2$, stated in Theorem 6.4, namely

Theorem 6.5.
Suppose that MIN-VC$_d$ *is hard to approximate within approximation ratio* $1+\varepsilon_d$. *Then, for* $\beta = 2$, MIN-VC$_{\alpha,\beta}$ *is hard to approximate within approximation ratio*

$$1+\varepsilon_\beta = 1+\frac{\varepsilon_d}{1+d\cdot\frac{(\zeta(\beta)-1)\cdot(d+1)^\beta-1}{2}} \quad .$$

We proceed to the case $\beta > 2$.

6.6.3 *Subcase* $\beta > 2$

For the case $2 < \beta < \beta_{max} = \inf\{x \mid \zeta(x-1)-2\zeta(x) \leqslant 0\} \approx 2.48$ we consider the following construction. The wheel of the embedding is constructed in such a way that the set W_1, i. e. the set of neighbors of degree 1 vertices, consists of all the wheel nodes of degree $\geqslant 3$. This yields $|W_1| = \sum_{j=3}^{\Delta}\left(\frac{e^\alpha}{j^\beta} - n_{j-1}\right)$ and $|W \setminus W_1| = \frac{e^\alpha}{2^\beta} - n_1 = \frac{e^\alpha}{2^\beta}$. We obtain

$$\begin{aligned}
|C_d| &\leqslant (1+\varepsilon_\beta)OPT_d + \varepsilon_\beta\left(\frac{1}{2}\cdot\frac{e^\alpha}{2^\beta} + \zeta(\beta)\,e^\alpha - n - e^\alpha - \frac{e^\alpha}{2^\beta}\right)\\
&= (1+\varepsilon_\beta)OPT_d + \varepsilon_\beta\cdot e^\alpha\left(\frac{1}{2^{\beta-1}} + \zeta(\beta) - \frac{1}{(d+1)^\beta} - 1 - \frac{1}{2^\beta}\right)\\
&\leqslant OPT_d\left(1+\varepsilon_\beta\left(1+(d+1)^\beta d\left(\frac{1}{2^{\beta+1}} + \zeta(\beta) - 1 - \frac{1}{2^\beta} - \frac{1}{(d+1)^\beta}\right)\right)\right) \quad .
\end{aligned}$$

If instead we choose $|W_1| = \sum_{j=4}^{\Delta}\left(\frac{e^\alpha}{j^\beta} - n_{j-1}\right)$, then we obtain $|W \setminus W_1| = \frac{e^\alpha}{2^\beta} - \underbrace{n_1}_{=0} + \frac{e^\alpha}{3^\beta} - n_2$. This yields

$$\begin{aligned}
|C_d| &\leqslant (1+\varepsilon_\beta)OPT_d\\
&\quad + \varepsilon_d\left(e^\alpha\left(\zeta(\beta) - 1 - \frac{1}{2^\beta} - \frac{1}{3^\beta}\right) + e^\alpha\,\frac{1+2^{-\beta}+3^{-\beta}}{2} - n + \frac{n_1+n_2}{2}\right)\\
&= OPT_d\left(1+\varepsilon_\beta\left(1+(d+1)^\beta\cdot d\left(\zeta(\beta) - \frac{1}{2} - \frac{1}{2^{\beta+1}} - \frac{1}{3^\beta\cdot 2}\right) - n + \frac{n_1+n_2}{2}\right)\right).
\end{aligned}$$

We obtain the following theorem.

Theorem 6.6.
Suppose that MIN-VC$_d$ *is NP-hard to approximate within approximation ratio* $1+\varepsilon_d$. *Then, for* $2 < \beta < \beta_{max} = \inf\{x \mid \zeta(x-1) - 2\zeta(x) \leqslant 0\} \approx 2.48$, MIN-VC$_{\alpha,\beta}$ *is hard to approximate within approximation ratio*

$$1 + \frac{\varepsilon_d}{1 + d(d+1)^\beta \left(\frac{1}{2^{\beta+1}} + \zeta(\beta) - 1 - \frac{1}{2^\beta} - \frac{1}{(d+1)^\beta} \right)} \ .$$

The next section covers the case when $0 < \beta \leqslant 1$. In particular, we will have to use techniques presented in Section 6.5.1 in order to construct a valid embedding reduction.

6.7 THE CASE $\beta \leqslant 1$

We consider now the case $0 < \beta \leqslant 1$. Again we construct a polynomial time reduction which embeds any d-bounded graph G_d into an (α, β)-PLG $G_{\alpha,\beta}$. Since in the case $0 < \beta \leqslant 1$, the nodes of G_d need to have high degree in the fully composed graph $G_{\alpha,\beta}$, we will first map the graph G_d to a $(d+2)$-bounded degree graph $\widetilde{G}_{d+2}$, which provides a perfect matching M. Then the edges of M are duplicated in order to increase the degree of vertices in $\widetilde{G}_{d+2}$. The reduction technique for this special case is described in Section 6.5.1. We start with the subcase $0 < \beta < 1$.

6.7.1 *Subcase* $0 < \beta < 1$

We are now ready to consider the subcase $0 < \beta < 1$. We start by giving an estimate of the cardinality of node intervals $[x\Delta, y\Delta]$ in an (α, β)-PLG $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$. Although the rounding errors in the case $\beta < 1$ can be of order $\Theta(|G_{\alpha,\beta}|)$, as pointed out by Aiello, Chung, and Lu in [ACL01, p.6], our estimates will enable us to choose the interval sizes appropriately and to obtain explicit lower bounds for MIN-VC$_{\alpha,\beta}$.

The following lemma provides both upper and lower bounds on the cardinality of the interval $[x\Delta, y\Delta]$.

Lemma 6.5 (Sizes of Intervals).
Let $0 < \beta < 1$ and let $G_{\alpha,\beta} = (V, E)$ be an (α, β)-PLG. For every $0 < x < y < 1$, the cardinality of the interval $[x\Delta, y\Delta] = \{ v \in V \mid x\Delta \leqslant d_{\alpha,\beta}(v) \leqslant y\Delta \}$ is in

$$\left[\frac{\Delta}{1-\beta} \left(y^{1-\beta} - x^{1-\beta} \right) - (y-x)\Delta - 1 \, , \, \frac{\Delta}{1-\beta} \left(y^{1-\beta} - x^{1-\beta} \right) + \left(\frac{1}{x^\beta} - \frac{1}{y^\beta} \right) \right] \ .$$

Proof. We first observe that

$$|[x\Delta, y\Delta]| = \sum_{j=x\Delta}^{y\Delta} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \in \left[\sum_{j=x\Delta}^{y\Delta} \frac{e^\alpha}{j^\beta} - (y-x)\Delta - 1 \, , \, \sum_{j=x\Delta}^{y\Delta} \frac{e^\alpha}{j^\beta} \right] \ .$$

Since, for $0 < \beta < 1$, $\frac{1}{\chi^\beta}$ is a convex function, we obtain

$$\begin{aligned}
\sum_{j=x\Delta}^{y\Delta} \frac{e^\alpha}{j^\beta} &\in \left[e^\alpha \cdot \int_{x\Delta}^{y\Delta} \chi^{-\beta} d\chi \, , \, e^\alpha \cdot \left(\int_{x\Delta}^{y\Delta} \chi^{-\beta} d\chi + \left(\frac{1}{(x\Delta)^\beta} - \frac{1}{(y\Delta)^\beta} \right) \right) \right] \\
&= \left[e^\alpha \cdot \left[\frac{\chi^{1-\beta}}{1-\beta} \right]_{x\Delta}^{y\Delta} \, , \, e^\alpha \cdot \left(\left[\frac{\chi^{1-\beta}}{1-\beta} \right]_{x\Delta}^{y\Delta} + \left(\frac{1}{(x\Delta)^\beta} - \frac{1}{(y\Delta)^\beta} \right) \right) \right] \\
&= \left[\frac{\Delta}{1-\beta} \left(y^{1-\beta} - x^{1-\beta} \right) \, , \, \frac{\Delta}{1-\beta} \left(y^{1-\beta} - x^{1-\beta} \right) + \left(\frac{1}{x^\beta} - \frac{1}{y^\beta} \right) \right] \ .
\end{aligned}$$

□

We want to choose $0 \leqslant x < y < z \leqslant 1$ in such a way that the vertices of $\widetilde{G}_{d+2}$ will be contained in the interval $[x\Delta, y\Delta]$ and the vertices of Γ will be contained in the interval $(y\Delta, z\Delta]$. The preceding Lemma 6.5 shows that, in order to achieve $|G_{\alpha,\beta}| = \mathcal{O}(n)$, we have to choose $y = \Omega(1)$. The next lemma shows that we can even choose $y = 1 - o(1)$ and $x = o(1)$, which then implies $|[x\Delta, y\Delta]| = (1 - o(1)) \cdot |G_{\alpha,\beta}|$.

Lemma 6.6.
Let $x = \frac{d+1}{\Delta}, y = \left(1 + \frac{1}{\Delta^{1-\beta}} \right)^{-\frac{1}{2-\beta}}$ and $z = 1$. Then we have that $|[x\Delta, y\Delta]| = (1 - o(1)) \cdot |G_{\alpha,\beta}|$. Furthermore, the following inequality regarding the interval size holds

$$\sum_{j=y\Delta+1}^{\Delta} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \cdot (j-2) \geqslant (d-1) \cdot n \ . \tag{6.7}$$

Proof. Due to the previous Lemma 6.5, we have that

$$|[x\Delta, y\Delta]| \geqslant \frac{\Delta}{1-\beta}\left(\left(1+\frac{1}{\Delta^{1-\beta}}\right)^{-\frac{1-\beta}{2-\beta}} - \left(\frac{d+1}{\Delta}\right)^{1-\beta}\right) - y\Delta.$$

We cannot apply Lemma 6.5 directly to the interval $[x\Delta, y\Delta]$, since the rounding error $(y-x)\Delta$ is of order $\Omega(\Delta) = \Omega(|G_{\alpha,\beta}|)$ in this case. Instead, we apply the Lemma 6.5 to the complement of $[x\Delta, y\Delta]$ in $G_{\alpha,\beta}$:

$$\begin{aligned}|[x\Delta, y\Delta]| &= |G_{\alpha,\beta}| - |(y\Delta, \Delta]| - \sum_{j=1}^{d}\left\lfloor\frac{e^\alpha}{j^\beta}\right\rfloor \\ &\geqslant |G_{\alpha,\beta}| - \frac{\Delta}{1-\beta}\underbrace{\left(1-\left(1+\frac{1}{\Delta^{1-\beta}}\right)^{-\frac{1-\beta}{2-\beta}}\right)}_{=o(1)} - \left(1+\frac{1}{\Delta^{1-\beta}}\right)^{\frac{\beta}{2-\beta}} + 1 - \Theta(e^\alpha).\end{aligned}$$

Since $|G_{\alpha,\beta}| = \Theta(\Delta)$, we obtain $|[x\Delta, y\Delta]| = (1-o(1))\cdot|G_{\alpha,\beta}|$. Now we show that for this choice of x and y, the main inequality 6.7 of this lemma holds as well. We start from the following inequality:

$$\sum_{j=y\Delta+1}^{\Delta}\left\lfloor\frac{e^\alpha}{j^\beta}\right\rfloor(j-2) \geqslant (d-1)n \ .$$

This is transformed into

$$\begin{aligned}\sum_{j=y\Delta+1}^{\Delta}\left\lfloor\frac{e^\alpha}{j^\beta}\right\rfloor(j-2) &\geqslant \sum_{j=y\Delta+1}^{\Delta}\frac{e^\alpha}{j^\beta}(j-2) - \underbrace{(\Delta - y\Delta - 1)}_{=(1-y)\Delta-1=o(\Delta)} \\ &= \sum_{j=y\Delta}^{\Delta}\frac{e^\alpha}{j^\beta}(j-2) - \frac{e^\alpha}{(y\Delta)^\beta}(y\Delta-2) - o(\Delta) \\ &\geqslant e^\alpha\left(\int_{y\Delta}^{\Delta}\chi^{1-\beta}d\chi - 2\int_{y\Delta}^{\Delta}\chi^{-\beta}d\chi - 2\left(\frac{1}{y^\beta}-1\right)\right) - \frac{e^\alpha}{(y\Delta)^\beta}(y\Delta-2) - o(\Delta).\end{aligned}$$

Now we are going to use the fact that $2(y^{-\beta}-1)\,e^{\alpha} = o(\Delta)$ and have

$$\begin{aligned}
\sum_{j=y\Delta+1}^{\Delta} \left\lfloor \frac{e^{\alpha}}{j^{\beta}} \right\rfloor (j-2) &\geqslant e^{\alpha} \left(\left[\frac{\chi^{2-\beta}}{2-\beta} \right]_{y\Delta}^{\Delta} - 2 \left[\frac{\chi^{1-\beta}}{1-\beta} \right]_{y\Delta}^{\Delta} \right) - \frac{e^{\alpha}}{(y\Delta)^{\beta}}(y\Delta-2) - o(\Delta) \\
&= e^{\alpha} \underbrace{\left(\frac{\Delta^{2-\beta}}{2-\beta} - \frac{(y\Delta)^{2-\beta}}{2-\beta} - \frac{2\Delta^{1-\beta}}{1-\beta} + \frac{2(y\Delta)^{1-\beta}}{1-\beta} \right)}_{1=z=\left(1+\frac{1}{\Delta^{1-\beta}}\right)^{\frac{1}{2-\beta}} y} - \underbrace{\frac{e^{\alpha}}{(y\Delta)^{\beta}}(y\Delta-2)}_{=\Theta(y^{1-\beta}\Delta)} - o(\Delta) \\
&= e^{\alpha} \left(\frac{\left(1+\frac{1}{\Delta^{1-\beta}}-1\right) y^{2-\beta}\Delta^{2-\beta}}{2-\beta} - \Theta\left(\Delta^{1-\beta}\right) \right) - \Theta(\Delta) \ .
\end{aligned}$$

Now using the observation that $e^{\alpha}\,\Delta^{1-\beta} = e^{\alpha}\,e^{\alpha\frac{1-\beta}{\beta}} = \Delta$, we finally yield

$$\sum_{j=y\Delta+1}^{\Delta} \left\lfloor \frac{e^{\alpha}}{j^{\beta}} \right\rfloor (j-2) \geqslant e^{\alpha}\,\frac{y^{2-\beta}\Delta^{1+\beta}}{2-\beta} - \Theta\left(\Delta\right) = e^{\alpha}\,\frac{y^{2-\beta}\Delta^{1+\beta}}{2-\beta}(1-o(1)) \ .$$

Therefore, $\sum_{j=y\Delta+1}^{\Delta} \left\lfloor \frac{e^{\alpha}}{j^{\beta}} \right\rfloor \cdot (j-2) = \omega(\Delta)$. And since $(d-1)n = \mathcal{O}(n) = \mathcal{O}(\Delta)$, inequality 6.7 holds and the lemma follows. □

The next lemma shows for the size of the fully composed graph that $|G_{\alpha,\beta}| = (1 + o(1)) \cdot n$.

Lemma 6.7.
Let $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ be a power law graph and $\alpha = \min\{\,\alpha' \mid |[x\Delta, y\Delta]| \geqslant n\,\}$, where $x = \frac{d+1}{\Delta}$ and $y = \left(1+\frac{1}{\Delta^{1-\beta}}\right)^{-\frac{1}{2-\beta}}$. Then, the size of the interval it holds that $|[x\Delta, y\Delta]| = (1 + o(1)) \cdot n$.

Proof. For $\alpha = \min\{\,\alpha' \mid |[x\Delta, y\Delta]| \geqslant n\,\}$, we have that

$$|[x\Delta, y\Delta]| = (1-\tau(\alpha)) \cdot \frac{\Delta}{1-\beta} \geqslant n \ ,$$

where $\tau(\alpha) = o(1)$. Hence

$$\begin{aligned}\alpha &= \min\left\{\alpha' \,\middle|\, \alpha' \geqslant \beta \cdot \left(\ln\left(\frac{1-\beta}{1-\tau(\alpha')}\right) + \ln(n)\right) \text{ and } \frac{\Delta}{1-\beta} \in \mathbb{Z}\right\} \\ &= \beta \cdot \left(\ln\left(\frac{1-\beta}{1-o(1)}\right) + \ln(n)\right) + o(1) \ ,\end{aligned}$$

which implies that $|G_{\alpha,\beta}| = \frac{\Delta}{1-\beta} = n \cdot \frac{1-\beta}{1-o(1)} \cdot \frac{e^{o(1)}}{1-\beta} = (1+o(1))n$. □

We obtain a polynomial time reduction from MIN-VC$_d$ to MIN-VC$_{\alpha,\beta}$ for the case $0 < \beta < 1$ in algorithm `Reduction`$_{\beta\leqslant 1}$ on page 117.

RESULTING LOWER BOUND. Let us now give an estimate of the resulting lower bound for this reduction. Suppose that MIN-VC$^{PM}_{d+2}$ is NP-hard to approximate within approximation ratio $1+\varepsilon_0$. Let $G_d = (V_d, E_d)$ be the d-bounded degree graph and $G_{\alpha,\beta} = R(G_d)$. Let $|V_d| = n$ and let W denote the wheel in the composed graph $G_{\alpha,\beta}$. We have $|W| \leqslant c \cdot n$. Let Γ be the neighborhood of G_d in $G_{\alpha,\beta}$ with $|\Gamma| = \gamma = o(n)$.

Suppose that $\mathcal{A}$ is a polynomial time approximation algorithm for MIN-VC in power law graphs with an approximation ratio $1+\varepsilon$. On input $G_{\alpha,\beta} = R(G_d)$, algorithm $\mathcal{A}$ constructs a vertex cover $C = C_d \cup C_W$ with $C_d = C \cap V_d$ and $C_W = C \cap W$. Since, in our construction, no vertex from W is able to cover any edge in G_d, C_d is a vertex cover of G_d.

Let OPT denote a minimum cost vertex cover of $G_{\alpha,\beta}$. Let $OPT_d = OPT \cap V_d$ and $OPT_W = OPT \cap W$. Then $|C| \leqslant (1+\varepsilon) \cdot OPT$, $|C_d| > n/d$ and

$$\begin{aligned}|OPT| = |OPT_d \cup OPT_W| &\leqslant |OPT_d| + |OPT_W| + |\Gamma| \\ &\leqslant (1+o(1)) \cdot (|OPT_d \cup OPT_W|) = (1+o(1)) \cdot |OPT| \ .\end{aligned}$$

Hence the approximation algorithm $\mathcal{B}$, which on input $G_{\alpha,\beta}$ first computes the cover $C = \mathcal{A}(G_{\alpha,\beta})$ and then replaces C_W by the union of Γ and an optimum vertex cover for $W \setminus \Gamma$ (which can be computed efficiently by dynamic programming due to our construction), has approximation ratio $(1+\varepsilon_0) \cdot (1+o(1))$ for the instances $G_{\alpha,\beta} = G_d \cup W = R(G_d)$ of MIN-VC.

Algorithm 6.4 : `Reduction`$_{\beta\leqslant 1}$

Input : $G_d = (V, E)$ a d-bounded graph with $V = \{v_1, \ldots, v_n\}$ such that $2 \leqslant d_{G_d}(v_1) \leqslant \ldots \leqslant d_{G_d}(v_n) \leqslant d$

Output : (α, β)-PLG $G_{\alpha,\beta} = (V_{\alpha,\beta}, E_{\alpha,\beta})$ with $V_{\alpha,\beta} = V \cup W$

(1) **let** $\widetilde{G}_{d+2} = R_{PM}(G_d) = (\widetilde{V}, \widetilde{E})$; `/* Generate the Perfect Matching Graph */`
let $\widetilde{M} = \{e_1, \ldots, e_{2n}\}$ be a perfect matching in $\widetilde{G}_{d+2}$;

(2) **let** $\alpha := \min\left\{\alpha' \,\middle|\, |[x\Delta, y\Delta]| \geqslant \left|\widetilde{G}_{d+2}\right|\right\}$; `/* Choose α, x and y */`
let $x := \frac{d+1}{\Delta}$ and $y := \left(1 + \frac{1}{\Delta^{1-\beta}}\right)^{-\frac{1}{2-\beta}}$;

(3) Assign degrees $d_{\alpha,\beta}(v_i)$ to the nodes v_i of $\widetilde{G}_{d+2}$
such that $x\Delta \leqslant d_{\alpha,\beta}(v_i) \leqslant y\Delta$, respecting the (α, β) power law;
let $d_r(v_i) = d_{\alpha,\beta}(v_i) - d_{\widetilde{G}_{d+2}}(v_i)$;
for $i = 1, \ldots, 2n$ **do**
  let $e_i = \{v_{i_1}, v_{i_2}\}$;
  Replace e_i by $\min\{d_r(v_{i_1}), d_r(v_{i_2})\} - 1$ parallel edges;
  Update $d_r(v_{i_1}), d_r(v_{i_2})$ accordingly;
let $\mu(G_d)$ be the resulting multigraph;

(4) $n_j :=$ ♯nodes v_i with $d_{\alpha,\beta}(v_i) = j$ in $\widetilde{G}_{d+2}$ $(j = 2, \ldots, \Delta)$;
$V_{\alpha,\beta} := \widetilde{V}_{d+2} \cup W$ $(W = \bigcup_{j=1}^{\Delta} W_{(j)}$ with
$W_{(j)} = \left\{w_{j,l} \,\middle|\, 1 \leqslant l \leqslant \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor - n_{j-1}\right\}$ $(j = 1, \ldots, \Delta)$;

(5) Generate edges $\{w_{j,l}, w_{j,l+1}\}$ $(j = 2, \ldots, \Delta$ and $l = 1, \ldots, |W_{(j)}|)$;
Generate edges $\{w_{j,|W_{(j)}|}, w_{j+1,1}\}$ and one edge $\{w_{\Delta,1}, w_{2,1}\}$ $(j = 2, \ldots, \Delta)$;
let $d_r(w_{j,l}) := j - 2$ $(j = 2, \ldots, \Delta, 1 \leqslant l \leqslant |W_{(j)}|)$;

(6) **for** $(c = 1, j = y\Delta + 1; c \leqslant 4n; j{+}{+})$ **do**
  for $(l = 1; l < |W_{(j)}| \wedge c \leqslant 4n;\ )$ **do**
    Generate $\min\{d_r(v_c), d_r(w_{j,l})\}$ parallel edges between v_c and $w_{j,l}$;
    Update $d_r(v_c), d_r(w_{j,l})$ accordingly;
    if $d_r(w_{j,l}) = 0$ **then** $j{+}{+}$;
    if $d_r(v_c) = 0$ **then** $c{+}{+}$;

(7) Connect the nodes $w_{1,l}$ to $\bigcup_{y\Delta<j\leqslant\Delta} W_{(j)}$; `/* c.f. step (6) of Reduction`$_{\beta>1}$ `*/`

(8) Apply algorithm `Fill_Wheel`;
let $E_{\alpha,\beta}$ be the set of edges generated in steps (3)-(8);
return $(V_{\alpha,\beta}, E_{\alpha,\beta})$;

We will now show that algorithm $\mathcal{B}$ has also a similar approximation ratio for a slightly modified optimization problem:

Problem 5 ($\widehat{\text{MIN-VC}}$).

INPUT: d-bounded degree graph G_d.

OUTPUT: Vertex cover C for $R(G_d) = G_d \cup W$ such that $\Gamma \subseteq C$.

OBJECTIVE: Minimize $|C|$.

Let $\widehat{OPT}(G_d)$ denote an optimum solution for instance G_d of this modified optimization problem. Furthermore let $OPT(G_d)$ and $OPT(W \setminus \Gamma)$ denote minimum cost vertex covers for G_d and the graph $W \setminus \Gamma$, respectively. Then

$$\widehat{OPT}(G_d) = OPT(G_d) \cup \Gamma \cup OPT(W \setminus \Gamma) \ .$$

We observe that if $OPT = OPT_d \cup OPT_W$ is an optimum vertex cover for $G_{\alpha,\beta} = R(G_d)$, then

$$|OPT_d| \leqslant |OPT(G_d)| + |\Gamma| \ .$$

We have $|\widehat{OPT}(G_d)| \leqslant |OPT| + |\Gamma| + \mathcal{O}(1) = (1 + o(1)) \cdot |OPT|$. We show that algorithm $\mathcal{B}'$ has approximation ratio $(1 + o(1)) \cdot (1 + \varepsilon)$ for the modified optimization problem. Then we can conclude that

$$\begin{aligned} |C_d \cup \Gamma \cup OPT(W \setminus \Gamma)| &\leqslant (1 + o(1)) \cdot (1 + \varepsilon) \cdot \widehat{OPT}(G_d) \\ &= (1 + o(1)) \cdot (1 + \varepsilon) \cdot |OPT(G_d) \cup \Gamma \cup OPT(W \setminus \Gamma)| \ , \end{aligned}$$

which yields

$$|C_d| \leqslant (1 + o(1)) \cdot (1 + \varepsilon) \cdot |OPT(G_d)| + \varepsilon \cdot |\Gamma \cup OPT(W \setminus \Gamma)| \ .$$

Now since $|\Gamma \cup \mathsf{OPT}(W \setminus \Gamma)| \leqslant |W| \leqslant c \cdot n$ and $|\mathsf{OPT}(G_d)| \geqslant n/d$, we obtain $|\Gamma \cup \mathsf{OPT}(W \setminus \Gamma)| \leqslant c \cdot d \cdot |\mathsf{OPT}(G_d)|$ and therefore

$$|C_d| \leqslant (1 + o(1)) \cdot (1 + \varepsilon + \varepsilon \cdot c \cdot d) \cdot |\mathsf{OPT}(G_d)| \ ,$$

and thus $\varepsilon \cdot (1 + c \cdot d) \geqslant (1 - o(1)) \cdot \varepsilon_0$. For our choice of the parameters x, y, z we obtain $c = o(1)$, and thus the following theorem holds.

Theorem 6.7.
If MIN-VC$_d$ *is hard to approximate within approximation ratio* $1 + \varepsilon_d$*, then for* $0 < \beta < 1$*,* MIN-VC$_{\alpha,\beta}$ *is hard to approximate within approximation ratio* $1 + \frac{\varepsilon_d}{1+2d}$.

Now we are going to consider the case $\beta = 1$.

6.7.2 *Subcase* $\beta = 1$

The case $\beta = 1$ differs from the case $0 < \beta < 1$ in how we choose the intervals $[x\Delta, y\Delta]$ and $(y\Delta, z\Delta]$. Nevertheless, we will obtain the same lower bound as in the case $0 < \beta < 1$. We start with the proof of the following lemma.

Lemma 6.8 (Sizes of Intervals)**.**
Let $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ *be an* (α, β)*-PLG with* $\beta = 1$*. Then for all* $0 < x < y \leqslant 1$*, the size of the interval* $[x\Delta, y\Delta] = \{v \in V(G_{\alpha,\beta}) \mid x\Delta \leqslant d(v) \leqslant y\Delta\}$ *satisfies*

$$|[x\Delta, y\Delta]| \in \left[(\ln(y) - \ln(x) - (y - x + 1)) \cdot e^{\alpha} \, , \, e^{\alpha} \cdot (\ln(y) - \ln(x)) + \left(\frac{1}{x} - \frac{1}{y} \right) \right] \ .$$

Proof. First we give a bound for the rounding error of the sum $\sum_{j=x\Delta}^{y\Delta} \left\lfloor \frac{e^{\alpha}}{j} \right\rfloor$, namely

$$\sum_{j=x\Delta}^{y\Delta} \frac{e^{\alpha}}{j} - (y - x + 1) \cdot \Delta \leqslant \sum_{j=x\Delta}^{y\Delta} \left\lfloor \frac{e^{\alpha}}{j} \right\rfloor \leqslant \sum_{j=x\Delta}^{y\Delta} \frac{e^{\alpha}}{j} \ .$$

For the sum $\sum_{j=x\Delta}^{y\Delta} j^{-1}$, we get the following bounds:

$$\begin{aligned}\sum_{j=x\Delta}^{y\Delta} \frac{1}{j} &\in \left[\int_{x\Delta}^{y\Delta} \chi^{-1}\,d\chi\,,\ \int_{x\Delta}^{y\Delta} \chi^{-1}\,d\chi + \left(\frac{1}{x\Delta} - \frac{1}{y\Delta}\right)\right] \\ &= \left[\ln(y\Delta) - \ln(x\Delta), \ln(y\Delta) - \ln(x\Delta) + \left(\frac{1}{x\Delta} - \frac{1}{y\Delta}\right)\right] \\ &= \left[\ln(y) - \ln(x)\,,\ \ln(y) - \ln(x) + \left(\frac{1}{x\Delta} - \frac{1}{y\Delta}\right)\right]\ ,\end{aligned}$$

and thus the lemma follows. □

In the case $\beta < 1$ we have mapped the graph G_d to a subinterval $[x\Delta, y\Delta) = \{v \in V(G_{\alpha,\beta}) \mid x\Delta \leqslant d(v) < y\Delta\}$, where $0 < x < y < 1$ and x, y are constant. However, in the case $\beta = 1$ the size of such an interval is $\Theta(e^{\alpha})$ which is $o(|\mathcal{G}_{\alpha,1}|)$. This means we have to choose the interval bounds in a different way.

Lemma 6.9.
Let $G_{\alpha,\beta} = (V, E)$ be an $(\alpha, 1)$-PLG. For $0 \leqslant c' < c \leqslant 1$ and parameters $x = e^{-(1-c')\alpha}, y = e^{-(1-c)\alpha}$ with c, c' being constant, the size of the set $[x\Delta, y\Delta] = \{v \in V \mid x\Delta \leqslant d(v) \leqslant y\Delta\}$ satisfies

$$|[x\Delta, y\Delta)| = (1 - o(1))(c - c') \cdot \alpha\, e^{\alpha} = \Theta(\alpha\, e^{\alpha})\ .$$

Proof. Due to the preceding Lemma 6.8, the following holds

$$\begin{aligned}|[x\Delta, y\Delta]| &\geqslant e^{\alpha} \cdot (\ln(y) - \ln(x) - (y - x + 1)) \\ &= e^{\alpha} \cdot ((1 - c') - (1 - c))\alpha - \mathcal{O}(1)) \\ &= (1 - o(1)) \cdot (c - c')\alpha\, e^{\alpha}\ .\end{aligned}$$

□

The next lemma shows that if we choose α as small as possible such as to be able to embed G_d into the interval $[x\Delta, y\Delta]$, then we obtain $|[x\Delta, y\Delta]| = (1 + o(1))n$.

Lemma 6.10.
For $x = \frac{1}{e^{(1-c')\cdot\alpha}}, y = \frac{1}{e^{(1-c)\cdot\alpha}}$ *and* $\alpha = \min\{\,\alpha' \mid |[x\Delta, y\Delta)| \geqslant n\,\}$*, we get the following bounds on the size of the interval* $[x\Delta, y\Delta)$*.*

$$n \leqslant |[x\Delta, y\Delta)| \leqslant n + t(n) \ ,$$

where $t(n) = \lfloor y\Delta \rfloor - \lceil x\Delta \rceil + \mathcal{O}(1)$ *and especially* $t(n) = o(n)$*.*

Proof. The equation $t(n) = \lfloor y\Delta \rfloor - \lceil x\Delta \rceil + \mathcal{O}(1)$ follows directly from the choice of x and y. It remains to show that $t(n) = o(n)$. From $t(n) = o(|G_{\alpha,\beta}|)$ and $|[x\Delta, y\Delta)| = \Theta(|G_{\alpha,\beta}|)$, we obtain $t(n) = o(|[x\Delta, y\Delta)|)$. The inequality $n \leqslant |[x\Delta, y\Delta)| \leqslant n + t(n)$ then implies $n = \Theta(|[x\Delta, y\Delta)|)$, whence $t(n) = o(n)$. □

Finally we show that we can choose the set $\Gamma = (y\Delta, z\Delta]$ with $z = c''\Delta$, where $c'' = c + o(1)$.

Lemma 6.11.
Let $0 \leqslant c' < c < 1$ *be constants and* $c'' = c + \frac{1}{\alpha}$*, and let* $x = e^{(c'-1)\alpha}, y = e^{(c-1)\alpha}$ *and* $z = e^{(c''-1)\alpha}$*. Then* $|(y\Delta, z\Delta]| = o(\alpha\, e^{\alpha})$ *and*

$$\sum_{j=y\Delta+1}^{z\Delta} \left\lfloor \frac{e^{\alpha}}{j} \right\rfloor \cdot (j-2) = \omega\left(|G_{\alpha,\beta}|\right) \ .$$

Proof. Using the preceding Lemma 6.10, we have that

$$\begin{aligned} |(y\Delta, z\Delta]| &= |[y\Delta, z\Delta]| - \left\lfloor \frac{e^{\alpha}}{y\Delta} \right\rfloor \\ &\leqslant e^{\alpha}\cdot(c''-c)\cdot\alpha + e^{(1-c)\alpha} - e^{(1-c'')\alpha} \\ &= e^{\alpha} + e^{(1-c)\alpha} - e^{(1-c'')\alpha} \\ &= o\left(|G_{\alpha,\beta}|\right) \ . \end{aligned}$$

Furthermore,

$$\begin{aligned}
\sum_{j=y\Delta+1}^{z\Delta} \left\lfloor \frac{e^\alpha}{j} \right\rfloor \cdot (j-2) &\geqslant \sum_{j=y\Delta+1}^{z\Delta} \frac{e^\alpha}{j} \cdot (j-2) - (z-y)\Delta \cdot (z\Delta - 2) \\
&= \sum_{j=y\Delta}^{z\Delta} \frac{e^\alpha}{j} \cdot (j-2) - (z-y)\Delta \cdot (z\Delta - 2) - \frac{e^\alpha}{y\Delta} \\
&\geqslant e^\alpha \cdot ((z-y+1)\Delta - 2(\ln(z\Delta) - \ln(y\Delta))) \\
&\quad - (z-y)\Delta \cdot (z\Delta - 2) - \frac{e^\alpha}{y\Delta} - 2\left(\frac{1}{y\Delta} - \frac{1}{z\Delta}\right) \\
&= (1 - o(1)) \cdot e^{2\alpha} \\
&= \omega\left(\alpha e^\alpha\right) \quad .
\end{aligned}$$

□

This means, for $\beta = 1$, we basically obtain the same polynomial time reduction from MIN-VC$_d$ to MIN-VC$_{\alpha,\beta}$ as in the case $0 < \beta < 1$. The only difference is how we choose the parameters x, y, z. Altogether, we obtain the following result.

Theorem 6.8.
If MIN-VC$_d$ *is hard to approximate within ratio* $1 + \varepsilon_d$*, then for* $\beta = 1$, MIN-VC$_{\alpha,\beta}$ *is hard to approximate within approximation ratio* $1 + \frac{\varepsilon_d}{1+2d}$.

We investigate now the phase transition point for $\beta = 1$. This is done by defining the parameter $\beta = \beta_f = 1 \pm \frac{1}{f(n)}$ as being a function only dependent on the number of nodes n, and thus converging to 1 from below (above) as $n \to \infty$. We start with the case $\beta = 1 - \frac{1}{f(n)}$.

6.8 THE FUNCTIONAL CASE $\beta = 1 - \frac{1}{f(n)}$

First, let us give a precise description of the model of an (α, β_f)-PLG for $\beta_f = 1 - \frac{1}{f(n)}$.

Definition 6.5 ((α, β_f)-PLG for $\beta_f = 1 - \frac{1}{f(n)}$)**.**
Let $f(n)$ be a monotone increasing unbounded function. For $\beta_f = 1 - \frac{1}{f(n)}$, an (α, β_f)-PLG $G_{\alpha,\beta_f} \in \mathcal{G}_{\alpha,\beta_f}$ is a multigraph with n nodes which has the following properties:

(1) The maximum degree of G_{α,β_f} is $\Delta_f = \left\lfloor e^{\alpha/\beta_f} \right\rfloor$.

(2) There are $\left\lfloor \frac{e^\alpha}{j^{\beta_f}} \right\rfloor$ nodes of degree j, with $j = 1, \ldots, \Delta_f$.

Thus, in the above definition of an (α, β_f)-PLG, the number of vertices in a graph $G_{\alpha,\beta_f} \in \mathcal{G}_{\alpha,\beta_f}$ satisfies the equation $n = \sum_{j=1}^{\Delta_f} \left\lfloor \frac{e^\alpha}{j^{\beta_f}} \right\rfloor$. In this section we will show that the approximation lower bound for MIN-VC$_{\alpha,1}$ also holds for the functional case $\beta = 1 - \frac{1}{f(n)}$. We achieve this by showing that the quantities of the crucial parameters (maximum degree, sizes of intervals) of (α, β_f)-PLG in the functional case converge to those in the case $\beta = 1$ and general (α, β)-PLG. Let us start by giving an outline of the main steps. First we ask when the single terms $\frac{1}{i^{1-f(n)^{-1}}}$ that appear in the sum $\sum_{j=1}^{\Delta_f} \left\lfloor \frac{e^\alpha}{j^{\beta_f}} \right\rfloor$ converge to the terms $\frac{1}{i}$. The differences between the single terms in the functional and the non-functional case, i. e. the *local errors*, are

$$\frac{1}{i^{\frac{f(n)-1}{f(n)}}} - \frac{1}{i} = \frac{i^{\frac{1}{f(n)}} - 1}{i} \leqslant \frac{n^{\frac{1}{f(n)}} - 1}{i} \ .$$

We have that $\log\left(n^{1/f(n)}\right) = \frac{\log(n)}{f(n)}$ and hence, for $f(n) = \omega(\log n)$, the nominator converges to 0. Another estimate of the local error can be obtained by the following bounds.

$$\frac{i^{1/f(n)} - 1}{i} \leqslant \frac{\Delta^{1/f(n)} - 1}{i} = \left(\frac{n}{f(n)}\right)^{\frac{1}{f(n)}} - 1 \ .$$

In the next step, we are going to deal with the *global error*, i. e. we consider the sum

$$\sum_{i=1}^{\Delta} \frac{e^\alpha}{i^\beta} = e^\alpha \cdot \sum_{i=1}^{e^{\alpha \cdot \frac{f(n)}{f(n)-1}}} \frac{1}{i^{\frac{f(n)-1}{f(n)}}} \stackrel{!}{=} n \ .$$

We will show that this sum differs from the according sum of the terms $1/i$ in the non-functional case by an amount of $\alpha e^{\alpha} \frac{f(n)}{f(n)-1} \left(e^{\frac{\alpha}{f(n)-1}} - 1\right)$, which is a lower oder term provided $f(n) = \omega(\alpha)$.

Finally, we will also give bounds on the rounding error, which results when we replace the terms $\lfloor e^{\alpha}/j^{\beta} \rfloor$ by their fractional counterparts.

DETAILED DESCRIPTION. Let us now give a detailed description of the steps outlined above. We let $\Delta_f = \left\lfloor e^{\alpha \cdot \frac{f(n)}{f(n)-1}} \right\rfloor$ and $\Delta = \lfloor e^{\alpha} \rfloor$. Thus, we have that $\Delta_f = \left\lfloor e^{\alpha} \cdot e^{\frac{\alpha}{f(n)-1}} \right\rfloor = (1 + o(1)) \cdot \Delta$, provided that f satisfies $f(n) = \omega(\alpha)$. In the case $\beta_f = 1 - \frac{1}{f(n)}$, we have $n = \sum_{j=1}^{\Delta_f} \left\lfloor \frac{e^{\alpha}}{j^{\beta_f}} \right\rfloor$ and we want to give upper and lower bounds for this term. Since $\Delta_f = (1 + o(1))\Delta$, we obtain

$$\sum_{j=1}^{\Delta} \frac{e^{\alpha}}{j^{\beta}} - (1 + o(1))\Delta \leqslant \sum_{j=1}^{\Delta_f} \frac{e^{\alpha}}{j^{\beta}} - \Delta_f \leqslant \sum_{j=1}^{\Delta_f} \left\lfloor \frac{e^{\alpha}}{j^{\beta}} \right\rfloor \leqslant \sum_{j=1}^{\Delta_f} \frac{e^{\alpha}}{j^{\beta}} \ .$$

The right-hand side of this inequality, i. e. the upper bound, can be further bounded and rearranged as follows.

$$\begin{aligned}
\sum_{j=1}^{\Delta_f} \frac{e^{\alpha}}{j^{\beta}} &= \sum_{j=1}^{\Delta} \frac{e^{\alpha}}{j^{\beta}} + (\Delta_f - \Delta) \cdot \frac{e^{\alpha}}{\Delta^{\beta}} & &= \sum_{j=1}^{\Delta} \frac{e^{\alpha}}{j^{\beta}} + o(1) \cdot \Delta \cdot \frac{e^{\alpha}}{\Delta^{1-\frac{1}{f(n)}}} \\
&= \sum_{j=1}^{\Delta} \frac{e^{\alpha}}{j^{\beta}} + o(1) \cdot e^{\alpha} \cdot \Delta^{\frac{1}{f(n)}} & &= \sum_{j=1}^{\Delta} \frac{e^{\alpha}}{j^{\beta}} + o(1) \cdot e^{\alpha} \cdot e^{\frac{\alpha}{f(n)}} \\
&= \sum_{j=1}^{\Delta} \frac{e^{\alpha}}{j^{\beta}} + o(1) \cdot (1 + o(1)) \cdot e^{\alpha} \ ,
\end{aligned}$$

where the last equality holds again due to the fact that $f(n) = \omega(\alpha)$. Thus we obtain the following lemma.

Lemma 6.12.
If $f(n) = \omega(\alpha)$, *then*

$$\sum_{j=1}^{\Delta} \frac{e^{\alpha}}{j^{\beta}} - (1 + o(1))\Delta \leqslant \sum_{j=1}^{\Delta_f} \left\lfloor \frac{e^{\alpha}}{j^{\beta}} \right\rfloor = n \leqslant \sum_{j=1}^{\Delta} \frac{e^{\alpha}}{j^{\beta}} + o(1) \cdot \Delta \ .$$

We also need similar bounds for the sizes of intervals. It turns out that we can restrict ourselves to intervals $[a(n), b(n)]$ in (α, β)-PLG where $b(n) \leqslant \lfloor e^{\alpha} \rfloor$, instead of $b(n) \leqslant \Delta_f = \left\lfloor e^{\alpha \cdot \frac{f(n)}{f(n)-1}} \right\rfloor$ in the functional case.

Lemma 6.13.
Suppose $f(n) = \omega(\alpha)$. *Let* $a, b \colon \mathbb{N} \to \mathbb{N}$ *such that for all* n, $1 \leqslant a(n) < b(n) \leqslant \lfloor e^{\alpha} \rfloor$. *Then*

$$\sum_{j=a(n)}^{b(n)} \frac{e^{\alpha}}{j^{\beta}} - (b(n) - a(n) + 1) \leqslant |[a(n), b(n)]| = \sum_{j=a(n)}^{b(n)} \left\lfloor \frac{e^{\alpha}}{j^{\beta}} \right\rfloor \leqslant \sum_{j=a(n)}^{b(n)} \frac{e^{\alpha}}{j^{\beta}} \ .$$

The next lemma gives the desired bounds for the sizes of intervals in the functional case $\beta_f = 1 - \frac{1}{f(n)}$. The upper and lower bounds are sums of terms $\frac{e^{\alpha}}{j}$ instead of $\frac{e^{\alpha}}{j^{\beta_f}}$. Afterwards we will use this result to show that we can actually choose the same parameters $x\Delta, y\Delta, z\Delta$ as in the case $\beta = 1$.

Lemma 6.14 (Convergence of Sizes of Intervals)**.**
For each pair of functions $a, b \colon \mathbb{N} \to \mathbb{N}$ *with* $1 \leqslant a(n) < b(n) \leqslant \Delta_f$, *we have*

$$\sum_{j=a(n)}^{b(n)} \frac{e^{\alpha}}{j} - \frac{n}{\log(n)} \leqslant |[a(n), b(n)]| \leqslant (1 + \varepsilon(n)) \cdot \sum_{j=a(n)}^{b(n)} \frac{e^{\alpha}}{j}, \tag{6.8}$$

where $\varepsilon(n) = n^{\frac{1}{f(n)-1}}$. *Especially* $\varepsilon(n) = o(1)$ *for* $f(n) = \omega(\log(\alpha))$, *which implies*

$$(1 - o(1)) \sum_{j=a(n)}^{b(n)} \frac{e^{\alpha}}{j} \leqslant |[a(n), b(n)]| \leqslant (1 + o(1)) \sum_{j=a(n)}^{b(n)} \frac{e^{\alpha}}{j} \ .$$

Proof. The second inequality in Equation 6.8 holds if for each j, $\frac{1}{j^{1-\frac{1}{f(n)}}} \leqslant (1 + \varepsilon(n)) \cdot \frac{1}{j}$, i. e. $j^{\frac{1}{f(n)}} \leqslant 1 + \varepsilon(n) \Leftrightarrow j \leqslant (1 + \varepsilon(n))^{f(n)} \Leftarrow \Delta_f \leqslant (1 + \varepsilon(n))^{f(n)}$. This last inequality

is equivalent to $e^{\frac{\alpha}{f(n)-1}} \leqslant 1+\varepsilon(n)$ and after taking the logarithm, we have $\frac{\alpha}{f(n)-1} \leqslant f(n)\cdot\ln(1+\varepsilon(n))$. Since $f(n)=\omega(\alpha)$, we obtain $\varepsilon(n) \geqslant e^{\frac{\alpha}{f(n)-1}}-1 = n^{\frac{1-o(1)}{f(n)-1}}$, hence for $\varepsilon(n)=n^{\frac{1}{f(n)-1}}$ we obtain

$$\sum_{j=a(n)}^{b(n)} \frac{e^\alpha}{j} - (b(n)-a(n)+1) \leqslant |[a(n),b(n)]| \leqslant (1+\varepsilon(n))\cdot \sum_{j=a(n)}^{b(n)} \frac{e^\alpha}{j} \ . \tag{6.9}$$

We have $b(n)-a(n)+1 \leqslant \Delta_f = (1+o(1))\cdot\Delta$. Now we consider inequality 6.9 for the special case when $a(n)=1$ and $b(n)=\Delta_f$. We obtain:

$$(1+o(1))\sum_{j=1}^{\Delta}\frac{e^\alpha}{j} - (1+o(1))\Delta = \sum_{j=1}^{\Delta_f}\frac{e^\alpha}{j} - (1+o(1))\Delta \leqslant n = |[1,\Delta_f]| \ ,$$

which implies $n=(1+o(1))\cdot\alpha e^\alpha$. □

Now let $x=\frac{1}{e^{(1-c')\alpha}}, y=\frac{1}{e^{(1-c)\alpha}}, z=\frac{1}{e^{(1-c'')\alpha}}$ with $0\leqslant c'<c<c''=c+\frac{1}{\alpha}$. Combining Lemma 6.14 with the proof of Lemma 6.8, we obtain

$$(c-c')\alpha e^\alpha - \frac{n}{\log(n)} \leqslant |[x\Delta,y\Delta]| \leqslant \left((c-c')\alpha e^\alpha + \frac{1}{x\Delta}-\frac{1}{y\Delta}\right)\cdot(1+o(1))$$

and

$$e^\alpha - \frac{n}{\log(n)} - \frac{e^\alpha}{(y\Delta)^{\beta_f}} \leqslant |(y\Delta,z\Delta]| \leqslant \left(e^\alpha+\frac{1}{z\Delta}-\frac{1}{y\Delta}\right)\cdot(1+o(1)) \ ,$$

which yields $|[x\Delta,y\Delta]| = (1\pm o(1))(c-c')\alpha e^\alpha$ and $|(y\Delta,z\Delta]| = o(\alpha e^\alpha)$. Now choose $c=1-\frac{1}{\alpha}$ and $c'=\frac{d+1}{\Delta}$. Then

$$\begin{aligned}
\sum_{j=y\Delta+1}^{\Delta}\left\lfloor\frac{e^\alpha}{j^{\beta_f}}\right\rfloor(j-2) &\geqslant \sum_{j=y\Delta}^{\Delta}\frac{e^\alpha}{j^{\beta_f-1}} - \frac{1-y}{2}\Delta^2 - 2(1+o(1))\,e^\alpha \\
&\geqslant e^\alpha\int_{y\Delta}^{z\Delta}\chi^{1/f(n)}d\chi - \left(1+o(1)-\frac{1}{e}\right)\frac{\Delta^2}{2} \\
&= e^\alpha\,\frac{f(n)}{f(n)+1}\left(\Delta^{1+\frac{1}{f(n)}}-(y\Delta)^{1+\frac{1}{f(n)}}\right) - \left(1+o(1)-\frac{1}{e}\right)\frac{\Delta^2}{2} \\
&= \frac{1-e^{-1}-o(1)}{2}\,e^{\alpha\left(2+\frac{1}{f(n)}\right)} \quad = \quad \omega\left(\left|G_{\alpha,\beta_f}\right|\right) \ .
\end{aligned}$$

Hence we can map G'_d to the interval $[x\Delta, y\Delta]$ and choose $\Gamma \subseteq (y\Delta, z\Delta]$ (note that we choose $\Delta = \lfloor e^{\alpha} \rfloor$ instead of $\Delta_f = \left\lfloor e^{\alpha/\beta_f} \right\rfloor$) and obtain the same hardness result as in the case $\beta \leqslant 1$ when β is a constant. We have the following result.

Theorem 6.9.
Suppose MIN-VC$_d$ *is hard to approximate within ratio* $1 + \varepsilon_d$. *Let* $f\colon \mathbb{N} \to \mathbb{N}$ *be a function such that* $f(n) = \omega(\log(n))$. *Then for* $\beta_f = 1 - \frac{1}{f(n)}$, *the problem* MIN-VC$_{\alpha,\beta_f}$ *is hard to approximate within approximation ratio* $1 + \frac{\varepsilon_d}{1+2d}$.

Lastly, we consider the second functional case when $\beta = 1 + \frac{1}{f(n)}$, that is, the model parameter β converges to 1 from above.

6.9 THE FUNCTIONAL CASE $\beta = 1 + \frac{1}{f(n)}$

It turns out that even in the functional case $\beta = 1 + \frac{1}{f(n)}$, we obtain the same hardness result as in the case $\beta \leqslant 1$. This is especially interesting since we have the phase transition at $\beta = 1$, which is also reflected by our hardness results for β being a constant. Again, our result is based on an estimate of the sizes of intervals $[a(n), b(n)]$.

Lemma 6.15.
Let $\beta_f = 1 + \frac{1}{f(n)}$ *with* $f(n) = \omega(\alpha)$, *and let* $\Delta_f = \left\lfloor e^{\alpha/\beta_f} \right\rfloor$. *Then for each* $j \in \{1, \dots, \Delta_f\}$,

$$\frac{1}{j^{\beta_f}} \in \left[\frac{1}{j} - \tau(n)\,,\, \frac{1}{j}\right] \quad ,$$

where $\tau(n) = \frac{2^{1/f(n)}-1}{2^{1+1/f(n)}}$. *Especially,* $\tau(n) \to 0$ *as* $n \to \infty$.

Proof. We have $\frac{1}{j^{\beta_f}} \in \left[\frac{1}{j} - t(j), \frac{1}{j}\right]$ for $t(j) = \frac{1}{j} - \frac{1}{j^{\beta}} = \frac{j^{\beta_f - 1}-1}{j^{\beta_f}}$. It suffices to show that $t(j) \leqslant \tau(n) = \frac{2^{1/f(n)}-1}{2^{1+1/f(n)}}$. Since for fixed n, the derivative of the function $x \mapsto \frac{x^{1/f(n)}-1}{x^{1+1/f(n)}} = x^{-1} - x^{-1-\frac{1}{f(n)}}$ is equal to $-x^{-2} + \left(1 + \frac{1}{f(n)}\right) x^{-2-\frac{1}{f(n)}} \leqslant 0$, the inequality $t(j) \leqslant t(2) = \tau(n)$ holds, and thus the lemma follows. □

If we combine this result with our techniques from previous sections, the resulting estimate for sizes of intervals is rather weak. For the number of nodes of the α, β-PLG, we obtain

$$\frac{\alpha\, e^{\alpha}}{1+\frac{1}{f(n)}} - \frac{e^{\left(1+\frac{f(n)}{f(n)+1}\right)\alpha+\mathcal{O}(1)}}{2^{1+1/f(n)}} \leqslant |[1,\Delta_f]| \leqslant (1\pm o(1))\alpha\, e^{\alpha} \quad .$$

We will make use of the following estimate of the local rounding errors.

Lemma 6.16.
For every $j \in \{1,\dots,\Delta_f\}$, *we have that* $\frac{1}{j^{1+\frac{1}{f(n)}}} \in \left[\frac{1}{n^{\frac{1}{f(n)}}}\cdot\frac{1}{j}\,,\,\frac{1}{j}\right]$.

Proof. We just observe that $\frac{1}{j^{1+\frac{1}{f(n)}}} = \frac{1}{j}\cdot\frac{1}{j^{1/f(n)}}$, and the function $x \mapsto \frac{1}{x^{1/f(n)}}$ is monotone decreasing. □

This gives the following estimate of sizes of intervals.

Lemma 6.17 (Sizes of Intervals)**.**
In the case $\beta_f = 1+\frac{1}{f(n)}$, *for any* $1 \leqslant a < b \leqslant \Delta_f = \left\lfloor e^{\alpha/\beta_f}\right\rfloor$, *the size of the interval* $[a,b] = \{v \in V(G_{\alpha,\beta}) \mid a \leqslant d(v) \leqslant b\}$ *is in*

$$\left[\frac{1}{n^{\frac{1}{f(n)}}}\cdot e^{\alpha}\cdot(\ln(b)-\ln(a)) - (b-a+1)\,,\, e^{\alpha}\cdot(\ln(b)-\ln(a)) + e^{\alpha}\cdot\left(\frac{1}{a}-\frac{1}{b}\right)\right] \quad .$$

Since for $f(n) = \omega(\ln(n))$, we have the convergence $n^{\frac{1}{f(n)}} \to 1$ as $n \to \infty$) and obtain the following estimates.

Corollary.
For $f(n) = \omega(\ln(n))$, *the number of nodes of an* (α,β_f)*-PLG* G_{α,β_f} *satisfies* $|[1,\Delta_f]| = (1\pm o(1))\cdot\alpha\, e^{\alpha}$. *Furthermore, for the parameters* $x = \frac{1}{e^{(1-c')\alpha}}, y = \frac{1}{e^{(1-c)\alpha}}, z = \frac{1}{e^{(1-c'')\alpha}}$ *with* $0 \leqslant c' < c < c'' = c+\frac{1}{\alpha}$, *we have that* $|[x\Delta_f, y\Delta_f]| = (1\pm o(1))\cdot|[1,\Delta_f]|$ *and* $|(y\Delta_f, z\Delta_f]| = (1\pm o(1)\, e^{\alpha} = o(|G_{\alpha,\beta_f}|)$.

Now we show that if we choose the parameters x, y, z such that $z = 1$, then the amount of node-degree in the interval $(y\Delta_f, z\Delta_f]$ suffices to connect all the nodes

from $\widetilde{G}_d$ as well as the degree 1 nodes. Namely, choose $c'' = 1, c = 1 - \frac{1}{\alpha}$. Then we obtain

$$\begin{aligned}
\sum_{j=y\Delta_f+1}^{\Delta_f} \left\lfloor \frac{e^\alpha}{j^{\beta_f}} \right\rfloor (j-2) &\geqslant \sum_{j=y\Delta_f}^{\Delta_f} \frac{e^\alpha}{j} \cdot \frac{1}{n^{1/f(n)}} \cdot (j-2) - (j-2) \\
&= \frac{e^\alpha}{n^{1/f(n)}} \cdot (1-y)\Delta_f - \frac{2\,e^\alpha}{n^{1/f(n)}} \cdot \sum_{j=y\Delta_f+1}^{\Delta_f} \frac{1}{j} \\
&= (1-o(1)) \cdot \left(1 - \frac{1}{e}\right) \cdot e^{\alpha\left(1+\frac{f(n)}{1+f(n)}\right)} \\
&= \omega\left(\left|G_{\alpha,\beta_f}\right|\right) \quad .
\end{aligned}$$

Finally we obtain the following result.

Theorem 6.10.
Suppose Min-VC$_d$ *is hard to approximate within ratio* $1 + \varepsilon_d$. *Let* $f\colon \mathbb{N} \to \mathbb{N}$ *be a function such that* $f(n) = \omega(\log(n))$. *Then for* $\beta_f = 1 + \frac{1}{f(n)}$, *the problem* Min-VC$_{\alpha,\beta_f}$ *is hard to approximate within approximation ratio* $1 + \frac{\varepsilon_d}{1+2d}$.

We conclude the present chapter with a short summary of the results, an outlook on further research and some bibliographic notes.

6.10 SUMMARY AND FURTHER RESEARCH

In this chapter, we have given explicit lower bounds for the approximability of Min-VC in connected (α, β)-PLG. It remains an important open question to close the gaps between inapproximability and approximability bounds of the underlying problems. We also believe that our results for the two functional cases $\beta = 1 \pm \frac{1}{f(n)}$ can be extended to hold for any $\beta_f = \beta \pm \frac{1}{f(n)}$ with $0 < \beta < \beta_{max} \approx 2.48$. It would also be interesting to study the approximation complexity of various network design problems on power law graphs, e. g. the *Steiner Tree Problem* and related problems.

6.11 BIBLIOGRAPHIC NOTES

The material and the results presented in this chapter are based on the following publication: Mikael Gast, Mathias Hauptmann, and Marek Karpinski. "Improved approximation lower bounds for vertex cover on power law graphs and some generalizations." In: *Computing Research Repository (CoRR) preprint arXiv:1210.2698 [cs.CC]* (Oct. 2012), pp. 1–26. arXiv: 1210.2698.

In particular, the proofs of our main theorems (Theorem 6.3 on page 100; Theorem 6.4 on page 109; Theorem 6.5 on page 110; Theorem 6.6 on page 111; Theorem 6.7 on page 119; Theorem 6.8 on page 122; Theorem 6.9 on page 127; Theorem 6.10 on page 129) also appeared in [GHK12a].

7 THE APPROXIMABILITY OF MINIMUM DOMINATING SET IN POWER LAW GRAPHS

7.1 INTRODUCTION

In the research of epidemic spreading of diseases across networks of travel routes and networks of social contacts [PV01; Eub+04a] or the broadcasting of information inside large wireless networks [SSŽ02], a natural question arises about how to efficiently place key nodes at key positions inside a network such as to reach and to effect all or most of the remaining nodes. Here, the feasibility of a solution also heavily depends on the number of key nodes needed in order to cover the whole network and thus this number is often tried to be minimized. Questions like these quickly resemble or are equivalent to classical NP-hard optimization problems, i. e. minimum covering and domination problems in the context of graph theory.

In this chapter we will focus on the MINIMUM DOMINATING SET (MIN-DS) problem on graphs. On a graph $G = (V, E)$ a *dominating set* is a subset of vertices $D \subseteq V$ such that every vertex in $V \setminus D$ is connected to D by some edge in E. The MIN-DS problem then asks for a dominating set of minimum cardinality $|D|$. There exist approximation algorithms that achieve an approximation ratio of $\ln n$, with n being the size of the graph. This is essentially tight due to the lower bound of $\ln n$ by Feige [Fei98] under the assumption that $\mathsf{NP} \notin \mathsf{DTIME}(n^{\mathcal{O}(\log \log n)})$.

For the case of power law graphs, the situation is not perfectly clear. One the one hand there is a result by Eubank et al. [Eub+04b] that gives a $(1 + o(1))$-approximation (a PTAS) for the special case of the so called *near-optimal* MIN-DS problem on a class of *bipartite* random power law graphs. On the other hand, there is a result of Shen et al. [She+12] that rules out the possibility of a PTAS for general (α, β)-power law graphs ((α, β)-PLG), by proving a constant factor approximation lower bound for the unconstrained MIN-DS problem. In particular, this leaves a large gap between the known upper bounds of $\ln n$ for arbitrary graphs and the known lower bounds on power law graphs.

In this chapter, we will close this gap (up to a constant factor) by showing that there exists no $\Omega(\ln n)$-approximation for MIN-DS on (α, β)-PLG, for $0 < \beta \leqslant 2$. Our results are based on a reduction from the SET COVER instances used in the reduction of Feige [Fei98] and also hold for the case of connected (α, β)-PLG.

7.2 PREVIOUS RESULTS

Let us first give a formal definition of the MINIMUM DOMINATING SET problem.

Problem 6 (MINIMUM DOMINATING SET (MIN-DS))**.**

INPUT: A graph $G = (V, E)$.

OUTPUT: A *dominating set* of G, that is, a set $D \subseteq V$ such that for all $v \in V$ either $v \in D$ or $D \cap N(v) \neq \emptyset$.

OBJECTIVE: Minimize $|D|$.

Now we are going to summarize known results in approximability and hardness of approximation of MIN-DS.

APPROXIMABILITY AND HARDNESS IN GENERAL GRAPHS. The MINIMUM DOMINATING SET problem is known to be L-reducible to the SET COVER problem and vice versa (see e. g. [Kan92]), which means that any result on the approximability of SET COVER provides the same result also for MIN-DS. The SET COVER problem is one of Karp's 21 NP-complete problems and was shown to be NP-complete in [Kar72]. The formal definition of the problem is as follows:

Problem 7 (SET COVER (SC))**.**

INPUT: A pair $(U, \mathcal{S})$, with a set of ground elements or *universe* $U = \{x_1, x_2, \ldots, x_n\}$ and a family of subsets $\mathcal{S} = \{S_1, S_2, \ldots, S_m\} \subseteq U$, where $\bigcup_{i=\{1,\ldots,m\}} S_i = U$.

OUTPUT: A *set cover* $C \subseteq \{1, 2, \ldots, m\}$ such that $\bigcup_{j \in C} S_j = U$.

OBJECTIVE: Minimize $|C|$.

In case of the *weighted* SET COVER problem, there is a *weight function* $w \colon \mathcal{S} \to \mathbb{R}_+$ and each subset $S_i \in \mathcal{S}$ is associated with a weight $w(S_i)$. We want to find a *minimum cost* set cover $C \subseteq \{1, 2, \ldots, m\}$, where the cost is defined as the total weight $\sum_{j \in C} w(S_j)$.

The SET COVER problem was one of the first problems for which approximation algorithm were designed and the corresponding approximation ratios were analyzed. In 1974 Johnson [Joh74] showed that the greedy algorithm, that repeatedly selects

the subset $S_i \in \mathcal{S}$ which covers the larges number of remaining uncovered elements, achieves an approximation ratio of $\ln n$. The same result was shown by Lovász [Lov75] using a linear programming relaxation of the problem and this approach was extended to the weighted case by Chvátal [Chv79]. A (tight) analysis of the lower order terms of the approximation ratio was provided by Srinivasan [Sri95] for the linear programming formulation, and by Slavík [Sla97] for the greedy algorithm.

The first inapproximability result for Set Cover was provided by Lund and Yannakakis and followed from work on *probabilistically checkable proof systems* (PCP) and the notion of *multi-prover interactive proofs* (see Section 4.1 and Section 4.2). In 1994, Lund and Yannakakis [LY94] showed that the Set Cover problem cannot be approximated within an approximation ratio of $1/4 \log n$ unless $\mathsf{NP} \subset \mathsf{DTIME}(n^{\mathcal{O}(\text{poly} \log n)})$, and that there exists no probabilistic algorithm with approximation ratio better than $1/2 \log n$, unless $\mathsf{NP} \subset \mathsf{ZTIME}(n^{\mathcal{O}(\text{poly} \log n)})$. The proof is based on a reduction from efficient *two-prover proof systems* for the class NP.

In order to obtain hardness results under weaker complexity assumptions, subsequent work was aimed at optimizing the proof systems used in the reduction. The focus was on reducing the *error* of the proof system, that is, reducing the probability of accepting false proofs while maintaining small values for the number of provers, the cardinality of the alphabet and the number of random bits used by the verifier. In 1995, Raz [Raz95; Raz98] provided an improved analysis of two-prover proof systems which implied that Set Cover cannot be approximated within $\log n/4$ unless $\mathsf{NP} \subset \mathsf{DTIME}(n^{\mathcal{O}(\log \log n)})$, and that there exists no probabilistic algorithm with approximation ratio better than $\log n/2$, unless $\mathsf{NP} \subset \mathsf{ZTIME}(n^{\mathcal{O}(\log \log n)})$. Later in the same year, Naor, Schulman, and Srinivasan [NSS95] provided an efficient derandomization of the probabilistic reduction and closed the gap between the results under the assumptions $\mathsf{NP} \subset \mathsf{DTIME}(n^{\mathcal{O}(\log \log n)})$ and $\mathsf{NP} \subset \mathsf{ZTIME}(n^{\mathcal{O}(\log \log n)})$. They showed that the Set Cover problem cannot be approximated within $\log n/2$ unless $\mathsf{NP} \subset \mathsf{DTIME}(n^{\mathcal{O}(\log \log n)})$.

The celebrated result of Feige [Fei98] finally closed the gap between the known $\ln n$ approximation ratio and the above hardness result of $\log n/2$. Feige showed that the upper bound is tight up to lower-order terms under the assumption that $\mathsf{NP} \nsubseteq \mathsf{DTIME}(n^{\mathcal{O}(\log \log n)})$, which means that the known upper approximation bounds are essentially best possible. The result is based on a new reduction from multi-prover

proof systems, which were designed specifically for this purpose. In Section 7.4.3 we will take a close look at the reduction due to Feige and extract certain bounds on the parameters of the construction that enable us to achieve first logarithmic lower bounds for the MIN-DS problem on power law graphs.

Note that it is not known if under the assumption $\mathsf{P} \neq \mathsf{NP}$ hardness of approximating SET COVER within a ratio of $\ln n$ can be shown. However, under this weaker assumption, Raz and Safra [RS97] ruled out the existence of an approximation algorithm with an approximation ratio better than $c \cdot \log n$ for some constant $c > 0$.

APPROXIMABILITY AND HARDNESS IN POWER LAW GRAPHS. In connection with the optimal placement of sensors for disease detection inside social networks, Eubank et al. [Eub+04a] studied near-optimal dominating set problems—namely $(1-\varepsilon)$-MINIMUM DOMINATING SET, where $\varepsilon > 0$ is a given constant denoting the fraction of nodes that can be neglected for the construction of an optimal dominating set—in bipartite random power law graphs. On the positive side, Eubank et al. [Eub+04a] found that for a class of bipartite random power law graphs the problem $(1-\varepsilon)$-MIN-DS admits a PTAS, i. e. they presented a simple greedy algorithm which achieves a $(1+o(1))$-approximation on these instances.

On the contrary, Ferrante, Pandurangan, and Park [FPP08] and Shen, Nguyen, and Thai [SNT10; She+12] studied the approximation hardness of MINIMUM VERTEX COVER (MIN-VC), MAXIMUM INDEPENDENT SET (MAX-IS) and MIN-DS in *combinatorial power law graphs* in the $G(\alpha, \beta)$ model and showed NP-hardness and APX-hardness for simple (α, β)-PLG and (α, β)-PLG multigraphs, respectively. In Table 7.1 we list some of these results, especially the previously best lower bound for MIN-DS in (α, β)-PLG and (α, β)-PLG multigraphs for $\beta > 0$ and $\beta > 1$, respectively.

7.3 OVERVIEW AND RESULTS

In this chapter we study the approximation complexity of MIN-DS in (α, β)-PLG. We give first logarithmic lower bounds for the approximability of this problem for the parameter range $0 < \beta \leqslant 2$. Our results are based on a reduction from the SET COVER problem combined with the logarithmic lower bound for SET COVER given by Feige

Table 7.1: Previously known lower bounds for the inapproximability of MAX-IS and MIN-DS in PLG under condition $P \neq NP$, MIN-VC under UGC in disconnected power law graphs with $\beta > 1$ due to Shen et al. [She+12].

Problem	(α, β)-PLG multigraphs	(α, β)-PLG
MAX-IS	$1 + \frac{1}{140(2\zeta(\beta)3^\beta - 1)} - \varepsilon$	$1 + \frac{1}{1120\zeta(\beta)3^\beta} - \varepsilon$
MIN-DS	$1 + \frac{1}{390(2\zeta(\beta)3^\beta - 1)}$	$1 + \frac{1}{3120\zeta(\beta)3^\beta}$
MIN-VC	$1 + \frac{2\left(1-(2+o_c(1))\frac{\log\log c}{\log c}\right)}{\left(\zeta(\beta)c^\beta + c^{\frac{1}{\beta}}\right)(c-1)}$	$1 + \frac{2-(2+o_c(1))\frac{\log\log c}{\log c}}{2\zeta(\beta)c^\beta(c+1)}$

[Fei98]. The previously known results were the constant factor lower bounds given by Shen et al. [She+12], which were based on reductions from the bounded degree MIN-DS. It was also shown in [She+12] that, for $\beta > 2$, MIN-DS in (α, β)-PLG is in APX. We improve on this result by giving new upper bounds on the approximation ratio of an algorithm based on the greedy algorithm for MIN-DS. In [She+12], membership of MIN-DS in (α, β)-PLG in APX was shown by constructing a lower bound for the optimum and an upper bound for the greedy solution separately. We obtain our new results by relating the cost and structure of an optimum solution to those of a greedy-based solution. This sophisticated analysis yields improved upper bounds for almost the whole range $\beta > 2$. Finally, we take a very close look at the phase transition at $\beta = 2$. Similar as in Section 6.8 and Section 6.9 we extend the power law model and consider the case when $\beta_f = 2 + \frac{1}{f(n)}$ is a function of the graph size n which converges to 2 from above. We obtain the following surprising result: For every function $f(n)$ with $f(n) = \omega(\log(n))$, i. e. when β_f converges fast enough, MIN-DS in (α, β_f)-PLG still provides a logarithmic approximation lower bound, and for every function $f(n)$ with $f(n) = o(\log(n))$, the problem is in APX. The summary of main results of this chapter is given in Table 7.2.

ORGANIZATION OF THE CHAPTER. The chapter is organized as follows. In Section 7.4 we give an outline of our methods and the embedding constructions on which our reductions are based. In Section 7.4.3 we take a close look at U. Feige's original reduction from 5OCC-MAX-E3SAT (5 OCCURRENCE MAXIMUM E3-SAT) to the SET COVER problem [Fei98] and the standard reduction from the SET COVER to the MINIMUM

Table 7.2: Summary of the main results of this chapter: Approximation lower bounds and approximation upper bounds for MIN-DS on (α, β)-PLG for certain ranges of the parameter β. The parameter d is chosen minimally such as to satisfy the constraint given in Theorem 7.5.

	Approximation Lower Bound
$0 < \beta < 1$	$\Omega\left(\ln(n) - \ln\left(\frac{1}{1-\beta}\right)\right)$
$\beta = 1$	$\Omega\left(\ln(n)\right)$
$1 < \beta < 2$	$\Omega\left(\ln(n) - \ln(\zeta(\beta))\right)$
$\beta = 2$	$\Omega\left(\ln(n) - \ln(\zeta(\beta))\right)$
$\beta = 2 + \frac{1}{f(n)}, f(n) = \omega(\log(n))$	$\Omega\left(\ln(n) - \ln(\zeta(\beta))\right)$
	Approximation Upper Bound
$\beta = 2 + \frac{1}{f(n)}, f(n) = o(\log(n))$	APX
$\beta > 2$	$\frac{\zeta(\beta)-1}{\zeta(\beta)-\sum_{j=1}^{d-1} j^{-\beta}}$
$\beta > 2.729$	$\frac{\zeta(\beta)-\frac{\zeta(\beta-1)}{2}}{1-\frac{\zeta(\beta-1)}{2}}$

DOMINATING SET problem. As a result of this section, we obtain sufficient information about the degree distribution of the resulting MIN-DS instances $G_{U,S}$. In Section 7.4.4 we describe our lower bound techniques together with the scaling method for the realization of a power law degree sequence.

Starting with Section 7.5, we present new lower bounds on the approximability of MIN-DS in (α, β)-PLG for the different ranges of the parameter β. The case $0 < \beta < 1$ is treated in Section 7.5, based on a precise rounding error analysis for the terms that determine the lower approximation bound. A similar analysis is used for the case $\beta = 1$ in Section 7.6. The Section 7.7 deals with the case $1 < \beta < 2$. We describe how to *rescale* the degree distribution of instances $G_{U,S}$ in order to embed them into an (α, β)-PLG $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$. We also apply our scaling technique for the case $\beta = 2$ in Section 7.8 together with a slightly different analysis.

In Section 7.9 we present new upper bounds for the case of $\beta > 2$ and provide a detailed comparison of the previous and new upper bounds in terms of the parameter β. In Section 7.10 we consider the functional case when the parameter $\beta_f = 2 + \frac{1}{f(n)}$ is a function of the graph size n which converges to 2 from above. Figure 7.1 shows

the global organization of the chapter, pointing to the different ranges and the phase transitions dependent on the model parameter β.

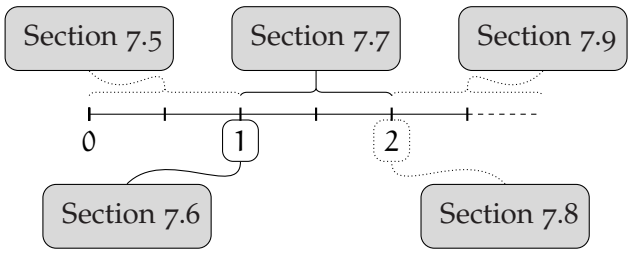


Figure 7.1: Chapter guide with respect to the phase transitions and different ranges of the model parameter β.

7.4 OUTLINE OF THE METHOD

Let us give an outline of the methods and constructions used in this chapter.

INTERVALS AND VOLUMES. In the following we will introduce notations of *intervals* of nodes inside an power law graph $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ and of the *volume* of such an interval. From [ACL00; ACL01] we have that in an (α,β)-PLG $G_{\alpha,\beta} = (V_{\alpha,\beta}, E_{\alpha,\beta})$ with maximum degree $\Delta = \left\lfloor e^{\alpha/\beta} \right\rfloor$ the number of nodes $|V_{\alpha,\beta}| = n = \sum_{i=1}^{\Delta} \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor$ and the number of edges $|E_{\alpha,\beta}| = m = \frac{1}{2}\sum_{i=1}^{\Delta} \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor$ satisfy

$$n \approx \begin{cases} \zeta(\beta)\, e^\alpha & \text{if } \beta > 1 \\ \alpha\, e^\alpha & \text{if } \beta = 1 \\ \frac{e^{\frac{\alpha}{\beta}}}{1-\beta} & \text{if } 0 < \beta < 1 \end{cases} \quad \text{and} \quad m \approx \begin{cases} \frac{1}{2}\zeta(\beta-1)\, e^\alpha & \text{if } \beta > 2 \\ \frac{1}{4}\alpha\, e^\alpha & \text{if } \beta = 2 \\ \frac{1}{2}\frac{e^{\frac{2\alpha}{\beta}}}{2-\beta} & \text{if } 0 < \beta < 2 \end{cases}$$

An *interval* of nodes in $G_{\alpha,\beta}$ is a set $[a,b] = \{v \in V \mid a \leqslant d(v) \leqslant b\}$, where $1 \leqslant a \leqslant b \leqslant \Delta = \left\lfloor e^{\alpha/\beta} \right\rfloor$. Furthermore, let $|[a,b]|$ be the number of nodes inside the interval $[a,b]$. For the *volume* of an interval $[a,b]$ we define $\mathrm{vol}([a,b]) = \sum_{j=a}^{b} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \cdot j$, i. e. the sum of node degrees of nodes inside of the interval.

7.4.1 *Lower Bound Technique*

In order to obtain logarithmic approximation lower bounds for the MINIMUM DOMINATING SET problem in (α, β)-*power law graphs*, we construct reductions from MIN-DS in graphs, which is basically as hard to approximate as the SET COVER problem. It is well known [PM81; BM84; Kan92] that SET COVER instances $(U, \mathcal{S})$ with universe U and set system $\mathcal{S}$ can be translated into instances $G_{U,\mathcal{S}}$ of MIN-DS in graphs, where $V_{U,\mathcal{S}} = I \cup U$, $\{i, j\} \in E_{U,\mathcal{S}}, \forall i, j \in I$ and $\{i, u\} \in E_{U,\mathcal{S}}, \forall i \in I, u \in S_i$. Consider a feasible solution $C = \{S_i \mid i \in D\}, D \subseteq I$, for the SET COVER problem, then D is a dominating set for $G_{U,\mathcal{S}}$ with $|D| = |C|$. This can be seen as follows: for each $u \in U$ there is an $i \in I$ such that $u \in S_i$ and since $\{i, u\} \in E_{U,\mathcal{S}}$, u is dominated by i. Furthermore, since D is nonempty, each $i \in I$ is adjacent to some vertex $j \in D$ in $G_{U,\mathcal{S}}$. Conversely, let D be a dominating set for $G_{U,\mathcal{S}}$ with some $u \in D \cap U$, i. e. not all of the dominating set vertices in D correspond to sets in $\mathcal{S}$. It is possible to construct a dominating set D' such that $|D'| \leqslant |D|$ and $D' \subseteq I$ by replacing each $u \in D \cap U$ by a neighbor $i \in I$ of u in $G_{U,\mathcal{S}}$. Now $C = \{S_i \mid i \in D'\}$ is a feasible solution of the SET COVER problem with $|C| = |D'| \leqslant |D|$.

Our reductions map those graphs $G_{U,\mathcal{S}}$, which are stemming from SET COVER instances $(U, \mathcal{S})$ to (α, β)-power law graphs $G_{\alpha,\beta}$. In this construction, nodes of the graph $G_{U,\mathcal{S}}$ are connected to a set Γ of degree 2 nodes, and those are again connected to the residual graph $G_{\alpha,\beta} \setminus (G_{U,\mathcal{S}} \cup \Gamma)$. The set Γ enforces any *reasonable* dominating set in the fully composed graph $G_{\alpha,\beta}$ to contain a dominating set of the graph $G_{U,\mathcal{S}}$.

Another important property of our constructions is that the residual graph $G_{\alpha,\beta} \setminus (G_{U,\mathcal{S}} \cup \Gamma)$ contains a sufficiently small set X of vertices which dominate every node in $G_{\alpha,\beta} \setminus G_{U,\mathcal{S}}$. It is precisely this property which enables us to obtain logarithmic lower bounds for the approximability of MIN-DS in (α, β)-PLG, instead of the previously known constant lower bounds due to Shen et al. [She+12].

The next elementary step in this construction is the implementation of the *power law degree distribution*. Therefore, we need to know the degree distribution in the graph $G_{U,\mathcal{S}}$. In Section 7.4.3 we will take a close look at Feige's original construction and obtain upper and lower bounds for the degrees of nodes in the graph $G_{U,\mathcal{S}}$, where $(U, \mathcal{S})$ is a SET COVER instance in the construction. We apply our construction only to those SET COVER instances $(U, \mathcal{S}) = F_{SC}(\varphi)$, where φ is a 5OCC-MAX-E3SAT instance

and F_{SC} is the reduction from [Fei98]. We show that the MIN-DS instances $G_{U,\mathcal{S}}$ have the following property: There exist constants $0 < a < b < 1$ such that for every SET COVER instance $(U, \mathcal{S})$ with $(U, \mathcal{S}) = F_{SC}(\varphi)$, the node degrees of all vertices in $G_{U,\mathcal{S}}$ are contained in the interval $\left[N^a, N^b\right]$, where N is the number of vertices of $G_{U,\mathcal{S}}$.

7.4.2 *Embedding Technique*

We construct a map which embeds every graph $G_{U,\mathcal{S}}$—where $(U, \mathcal{S})$ is a SET COVER instance from Feige's hardness result—into an (α, β)-PLG $G_{\alpha,\beta}$. Let $G_{U,\mathcal{S}} = (V_{U,\mathcal{S}}, E_{U,\mathcal{S}})$ with $|V_{U,\mathcal{S}}| = N$. The graphs $G_{U,\mathcal{S}}$ have the following property: There exist constants $0 < a < b < 1$ such that for all $v \in V_{U,\mathcal{S}}$, $N^a \leqslant d_{U,\mathcal{S}}(v) \leqslant N^b$. The power law graph $G_{\alpha,\beta} = (V_{\alpha,\beta}, E_{\alpha,\beta})$ has the vertex set $V_{\alpha,\beta} = V_{U,\mathcal{S}} \cup X \cup \Gamma \cup V_1 \cup W$, where $X \subseteq [x\Delta, y\Delta] = \{v \in V_{\alpha,\beta} \mid x\Delta \leqslant d_{\alpha,\beta}(v) \leqslant y\Delta\}$ is the set of nodes chosen to be able to dominate all the nodes in $G_{\alpha,\beta} \setminus G_{U,\mathcal{S}}$, V_1 is the set of degree 1 nodes and W the set of remaining nodes of the targeted degree sequence. $G_{\alpha,\beta}$ is constructed such that each node in $V_{U,\mathcal{S}}$ has precisely one neighbor in $\Gamma \subseteq W$, and every $u \in \Gamma$ has precisely one neighbor in $V_{U,\mathcal{S}}$. This is motivated by the connectivity properties of random graphs $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ in the $G(\alpha, \beta)$ model, i. e. the expected number of edges between subgraphs induced by two vertex subsets, as discussed in Section 6.4.2 in Chapter 6. Furthermore, each node $w \in W$ is adjacent to precisely one node in X and every degree 1 node is adjacent to a node in X, whereas each $v \in X$ has at least one degree 1 neighbor. Thus, the set X is chosen to be able to dominate every vertex in W and all the degree 1 nodes in V_1, i. e. all the vertices of the residual graph $G_{\alpha,\beta} \setminus G_{U,\mathcal{S}}$. The construction and partition of the vertex set described above is shown in Figure 7.2.

In order to be able to monitor the current status of implementing the power law degree distribution inside the graph $G_{\alpha,\beta}$, we keep track of the *residual degrees* $d_r(\cdot)$ of nodes in $X \cup W \cup V_1$. We are now ready to formulate the algorithm for the embedding construction of our reduction (see algorithm `Construct_PLG` on page 142).

The last two steps of algorithm `Construct_PLG` are calling the procedure `Fill_Wheel` on the sets which may still have residual degrees after the completion of the previous steps. The procedure `Fill_Wheel` gets as an input a set of nodes V with residual degrees $d_r(v) > 0, \forall v \in V$ and generates the missing edges degree wise in a cyclic

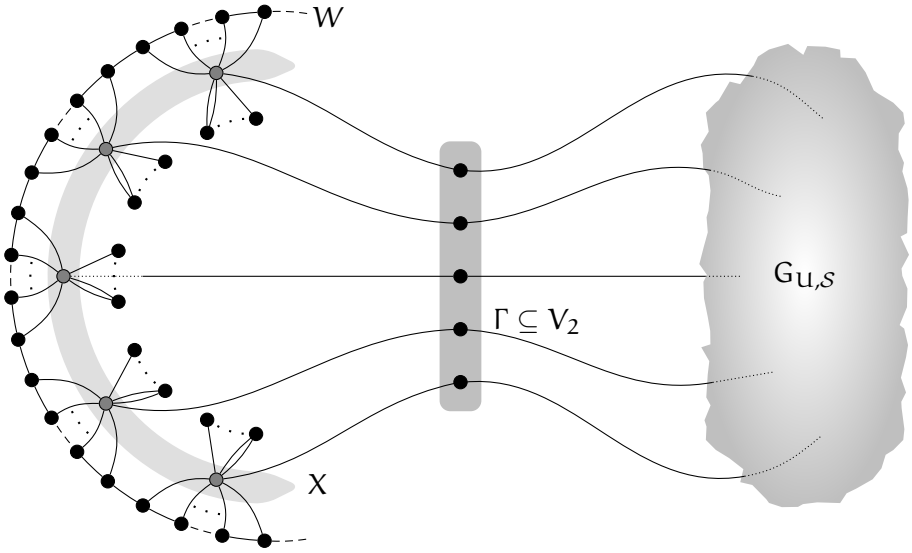


Figure 7.2: The main construction for the embedding of a MIN-DS (SET COVER) instance $G_{U,\mathcal{S}}$ into a (α, β)-PLG. In the resulting graph the nodes $\bullet \in X$ are dominating the sets $W \cup V_1$, separating the dominating set in $G_{U,\mathcal{S}}$ from the dominating set in $G_{\alpha,\beta} \setminus G_{U,\mathcal{S}}$.

order. Let $v_{j,1}, \ldots, v_{j,n_j}$ be the nodes with target degree $d_{\alpha,\beta}(v_{j,1}) = j$ in the set V, then, in every stage of the construction, for every $j \in \{1, \ldots, \Delta\}$, $d_r(v_{j,1}) \leqslant \ldots \leqslant d_r(v_{j,n_j})$ and $d_r(v_{j,n_j}) - d_r(v_{j,1}) \leqslant 1$ (for a more detailed description see also Section 6.7 and algorithm `Fill_Wheel` on page 105).

Figure 7.3 shows an example of how intervals of uniform residual degrees are filled by the procedure `Fill_Wheel` and how the problems of uneven interval-lengths and uneven residual degrees are resolved at the borders of the intervals.

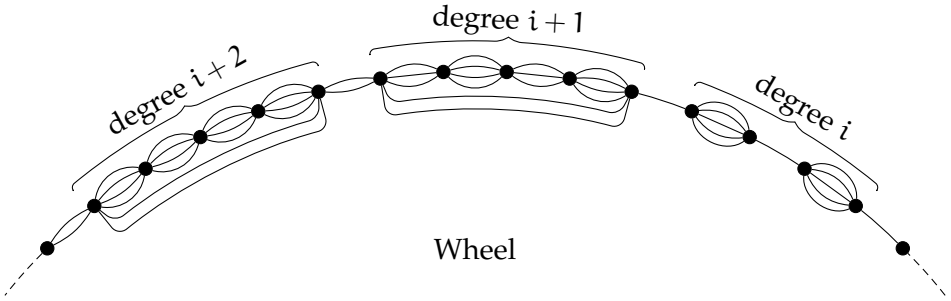


Figure 7.3: Procedure `Fill_Wheel` realizes the residual degrees on the wheel nodes in W and X.

Algorithm 7.1 : `Construct_PLG`

Input : $G_{U,\mathcal{S}} = (V_{U,\mathcal{S}}, E_{U,\mathcal{S}})$ with $|V_{U,\mathcal{S}}| = N$.
Output : Power law graph $G_{\alpha,\beta} = (V_{\alpha,\beta}, E_{\alpha,\beta})$ with $V_{\alpha,\beta} = V_{U,\mathcal{S}} \cup X \cup W \cup V_1 \cup V_2$, $|V_{\alpha,\beta}| = n$ and $E_{U,\mathcal{S}} \subseteq E_{\alpha,\beta}$.

choose α, x, y such that $|[x\Delta, y\Delta]| \geqslant n$ and $|[N^a, N^b]| \geqslant N$;
set $X := [x\Delta, y\Delta]$, $W := [3, \Delta] \setminus (V_{U,\mathcal{S}} \cup X)$ and $\Gamma := \emptyset$;
set $V_{\alpha,\beta} := V_{U,\mathcal{S}} \cup X \cup W \cup V_1 \cup V_2$;
for $i = 1, \ldots, N(= |V_{U,\mathcal{S}}|)$ **do**
  map $s_i \in V_{U,\mathcal{S}}$ with $t_i \in V_2 \setminus \Gamma$ and **set** $E_{\alpha,\beta} := E_{\alpha,\beta} \cup \{s_i, t_i\}, \Gamma := \Gamma \cup \{t_i\}$;
  choose $v \in X$ with maximum $d_r(v) > 0$ and **set** $E_{\alpha,\beta} := E_{\alpha,\beta} \cup \{t_i, v\}$;
  update $d_r(t_i)$ and $d_r(v)$;
foreach $u \in V_1 \cup V_2, d_r(u) > 0$ **do**
  choose $v \in X$ with maximum $d_r(v) > 0$ and **set** $E_{\alpha,\beta} := E_{\alpha,\beta} \cup \{u, v\}$;
  update $d_r(t)$ and $d_r(v)$;
foreach $w \in W$ **do**
  choose $v \in X$ with maximum $d_r(v) > 0$ and **set** $E_{\alpha,\beta} := E_{\alpha,\beta} \cup \{w, v\}$;
  update $d_r(w)$ and $d_r(v)$;
`Fill_Wheel`(W); `/* realizes residual degrees on` W `and` X `*/`
`Fill_Wheel`(X);
return $G_{\alpha,\beta} = (V_{\alpha,\beta}, E_{\alpha,\beta})$;

Since $X \subseteq [x\Delta, y\Delta]$ and $x\Delta$ and $y\Delta$ are chosen such that the number of edges $\text{vol}([x\Delta, y\Delta]) = \sum_{j=x\Delta}^{\Delta} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \cdot j$ suffices to connect X to all nodes in $V_{\alpha,\beta} \setminus X$, we may have residual degrees for some $v \in X$ and call `Fill_Wheel`(X). Furthermore, we call `Fill_Wheel`(W) since we have that all $w \in W$ are connected only via a single edge to the set X and $w \in W$ were chosen to have targeted degrees in the interval $[3, \Delta]$, i. e. $d_{\alpha,\beta}(w) \in [3, \Delta]$ and thus $d_r(w) \in [2, \Delta - 1]$.

Depending on the parameter β, we will show how to choose x and y in such a way that the set X becomes sufficiently small to be included in any feasible dominating set and such that the size of X is bounded from above by a lower bound of the size of any dominating set in $G_{U,\mathcal{S}}$. Hence, any dominating set D' in $G_{\alpha,\beta}$ can be efficiently transformed into a dominating set D of size $|D| \leqslant |D'|$ such that $D = D_{U,\mathcal{S}} \cup X$, where $D_{U,\mathcal{S}} \subseteq V_{U,\mathcal{S}}$ is a dominating set of $G_{U,\mathcal{S}}$.

7.4.3 *Feige's Lower Bound for Set Cover*

The starting point for establishing our new lower bounds is the logarithmic lower bound of Feige [Fei98] for the approximability of the SET COVER problem. For each SET COVER instance $(U, \mathcal{S})$, we embed the associated MIN-DS instance $G_{U,\mathcal{S}}$ into an (α, β)-PLG $G_{\alpha,\beta}$. In order to implement the power law degree distribution, we need to know the degree distribution of the graph $G_{U,\mathcal{S}}$. Therefore we briefly review Feige's construction.

Feige constructs a k-prover proof system for the problem 5OCC-MAX-E3SAT. Consider a 3CNF formula φ with n variables such that each variable occurs at most 5 times in φ. One can assume that either the formula is satisfiable, or no assignment satisfies more than an ε-fraction of the clauses simultaneously. The k-prover proof system works as follows: It chooses k codewords of length $l = \Theta(\log \log n)$, weight $l/2$ and pairwise *Hamming distance* $\geqslant l/3$. The verifier picks l clauses $C_1, \ldots, C_l$ from φ independently uniformly at random. Independently, from each such clause C_i it picks one variable x_i of C_i uniformly at random. For each $1 \leqslant i \leqslant k$, the verifier sends to the prover i those $l/2$ clauses C_j for which the associated bit of prover i's codeword is 1 and those $l/2$ variables x_j for which the associated bit of prover i's codeword is 0. The provers return their answers, and based on this the verifier determines its output. The construction of the associated SET COVER instances makes use of some combinatorial building blocks called *partition systems*.

According to Feige [Fei98], a partition system $B(m, L, k, d)$ consists of a ground set B of cardinality $|B| = m$ and L partitions $p_1, \ldots, p_L$ of B into k disjoint subsets $p_{j,h} \subset B$. The defining property of these partition systems is that each cover of B by subsets $p_{j,h}$ which uses sets from pairwise different partitions must consist of at least d subsets. Feige gives a randomized construction of such partition systems with $L \approx (\log m)^c$, k being any number smaller than $\ln(m/3) \cdot \ln \ln(m)$ and $d = (1 - f(k)) \cdot k \cdot \ln(m)$, with some function $f(k)$ with $f(k) \longrightarrow 0$ as $k \longrightarrow \infty$. That construction yields partitions for which w.h.p. all the sets have the same size.

We show that the same result is obtained by making use of random permutations. But now in this case, for each partition p_j, the sets $p_{j,h}$ always have the same size m/k (provided $k < m$). Namely, choose a random permutation $\pi_j \in_R S_m$ and let $p_{j,h} = \{\pi_j((h-1)m/k + 1), \ldots, \pi_j(k \cdot m/k)\}$. Suppose now, we cover B with d subsets

$p_{j_1,h_1}, \ldots, p_{j_d,h_d}$ from pairwise different partitions. Then for a given point $v \in B$, the probability that v is covered by at least one of these subsets is

$$\begin{aligned}
&P(\text{point } v \in B \text{ is covered by at least one of these } d \text{ sets}) \\
&= 1 - \prod_{i=1}^{d} P\left(v \text{ is not in position } 1, \ldots, m/k \text{ in permutation } \pi_j\right) \\
&= 1 - \left(\frac{\binom{m-1}{m/k} \cdot \left(\frac{m}{k}\right)! \cdot \left(m - \frac{m}{k}\right)!}{m!} \right)^d \\
&= 1 - \left(\frac{(m-1)! \cdot \left(m - \frac{m}{k}\right)!}{\left(m - 1 - \frac{m}{k}\right)! \cdot m!} \right)^d \\
&= 1 - \left(\frac{m \cdot \left(1 - \frac{1}{k}\right)}{m} \right)^d = 1 - \left(1 - \frac{1}{k}\right)^d .
\end{aligned}$$

This is precisely the property of the randomized construction which has been used in the analysis of the construction. So, from now on, we assume that all sets of a partition p_j have the same size m/k.

RESULTING SET COVER INSTANCES ([FEI98]). For a given 5OCC-MAX-E3SAT formula φ with n variables and together with the property that either φ is satisfiable or no assignment satisfies more than an ε fraction of the clauses, an SET COVER instance $(U, \mathcal{S})$ is constructed as follows:

- $\mathcal{R}$ is the set of random strings used by the verifier in the k-prover proof system. The number of random strings is $|\mathcal{R}| = R = (5n)^l$.
- $|U| = mR$ with $m = (5n)^{\frac{2l}{\varepsilon}}$, hence $|U| = (5n)^{l\left(1+\frac{2}{\varepsilon}\right)}$.
- For each $r \in \mathcal{R}$, $B_r(m, L, k, d)$ is a partition system with $L = 2^l$.
- $Q = n^{l/2} \cdot \left(\frac{5n}{3}\right)^{l/2}$ is the number of different queries the verifier may ask to a prover.
- $\mathcal{S}$ contains for every triplet (q, a, i) a set $S_{q,a,i}$, where q is a query, i is (the index of) a prover and a is the prover's answer. The set $S_{q,a,i}$ is defined as $S_{q,a,i} = \bigcup_{r:\,(q,i)\in r} B(r, a_r, i)$.

Hence, the number of sets in $\mathcal{S}$ is $Q \cdot k$ and each set is of cardinality $\sqrt{R} \cdot m/k$. We have to give an estimate for the number of sets in which a point, i. e. an element of U, occurs. For each prover i, for each query q, each point in B_r with $|B_r| = m$ occurs in 2^l sets $S_{q,a,i}$. Hence, the total degree of points, which equals the number of occurrences of this point in sets of $\mathcal{S}$, is $2^l \cdot Q$.

FROM SET COVER TO DOMINATING SET. Let $(U, \mathcal{S})$ denote a SET COVER instance with $U = \{u_1, \ldots, u_{|U|}\}$ and $\mathcal{S} = \{S_1, \ldots, S_{|\mathcal{S}|}\}$. Let $G_{U,\mathcal{S}}$ be the undirected graph with set of vertices $V_{U,\mathcal{S}} = U \cup \mathcal{S}$ and set of edges $E_{U,\mathcal{S}} = \{\{S_i, u_j\} \mid u_j \in S_i\} \cup \{\{S_j, S_l\} \mid S_j \cap S_l \neq \emptyset\}$. We observe that each set cover $\mathcal{C} \subseteq \mathcal{S}$ is a dominating set in $G_{U,\mathcal{S}}$. On the other hand, let $D \subseteq V_{U,\mathcal{S}}$ be a dominating set in $G_{U,\mathcal{S}}$ with $D = D_U \cup D_{\mathcal{S}}, D_U = D \cap U$ and $D_{\mathcal{S}} = D \cap \mathcal{S}$. If we replace each $u_i \in D_U$ by an arbitrary set S_j with $u_i \in S_j$, the resulting set D' is a dominating set with $D_{\mathcal{S}} \subseteq D' \subseteq \mathcal{S}$ and $|D'| \leqslant |D|$. Hence, in this sense, we can say that dominating sets in $G_{U,\mathcal{S}}$ correspond to set covers $\mathcal{C}$ for $U, \mathcal{S}$.

In the construction, the parameter l satisfies $l = \Theta(\log\log n)$. If $N_0 = |U| + |\mathcal{S}|$ is the number of nodes of $G_{U,\mathcal{S}}$, then—up to logarithmic factors, $N_0 \approx n^l + n^{l(1+2/\varepsilon)}$, the degree of element nodes $u \in U$ is $d(u) \approx n^l$, each set contains $n^{l(1/2+2/\varepsilon)}$ elements and there are $\approx n^l$ sets. The degree of set nodes in $G_{U,\mathcal{S}}$ is bounded by the sum of the cardinality of that set and the number of sets in the instance $(U, \mathcal{S})$, which is $|\mathcal{S}| \approx n^{l(1/2+2/\varepsilon)}$. Hence we obtain the following result that is used in the sequel.

Lemma 7.1.
Let F_{SC} denote the reduction from 5OCC-MAX-E3SAT *to the* SET COVER *problem, and for a given* SET COVER *instance $(U, \mathcal{S}) = f(\varphi)$ let $G_{U,\mathcal{S}}$ be the associated* MIN-DS *instance as described above. If N_0 is the number of nodes of $G_{U,\mathcal{S}}$, then for every node v in $G_{U,\mathcal{S}}$, the node degree of v in $G_{U,\mathcal{S}}$ satisfies $N_0^a \leqslant d_{U,\mathcal{S}}(v) \leqslant N_0^b$, where $0 < a < b < 1$ and*

$$b = (1 + o(1)) \cdot \frac{\frac{1}{2} + \frac{2}{\varepsilon}}{1 + \frac{2}{\varepsilon}} = (1 + o(1)) \cdot \frac{\varepsilon + 4}{2\varepsilon + 4} .$$

7.4.4 *Reduction Construction and Scaling*

With these notions and results at hand, we will now describe our reduction to achieve new logarithmic lower bounds for approximability of the MINIMUM DOMINATING SET

problem in (α, β)-PLG. We distinguish several cases depending on the range of the parameter β. For the cases $0 < \beta < 1$, $1 < \beta < 2$ and $\beta = 2$ our construction involves *rescaling* of the instances $G_{U,\mathcal{S}}$, which has the effect of shifting the degree interval $\left[N^a, N^b\right]$ towards the left end of the full interval $[1, \Delta]$. It turns out that for the case $\beta = 1$ we can omit the scaling and directly implement the power law distribution.

BOUNDS ON OPTIMA IN $G_{U,\mathcal{S}}$. Let $(U, \mathcal{S})$ be an instance of the SET COVER problem which is an image $(U, \mathcal{S}) = F_{SC}(\varphi, \varepsilon)$ of some 5OCC-MAX-E3SAT instance φ under Feige's reduction F_{SC} with parameter $\varepsilon > 0$. Suppose the number of nodes of $G_{U,\mathcal{S}}$ is N_0. Let $\mathsf{OPT}(G_{U,\mathcal{S}})$ denote a minimum cardinality dominating set of $G_{U,\mathcal{S}}$. Then

$$|\mathsf{OPT}(G_{U,\mathcal{S}})| \leqslant k \cdot N_0^{\frac{\varepsilon}{2+\varepsilon}}$$

or

$$|\mathsf{OPT}(G_{U,\mathcal{S}})| \geqslant (1-\varepsilon) \cdot k \cdot N_0^{\frac{\varepsilon}{2+\varepsilon}} \cdot \frac{\varepsilon}{2+\varepsilon} \cdot 2^{-\frac{\varepsilon}{2+\varepsilon}} \cdot (\ln(N_0) - \mathcal{O}(1)) \ ,$$

where k is the number of provers in Feige's k-prover proof system. Recall that the 3CNF formula φ with $F_{SC}(\varphi) = (U, \mathcal{S})$ is either satisfiable, or no assignment satisfies more than an ε fraction of its clauses. Furthermore, as a result of Lemma 7.1, the node degrees in $G_{U,\mathcal{S}}$ are contained in the interval $\left[N_0^a, N_0^b\right]$, with $0 < a < b < 1$ being constant.

SCALING. In the three cases $0 < \beta < 1$, $1 < \beta < 2$ and $\beta = 2$, it turns out that we have to rescale the degrees of nodes in $G_{U,\mathcal{S}}$ in order shift the interval associated to $G_{U,\mathcal{S}}$ towards the left end of the full interval. This enables us to get a feasible lower bound for the size of a dominating set in $G_{U,\mathcal{S}}$ and, at the same time, prevents overlapping of the intervals $[N^a, N^b]$ and $[x\Delta, y\Delta]$. For this purpose, we replace $G_{U,\mathcal{S}}$ by the graph $G^d_{U,\mathcal{S}}$ which consists of N_0^{d-1} disjoint copies of the graph $G_{U,\mathcal{S}}$ (cf. Figure 7.4).

Here, d is a parameter of our construction. The graph $G^d_{U,\mathcal{S}}$ has the following properties:

- The number of nodes is $N := N_0^d$.

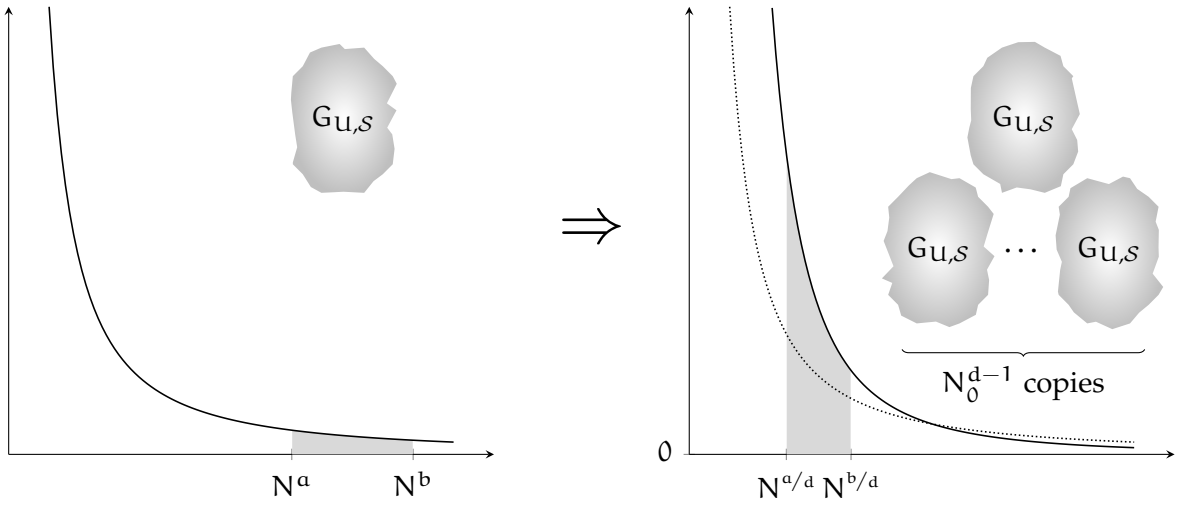


Figure 7.4: Scaling of the graph $G_{u,\mathcal{S}}$ by replacing the graph $G_{u,\mathcal{S}}$ with the graph $G^d_{u,\mathcal{S}}$, which consists of N_0^{d-1} disjoint copies of the graph $G_{u,\mathcal{S}}$.

- The node degrees are contained in the interval $\left[N^{a/d}, N^{b/d}\right]$.
- Let $\mathsf{OPT}(G^d_{u,\mathcal{S}})$ denote an optimum dominating set of $G_{u,\mathcal{S}}$. Then

$$|\mathsf{OPT}(G^d_{u,\mathcal{S}})| \leqslant N^{\frac{d-1}{d}} \cdot k \cdot N^{\frac{1}{d}\frac{\varepsilon}{2+\varepsilon}} = k \cdot N^{\frac{1}{d}\left(d-1+\frac{\varepsilon}{2+\varepsilon}\right)}$$

or

$$\begin{aligned}|\mathsf{OPT}(G^d_{u,\mathcal{S}})| &\geqslant (1-\varepsilon) \cdot k \cdot N^{\frac{1}{d}\frac{\varepsilon}{2+\varepsilon}} \cdot \frac{\varepsilon}{2+\varepsilon} \cdot 2^{-\frac{\varepsilon}{2+\varepsilon}} \left(\ln\left(N^{\frac{1}{d}}\right) - \mathcal{O}(1)\right) N^{\frac{d-1}{d}} \\ &= k \cdot \frac{\varepsilon(1-\varepsilon)}{2+\varepsilon} \cdot 2^{-\frac{\varepsilon}{2+\varepsilon}} \cdot N^{\frac{1}{d}\left(d-1+\frac{\varepsilon}{2+\varepsilon}\right)} \left(\ln\left(N^{\frac{1}{d}}\right) - \mathcal{O}(1)\right) \quad .\end{aligned}$$

CONSTRUCTION OF $G_{\alpha,\beta}$. Let us now describe how the fully composed graph $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ is constructed. We choose α and the parameters d, x, y such as to satisfy the following three constraints:

1. $\left|\left[N^{a/d}, N^{b/d}\right]\right| \geqslant N$.
2. $|[x\Delta, y\Delta]| = o\left(N^{\frac{d-1}{d}}\right)$,
   where $N^{\frac{d-1}{d}}$ is a lower bound for the size of an optimum dominating set in $G_{u,\mathcal{S}}$.

3. $\sum_{j=x\Delta}^{y\Delta} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \cdot j = \mathrm{vol}(|x\Delta, y\Delta|) \geqslant \zeta(\beta) \cdot e^\alpha$,
   i. e. the total volume of the set $[x\Delta, y\Delta]$ is large enough such that $[x\Delta, y\Delta]$ can dominate the wheel W as well as all the degree 2 nodes, which are matched to nodes in the graph $G_{U,\mathcal{S}}$.

Constraint 1 is implied by the following stronger constraint (1′):

(1′) $\frac{e^\alpha}{N^{\frac{b\beta}{d}}} \geqslant N.$

In all of the following cases, we work with constraint (1′) instead of 1 and obtain the following bound for the parameter α:

$$e^\alpha \geqslant N^{1+\frac{b\beta}{d}} \quad .$$

In order to minimize the value of the parameter α, and thereby the overall graph size, we choose $e^\alpha = N^{1+\frac{b\beta}{d}}$.

7.5 THE CASE $0 < \beta < 1$

Let us now consider the case $0 < \beta < 1$. Here, we will make use of the scaling technique described above. Furthermore, in this case we have to choose parameters x, y of the interval $X = [x\Delta, y\Delta]$ carefully in order to obtain a logarithmic lower bound. The next lemma provides an estimate for the size of the interval $|[x\Delta, y\Delta]|$ and the volume $\mathrm{vol}([x\Delta, y\Delta])$.

Lemma 7.2.
Let $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ be an (α, β)-PLG with $0 < \beta < 1$. Then, for all $0 < x < y \leqslant 1$, the size and the volume of the interval $[x\Delta, y\Delta]$ satisfies

$$|[x\Delta, y\Delta]| \in \left[\frac{\Delta}{1-\beta}\left(1 - x^{1-\beta}\right) - \left(\frac{1}{x^\beta} - 1\right) - (2-x)\Delta,\ \frac{\Delta}{1-\beta}\left(1 - x^{1-\beta}\right) \right]$$

and

$$\mathrm{vol}([x\Delta, y\Delta]) \geqslant \Delta^2 \left(\frac{1 - x^{2-\beta}}{2-\beta} - \frac{1}{2} + \frac{x^2}{2} \right) - \Delta \left(1 - x^{1-\beta} - \frac{1}{2} + \frac{x}{2} \right) \quad .$$

Proof. Regarding the requirement of constraint 2, we have

$$|[x\Delta, y\Delta]| \in \left[\sum_{j=x\Delta}^{y\Delta} \frac{e^\alpha}{j^\beta} - (y - x + 1)\Delta, \sum_{j=x\Delta}^{y\Delta} \frac{e^\alpha}{j^\beta} \right] ,$$

where

$$\begin{aligned}
\sum_{j=x\Delta}^{y\Delta} \frac{e^\alpha}{j^\beta} &\in \left[e^\alpha \int_{x\Delta}^{y\Delta} \frac{1}{j^\beta}\, dj - e^\alpha \left(\frac{1}{(x\Delta)^\beta} - \frac{1}{(y\Delta)^\beta} \right), e^\alpha \int_{x\Delta}^{y\Delta} \frac{1}{j^\beta}\, dj \right] \\
&= \left[e^\alpha \left[\frac{j^{1-\beta}}{1-\beta} \right]_{x\Delta}^{y\Delta} - \frac{e^\alpha}{\Delta^\beta} \left(\frac{1}{x^\beta} - \frac{1}{y^\beta} \right), e^\alpha \left[\frac{j^{1-\beta}}{1-\beta} \right]_{x\Delta}^{y\Delta} \right] \\
&= \left[\frac{e^\alpha \Delta^{1-\beta}}{1-\beta} \left(y^{1-\beta} - x^{1-\beta} \right) - \left(\frac{1}{x^\beta} - \frac{1}{y^\beta} \right), \frac{e^\alpha \Delta^{1-\beta}}{1-\beta} \left(y^{1-\beta} - x^{1-\beta} \right) \right] \\
&= \left[\frac{\Delta}{1-\beta} \left(y^{1-\beta} - x^{1-\beta} \right) - \left(\frac{1}{x^\beta} - \frac{1}{y^\beta} \right), \frac{\Delta}{1-\beta} \left(y^{1-\beta} - x^{1-\beta} \right) \right] .
\end{aligned}$$

In order to fulfill the volume requirement of constraint 3, we also take into account the rounding error resulting when we replace the sum $\sum_{x\Delta}^{y\Delta} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor$ by $\sum_{x\Delta}^{y\Delta} \frac{e^\alpha}{j^{\beta-1}}$. The sum of node degrees of nodes in $[x\Delta, y\Delta]$ is

$$\mathrm{vol}([x\Delta, y\Delta]) = \sum_{x\Delta}^{y\Delta} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \cdot j \in \left[\sum_{x\Delta}^{y\Delta} \frac{e^\alpha}{j^{\beta-1}} - \underbrace{\left(\frac{y\Delta(y\Delta - 1)}{2} - \frac{x\Delta(x\Delta - 1)}{2} \right)}_{\text{rounding error}}, \sum_{x\Delta}^{y\Delta} \frac{e^\alpha}{j^{\beta-1}} \right] ,$$

where

$$\begin{aligned}
\sum_{x\Delta}^{y\Delta} \frac{e^\alpha}{j^{\beta-1}} &\in \left[e^\alpha \int_{x\Delta}^{y\Delta} j^{1-\beta}\, dj - e^\alpha \left((y\Delta)^{1-\beta} - (x\Delta)^{1-\beta} \right), e^\alpha \int_{x\Delta}^{y\Delta} j^{1-\beta}\, dj \right] \\
&= \left[e^\alpha \left[\frac{j^{2-\beta}}{2-\beta} \right]_{x\Delta}^{y\Delta} - \Delta \left(y^{1-\beta} - x^{1-\beta} \right), e^\alpha \left[\frac{j^{2-\beta}}{2-\beta} \right]_{x\Delta}^{y\Delta} \right] \\
&= \left[\frac{\Delta^2}{2-\beta} \left(y^{2-\beta} - x^{2-\beta} \right) - \Delta \left(y^{1-\beta} - x^{1-\beta} \right), \frac{\Delta^2}{2-\beta} \left(y^{2-\beta} - x^{2-\beta} \right) \right] .
\end{aligned}$$

We choose $y = 1$ and obtain

$$|[x\Delta, \Delta]| \in \left[\frac{\Delta}{1-\beta}\left(1-x^{1-\beta}\right) - \left(\frac{1}{x^\beta} - 1\right) - (2-x)\Delta,\ \frac{\Delta}{1-\beta}\left(1-x^{1-\beta}\right)\right] .$$

The volume of that interval is then estimated as

$$\begin{aligned}
\operatorname{vol}([x\Delta, \Delta]) &\geqslant \frac{\Delta^2}{2-\beta}\left(1-x^{2-\beta}\right) - \Delta\left(1-x^{1-\beta}\right) - \left(\frac{\Delta(\Delta+1)}{2} - \frac{x^2\Delta^2 - x\Delta}{2}\right) \\
&= \frac{\Delta^2}{2-\beta}\left(1-x^{2-\beta}\right) - \frac{\Delta^2}{2} + \frac{x^2}{2}\Delta^2 - \Delta\left(1-x^{1-\beta} - \frac{1}{2} + \frac{x}{2}\right) \\
&= \Delta^2\left(\frac{1-x^{2-\beta}}{2-\beta} - \frac{1}{2} + \frac{x^2}{2}\right) - \Delta\left(1-x^{1-\beta} - \frac{1}{2} + \frac{x}{2}\right) . \qquad (7.1)
\end{aligned}$$

which finishes the proof. □

We use the scaling technique with scaling parameter d, hence, we want to choose α such that $e^\alpha \geqslant N^{\frac{d+b\beta}{d}}$. Since $N^{\frac{d-1}{d}}$ is a lower bound for the optimum in $G^d_{U,\mathcal{S}}$, we have $N^{\frac{d-1}{d}} = e^{\frac{d-1}{d+b\beta}\cdot\alpha} = e^{(1-\delta)\alpha}$, where we can choose $1-\delta$ arbitrary close to 1. The size of the interval $[x\Delta, \Delta]$ is of order $\Delta(1-x^{1-\beta})$, hence we want to choose x such that $\Delta(1-x^{1-\beta}) = e^{\alpha/\beta}\cdot e^p$ with $\frac{\alpha}{\beta}\cdot p < (1-\delta)\alpha$, i. e. $p < (1-\delta)\beta$. So suppose we choose x such that $p = (1-\delta')\beta$, where $1-\delta'$ can be chosen arbitrary close to 1. Furthermore, the interval $[x\Delta, \Delta]$ needs to provide sufficient volume to dominate the rest of the graph, i. e. (using our volume estimate of Equation 7.1) we require that

$$\Delta^2\left(\frac{1}{2-\beta} - \frac{1}{2} - x^{2-\beta}\left(\frac{1}{2-\beta} - \frac{x^\beta}{2}\right)\right) > \Delta .$$

This yields the requirement $\frac{1}{2-\beta} - \frac{1}{2} - x^{2-\beta}\left(\frac{1}{2-\beta} - \frac{x^\beta}{2}\right) > \frac{1}{\Delta}$, which is implied by

$$1 - \frac{1}{\Delta\left(\frac{1}{2-\beta} - \frac{1}{2}\right)} > x^2 .$$

Combining this with the upper bound requirement for the size of the interval, we obtain

$$\left(1 - \frac{1-\beta}{e^{\alpha\left(\frac{1}{\beta}-(1-\delta')\right)}}\right)^{\frac{1}{1-\beta}} \leqslant x < \left(1 - \frac{1}{\left(\frac{1}{2-\beta} - \frac{1}{2}\right) \cdot e^{\alpha/\beta}}\right)^{1/2} . \tag{7.2}$$

We observe that $\frac{1}{1-\beta} > 1 > \frac{1}{2}$ for $\beta \in (0,1)$, and furthermore $\frac{\alpha}{\beta} - (1-\delta')\alpha < \frac{\alpha}{\beta}$. Hence we can choose x such that Equation 7.2 holds. Thus for this choice of x we have $|[x\Delta, \Delta]| = o\left(N^{\frac{d-1}{d}}\right)$ and $\mathrm{vol}([x\Delta, \Delta]) \geqslant |G_{\alpha,\beta}|$, fulfilling the constraints 2 and 3. We have $\mathsf{OPT}(G_{\alpha,\beta}) = (1 + o(1))\mathsf{OPT}(G^d_{U,S})$, and furthermore $N = \left(|G_{\alpha,\beta}| \cdot (1-\beta)\right)^{\frac{d\beta}{d+b\beta}}$. Altogether we obtain the following result.

Theorem 7.1.
For $0 < \beta < 1$, the Minimum Dominating Set *problem on α, β-power law graphs is hard to approximate within*

$$\frac{(1-\varepsilon)\varepsilon}{2+\varepsilon} \cdot \left(\frac{1}{2}\right)^{\frac{\varepsilon}{2+\varepsilon}} \cdot \left(\frac{\beta}{d+b\beta} \cdot \left(\ln(|G_{\alpha,\beta}|) - \ln\left(\frac{1}{1-\beta}\right)\right) - \mathcal{O}(1)\right) .$$

7.6 THE CASE $\beta = 1$

In the case $\beta = 1$, we can omit the scaling and directly embed the graph $G_{U,S}$ into a PLG $G_{\alpha,\beta}$. We proof the following lemma.

Lemma 7.3.
Let $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ be an (α,β)-PLG with $\beta = 1$. Then, for all $0 < x < y \leqslant 1$, the size and the volume of the interval $[x\Delta, y\Delta]$ satisfies

$$|[x\Delta, y\Delta]| \in \left[e^{\alpha} \ln\left(\frac{1}{x}\right) - \left(\frac{1}{x} - 1\right),\ e^{\alpha} \ln\left(\frac{1}{x}\right)\right]$$

and

$$\mathrm{vol}([x\Delta, y\Delta]) \in \left[\Delta^2\left(\frac{1}{2} - x + \frac{x^2}{2}\right) - \frac{1-x}{2}\Delta,\ (1-x)\Delta^2\right] .$$

Proof. For a given $x \in [0,1]$, the size of the interval $[x\Delta,\Delta] = \{v \in V(G_{\alpha,\beta}) \mid x\Delta \leqslant d_{\alpha,\beta}(v) \leqslant \Delta\}$ satisfies

$$\begin{aligned}|[x\Delta,\Delta]| &\in \left[\sum_{x\,e^\alpha}^{e^\alpha} \frac{e^\alpha}{j} - (1-x)\,e^\alpha,\ \sum_{x\,e^\alpha}^{e^\alpha} \frac{e^\alpha}{j}\right] \\ &\subseteq \left[e^\alpha\,(\ln(e^\alpha) - \ln(x\,e^\alpha)) - e^\alpha(\frac{1}{x}-1)\cdot\frac{1}{e^\alpha},\ e^\alpha\cdot\ln\left(\frac{1}{x}\right)\right] \\ &= \left[e^\alpha \ln\left(\frac{1}{x}\right) - \left(\frac{1}{x}-1\right),\ e^\alpha \ln\left(\frac{1}{x}\right)\right] \quad .\end{aligned}$$

The volume $\sum_{x\Delta}^{\Delta}\left\lfloor\frac{e^\alpha}{j}\right\rfloor\cdot j$ of that interval is

$$\begin{aligned}\mathrm{vol}([x\Delta,\Delta]) &\in \left[\sum_{x\Delta}^{\Delta} e^\alpha - j,\ \sum_{x\Delta}^{\Delta} e^\alpha\right] \\ &\subseteq \left[e^\alpha(1-x)\Delta - \left(\frac{\Delta(\Delta+1)}{2} - \frac{x\Delta(x\Delta+1)}{2}\right),\ e^\alpha(1-x)\Delta\right] \\ &= \left[\Delta^2\left(\frac{1}{2}-x+\frac{x^2}{2}\right) - \frac{1-x}{2}\Delta,\ (1-x)\Delta^2\right] \quad .\end{aligned}$$

□

Hence for every $x < 1$ being bounded away from 1, the volume of the interval $[x\Delta,\Delta]$ is $\omega(|G_{\alpha,1}|)$. Recall that in order to achieve $N_0 \leqslant \left|\left[N_0^a, N_0^b\right]\right|$, it suffices to choose α sufficiently large such that $N_0 \leqslant \frac{e^\alpha}{N_0^{b\beta}} = \frac{e^\alpha}{N_0^b}$. Hence suppose we have $N_0^{1+b} = e^\alpha$. This implies $\frac{e^\alpha}{N_0^b} = e^{\alpha\cdot\frac{1}{1+b}}$. Thus it suffices to choose x such that $\ln\left(\frac{1}{x}\right) = o\left(e^{\alpha\cdot\frac{b}{1+b}}\right)$.

The size of the PLG is $|G_{\alpha,\beta}| = \alpha\,e^\alpha$, and from $N_0^{1+b} = e^\alpha$, we obtain $N_0 = e^{\frac{\alpha}{1+b}} = \left(\frac{|G_{\alpha,\beta}|}{\ln(G_{\alpha,\beta})}\right)^{\frac{1}{1+b}}$. Hence, we obtain the following lower bound for the case $\beta = 1$.

Theorem 7.2.
For $\beta = 1$, the Minimum Dominating Set *problem on α,β-power law graphs is hard to approximate within*

$$\frac{(1-\varepsilon)\varepsilon}{2+\varepsilon}\cdot\left(\frac{1}{2}\right)^{\frac{\varepsilon}{2+\varepsilon}}\cdot\left(\frac{(1-o(1))\ln(|G_{\alpha,\beta}|)}{1+b} - \mathcal{O}(1)\right) \quad .$$

7.7 THE CASE $1 < \beta < 2$

We consider the case $1 < \beta < 2$ and the following lemma.

Lemma 7.4.
Let $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ be an (α, β)-PLG with $1 < \beta < 2$. Then, for all $0 < x < y \leqslant 1$, the size and the volume of the interval $[x\Delta, y\Delta]$ satisfies

$$|[x\Delta, y\Delta]| \in \left[\Delta(y-x)\left(\frac{1}{y^\beta} - 1\right), \Delta\frac{y-x}{x^\beta}\right]$$

and

$$\mathrm{vol}([x\Delta, y\Delta]) \geqslant \frac{\beta - 2x^{2-\beta} + (2-\beta)x^2}{2\cdot(2-\beta)} - o(1) \ .$$

Proof. For $1 < \beta < 2$, we have for the size of the interval $[x\Delta, y\Delta]$ that

$$\begin{aligned} |[x\Delta, y\Delta]| &\in \left[\frac{e^\alpha}{\Delta^\beta}(y-x)\Delta\frac{1}{y^\beta} - (y-x)\Delta,\ \frac{e^\alpha}{\Delta^\beta}(y-x)\Delta\frac{1}{x^\beta}\right] \\ &= \left[\Delta(y-x)\left(\frac{1}{y^\beta} - 1\right), \Delta\frac{y-x}{x^\beta}\right] \ . \end{aligned}$$

The volume $\mathrm{vol}(|x\Delta, y\Delta|) = \sum_{j=x\Delta}^{y\Delta} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \cdot j$ can be estimated as follows:

$$\begin{aligned} \mathrm{vol}(|x\Delta, y\Delta|) &\geqslant e^\alpha \sum_{j=x\Delta}^{y\Delta} j^{1-\beta} - r_\beta \\ &= (1-o(1))\, e^\alpha \cdot \int_{x\Delta}^{y\Delta} j^{1-\beta}\, dj - r_\beta \\ &= (1-o(1))\, e^\alpha \cdot \left[\frac{j^{2-\beta}}{2-\beta}\right]_{x\Delta}^{y\Delta} - r_\beta \\ &= (1-o(1))\, e^\alpha \cdot e^{\alpha\frac{2-\beta}{\beta}} \cdot \frac{y^{2-\beta} - x^{2-\beta}}{2-\beta} - r_\beta \\ &= (1-o(1))\Delta^2 \cdot \frac{y^{2-\beta} - x^{2-\beta}}{2-\beta} - r_\beta \ , \end{aligned}$$

where $r_\beta = \frac{\Delta^2(y^2)-x^2)}{2} + \frac{\Delta(y+x)}{2}$ is an upper bound for the rounding error. Hence, we obtain $\mathrm{vol}([x\Delta, y\Delta]) = \omega(|G_{\alpha,\beta}|)$ provided we choose x and y in such a way that $\frac{y^{2-\beta}-x^{2-\beta}}{2-\beta} - r_\beta > 0$. Let us choose $y = 1$. Then we have

$$\frac{y^{2-\beta}-x^{2-\beta}}{2-\beta} - r_\beta = \frac{1-x^{2-\beta}}{2-\beta} - \frac{1-x^2}{2} - o(1) = \frac{\beta - 2x^{2-\beta} + (2-\beta)x^2}{2\cdot(2-\beta)} - o(1) \ .$$

□

Now, we want to choose $x \in (0,1)$ such that $\beta - 2x^{2-\beta} + (2-\beta)x^2 > 0$. This inequality holds for $x < (\beta/2)^{\frac{1}{2-\beta}}$, since $\beta/2 < 1$.

For our choice of α, we have that $N^{\frac{d-1}{d}} = e^{\alpha\cdot\frac{d-1}{d+b\beta}}$, and hence constraint 2 holds if the following constraint is satisfied:

$$\Delta \cdot \frac{y-x}{x^\beta} = \frac{y-x}{x^\beta}\cdot e^{\frac{\alpha}{\beta}} = o\left(e^{\alpha\cdot\frac{d-1}{d+b\beta}}\right) \ .$$

Hence, for our choice of $y = 1$ and $x < (\beta/2)^{\frac{1}{2-\beta}}$, this last constraint is satisfied if $\frac{\alpha}{\beta} < \alpha\cdot\frac{d-1}{d+b\beta}$, i.e. $d > \frac{(b+1)\beta}{\beta-1}$.

RESULTING LOWER BOUND. Since the parameter α is chosen such that $e^\alpha = N^{1+\frac{b\beta}{d}}$, we have $\left|G_{\alpha,\beta}\right| = \zeta(\beta)\cdot N^{1+\frac{b\beta}{d}}$ and thus obtain the following bounds on the size of an optimum dominating set for $G_{\alpha,\beta}$. Let $\frac{|G_{\alpha,\beta}|}{\zeta(\beta)} = \phi$, then

$$\begin{aligned}|\mathsf{OPT}(G_{\alpha,\beta})| &\leqslant \left(\phi^{\frac{d}{d+b\beta}}\right)^{\frac{d-1}{d}} k\left(\phi^{\frac{d}{d+b\beta}}\right)^{\frac{1}{d}\cdot\frac{\varepsilon}{2+\varepsilon}} \\ &= k\left(\phi^{\frac{d-1+\frac{\varepsilon}{2+\varepsilon}}{d+b\beta}}\right)\end{aligned}$$

or

$$|\mathsf{OPT}(G_{\alpha,\beta})| \geqslant k\left(\phi^{\frac{d-1+\frac{\varepsilon}{2+\varepsilon}}{d+b\beta}}\right)\frac{(1-\varepsilon)\varepsilon}{2+\varepsilon}\left(\frac{1}{2}\right)^{\frac{\varepsilon}{2+\varepsilon}}\left(\ln\left(\phi^{\frac{d}{d+b\beta}\frac{1}{d}}\right) - \mathcal{O}(1)\right) \ .$$

Altogether, we obtain the following theorem.

Theorem 7.3.
For $1 < \beta < 2$, the Minimum Dominating Set *problem on α, β-power law graphs is hard to approximate within*

$$\frac{(1-\varepsilon)\cdot\varepsilon}{2+\varepsilon}\cdot\left(\frac{1}{2}\right)^{\frac{\varepsilon}{2+\varepsilon}}\cdot\frac{\ln\left(\left|G_{\alpha,\beta}\right|\right)-\ln(\zeta(\beta))}{d+b\beta} \quad .$$

7.8 THE CASE $\beta = 2$

For the case $\beta = 2$, we state a slightly different version of Lemma 7.4.

Lemma 7.5.
Let $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ be an (α, β)-PLG with $\beta = 2$. Then, for all $0 < x < y \leqslant 1$, the size and the volume of the interval $[x\Delta, y\Delta]$ satisfies

$$|[x\Delta, y\Delta]| \in \left[\sqrt{e^{\alpha}}\cdot\frac{y-x}{y^{\beta}}, \sqrt{e^{\alpha}}\cdot\frac{y-x}{x^{\beta}}\right]$$

and

$$\mathrm{vol}([x\Delta, y\Delta]) = (1-o(1))\, e^{\alpha} \ln\left(\frac{1}{x}\right) \quad .$$

Proof. We give an estimate of the size of the interval $[x\Delta, y\Delta]$ and of the volume of that interval. We have that

$$|[x\Delta, y\Delta]| \in \left[\Delta\frac{y-x}{y^{\beta}}, \Delta\frac{y-x}{x^{\beta}}\right] = \left[\sqrt{e^{\alpha}}\cdot\frac{y-x}{y^{\beta}}, \sqrt{e^{\alpha}}\cdot\frac{y-x}{x^{\beta}}\right] \quad .$$

The volume of the interval $[x\Delta, y\Delta]$ is

$$\begin{aligned}
\mathrm{vol}([x\Delta, y\Delta]) &= (1-o(1)) \sum_{j=x\Delta}^{y\Delta} \frac{e^{\alpha}}{j^{\beta}}\cdot j \\
&= (1-o(1))\, e^{\alpha}\,(\ln(y\Delta) - \ln(x\Delta)) \\
&= (1-o(1))\, e^{\alpha}\left(\ln\left(\frac{1}{x}\right) - \ln\left(\frac{1}{y}\right)\right) \quad .
\end{aligned}$$

We choose $y = 1$ and obtain

$$\mathrm{vol}([x\Delta, y\Delta]) = (1 - o(1)) \sum_{j=x\Delta}^{y\Delta} \frac{e^{\alpha}}{j^{\beta}} \cdot j = (1 - o(1))\, e^{\alpha} \left(\ln\left(\frac{1}{x}\right) - 0 \right) \quad . \tag{7.3}$$

□

In Equation 7.3, in order to satisfy constraint 3, we want to choose x such that $\ln\left(\frac{1}{x}\right) \geqslant \zeta(\beta)$, i. e. $x \leqslant \frac{1}{e^{\zeta(\beta)}}$. Then the volume of the interval $[x\Delta, \Delta]$ suffices to dominate the rest of the graph, where the size of the interval $[x\Delta, \Delta]$ satisfies $|[x\Delta, \Delta]| \in \left[\Delta\frac{1-x}{1}, \Delta\frac{1-x}{x^{\beta}}\right]$. Now, for the setting of the interval $[N^{a/d}, N^{b/d}]$, we do not want to occupy space that is needed for our embedding and the choice of the interval $[x\Delta, \Delta]$. Therefore, the two intervals $[x\Delta, \Delta]$ and $[N^{a/d}, N^{b/d}]$ need to be node disjoint and hence we want to choose the scaling parameter d such that $N^{b/d} < x\Delta$.

For $x = \frac{1}{e^{\zeta(\beta)}}$, we have $x\Delta = e^{\alpha/\beta - \zeta(\beta)}$. Furthermore, the size N of the graph $G^d_{U,\mathcal{S}}$ satisfies $N = |G^d_{U,\mathcal{S}}| \leqslant e^{\alpha\frac{d}{d+b\beta}}$. This yields the following bound for the scaling parameter d:

$$N^{b/d} < x\Delta \iff e^{\alpha b \cdot \frac{1}{d+b\beta}} < e^{\alpha/\beta - \zeta(\beta)} \iff d > \frac{\alpha \cdot b}{\alpha/\beta - \zeta(\beta)} - b\beta \quad .$$

RESULTING LOWER BOUND. Constraint (1′) yields the following bound for the size of the power law graph: $e^{\alpha} \geqslant N^{1+\frac{b\beta}{d}}$. Thus we choose $e^{\alpha} = N^{1+\frac{b\beta}{d}}$ which implies $\left|G_{\alpha,\beta}\right| = \zeta(\beta) \cdot N^{1+\frac{b\beta}{d}}$. Thus, we obtain the following bounds for the size of an optimum dominating set for $G_{\alpha,\beta}$. Let $\frac{|G_{\alpha,\beta}|}{\zeta(\beta)} = \phi$, then

$$|\mathsf{OPT}(G_{\alpha,\beta})| \leqslant \left(\phi^{\frac{d}{d+b\beta}}\right)^{\frac{d-1}{d}} k \left(\phi^{\frac{d}{d+b\beta}}\right)^{\frac{1}{d}\frac{\varepsilon}{2+\varepsilon}} = k\left(\phi^{\frac{d-1+\frac{\varepsilon}{2+\varepsilon}}{d+b\beta}}\right)$$

or

$$|\mathsf{OPT}(G_{\alpha,\beta})| \geqslant k\left(\phi^{\frac{d-1+\frac{\varepsilon}{2+\varepsilon}}{d+b\beta}}\right) \frac{(1-\varepsilon)\varepsilon}{2+\varepsilon} \left(\frac{1}{2}\right)^{\frac{\varepsilon}{2+\varepsilon}} \left(\ln\left(\phi^{\frac{d}{d+b\beta}\frac{1}{d}}\right) - \mathcal{O}(1)\right) \quad .$$

Hence, we obtain the following result.

Theorem 7.4.
For $\beta = 2$, the Minimum Dominating Set *problem on α, β-power law graphs is hard to approximate within*

$$\frac{(1-\varepsilon)\varepsilon}{2+\varepsilon} \cdot \left(\frac{1}{2}\right)^{\frac{\varepsilon}{2+\varepsilon}} \cdot \frac{\ln\left(\left|G_{\alpha,\beta}\right|\right) - \ln(\zeta(\beta))}{d + b\beta} \ .$$

We are now going to discuss the case $\beta > 2$ and present improved approximation upper bounds for this case.

7.9 NEW UPPER BOUNDS FOR $\beta > 2$

For $\beta > 2$, the Min-DS problem on (α, β)-PLG is in APX. This was already observed by Shen et al. [She+12]. They showed that in that case, there exists an efficient approximation algorithm with approximation ratio $(\zeta(\beta) - 1/2)/(\zeta(\beta) - \sum_{j=1}^{t_0} 1/j^\beta)$ for some $t_0 = \mathcal{O}(1)$. In this section we will give an explicit upper bound, based on our techniques of estimating sizes and volumes of intervals in (α, β)-PLG.

Lemma 7.6.
If $\mathrm{vol}([x\Delta, \Delta]) = \sum_{j=x\Delta}^{\Delta} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \cdot j < \lfloor e^\alpha \rfloor$, *then* $|[x\Delta, \Delta]|$ *is a lower bound on the size of a dominating set in* $G_{\alpha,\beta}$.

Proof. Let D be a dominating set in $G_{\alpha,\beta}$, and let $D_1 = D \cap [x\Delta, \Delta]$ and $D_2 = D \setminus D_1$. Suppose $|D_2| < |[x\Delta, \Delta] \setminus D_1|$. Since $\forall v \in D_2, u \in [x\Delta, \Delta] \setminus D_1$ we have $d_{\alpha,\beta}(v) < d_{\alpha,\beta}(u)$, this implies $\mathrm{vol}(D_2) < \mathrm{vol}([x\Delta, \Delta] \setminus D_1)$ and thus $\mathrm{vol}(D) < \mathrm{vol}([x\Delta, \Delta]) < \lfloor e^\alpha \rfloor$, a contradiction. □

The lower bound on the size of a dominating set in $G_{\alpha,\beta}$ given in the following Lemma 7.7 was also used in a somewhat different formulation by Shen et al. [She+12].

Lemma 7.7.
If $\mathrm{vol}([x\Delta, \Delta]) = \sum_{j=x\Delta}^{\Delta} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor \cdot j < \sum_{j=1}^{x\Delta-1} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor$, *then* $|[x\Delta, \Delta]|$ *is a lower bound on the size of a dominating set in* $G_{\alpha,\beta}$.

Proof. Suppose that $\mathrm{vol}([x\Delta, \Delta]) < |[1, x\Delta - 1]|$ and that D, D_1, D_2 are the same as in the proof of Lemma 7.6. Again we obtain $\mathrm{vol}(D_2) < \mathrm{vol}([x\Delta, \Delta] \setminus D_1)$, which implies

$\mathrm{vol}(D) < \mathrm{vol}([x\Delta, \Delta]) < |[1, x\Delta - 1]$. Thus the volume of D is not sufficient to dominate the subset $[1, x\Delta - 1]$, a contradiction. □

In order to demonstrate the power and the limitations of this lower bound, we want to determine the value $\min\{\, x \mid \mathrm{vol}([x\Delta, \Delta]) < \lfloor e^{\alpha} \rfloor \,\}$.

In the case $\beta > 2$, we consider the following estimates of sizes of intervals and the node degree, i. e. the volume, that is available in such intervals.

$$\begin{aligned}
\sum_{x\Delta}^{y\Delta} \frac{1}{j^{\beta-1}} &\in \left[\int_{x\Delta}^{y\Delta} \frac{1}{j^{\beta-1}}\, dj - \left(\frac{1}{(x\Delta)^{\beta-1}} - \frac{1}{(y\Delta)^{\beta-1}} \right), \int_{x\Delta}^{y\Delta} \frac{1}{j^{\beta-1}}\, dj \right] \\
&= \left[\left[\frac{j^{2-\beta}}{2-\beta} \right]_{x\Delta}^{y\Delta} - \left(\frac{1}{(x\Delta)^{\beta-1}} - \frac{1}{(y\Delta)^{\beta-1}} \right), \left[\frac{j^{2-\beta}}{2-\beta} \right]_{x\Delta}^{y\Delta} \right] \\
&= \left[\frac{y^{2-\beta} - x^{2-\beta}}{2-\beta} \Delta^{2-\beta} - \left(\left(x^{1-\beta} - y^{1-\beta} \right) \Delta^{1-\beta} \right), \frac{y^{2-\beta} - x^{2-\beta}}{2-\beta} \Delta^{2-\beta} \right] .
\end{aligned}$$

For the size of the interval $|[x\Delta, y\Delta]| = \sum_{x\Delta}^{y\Delta} \frac{e^{\alpha}}{j^{\beta}}$ we get

$$\begin{aligned}
|[x\Delta, y\Delta]| &\in e^{\alpha} \left[\frac{\Delta^{1-\beta}}{1-\beta} \left(y^{1-\beta} - x^{1-\beta} \right) - \Delta^{-\beta} \left(\frac{1}{x^{\beta}} - \frac{1}{y^{\beta}} \right), \frac{\Delta^{1-\beta}}{1-\beta} \left(y^{1-\beta} - x^{1-\beta} \right) \right] \\
&= e^{\alpha} \cdot e^{\alpha \frac{1-\beta}{\beta}} \left[\frac{x^{1-\beta} - y^{1-\beta}}{\beta - 1} - \frac{1}{\Delta} \left(\frac{1}{x^{\beta}} - \frac{1}{y^{\beta}} \right), \frac{x^{1-\beta} - y^{1-\beta}}{\beta - 1} \right] \\
&= \left[\Delta \frac{x^{1-\beta} - y^{1-\beta}}{\beta - 1} - \left(\frac{1}{x^{\beta}} - \frac{1}{y^{\beta}} \right), \Delta \frac{x^{1-\beta} - y^{1-\beta}}{\beta - 1} \right] .
\end{aligned}$$

We consider the case $x = 2/\Delta, y = 1$. We obtain $|[2, \Delta]| = \zeta(\beta)\, e^{\alpha} - e^{\alpha} = (\zeta(\beta) - 1)\, e^{\alpha}$ and

$$\sum_{j=2}^{\Delta} \frac{e^{\alpha}}{j^{\beta}} \cdot j \in \left[e^{\alpha} \left(\frac{\left(\frac{\Delta}{2}\right)^{\beta-2} - 1}{\beta - 2} \Delta^{2-\beta} - \left(\left(\frac{\Delta}{2} \right)^{\beta-1} - 1 \right) \Delta^{1-\beta} \right), e^{\alpha} \frac{\left(\frac{\Delta}{2}\right)^{\beta-2} - 1}{\beta - 2} \Delta^{2-\beta} \right] .$$

For the interval $[d, \Delta]$ we obtain

$$\sum_{j=d}^{\Delta} \left\lfloor \frac{e^{\alpha}}{j^{\beta}} \right\rfloor j \leqslant \Delta^2 \frac{\left(\frac{\Delta}{d}\right)^{\beta-2} - 1}{\beta - 2} = \frac{1}{\beta - 2} \left(\frac{e^{\frac{2\alpha}{\beta}}\, e^{\alpha \frac{\beta-2}{\beta}}}{d^{\beta-2}} - e^{\frac{2\alpha}{\beta}} \right) = (1 - o(1)) \frac{e^{\alpha}}{d^{\beta-2}(\beta - 2)} .$$

We obtain the following estimate for the size of the interval $[1, d-1]$:

$$\begin{aligned}
|[1, d-1]| &= \sum_{j=1}^{d} \left\lfloor \frac{e^{\alpha}}{j^{\beta}} \right\rfloor \geqslant \sum_{j=1}^{d} \frac{e^{\alpha}}{j^{\beta}} - (d-1) \\
&\geqslant e^{\alpha} \cdot \left(\int_{1}^{d-1} j^{-\beta} \, dj - \left(1 - \frac{1}{(d-1)^{\beta}} \right) \right) - (d-1) \\
&= e^{\alpha} \cdot \left(\frac{(d-1)^{1-\beta} - 1}{1-\beta} - \left(1 - \frac{1}{(d-1)^{\beta}} \right) \right) - (d-1) \\
&= e^{\alpha} \cdot \left(\frac{1 - \frac{1}{(d-1)^{\beta-1}}}{\beta - 1} - 1 + \frac{1}{(d-1)^{\beta}} \right) - (d-1) \ .
\end{aligned}$$

Hence, we want to determine the smallest $d \geqslant 2$ such that

$$\frac{1}{d^{\beta-2} \cdot (\beta-2)} < \frac{(d-1)^{\beta} - (d-1) - (\beta-1)(d-1)^{\beta} + \beta - 1}{(\beta-1)(d-1)^{\beta}} \ .$$

We observe that $1 + \frac{1}{2^{\beta}} > \frac{1}{3^{\beta-2}(\beta-2)}$ for $\beta \geqslant \beta_2 \approx 2.48$, $1 + \frac{1}{2^{\beta}} + \frac{1}{3^{\beta}} > \frac{1}{4^{\beta-2}(\beta-2)}$ for $\beta \geqslant \beta_3 \approx 2.44$ and $1 + \frac{1}{2^{\beta}} + \frac{1}{3^{\beta}} + \frac{1}{4^{\beta}} > \frac{1}{5^{\beta-2}(\beta-2)}$ for $\beta \geqslant \beta_4 \approx 2.40$. This gives the following upper bounds for the approximability of MIN-DS on (α, β)-PLG for $\beta > 2$.

Lemma 7.8.
For $k \in \{2, 3, 4\}$ *let* $\beta_k = \min \left\{ \beta \;\middle|\; \sum_{j=1}^{k} \frac{1}{j^{\beta}} > \frac{1}{k^{\beta-2}(\beta-2)} \right\}$. *Then we have* $\beta_2 \approx 2.48$, $\beta_3 \approx 2.44$ *and* $\beta_3 \approx 2.40$. *For* $k \in \{2, 3, 4\}$, *for* $\beta \geqslant \beta_k$, *the* MINIMUM DOMINATING SET *problem in* (α, β)*-PLG is hard to approximate within approximation ratio* $\left(\zeta(\beta) - \frac{1}{2}\right) \cdot (\beta - 2) \cdot (k+1)^{\beta-2}$.

7.9.1 *Improved Analysis*

We will now significantly improve the analysis based on the lower bounds from Lemma 7.6. Instead of just giving upper and lower bounds on the size of an optimum dominating set and a greedy solution separately, we will explicitly relate upper and lower bound to each other.

Let $G_{\alpha,\beta}$ be an (α,β)-PLG with $\beta > 2$. Let W be the set of neighbors of degree 1 nodes of degree at least 2 in $G_{\alpha,\beta}$ and let M be the set of degree 1 nodes in $G_{\alpha,\beta}$ which are adjacent to another degree 1 node. Let $R = V \setminus (W \cup \{v \in V \mid d_{\alpha,\beta}(v) = 1\})$.

Then there exists some $c = c_\beta > 0$ not depending on α such that $|W| \geqslant c \cdot e^\alpha$. This implies $|R| \leqslant (\zeta(\beta) - c - 1)\, e^\alpha$.

Lemma 7.9.
If $G_{\alpha,\beta}$ is a connected (α,β)-PLG with $\beta > 2$ and W and R are defined as above, then there exists an optimum dominating set OPT in $G_{\alpha,\beta}$ with $OPT = OPT_R \cup W \cup M'$, where OPT_R is an optimum dominating set for the induced subgraph $G_{\alpha,\beta}[R]$ on R and $M' \subset M$ is of cardinality $|M'| = \frac{|M|}{2}$.

The maximum degree in $G_{\alpha,\beta}[R]$ is at most Δ. We consider the dominating set $D = W \cup D_{Gr} \cup M'$ where D_{Gr} is a dominating set for $G_{\alpha,\beta}[R]$ constructed by the greedy algorithm and $M' \subset M$ is a subset of size $\frac{|M|}{2}$ dominating M. The approximation ratio is at most

$$\frac{\ln(\Delta+1) \cdot |OPT_R| + |W| + \frac{|M|}{2}}{|OPT_R| + |W| + \frac{|M|}{2}} \leqslant \frac{\frac{\alpha}{\beta} \cdot |OPT_R| + |W| + \frac{|M|}{2}}{|OPT_R| + |W| + \frac{|M|}{2}} \quad .$$

We can further improve this bound as follows. Since $R = V \setminus (W \cup V_1)$ and $|OPT_R| \leqslant |R|$, the approximation ratio is at most

$$\max \left\{ \frac{r \cdot |OPT_R| + |W| + \frac{|M|}{2}}{|OPT_R| + |W| + \frac{|M|}{2}} \;\middle|\; \begin{array}{l} |OPT_R| \leqslant |R|, \\ r = \min\left\{\frac{\alpha}{\beta}, \frac{|R|}{|OPT_R|}\right\} \end{array} \right\} \quad .$$

CASE 1: $\left(r = \frac{\alpha}{\beta}\right)$ This means that $\frac{\alpha}{\beta} \leqslant \frac{|R|}{|OPT_R|}$, i. e. $|OPT_R| \leqslant \frac{\beta}{\alpha} \cdot |R|$. The upper bound for the approximation ratio is monotone increasing in $|OPT_R|$, hence it is bounded by

$$\frac{\frac{\alpha}{\beta} \cdot \frac{\beta}{\alpha} \cdot |R| + |W| + \frac{|M|}{2}}{\frac{\beta}{\alpha} \cdot |R| + |W| + \frac{|M|}{2}} = \frac{|R| + |W| + \frac{|M|}{2}}{\frac{\beta}{\alpha} \cdot |R| + |W| + \frac{|M|}{2}} \quad .$$

CASE 2: $\left(r = \frac{|R|}{|\mathsf{OPT}_R|} < \frac{\alpha}{\beta}\right)$ Here, we have $|\mathsf{OPT}_R| > \frac{\beta \cdot |R|}{\alpha}$ and obtain

$$\frac{r \cdot |\mathsf{OPT}_R| + |W| + \frac{|M|}{2}}{|\mathsf{OPT}_R| + |W| + \frac{|M|}{2}} = \frac{|R| + |W| + \frac{|M|}{2}}{|\mathsf{OPT}_R| + |W| + \frac{|M|}{2}} \leqslant \frac{|R| + |W| + \frac{|M|}{2}}{\frac{\beta}{\alpha} \cdot |R| + |W| + \frac{|M|}{2}} \quad .$$

Now we need to construct an upper bound for the term $\frac{|R|+|W|+\frac{|M|}{2}}{\frac{\beta}{\alpha}\cdot|R|+|W|+\frac{|M|}{2}}$. Recall that the volume of a set of nodes $U \subseteq V$ is defined as $\mathrm{vol}(U) = \sum_{u \in U} d_{\alpha,\beta}(u)$. We consider the following two cases.

CASE 1: $(\zeta(\beta - 1) - 1 < 1)$ In this case, the volume of nodes of degree at least 2 does not suffice to dominate all the degree 1 nodes. Hence in this case, $M \neq \emptyset$. We obtain the following lower bound for the cardinality of M.

$$|M| \geqslant e^{\alpha} - (\zeta(\beta - 1) - 1)\, e^{\alpha} = (2 - \zeta(\beta - 1))\, e^{\alpha} \quad .$$

Nevertheless, we will use the upper bound $|R| \leqslant (\zeta(\beta) - 1)\, e^{\alpha}$. Since the term

$$\frac{|R| + |W| + \frac{|M|}{2}}{\frac{\beta}{\alpha} \cdot |R| + |W| + \frac{|M|}{2}}$$

is monotone increasing in $|R|$, we obtain an approximation ratio of

$$\rho(\beta) = \frac{|R| + |W| + \frac{|M|}{2}}{\frac{\beta}{\alpha} \cdot |R| + |W| + \frac{|M|}{2}} \leqslant \frac{(\zeta(\beta) - 1)\, e^{\alpha} + \frac{(2 - \zeta(\beta-1))\, e^{\alpha}}{2}}{\frac{\beta}{\alpha} \cdot (\zeta(\beta) - 1)\, e^{\alpha} + \frac{(2 - \zeta(\beta-1))\, e^{\alpha}}{2}} = \frac{\zeta(\beta) - \frac{\zeta(\beta-1)}{2}}{1 - \frac{\zeta(\beta-1)}{2}} \quad .$$

In Figure 7.5 we plot the above approximation ratio in comparison to the ratio $\frac{\zeta(\beta)-1/2}{\zeta(\beta)-1}$ of Shen et al. [She+12] for values of the parameter $\beta \geqslant 2.75$.

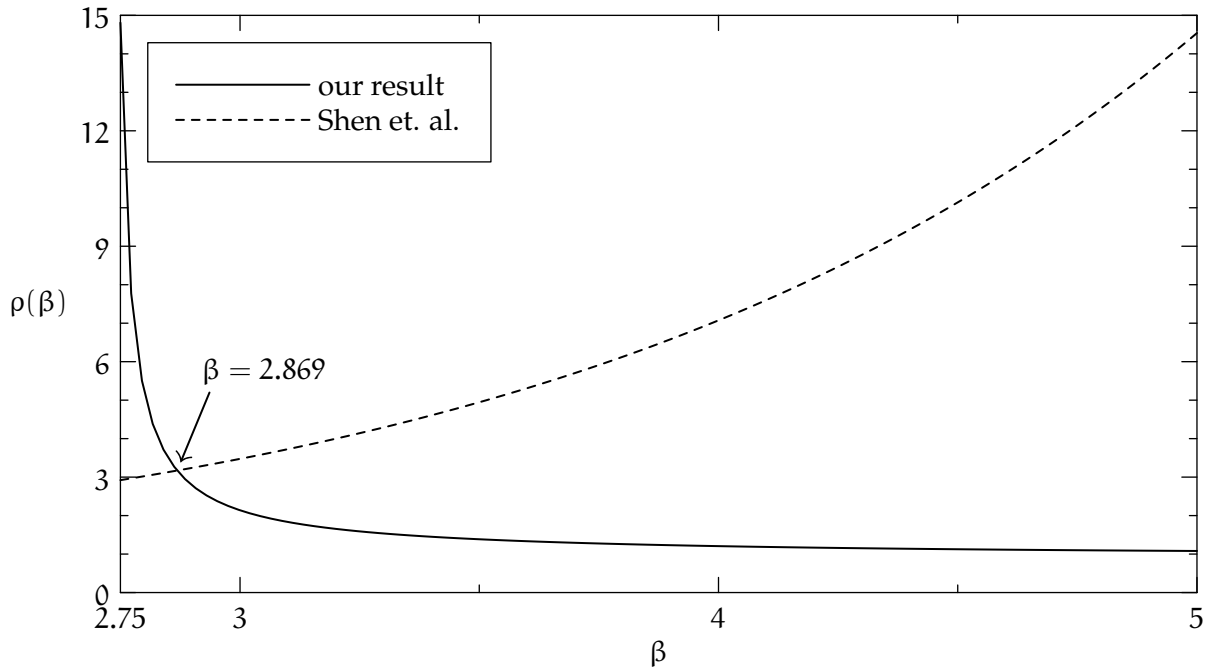


Figure 7.5: Plot of the approximation ratios $\frac{\zeta(\beta)-\frac{\zeta(\beta-1)}{2}}{1-\frac{\zeta(\beta-1)}{2}}$ (our result) in comparison to $\frac{\zeta(\beta)-\frac{1}{2}}{\zeta(\beta)-1}$ (Shen et al. [She+12]) for $\beta \geqslant 2.75$.

CASE II: $(\zeta(\beta-1)-1 \geqslant 1)$ In this case, the volume of the nodes of degree at least 2 suffices to dominate the degree 1 nodes. Now, we construct a lower bound for $|W|$ as follows:

$$\begin{aligned} |W| &\geqslant \min\{\, |[d,\Delta]| \mid \mathrm{vol}([d,\Delta]) > e^{\alpha}\,\} \\ &= \min\left\{ \left(\zeta(\beta) - \sum_{j=1}^{d-1} \frac{1}{j^{\beta}}\right) e^{\alpha} \;\middle|\; \left(\zeta(\beta-1) - \sum_{j=1}^{d-1} \frac{1}{j^{\beta-1}}\right) e^{\alpha} > e^{\alpha} \right\} . \end{aligned}$$

Hence, in this case, the approximation ratio is bounded by

$$\rho'(\beta) = \frac{\zeta(\beta)-1}{\frac{\beta}{\alpha}\cdot|[1,d-1]|+|[d,\Delta]|} = \frac{\zeta(\beta)-1}{\zeta(\beta)-\sum_{j=1}^{d-1}\frac{1}{j^{\beta}}} ,$$

where $d = \min\{\, d' \mid \mathrm{vol}([d',\Delta]) > e^{\alpha}\,\}$.

Altogether, we obtain the following theorem.

Theorem 7.5.
For $2 < \beta \leqslant 2.729$, the Min-DS *problem on (α, β)-PLG is approximable within approximation ratio $\rho'(\beta)$ and for $\beta > 2.729$ within approximation ratio $\rho(\beta)$, where $d = \min\{\, d' \mid \mathrm{vol}([d', \Delta]) > e^{\alpha}\,\}$ and*

$$\rho'(\beta) = \frac{\zeta(\beta) - 1}{\zeta(\beta) - \sum_{j=1}^{d-1} \frac{1}{j^{\beta}}} \quad \textit{and} \quad \rho(\beta) = \frac{\zeta(\beta) - \frac{\zeta(\beta-1)}{2}}{1 - \frac{\zeta(\beta-1)}{2}}$$

In Figure 7.6 we present a plot of the above approximation ratios $\rho(\beta)$ and $\rho'(\beta)$ in the valid ranges for certain choices of the parameter d.

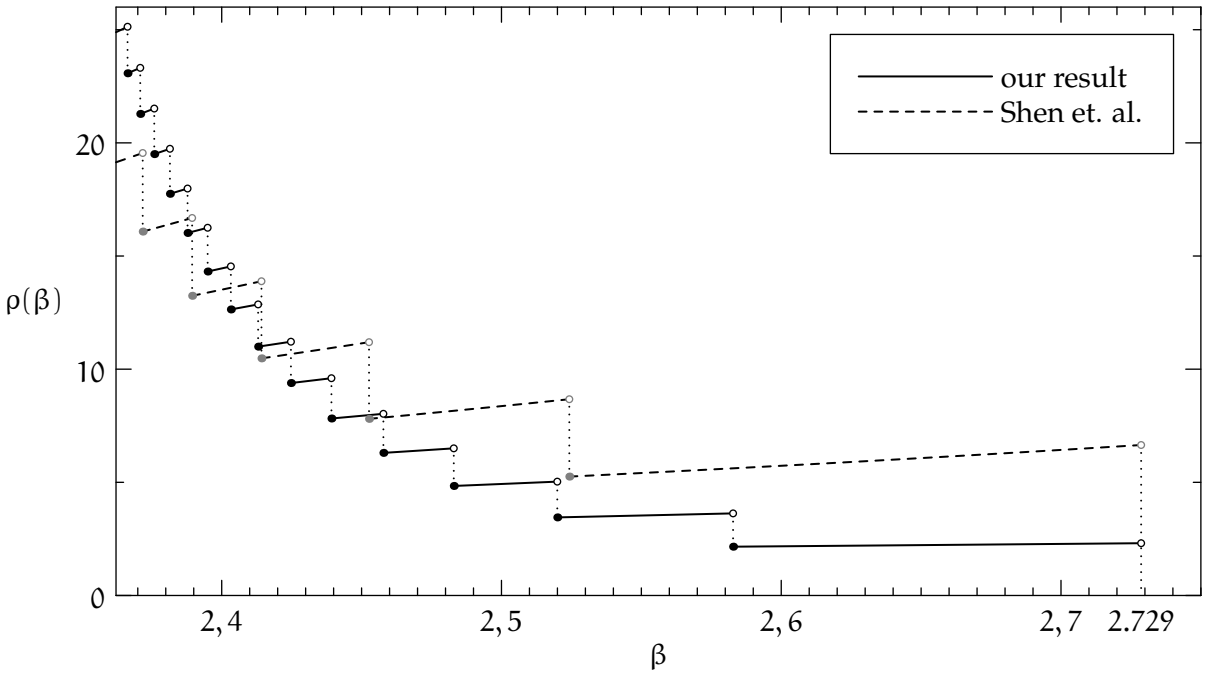


Figure 7.6: Comparison of our approximation ratio to the previous approximation ratio of Shen et al. [She+12]

In what follows, we are going to analyze the functional dependencies of the parameter β at the phase transition point $\beta = 2$.

7.10 THE FUNCTIONAL CASE $\beta_f = 2 + \frac{1}{f(n)}$

We consider now the case when the parameter β is a function of the size n of the power law graph, converging to 2 from above. In the preceding sections we have shown that for $\beta \leqslant 2$, there is a logarithmic lower bound for the approximability of the MINIMUM DOMINATING SET problem in (α, β)-PLG. On the other hand, for $\beta > 2$ the problem is in APX (cf. Shen et al. [She+12] and the previous Section 7.9). Thus we may now have a closer look at this phase transition at $\beta = 2$. Similar as in the previous Chapter 6, we consider the case when the value of the parameter β is a function of the size n of the power law graph such that this function converges to 2 from above. Surprisingly we will obtain a very tight phase transition of the computational complexity of the problem, depending on the convergence rate of the function. Let us first give a precise description of the model.

Definition 7.1 ((α, β_f)-PLG for $\beta_f = 2 + 1/f(n)$)**.**
Let $f\colon \mathbb{N} \to \mathbb{N}$ be a monotone increasing unbounded function. For $\beta_f = 2 + \frac{1}{f(n)}$, an (α, β_f)-PLG is an undirected multigraph G_{α,β_f} with n nodes and maximum degree $\Delta_f = \left\lfloor e^{\alpha/\beta_f} \right\rfloor$ such that for $j = 1, \ldots, \Delta_f = \left\lfloor e^{\alpha/\beta_f} \right\rfloor$, the number of nodes of degree j in G_{α,β_f} equals $\left\lfloor \frac{e^\alpha}{j^{\beta_f}} \right\rfloor$.

Especially, this means that the total number of nodes in G_{α,β_f} is $\sum_{j=1}^{\Delta_f} \left\lfloor \frac{e^\alpha}{j^{2+1/f(n)}} \right\rfloor = n$.

In order to study the computational complexity of the MINIMUM DOMINATING SET problem in (α, β_f)-power law graphs, we consider two cases for $\beta_f = 2 + f(n)^{-1}$, namely, $f(n) = \omega(\log(n)$ and $f(n) = o(\log(n))$. We begin with the case $f(n) = \omega(\log(n)$.

7.10.1 *The Case* $f(n) = \omega(\log(n))$

For the case $f(n) = \omega(\log(n))$, we prove the following theorem regarding the inapproximability of MIN-DS in (α, β_f)-power law graphs.

Theorem 7.6.
For $\beta_f = 2 + f(n)^{-1}$ *with* $f(n) = \omega(\log(n))$, *the* MIN-DS *problem on* (α, β_f)*-PLG is hard to approximate within*

$$\frac{(1-\varepsilon)\cdot\varepsilon}{2+\varepsilon} \cdot \left(\frac{1}{2}\right)^{\frac{\varepsilon}{2+\varepsilon}} \cdot \frac{\ln\left(\left|\mathcal{G}_{\alpha,\beta}\right|\right) - \ln(\zeta(\beta))}{d + b\beta} \quad .$$

Proof. The proof of the above theorem is structured as follows. First, we will first prove two auxiliary lemmas regarding the convergence of terms $j^{-\beta_f}$ and the size and volume of the interval $X = [x\Delta_f, y\Delta_f]$. Then we consider the requirements 1–3 and show how to choose the parameters α, d, x, y in order to satisfy the requirements.

CONVERGENCE OF TERMS $j^{-\beta_f}$. We start by showing the convergence of the terms $j^{-\beta_f}$, i. e. we prove the following auxiliary lemma.

Lemma 7.10 (Convergence)**.**
Let $\beta_f = 1 + \frac{1}{f(n)}$ *with* $f(n) = \omega(\alpha)$, *and let* $\Delta_f = \left\lfloor e^{\alpha/\beta_f} \right\rfloor$. *Then for each* $j \in \{1, \dots, \Delta_f\}$,

$$j^{-\beta_f} \in \left[\frac{1}{n^{1/f(n)}} \cdot \frac{1}{j^2}, \frac{1}{j^2}\right] \quad .$$

Proof of Lemma 7.10. First we give an additive bound for the terms $j^{-\beta_f}$.

$$\frac{1}{j^{\beta_f}} = \frac{1}{j^{2+\frac{1}{f(n)}}} \in \left[\frac{1}{j^2} - \tau(n), \frac{1}{j^2}\right] \quad ,$$

where

$$\begin{aligned}\tau(n) &= \max\left\{\left.\frac{1}{j^2} - \frac{1}{j^{2-\frac{1}{f(n)}}}\right| j = 1, \dots, \Delta_f\right\} \\ &= \max\left\{\left.\frac{j^{\frac{1}{f(n)}} - 1}{j^{2+\frac{1}{f(n)}}}\right| j = 1, \dots, \Delta_f\right\} \quad .\end{aligned}$$

We consider the function $x \mapsto h(x) := \frac{x^{\frac{1}{f(n)}}-1}{x^{2+\frac{1}{f(n)}}} = x^{-2} - x^{-2-\frac{1}{f(n)}}$. Its derivative is $\frac{d}{dx}h(x) = \frac{d}{dx}\frac{x^{\frac{1}{f(n)}}-1}{x^{2+\frac{1}{f(n)}}} = -2x^{-3} + \left(2+\frac{1}{f(n)}\right)x^{-3-\frac{1}{f(n)}}$. The condition $h(x) < 0$ is equivalent to $1 + \frac{1}{2f(n)} < x^{\frac{1}{f(n)}}$. We observe that the derivative attains its maximum at $x = 2$. We have

$$h'(2) < 0 \iff \left(1+\frac{1}{2f(n)}\right)^{f(n)} < 2 \ .$$

We observe that $\lim_{n\to\infty}\left(1+\frac{1}{2f(n)}\right)^{f(n)} = e^{1/2} < 2$. Thus we obtain

$$\tau(n) = \frac{2^{1/f(n)}-1}{2^{2+1/f(n)}} \ .$$

Now we give a multiplicative bound as follows. We have

$$\frac{1}{j^{\beta_f}} = \frac{1}{j^2}\cdot j^{2-\beta_f} = \frac{1}{j^2}\cdot\frac{1}{j^{1/f(n)}} \in \left[\frac{1}{n^{1/f(n)}}\cdot\frac{1}{j^2}, \frac{1}{j^2}\right] \ .$$

□

SIZES OF INTERVALS. Let us now give sufficiently precise estimates of sizes and volumes of intervals in the functional case. We prove the following lemma.

Lemma 7.11 (Size and Volume of the Intervals)**.**
Let $\beta_f = 2 + \frac{1}{f(n)}$ *and* $X = [x\Delta_f, y\Delta_f]$. *We have the following bounds on the size and the volume of the interval:*

$$|[x\Delta_f, y\Delta_f]| \in$$
$$\left[e^{\alpha\frac{f(n)+1}{2f(n)+1}}\cdot\left(\frac{1}{x}-\frac{1}{y}\right) - (y-x)\Delta_f,\ e^{\alpha\frac{f(n)+1}{2f(n)+1}}\cdot\left(\frac{1}{x}-\frac{1}{y}\right) + e^{\alpha\frac{1}{2f(n)+1}}\cdot\left(\frac{1}{x^2}-\frac{1}{y^2}\right)\right]$$

and

$$\operatorname{vol}([x\Delta_f, y\Delta_f]) \in$$
$$\left[\frac{e^{\alpha}(\ln(y)-\ln(x))}{n^{\frac{1}{f(n)}}} - \frac{(y^2-x^2)\Delta_f^2 + (x+y)\Delta_f}{2},\ e^{\alpha}(\ln(y)-\ln(x)) + e^{\alpha}\left(\frac{1}{x\Delta_f} - \frac{1}{y\Delta_f}\right)\right].$$

Proof of Lemma 7.11. For $\beta = 2$, our technique based on integration yields the following estimate of sizes of intervals:

$$\sum_{j=x\Delta}^{y\Delta} \frac{1}{j^2} \in \left[\int_{x\Delta}^{y\Delta} j^{-2}\,dj,\ \int_{x\Delta}^{y\Delta} j^{-2}\,dj + \frac{1}{(x\Delta)^2} - \frac{1}{(y\Delta)^2}\right]$$
$$= \left[\frac{1}{x\Delta} - \frac{1}{y\Delta},\ \frac{1}{x\Delta} - \frac{1}{y\Delta} + \frac{1}{(x\Delta)^2} - \frac{1}{(y\Delta)^2}\right],$$

and with $\Delta = e^{\alpha/2}$, we get for the size of the interval $|[x\Delta, y\Delta]| = \sum_{j=x\Delta}^{y\Delta} \frac{e^{\alpha}}{j^2}$

$$|[x\Delta, y\Delta]| \in \left[e^{\alpha/2}\cdot\left(\frac{1}{x} - \frac{1}{y}\right),\ e^{\alpha/2}\cdot\left(\frac{1}{x} - \frac{1}{y}\right) + \frac{1}{x^2} - \frac{1}{y^2}\right].$$

We combine this with the multiplicative bound and obtain the following estimate of the size of intervals in the case $\beta_f = 2 + \frac{1}{f(n)}$.

$$|[x\Delta_f, y\Delta_f]| = \sum_{j=x\Delta_f}^{y\Delta_f} \left\lfloor \frac{e^{\alpha}}{j^{\beta_f}} \right\rfloor$$
$$\in \left[e^{\alpha\frac{1+\frac{1}{f(n)}}{2+\frac{1}{f(n)}}}\left(\frac{1}{x} - \frac{1}{y}\right) - (y-x)\Delta_f,\ e^{\alpha\frac{1+\frac{1}{f(n)}}{2+\frac{1}{f(n)}}}\left(\frac{1}{x} - \frac{1}{y}\right) + e^{\alpha\left(1-\frac{1}{1+\frac{1}{2f(n)}}\right)}\left(\frac{1}{x^2} - \frac{1}{y^2}\right)\right]$$
$$= \left[e^{\alpha\frac{f(n)+1}{2f(n)+1}}\left(\frac{1}{x} - \frac{1}{y}\right) - (y-x)\Delta_f,\ e^{\alpha\frac{f(n)+1}{2f(n)+1}}\left(\frac{1}{x} - \frac{1}{y}\right) + e^{\alpha\frac{1}{2f(n)+1}}\left(\frac{1}{x^2} - \frac{1}{y^2}\right)\right].$$

Especially, we obtain the following estimate on the size of G_{α,β_f}:

$$|[1,\Delta_f]| \in \left[e^\alpha - e^{\alpha\frac{f(n)+1}{2f(n)+1}} - e^{\alpha\frac{f(n)}{2f(n)+1}} + 1,\ e^\alpha - e^{\alpha\frac{f(n)+1}{2f(n)+1}} + e^{\alpha\frac{1}{2f(n)+1}}\, e^{\alpha\frac{2f(n)}{2f(n)+1}} - e^{\alpha\frac{1}{2f(n)+1}}\right]$$
$$= [(1-o(1))\, e^\alpha,\ (2-o(1))\, e^\alpha] \quad .$$

This estimate can be refined as follows:

$$\begin{aligned}\sum_{j=1}^{\Delta_f}\left\lfloor\frac{e^\alpha}{j^{\beta_f}}\right\rfloor &\in \left[\sum_{j=1}^{\Delta_f}\frac{e^\alpha}{j^{\beta_f}} - \Delta_f,\ \sum_{j=1}^{\Delta_f}\frac{e^\alpha}{j^{\beta_f}}\right]\\ &\subseteq \left[\frac{1}{n^{1/f(n)}}\cdot\sum_{j=1}^{\Delta_f}\frac{e^\alpha}{j^2} - \Delta_f,\ \sum_{j=1}^{\Delta_f}\frac{e^\alpha}{j^2}\right]\\ &\subseteq [(1-o(1))\cdot\zeta(2)\, e^\alpha,\ \zeta(2)\, e^\alpha] \quad ,\end{aligned}$$

where the last inclusion holds for $f(n) = \omega(\log(\alpha))$. The volume of the interval $[x\Delta_f, y\Delta_f]$ can be estimated as follows:

$$\begin{aligned}\mathrm{vol}([x\Delta_f, y\Delta_f]) &= \sum_{x\Delta_f}^{y\Delta_f}\left\lfloor\frac{e^\alpha}{j^{\beta_f}}\right\rfloor\cdot j\\ &\in \left[\sum_{x\Delta_f}^{y\Delta_f}\frac{e^\alpha}{j^{\beta_f-1}} - (x\Delta_f + (x\Delta_f+1) + \ldots + y\Delta_f),\ \sum_{x\Delta_f}^{y\Delta_f}\frac{e^\alpha}{j^{\beta_f-1}}\right]\\ &= \left[\sum_{x\Delta_f}^{y\Delta_f}\frac{e^\alpha}{j^{\beta_f-1}} - \frac{(y^2-x^2)\Delta_f^2 + (x+y)\Delta_f}{2},\ \sum_{x\Delta_f}^{y\Delta_f}\frac{e^\alpha}{j^{\beta_f-1}}\right] \quad .\end{aligned}$$

Since $j^{\beta_f-1} = j^{1+\frac{1}{f(n)}}, j = x\Delta_f, y\Delta_f$, we use Lemma 6.17 in Section 6.9 from the previous chapter and obtain that the volume $\mathrm{vol}([x\Delta_f, y\Delta_f])$ is within the interval

$$\left[\frac{e^\alpha(\ln(y)-\ln(x))}{n^{\frac{1}{f(n)}}} - \frac{(y^2-x^2)\Delta_f^2 + (x+y)\Delta_f}{2},\ e^\alpha(\ln(y)-\ln(x)) + e^\alpha\left(\frac{1}{x\Delta_f} - \frac{1}{y\Delta_f}\right)\right] \quad .$$

□

We are now well prepared to compute the parameters α, d, x, y of our embedding $G_{U,\mathcal{S}} \mapsto G_{\alpha,\beta_f}$ for the functional case $\beta_f = 2 + \frac{1}{f(n)}$, $f(n) = \omega(\log(n))$, and to finish up the proof of Theorem 7.6. Recall that we want to choose the values of the parameters α, d, x, y such as to meet the following requirements:

1. $\left|\left[N^{a/d}, N^{b/d}\right]\right| \geqslant N$.
2. $|[x\Delta_f, y\Delta_f]| = o\left(N^{\frac{d-1}{d}}\right)$, where $N^{\frac{d-1}{d}}$ is a lower bound for the size of an optimal dominating set in $G_{U,\mathcal{S}}$.
3. $\sum_{j=x\Delta_f}^{y\Delta_f} \left\lfloor \frac{e^\alpha}{j^{\beta_f}} \right\rfloor \cdot j = \mathrm{vol}\left([x\Delta_f, y\Delta_f]\right) \geqslant \zeta(2) \cdot e^\alpha$.

Considering constraint 1, we want to get an estimate for the value $\left|\left[N^{a/d}, N^{b/d}\right]\right|$. Note that $e^{\alpha \cdot \frac{1}{2f(n)-1}} \cdot \Delta_f^2 = e^{\alpha \cdot \frac{f(n)+1}{2f(n)+1}} \cdot \Delta_f = e^\alpha$. Thus our estimate of interval sizes yields

$$\left|\left[N^{\frac{a}{d}}, N^{\frac{b}{d}}\right]\right| \in \left[e^\alpha\left(\frac{1}{N^{\frac{a}{d}}} - \frac{1}{N^{\frac{b}{d}}}\right) - \left(N^{\frac{b}{d}} - N^{\frac{a}{d}}\right), e^\alpha\left(\frac{1}{N^{\frac{a}{d}}} - \frac{1}{N^{\frac{b}{d}}}\right) + e^\alpha\left(\frac{1}{N^{\frac{2a}{d}}} - \frac{1}{N^{\frac{2b}{d}}}\right)\right]$$

In order to satisfy constraint 1, for a given d, we want to choose α such that

$$\left|\left[N^{a/d}\right]\right| \geqslant e^\alpha\left(\frac{1}{N^{\frac{a}{d}}} - \frac{1}{N^{\frac{b}{d}}}\right) - \left(N^{\frac{b}{d}} - N^{\frac{a}{d}}\right)$$

$$\iff \qquad e^\alpha\left(N^{\frac{b-a}{d}} - 1\right) - \left(1 - N^{\frac{a-b}{d}}\right) \geqslant N^{1+\frac{b}{d}} \ .$$

Hence we choose

$$e^\alpha \approx N^{1+\frac{b}{d}} \quad \iff \quad \alpha \approx \left(1 + \frac{b}{d}\right) \cdot \ln(N) \ .$$

If we now choose $d > \frac{(b+1)\beta_f}{\beta_f - 1}$, then the constraint 2 holds, and for $y = 1$ and $x > 0$ such that $x\Delta_f > N^{b/d}$, constraint 3 holds as well. Thus, we obtain asymptotically the same approximation hardness result as for the case $\beta = 2$ and the theorem follows. □

Let us now consider the case when $f(n) = o(\log(n))$.

7.10.2 *The Case* $f(n) = o(\log(n))$.

In the case $f(n) = o(\log(n))$, the hardness of MIN-DS shows a surprising phase transition and we yield the following theorem.

Theorem 7.7.
For $\beta_f = 2 + f(n)^{-1}$ *with* $f(n) = o(\log(n))$, *the* MIN-DS *problem on* (α, β_f)*-PLG is in APX.*

Proof. We consider the case when $f(n)$ is a "*slowly growing*" function, namely $f(n) = o(\log(n))$. In that case, $n^{1/f(n)} \longrightarrow \infty$ as $n \longrightarrow \infty$. For $x\Delta_f \leqslant j \leqslant y\Delta_f$, we obtain

$$\frac{1}{j^{1+\frac{1}{f(n)}}} = \frac{1}{j} \cdot \frac{1}{j^{\frac{1}{f(n)}}} \leqslant \frac{1}{j} \cdot \frac{1}{(x\Delta_f)^{\frac{1}{f(n)}}} = \frac{1}{j} \cdot \frac{1}{x^{\frac{1}{f(n)}}} \cdot \frac{1}{e^{\alpha \cdot \frac{1}{2f(n)+1}}} \ ,$$

and therefore

$$\mathrm{vol}([x\Delta_f, \Delta_f]) \leqslant e^{\alpha} \cdot \ln\left(\frac{1}{x}\right) \cdot \frac{1}{x^{\frac{1}{f(n)}}} \cdot \frac{1}{e^{\alpha \cdot \frac{1}{2f(n)+1}}} \ ,$$

which yields the requirement $\frac{\ln(1/x)}{x^{1/f(n)}} \geqslant c \cdot e^{\alpha \cdot \frac{1}{2f(n)+1}}$. This is equivalent to

$$\ln\ln\left(\frac{1}{x}\right) + \frac{1}{f(n)} \cdot \ln\left(\frac{1}{x}\right) \geqslant \ln(c) + \frac{\alpha}{2f(n)+1} \ ,$$

which means the following: In order to dominate the remaining vertices of the graph with vertices from $[x\Delta_f, \Delta_f]$, we have to choose $\ln(1/x) \geqslant \alpha/2$, i. e. $1/x \geqslant e^{\alpha/2}$. This gives the following lower bound for the size of that interval:

$$\begin{aligned} |[x\Delta_f, \Delta_f]| &\geqslant e^{\alpha \frac{f(n)+1}{2f(n)+1}} \left(e^{\frac{\alpha}{2}} - 1\right) - \left(1 - \frac{1}{e^{\frac{\alpha}{2}}}\right) e^{\frac{\alpha}{2+\frac{1}{f(n)}}} \\ &\geqslant (1 - o(1))\, e^{\frac{\alpha}{2}\left(1+\frac{f(n)+1}{f(n)+1/2}\right)} \ . \end{aligned}$$

This lower bound for the size of $[x\Delta_f, \Delta_f]$ converges to e^{α} as $n \to \infty$, which means there exists some $c > 0$ such that $|[x\Delta_f, \Delta_f]| \geqslant c \cdot |G_{\alpha,\beta_f}|$ in order to be a dominating set. Hence, each dominating set in G_{α,β_f} is of cardinality at least $c \cdot |G_{\alpha,\beta_f}|$ and thus we obtain the result. □

7.11 SUMMARY AND FURTHER RESEARCH

In this chapter we studied the approximation complexity of MIN-DS in (α, β)-PLG. The main result of this chapter is a first logarithmic lower bound for the approximability of MIN-DS for the parameter range $0 < \beta \leqslant 2$. The result is based on a reduction from the SET COVER problem combined with the logarithmic lower bound for SET COVER given by Feige [Fei98]. For $\beta > 2$, we presented an improvement of the approximation upper bounds given by Shen et al. [She+12] for the greedy algorithm for MIN-DS. We obtain our new results by relating the cost and structure of an optimum solution to those of a greedy-based solution. This sophisticated analysis yields improved upper bounds for almost the whole range $\beta > 2$. Finally, we investigated the phase transition at $\beta = 2$. For this purpose, we extended the power law model and considered the case when $\beta_f = 2 + \frac{1}{f(n)}$ is a function of the graph size n which converges to 2 from above. Surprisingly, for every function $f(n)$ with $f(n) = \omega(\log(n))$, i. e. when β_f converges fast enough, MIN-DS in (α, β_f)-PLG still provides a logarithmic approximation lower bound. For every function $f(n)$ with $f(n) = o(\log(n))$, the problem is in APX.

The further improvements on both lower and upper approximation bounds are important open questions in the area, especially the upper approximation bounds for $\beta \leqslant 2$. Another interesting problem concerns the approximability of PLG optimization problems on random or quasi-random instances.

7.12 BIBLIOGRAPHIC NOTES

The material and the results presented in this chapter are based on the following publication: Mikael Gast, Mathias Hauptmann, and Marek Karpinski. "Inapproximability of dominating set in power law graphs." In: *Computing Research Repository (CoRR) preprint arXiv:1212.3517 [cs.CC]; also submitted to Theoretical Computer Science* (Dec. 2012), pp. 1–23. arXiv: 1212.3517.

The approximation lower bound for SET COVER presented in Section 7.4.3 is due to Feige [Fei98].

Regarding our results, the proofs of the main theorems (Theorem 7.1 on page 151; Theorem 7.2 on page 152; Theorem 7.3 on page 154; Theorem 7.4 on page 156;Theorem 7.5 on page 162; Theorem 7.6 on page 164; Theorem 7.7 on page 170) also appeared in [GHK12b].

8 APPROXIMATION UPPER BOUNDS OF MINIMUM VERTEX COVER IN POWER LAW GRAPHS

8.1 INTRODUCTION

Apart from having certain structural properties, such as a power law degree distribution, high clustering coefficient, small-world characteristics and self similarity, there exists practical evidence that combinatorial optimization in PLG is easier than in general graphs [PL01; GMS03; Eub+04b; KSG06]. For example, Park and Lee [PL01] showed for the efficient placement of filters in *route-based distributed packet filtering* on power-law Internet topologies, that the greedy heuristic for the MIN-VC problem generally outperforms the constant-factor approximation algorithm.

A natural question now arises whether these observations and results can be translated into provable guarantees for a more general class of power law graphs. In order to get a positive answer to this question, one has to identify structural properties of power law graphs that allow for the design of efficient algorithms or for approximation algorithms with better approximation ratios.

In this chapter we study the approximability of the MIN-VC problem in the random power law graph model of Aiello, Chung, and Lu [ACL01]. Let us first repeat the formal definition of the MINIMUM VERTEX COVER problem.

Problem 8 (MINIMUM VERTEX COVER (MIN-VC))**.**

INPUT: A graph $G = (V, E)$.

OUTPUT: A *vertex cover* of G, i. e. a set $C \subseteq V$ such that each edge $\{u, v\} \in E$ has at least one endpoint in C.

OBJECTIVE: Minimize $|C|$.

The problem is known to be NP-complete due to Karp's original proof [Kar72] and APX-complete due to Papadimitriou and Yannakakis [PY91]. Moreover, Dinur and Safra [DS05] showed that MIN-VC cannot be approximated within a factor of 1.3606, unless $P = NP$, and Khot and Regev [KR08] conjectured the inapproximability within $2 - \varepsilon$ for any $\varepsilon > 0$ as long as the Unique Games Conjecture (UGC) holds true.

8.2 OVERVIEW AND RESULTS

In this chapter, we show that the MIN-VC problem can be approximated with an expected approximation ratio < 2 in random (α, β)-power law graphs. The formal definition of random power law graphs $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ in the $G(\alpha, \beta)$ model is given in Section 2.5.3. We prove the following theorem.

Theorem 8.1.
For $\beta > 2$, there exists a polynomial time algorithm which approximates the MINIMUM VERTEX COVER *problem in random power law graphs $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ in the $G(\alpha, \beta)$ model with an expected approximation ratio of*

$$\rho = 2 - \frac{\zeta(\beta) - 1 - \frac{1}{2^\beta}}{2^\beta \zeta(\beta - 1)\zeta(\beta)} \ ,$$

where $\zeta(\beta) = \sum_{i=1}^{\infty} \frac{1}{i^\beta}$ is the Riemann zeta function.

We also give a refined analysis for the case of the parameter $\beta > 2$ and obtain the following improvement in the limit of large graph sizes.

Theorem 8.2.
For $\beta > 2$, the MINIMUM VERTEX COVER *problem in random power law graphs $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ in the $G(\alpha, \beta)$ model can be approximated with expected asymptotic approximation ratio of*

$$\rho' = 2 - \frac{\left(\zeta(\beta) - 1 - \frac{1}{2^\beta}\right) \zeta(\beta - 1)}{\zeta(\beta - 1)\zeta(\beta)} \left[1 - \left(\frac{\zeta(\beta - 1) - \left(1 + \frac{1}{2^{\beta-1}}\right)}{\zeta(\beta - 1)} \right)^3 \right] \ .$$

The improvement for the asymptotic approximation ratio ρ' is due to a refined estimation of a lower bound on the size of a half-integral solution for the LP-relaxation formulation of MIN-VC. In Figure 8.1 the two upper bounds ρ and ρ' of Theorem 8.1 and Theorem 8.2 are shown as functions of the parameter $\beta > 2$.

ORGANIZATION OF THE CHAPTER. The chapter is organized as follows. In Section 8.3 we give some background on the MIN-VC problem and briefly describe the LP-relaxation and the *half-integrality property* of MIN-VC used in the theorem by

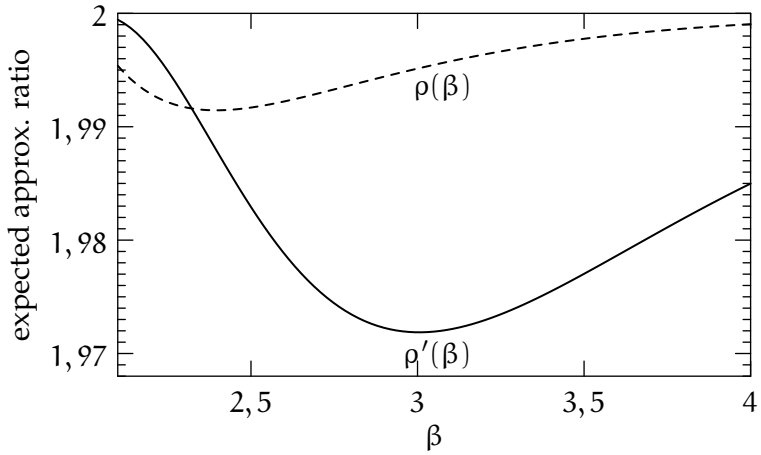


Figure 8.1: Comparison of first (- - -) and second (—) analysis in terms of the parameter β, for $\beta > 2$. Particularly, the approximation ratio ρ' yields an improvement over the ratio ρ for $\beta > 2.322$.

Nemhauser and Trotter [NT74] and a corollary that provides an approximation ratio of 2 for the problem. Section 8.5 presents our new approximation algorithm for MIN-VC in random (α, β)-PLGs. The algorithm basically consists of a deterministic rounding procedure on a half-integral solution for MIN-VC. We show that this rounding procedure yields an approximation ratio of $3/2$ on the subgraph induced by the low degree vertices of the power law graph and a 2-approximation on the residual subgraph. In Section 8.6, we show how to achieve a better approximation ratio for MIN-VC, given the deterministic rounding algorithm together with certain bounds on the size of a half-integral solution and the rounded integer solution. In Section 8.7 we construct upper and lower bounds on the expected size of the half-integral solution in the induced subgraph of low degree vertices and prove our first main theorem (Theorem 8.1). In Section 8.8, we present a refined analysis which consists of better estimates for the above upper and lower bounds in the limit of large graph sizes. We conclude the chapter by giving a short summary and further research in Section 8.9 and some bibliographic note in Section 8.10.

8.3 PRELIMINARIES

In this section we introduce the notions and notations that will be used in this chapter. Let us first introduce the LP-relaxation formulation proposed by Nemhauser and Trotter and the notion of the half-integrality property of MIN-VC. Later, we state the Nemhauser and Trotter Theorem (NTT) and show how to obtain good approximate solutions for MIN-VC through the half-integrality property.

8.3.1 *LP-Relaxation and Half-Integral Solutions*

One of the most well known results regarding the MIN-VC problem is the half-integrality property of the LP-relaxation of the classical integer programming formulation of MIN-VC. This was first described by Nemhauser and Trotter in their 1974 paper [NT74]. One year later, and only three years after the publication of Karp's famous list of NP-complete problems in 1972 [Kar72], Nemhauser and Trotter presented a reduction of the problem of finding a vertex cover in an arbitrary graph G to the problem of finding a vertex cover in a subgraph of G whose total weight does not exceed too much the cost of any of its vertex covers. The reduction—published in their seminal paper [NT75]—makes use of the half-integrality property of MIN-VC and additionally adds new structural properties to the MIN-VC problem in general. For example: the total weight of the graph G after applying the reduction of Nemhauser and Trotter can be used to better analyze the performance of approximate solutions, i. e. we may assume that the optimal solution OPT has cost $\mathsf{opt} \geqslant \frac{w(V)}{2}$ and are then able to compare this to an approximate solution.

Nemhauser and Trotter [NT75] considered the following LP-relaxation of the integer programming formulation of MIN-VC, which also applies to the *weighted* MIN-VC problem:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{n} w_i x_i, \\
\text{subject to} \quad & x_i + x_j \geqslant 1, \quad \text{for each edge } \{v_i, v_j\} \in E, \\
& x_i \geqslant 0, \quad \text{for each vertex } v_i \in V.
\end{aligned}
$$

They proved the following theorem.

Theorem 8.3 (Nemhauser and Trotter [NT75])**.**
There exists a polynomial-time algorithm which gets as an input a graph $G = (V, E)$ *together with a weight function* $w\colon V \to \mathbb{Q}_+$ *and partitions the set of vertices* V *into three subsets* $V_1, V_{1/2}, V_0$*, such that*

(i) if C *is an* α*-approximate solution for* Min-VC *on the induced subgraph* $G[V_{1/2}]$*, then* $V_1 \cup C$ *is an* α*-approximate solution for* Min-VC *on* G*, for all* $\alpha > 1$*, and*

(ii) the weight of any vertex cover C' *in* $G[V_{1/2}]$ *is at least* $1/2 \cdot \sum_{v \in V_{1/2}} w(v) = 1/2 \cdot w(V_{1/2})$.

The first observation is that there always exists an optimal solution x for the above LP-relaxation which is *half-integral*, i. e. for all i, $x_i \in \{0, 1/2, 1\}$. Then the algorithm partitions the set of vertices into subsets $V_1, V_{1/2}, V_0 \subseteq V$, such that $v_i \in V_1$ if $x_i = 1$, $v_i \in V_{1/2}$ if $x_i = 1/2$ and $v_i \in V_0$ if $x_i = 0$ in the half-integral solution. We note that an optimal half-integral solution can be computed efficiently since this problem corresponds to solving the Min-VC problem on a bipartite graph $G' = (V \cup V', E')$ which is generated from the original graph[1]. Here, V' is a copy of the original vertex set V and $E' = \{\{v_i, v_j'\} \mid \{v_i, v_j\} \in E\}$. Furthermore, Nemhauser and Trotter showed that at least one optimal vertex cover in G contains the set V_1, that each vertex in V_0 has all its neighbors in V_1 and—moreover—that each cover in G has weight at least $w(V_1) + 1/2 \cdot w(V_{1/2})$. From this it follows that at least one optimal vertex cover in G consists of the set V_1 and an optimal cover in the subgraph $G[V_{1/2}]$ induced by $V_{1/2}$. Figure 8.2 shows the partition and the interconnections of the vertex set generated in the algorithmic step of Theorem 8.3.

Hochbaum et al. [Hoc+93] observed—as a direct corollary of Theorem 8.3—that an integer solution y, obtained by setting $y_i = 1$ for all vertices $v_i \in V_{1/2} \cup V_1$ and $y_i = 0$ for all $v_i \in V_0$, is a 2-approximate solution for the Min-VC problem in G. Our approximation algorithm for Min-VC will make use of a more sophisticated rounding procedure on a half-integral solution x in order to achieve an approximation ratio of $3/2$ on a large subset of low degre vertices and a 2-approximation on the residual graph.

1 This follows directly from the theorem of Kőnig [Kőn16; Kőn31], which states that the problems of computing maximal matchings and minimum vertex covers on bipartite graphs are equivalent and can be solved in polynomial time (see also Biggs, Lloyd, and Wilson [BLW86]).

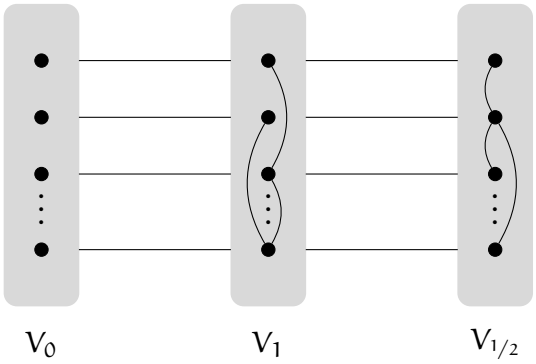


Figure 8.2: Partition of the vertex set V into a independent set V_0 connected to vertices of the set V_1 as generated in the algorithmic step of Theorem 8.3.

8.4 OUTLINE OF THE METHOD

Let us now give an outline of the methods used to obtain the main results of this chapter. That is, the design of an approximation algorithm with expected approximation ratio $2 - \frac{\zeta(\beta)-1-\frac{1}{2^\beta}}{2^\beta\zeta(\beta-1)\zeta(\beta)}$ for the Min-VC problem in (α,β)-PLG for $\beta > 2$, and, moreover, a refined analysis yielding an improved *asymptotic* approximation ratio in the limit.

The algorithm can be outlined as follows. On instance $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$, the algorithm starts with a half-integral solution $x\colon V \to \{0, 1/2, 1\}$ of the associated LP-relaxation of Min-VC and uses some deterministic rounding procedure to generate an integral solution $y\colon V \to \{0,1\}$. We show that for the set $V^* = \bigcup_{v\colon d(v)\in\{1,2\}} (\{v\} \cup N(v))$ of degree 1 and degree 2 vertices and their neighbors in $G_{\alpha,\beta}$, the rounding procedure satisfies $y(V^*) \leqslant 3/2 \cdot x(V^*)$, i.e. we prove that the rounding procedure achieves an approximation ratio of $3/2$ on the induced graph $G_{\alpha,\beta}[V^*]$ of the subset $V^* \subseteq V$. Furthermore, we show that the total value of the initial half-integral solution $x(V^*)$ on this subset is "large" in expectation and with respect to $G(\alpha,\beta)$, such that the contribution of the rounding procedure affects the global approximation ratio of Min-VC on $G_{\alpha,\beta}$ in a positive way. This last step involves a probabilistic analysis of the random graph model $G(\alpha,\beta)$ and aims to provide tight lower bounds on the expected value $\mathbb{E}[x(V^*)]$ and to construct an upper bound for the value $x(V)$.

Figure 8.3 shows an overview of the proof of our first main theorem (Theorem 8.1 on page 175). Especially, we also point to the lemmas and auxiliary theorems which are used in the various steps of the main proof.

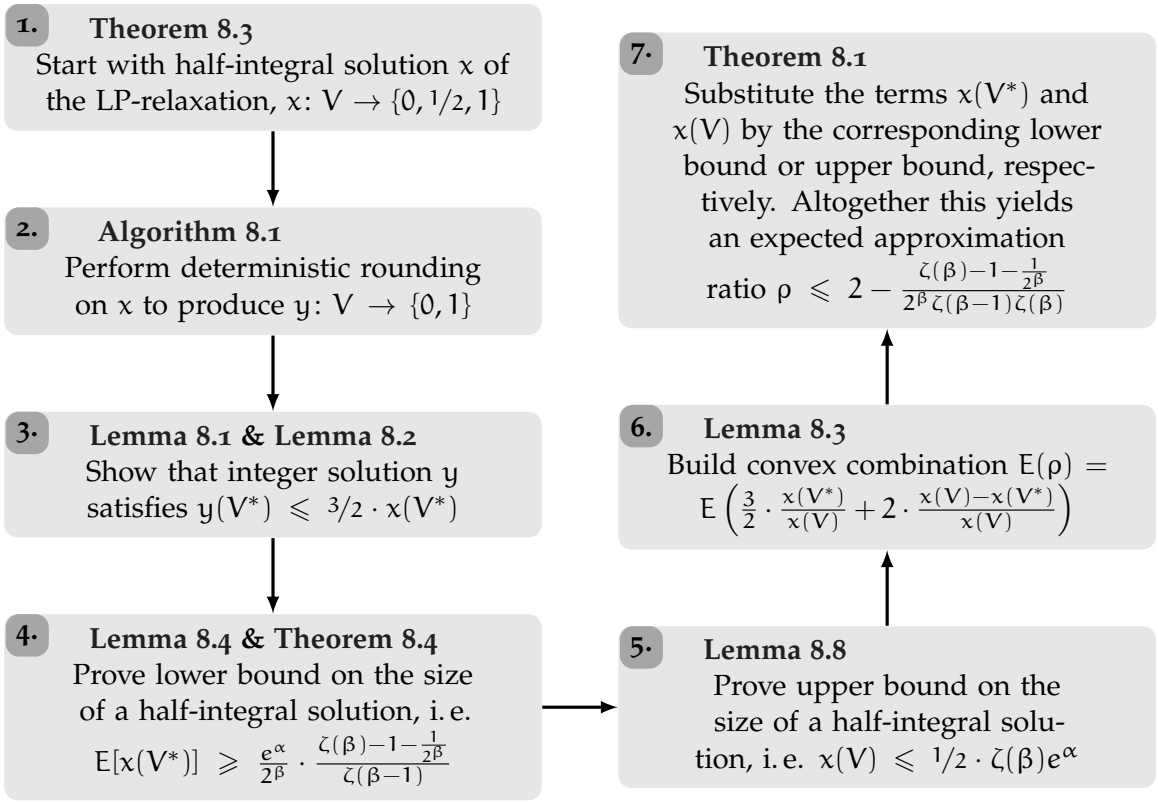


Figure 8.3: Overview of the proof for the first main theorem (Theorem 8.1) for the expected approximation ratio of MIN-VC in random (α, β)-PLG, where $\beta > 2$. The proof of the second main result in Theorem 8.2 goes along the same lines, except for a different estimate of the lower bound for $x(V^*)$ in step 4.

Next, we present a detailed description of the approximation algorithm outlined above.

8.5 APPROXIMATION ALGORITHM

In this section, we describe our deterministic rounding procedure on $G_{\alpha,\beta} = (V, E)$ for $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ (cf. algorithm `Deterministic_Rounding` on page 182). First, the algorithm processes all degree 1 and degree 2 nodes of the vertex subset $V' = L \cup N(L)$,

where $L = \{ v \in V \mid (d(v) = 2, x(v) = 1/2) \vee (d(v) = 1) \}$. On this subset, the algorithm provides a rounded integral solution y with $y(V') \leqslant 3/2 \cdot x(V')$. Furthermore, we show that $y(V^* \setminus V') \leqslant 4/3 \cdot x(V^* \setminus V')$ and $y(V \setminus V^*) \leqslant 2 \cdot x(V \setminus V^*)$ is achieved for the remaining subsets of vertices. Whenever the algorithm processes a vertex v, $y(v)$ is either set to 0 or 1 and the vertex itself is marked as `processed`.

The analysis of the algorithm `Deterministic_Rounding` is provided by the following Lemma 8.1 and Lemma 8.2. First, we will show that the algorithm indeed generates an integer solution on the set of degree 1 vertices and their high-degree neighbors, as well as on all degree 2 vertices u with $x(u) = 1/2$ in the initial half-integral solution and their corresponding neighborhood.

Lemma 8.1.
The assignment y generated by the algorithm `Deterministic_Rounding` is an integer solution and satisfies $y(u) = 1$ for all $u \in V'$ with $d(u) \geqslant 3$.

Proof. The y value of any high-degree neighbor of degree 1 vertices is set to 1 in step (1) of the algorithm.

Since either step (3) or (4) is processing every single degree 2 vertex $v \in V$ with $x(v) = 1/2$, there are no leftover vertices $v \in V'$ of degree 2 with fractional values $x(v)$.

Assume that there is a vertex $u \in V', d(u) \geqslant 3$ and $x(u) = y(u) = 1/2$. Then u has at least one degree 2 neighbor v_1 with $x(v_1) = 1/2$. Because of step (3) and (4) of the algorithm, v_1 must have been processed by another degree 2 vertex v_2, setting $y(v_1) = 1$. This again introduces another neighbor w of v_2 with $y(w) = 1$ and leads to the situation of a path uv_1v_2w described in step (2). In this case, the algorithm sets $y(u) = 1$ and thus we have a contradiction to the above assumption. □

The next lemma shows that the deterministic rounding algorithm achieves an approximation ratio of 3/2 on the induced subgraph $G_{\alpha,\beta}[V^*]$ where V^* is the set of vertices of degree 1 and 2 including their corresponding neighborhood, i.e. $V^* = \bigcup_{v:\, d(v) \in \{1,2\}} (\{v\} \cup N(v))$.

Lemma 8.2.
The assignment y generated by the algorithm `Deterministic_Rounding` satisfies $y(V^*) \leqslant 3/2 \cdot x(V^*)$.

Algorithm 8.1 : `Deterministic_Rounding`

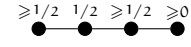
Input : $G_{\alpha,\beta} = (V, E), x\colon V \to \{0, 1/2, 1\}$.
Output : $y\colon V \to \{0, 1\}$.

forall the $v \in V$ **do**
  $y(v) := x(v)$; mark v as `unprocessed`;
  /* when $y(v)$ is changed in steps (0)-(5), v will be marked as processed */

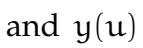
(0) compute $G'[L \cup N(L)]$, $L = \{ v \in V \mid (d(v) = 2, x(v) = 1/2) \vee (d(v) = 1) \}$;
(1) **forall the** $v \in V$ *with* $d(v) = 1$ **do**
  /* let u be the neighbor of v in $G_{\alpha,\beta}$ */
  set $y(v) = 0$; set $y(u) = 1$;
(2) **forall the** $P = uv_1v_2w \subset G'$ `unprocessed`, $d(u) \geqslant 3, d(v_1) = d(v_2) = 2$ **do**
  /* let u, w be the neighbors of v_1, v_2 in $G_{\alpha,\beta}$, respectively */
  set $y(u) = y(w) = y(v_1) = 1$; set $y(v_2) = 0$;
(3) **forall the** $v \in V'$ `unprocessed`, $d(v) = 2 \wedge \exists u \in N(v), d(u) \geqslant 3$ **do**
  /* let u, w be the neighbors of v in $G_{\alpha,\beta}$ */
(3.1) **else if** u `unprocessed`, w `processed` **then**
    set $y(v) = 0$; set $y(u) = 1$;
(3.2) **else if** *both* u, w `unprocessed` **then**
    set $y(v) = 0$; set $y(u) = y(w) = 1$; /* $x(u) \geqslant 1/2$ and $x(w) \geqslant 1/2$ */
(3.3) **if** *both* u, w `processed` **then**
    set $y(v) = 0$;
(3.4) **else if** u `processed`, w `unprocessed` **then**
    set $y(v) = 0$; set $y(w) = 1$; /* $y(u) = 1$ already set and $x(w) \geqslant 1/2$ */
(4) **forall the** $v \in V'$ `unprocessed`, $d(v) = 2$ **do**
(4.1) **else if** u `unprocessed`, w `processed` **then**
    set $y(v) = 0$; set $y(u) = 1$;
(4.2) **else if** *both* u, w `unprocessed` **then**
    set $y(v) = 0$; set $y(u) = y(w) = 1$; /* $x(u) \geqslant 1/2$ and $x(w) \geqslant 1/2$ */
(4.3) **if** *both* u, w `processed` **then**
    set $y(v) = 0$;
(4.4) **else if** u `processed`, w `unprocessed` **then**
    set $y(v) = 0$; set $y(w) = 1$; /* $y(u) = 1$ already set and $x(w) \geqslant 1/2$ */
(5) **forall the** $v \in V$ `unprocessed` **do**
  if $x(v) = 1/2$ **then**
    set $y(v) = 1$; /* $y(v) = \min\{1, 2 \cdot x(v)\}$ */

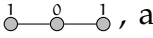
Proof. The algorithm partitions the graph induced by V^* into edge-disjoint subgraphs, namely stars, whose leaves are degree-1 vertices, and simple paths of length $\leqslant 4$. We show that for each such subgraph P_i, $y(P_i) \leqslant 3/2 \cdot x(P_i)$. Furthermore, we show that $y(v) = 1$ for each $v \in V^*$ which is contained in more than one such subgraph.

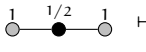
In step (1) of the algorithm all degree 1 vertices and their neighbors are processed.

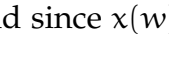
In step (2) the subgraphs are `unprocessed` paths P_i of length 3. Since the path $P_i =$  contains two disjoint edges $\{u, v_1\}, \{v_2, w\}$, $x(P_i) \geqslant 2$ and particularly $x(v_2) + x(w) \geqslant 1$. Therefore, $y(P_i) = 3 \leqslant 3/2 \cdot x(P_i)$ holds via mapping the paths of the form  (where the gray color indicates a `processed` vertex) and y restricted to P_i (denoted as $y \restriction P_i$) is a vertex cover for P_i.

In step (3) all paths $P_i =$  are processed, where at least one of u, w is of degree $\geqslant 3$. In cases (3.1)–(3.4) the algorithm considers all possible combinations of some of these nodes being already processed.

In case (3.1) u is marked `unprocessed`, w is already `processed` and $x(u) \geqslant 1/2$. The rounding algorithm sets $y(v) = 0$ and $y(u) = 1$, mapping  , again yielding a vertex cover $y \restriction P_i$ for P_i with $y(P_i) \leqslant x(P_i)$.

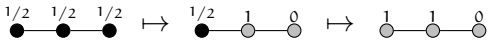
In case (3.2) we have that both u, w are marked as `unprocessed` and since $x(v) = 1/2$ we have that $x(u) \geqslant 1/2$ and $x(w) \geqslant 1/2$. The rounding algorithm sets $y(v) = 0, y(u) = y(w) = 1$, mapping  , and since $x(u) \geqslant 1/2$ and $x(w) \geqslant 1/2$ we have that $y(P_i) \leqslant 4/3 \cdot x(P_i)$.

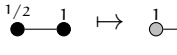
In case (3.3) both u, w are marked as `processed` and therefore $y(u) = y(w) = 1$, since u, w are adjacent to processed degree one or degree two vertices other than v. The algorithm sets $y(v) = 0$, mapping  . Hence $y \restriction P_i$ is a vertex cover for P_i with $y(P_i) \leqslant x(P_i)$.

In case (3.4) u is already `processed` and w is still marked `unprocessed`. Since $x(v) = 1/2$ we have that $x(w) \geqslant 1/2$. The rounding algorithm sets $y(v) = 0$ and $y(w) = 1$, mapping  , and since $x(w) \geqslant 1/2$ it yields a vertex cover $y \restriction P_i$ for P_i with $y(P_i) \leqslant x(P_i)$.

Step (4) considers all remaining unprocessed vertices of degree 2. If v is such a vertex with neighborhood $N(v) = \{u, w\}$, the sub-cases (4.1)–(4.4) are treated analogously to cases (3.1)–(3.4) and the mapping $x \mapsto y$ achieves $y(P_i) \leqslant 4/3 \cdot x(P_i)$ on the considered paths P_i.

After steps (0)–(4) of the algorithm there may still be some remaining high-degree vertices $u \in V^*, d(u) \geqslant 3$ with $x(u) = y(u) = 1/2$. These are treated separately (and rounded to $y(u) = 1$ together with all other vertices in $V \setminus (V' \setminus V^*)$) in step (5) of the algorithm. We have to argue that $y(V^*) \leqslant 3/2 \cdot x(V^*)$ still holds true.

We consider first the case that $u \in V', d(u) \geqslant 3$ and $x(u) = y(u) = 1/2$. Then u has a neighbor v of degree $\leqslant 2$ with $x(v) = 1/2$ and $y(v) = 1$, and since $y(u) = 1/2$ we have $d(v) = 2$. Let v_2 be the other neighbor of v, then $d(v_2) = 1$ (since otherwise the second neighbor w of v_2 would give rise to a path of length 3, containing also u and hence would have been processed in step (2)). But then locally on the set $\{u, v, v_2\}$ we have the mapping  with a local ratio of $4/3$.

Let us now assume $u \in V^* \setminus V', d(u) \geqslant 3$ and $x(u) = y(u) = 1/2$. Then every degree 2 neighbor v has $x(v) \neq 1/2$, hence $x(v) = 1$, and therefore $y(v) = 1$. We show that $v \notin V'$, i. e. that v was not processed by the algorithm and can be treated as a part of a subgraph disjoint to G' in $G_{\alpha,\beta}$. Let $w \in N(v)$ be the second neighbor of v besides u. Then $x(w) = 0$ since otherwise (in case $x(w) \geqslant 1/2$) we could decrease $x(v)$ from 1 to $1/2$ and would still have a feasible half-integral solution, which would contradict the optimality of x. Therefore $v, w \notin V'$, which means that v, w are not processed by the algorithm. Rounding $y(u) = 1$, mapping  , yields a vertex cover $y \restriction \{u, v\}$ with $y(\{u, v\}) \leqslant 4/3 \cdot x(\{u, v\})$.

We conclude that the assignment $y \colon V \mapsto \{0, 1\}$ is a vertex cover of $G_{\alpha,\beta}$ with $y(V^*) \leqslant 3/2 \cdot x(V^*)$ and $y(V \setminus V^*) \leqslant 2 \cdot x(V \setminus V^*)$. □

Now we are going to describe how to obtain the overall approximation ratio of the algorithm described and analysed above.

8.6 EXPECTED APPROXIMATION RATIO

The next task is to combine the above findings on the approximation ratio in the induced subgraphs $G_{\alpha,\beta}[V^*]$ and $G_{\alpha,\beta}[V \setminus V^*]$ to achieve an (expected) approximation ratio for MIN-VC on the whole initial graph $G_{\alpha,\beta}$.

The following lemma shows how to retrieve an expected approximation ratio of the algorithm `Deterministic_Rounding` for MIN-VC in random (α, β)-power law graphs $G_{\alpha,\beta}$.

Lemma 8.3.
If the rounding scheme $x \mapsto y$ satisfies $y(V^*) \leqslant 3/2 \cdot x(V^*)$ and $y(V \setminus V^*) \leqslant 2 \cdot x(V \setminus V^*)$, then this gives an approximation ratio of

$$\frac{y(V)}{OPT} \leqslant \frac{y(V)}{x(V)} \leqslant \frac{x(V^*)}{x(V)} \cdot \frac{3}{2} + \frac{x(V \setminus V^*)}{x(V)} \cdot 2 \ .$$

In order to apply Lemma 8.3 and to derive an expected approximation ratio for the algorithm, in the following we will give a lower bound on $\mathbb{E}[x(V^*)]$ and an upper bound on $x(V)$. The next lemma provides a first lower bound on $x(V^*)$ in terms of the number of high-degree vertices adjacent to vertices of degree 1 and 2.

Lemma 8.4.
Let $G_{\alpha,\beta}[V^*]$ be the subgraph of $G_{\alpha,\beta}$ induced by V^*. For every optimal half-integral solution x for the Min-VC *LP-relaxation, the size of the half-integral solution restricted to V^* is lower-bounded by the size of the high-degree neighborhood of degree 1 and degree 2 vertices:*

$$x(V^*) \geqslant \frac{1}{2} \cdot \left| \{ u \in V \mid d(u) \geqslant 3 \wedge \exists v \in N(u), d(v) \in \{1,2\} \} \right| \ .$$

Proof. Let $V^* = X \cup Y, X = \{ v \in V \mid d(v) \in \{1,2\} \}$ and $Y = \{ u \in V \mid d(u) \geqslant 3 \wedge \exists v \in N(u), d(v) \in \{1,2\} \}$. Let $E(X,Y)$ be the set of edges between the subsets X and Y. Choose some arbitrary function $f\colon Y \to E(X,Y)$ such that for every $u \in Y, f(u) = \{u, v\}$ for some $v \in X$ adjacent to u. $f(Y)$ consists of pairwise disjoint paths $Q_1, \ldots, Q_m$ of length $\leqslant 2$, such that each path contains one or two vertices from Y. This implies $x(V^*) \geqslant m \geqslant \frac{|Y|}{2}$. □

In the next sections, we are going to compute the explicit expected approximation ratios of the algorithm `Deterministic_Rounding` for the case when $\beta > 2$.

8.7 FIRST ANALYSIS FOR $\beta > 2$

In our first analysis, we will estimate the expected number of high-degree vertices adjacent to vertices of degree 1 or 2, which—combined with the preceding Lemma 8.4—gives a lower bound on $\mathbb{E}[x(V^*)]$. We will prove the following theorem:

Theorem 8.4.

$$\begin{aligned}\mathbb{E}[x(V^*)] &\geqslant \frac{1}{2}\cdot\mathbb{E}\left[\left|\{u\in V \mid d(u)\geqslant 3 \wedge \exists v\in N(u), d(v)\in\{1,2\}\}\right|\right]\\ &= \frac{1}{2}\cdot\sum_{u:\, d(u)\geqslant 3}\eta(u) \qquad (8.1)\\ &\geqslant \frac{e^{\alpha}}{2^{\beta}}\cdot\frac{\zeta(\beta)-1-\frac{1}{2^{\beta}}}{\zeta(\beta-1)}\ , \qquad (8.2)\end{aligned}$$

where $\eta(u)$ *is the probability that* $u\in V$ *has a neighbor in the set of vertices of degree* 1 *or* 2.

In order to provide bounds on the probability $\eta(u)$ for a vertex u of degree d of having a neighbor of degree 1 or 2, we consider how edges are generated in the random matching procedure of the distribution $G(\alpha,\beta)$:

In the random model for (α,β)-PLG described in Section 2.5.4, $d(u)$ copies of u are randomly matched with the copies of the remaining vertices $v\in V, v\neq u$. We use the following lower bound on $\eta(u)$.

Lemma 8.5.
For every u *with* $d(u)\geqslant 3$, $\eta(u)\geqslant \frac{1}{2^{\beta-1}\cdot\sum_{i=1}^{\Delta}\frac{1}{i^{\beta-1}}}$.

Proof. We will give an estimate on the probability that the first copy of u, in the random matching generation of $G_{\alpha,\beta}$, matches to a vertex of degree 2. We have that

$$\begin{aligned}\eta(u) &\geqslant \Pr(\text{the first copy of } u \text{ is neighbor of a degree 2 node})\\ &= \frac{2\cdot\#\text{deg-2-nodes}}{\left(\sum_{v\in V} d(v)\right)-1}\\ &\geqslant \frac{2\cdot\frac{e^{\alpha}}{2^{\beta}}}{\sum_{i=1}^{\Delta} i\cdot\frac{e^{\alpha}}{i^{\beta}}} = \frac{\frac{1}{2^{\beta-1}}}{\sum_{i=1}^{\Delta}\frac{1}{i^{\beta-1}}}\ ,\end{aligned}$$

where $\Delta = e^{\alpha/\beta}$ is the maximum degree of $G_{\alpha,\beta}$. □

In Equation 8.1 we substitute $\eta(u)$ by the bound given in Lemma 8.5 and obtain:

$$\begin{aligned}\mathbb{E}[x(V^*)] &\geqslant \frac{1}{2} \cdot \sum_{u\colon d(u)\geqslant 3} \eta(u) \\ &= \frac{1}{2} \cdot \left(\sum_{i=1}^{\Delta} \frac{e^\alpha}{i^\beta} - e^\alpha - \frac{e^\alpha}{2^\beta} \right) \cdot \frac{1}{2^{\beta-1} \cdot \sum_{i=1}^{\Delta} \frac{1}{i^{\beta-1}}} \\ &= \frac{e^\alpha}{2^\beta} \cdot \frac{\sum_{i=1}^{\Delta} \frac{1}{i^\beta} - 1 - \frac{1}{2^\beta}}{\sum_{i=1}^{\Delta} \frac{1}{i^{\beta-1}}} \ . \end{aligned} \tag{8.3}$$

We will now show that in Inequality 8.3 we can replace the two sums $\sum_{i=1}^{\Delta} \frac{1}{i^\beta}$ and $\sum_{i=1}^{\Delta} \frac{1}{i^{\beta-1}}$ by $\zeta(\beta)$ and $\zeta(\beta-1)$, respectively. We make use of the following lemma.

Lemma 8.6.
For any real numbers $A, B, a, b > 0$, $\frac{A}{B} \geqslant \frac{A+a}{B+b}$ if and only if $\frac{A}{B} \geqslant \frac{a}{b}$.

Recall that $\zeta(\beta) = \sum_{i=1}^{\infty} \frac{1}{i^\beta}$ is the Riemann zeta function. Therefore, in order to show

$$\mathbb{E}[x(V^*)] \geqslant \frac{e^\alpha}{2^\beta} \cdot \frac{\zeta(\beta) - 1 - \frac{1}{2^\beta}}{\zeta(\beta-1)} \ ,$$

we will show that the following inequality holds

$$\frac{\sum_{i=1}^{\Delta} \frac{1}{i^\beta} - 1 - \frac{1}{2^\beta}}{\sum_{i=1}^{\Delta} \frac{1}{i^{\beta-1}}} \geqslant \frac{\sum_{i=1}^{\Delta+1} \frac{1}{i^\beta} - 1 - \frac{1}{2^\beta}}{\sum_{i=1}^{\Delta+1} \frac{1}{i^{\beta-1}}} \ .$$

Due to Lemma 8.6, it is sufficient to show that there exists a Δ_0 such that for all $\Delta \geqslant \Delta_0$ the following holds

$$\frac{\sum_{i=1}^{\Delta} \frac{1}{i^\beta} - 1 - \frac{1}{2^\beta}}{\sum_{i=1}^{\Delta} \frac{1}{i^{\beta-1}}} \geqslant \frac{\frac{1}{(\Delta+1)^\beta}}{\frac{1}{(\Delta+1)^{\beta-1}}} = \frac{1}{\Delta+1} \ .$$

This is provided by the following lemma.

Lemma 8.7.
There exists a $\Delta_0 \geqslant 8$,such that for all $\Delta \geqslant \Delta_0$,

$$\frac{\sum_{i=1}^{\Delta} \frac{1}{i^\beta} - 1 - \frac{1}{2^\beta}}{\sum_{i=1}^{\Delta} \frac{1}{i^{\beta-1}}} \geqslant \frac{1}{\Delta+1} \ .$$

Proof. The above inequality is equivalent to

$$\begin{aligned} & \sum_{i=1}^{\Delta} \frac{1}{i^\beta} - 1 - \frac{1}{2^\beta} \geqslant \sum_{i=1}^{\Delta} \frac{1}{\Delta+1} \cdot \frac{1}{i^{\beta-1}} \\ \iff \quad & \sum_{i=1}^{\Delta} \left(\frac{1}{i^\beta} - \frac{1}{\Delta+1} \cdot \frac{1}{i^{\beta-1}} \right) \geqslant 1 + \frac{1}{2^\beta} \\ \iff \quad & \sum_{i=1}^{\Delta} \frac{\Delta+1-i}{(\Delta+1)i^\beta} \geqslant 1 + \frac{1}{2^\beta} \ . \end{aligned} \tag{8.4}$$

Suppose $\Delta \geqslant 8$, then the sum on the left-hand side of the Inequality 8.4 is bounded by the sum of the terms with indices $i = 1, 2, 4, 8$, that is

$$\begin{aligned} \sum_{i=1}^{\Delta} \frac{\Delta+1-i}{(\Delta+1)i^\beta} &\geqslant \frac{\Delta}{\Delta+1} + \frac{\Delta-1}{(\Delta+1)2^\beta} + \frac{\Delta-3}{(\Delta+1)4^\beta} + \frac{\Delta-7}{(\Delta+1)8^\beta} \\ &= \frac{\Delta 8^\beta + (\Delta-1)4^\beta + (\Delta-3)2^\beta + \Delta - 7}{(\Delta+1)8^\beta} \ . \end{aligned} \tag{8.5}$$

Using Inequality 8.5 and the fact that $1 + \frac{1}{2^\beta} = \frac{(\Delta+1)8^\beta + (\Delta+1)4^\beta}{(\Delta+1)8^\beta}$, in order to prove Inequality 8.4 it is sufficient to show the following:

$$\begin{aligned} & \frac{\Delta 8^\beta + (\Delta-1)4^\beta + (\Delta-3)2^\beta + \Delta - 7}{(\Delta+1)8^\beta} \geqslant \frac{(\Delta+1)8^\beta + (\Delta+1)4^\beta}{(\Delta+1)8^\beta} \\ \iff \quad & (\Delta-3)2^\beta + \Delta - 7 \geqslant 8^\beta + 2 \cdot 4^\beta. \end{aligned}$$

This is valid for $\Delta \geqslant \frac{8^\beta + 2 \cdot 4^\beta + 6 \cdot 2^\beta + 7}{1+2^\beta}$. Hence we choose $\Delta_0 = \left\lceil \frac{8^\beta + 2 \cdot 4^\beta + 6 \cdot 2^\beta + 7}{1+2^\beta} \right\rceil$. □

This completes the proof of Theorem 8.4. The next lemma provides an upper bound for $x(V)$, which is the size of a half-integral solution on the complete vertex set V.

Lemma 8.8.
Let $x(V)$ be the value of a half-integral solution on V. Then the following inequality holds:

$$x(V) \leqslant \frac{1}{2} \cdot \zeta(\beta)\, e^{\alpha} \quad .$$

Proof. In order to achieve the above upper bound for $x(V)$, we construct a feasible half-integral solution for $G_{\alpha,\beta}$ by setting $x(v) = 1/2$ for all $v \in V$ where $1/2 \cdot \sum_{v \in V} \leqslant 1/2 \cdot \zeta(\beta)\, e^{\alpha}$. □

Now let us restate the main Theorem 8.1 and finish the proof.

Theorem.
For $\beta > 2$, the algorithm `Deterministic_Rounding` *approximates the* Minimum Vertex Cover *problem in random power law graphs $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ in the $G(\alpha, \beta)$ model with an expected approximation ratio of*

$$\rho = 2 - \frac{\zeta(\beta) - 1 - \frac{1}{2^{\beta}}}{2^{\beta}\zeta(\beta-1)\zeta(\beta)} \quad .$$

Proof. The algorithm `Deterministic_Rounding` achieves an approximation ratio of $3/2$ for Min-VC in the subgraph $G_{\alpha,\beta}[V^*]$ and a ratio of 2 in $G_{\alpha,\beta}[V \setminus V^*]$. Therefore, the approximation ratio of the algorithm on $G_{\alpha,\beta}$ can be expressed in expectation as

$$\rho \leqslant \mathbb{E}\left[\frac{3}{2} \cdot \frac{x(V^*)}{x(V)} + 2 \cdot \frac{x(V) - x(V^*)}{x(V)}\right] = \mathbb{E}\left[2 - \frac{1}{2} \cdot \frac{x(V^*)}{x(V)}\right] \quad .$$

Due to Theorem 8.4 and Lemma 8.8, we have that $\mathbb{E}[x(V^*)] \geqslant \frac{1}{2} \cdot \frac{\left(\zeta(\beta) - 1 - \frac{1}{2^{\beta}}\right) e^{\alpha}}{2^{\beta-1}\zeta(\beta-1)}$ and $x(V) \leqslant \frac{1}{2} \cdot \zeta(\beta)\, e^{\alpha}$. This yields

$$\mathbb{E}\left[\frac{x(V^*)}{x(V)}\right] \geqslant \frac{\frac{1}{2} \cdot \frac{\left(\zeta(\beta) - 1 - \frac{1}{2^{\beta}}\right) e^{\alpha}}{2^{\beta-1}\zeta(\beta-1)}}{\frac{1}{2} \cdot \zeta(\beta)\, e^{\alpha}} = \frac{\zeta(\beta) - 1 - \frac{1}{2^{\beta}}}{2^{\beta-1}\zeta(\beta-1)\zeta(\beta)} \quad ,$$

and therefore

$$\rho \leqslant 2 - \frac{1}{2} \cdot \frac{\zeta(\beta) - 1 - \frac{1}{2^{\beta}}}{2^{\beta-1}\zeta(\beta-1)\zeta(\beta)} = 2 - \frac{\zeta(\beta) - 1 - \frac{1}{2^{\beta}}}{2^{\beta}\zeta(\beta-1)\zeta(\beta)} \quad .$$

□

In the next section, we give a better estimate of the lower bound for $\chi(V^*)$ in Lemma 8.4. The new estimate is based on the probability that a high degree vertex is adjacent to some vertex subset U. In particular, we will consider the case when U is the set of degree 1 and degree 2 vertices. The improved resulting lower bound, however, is only valid asymptotically as the graph size n approaches ∞.

8.8 REFINED ANALYSIS

We will now refine the analysis of algorithm `Deterministic_Rounding` on page 182 by giving a better estimate on the probability $\eta(u, U)$ of a high-degree node u being adjacent to a vertex in a given set $U \subseteq V$, i. e. a vertex of degree one or two. Again, we will first obtain a bound on the expected approximation ratio of the algorithm in terms of the partial sums $\sum_{i=1}^{\Delta} \frac{1}{i^{\beta}}$ and $\sum_{i=1}^{\Delta} \frac{1}{i^{\beta-1}}$ and then show that these can be replaced by $\zeta(\beta)$ and $\zeta(\beta-1)$, respectively.

Lemma 8.9.
For every u with $d(u) \geqslant 3$ and $U \subseteq V$,

$$\eta(u, U) \geqslant \frac{\sum_{i=1}^{\Delta} \frac{e^{\alpha}}{i^{\beta-1}} - e^{\frac{\alpha}{\beta}} + 1}{\sum_{i=1}^{\Delta} \frac{e^{\alpha}}{i^{\beta-1}}} \left[1 - \left(\frac{\sum_{i=1}^{\Delta} \frac{e^{\alpha}}{i^{\beta-1}} - d(U) - 6 + 1}{\sum_{i=1}^{\Delta} \frac{e^{\alpha}}{i^{\beta-1}} - 6 + 1} \right)^3 \right] .$$

Proof. For a given set U of vertices from $G_{\alpha,\beta}$ we let $d(U) = \sum_{v \in U} d(v)$. Furthermore let $\eta(u, U)$ be the probability that u is connected to at least one node in U. Recall how $G_{\alpha,\beta}$ is constructed in the $G(\alpha, \beta)$ model. For each node u of degree $d(u)$, $d(u)$ copies are introduced and then a random perfect matching on the set L of all these copies is constructed. This random perfect matching is generated by first choosing a neighbor for the first node in L uniformly at random, then choosing a neighbor for the next unoccupied node in L uniformly at random, and so on. This distribution does not

depend on the order of nodes in L. Thus, we may assume that copies corresponding to vertex u are the first $d(u)$ nodes in L. We obtain

$$\begin{aligned}
\eta(u, U) &= \Pr(u \text{ matches to } U) \\
&= \sum_{j=1}^{d(u)} \Pr(j\text{-th copy for node } u \text{ is the first one matching to } U) \\
&= \sum_{j=1}^{d(u)} \Pr(\text{copies } 1, \ldots, j-1 \text{ not connected to } U \text{ and copy } j \text{ conn. to } U) \\
&= \sum_{j=1}^{d(u)} \Pr\left(\begin{array}{l} \text{copies } 1, \ldots, j-1 \text{ neither connected to } U \text{ nor} \\ \text{to any copies of } u \text{ and copy } j \text{ connected to } U \end{array} \right) \\
&= \sum_{j=1}^{d(u)} \frac{d(U)}{\sum_{i=1}^{\Delta} \frac{e^{\alpha}}{i^{\beta-1}} - 2(j-1) - 1} \prod_{k=1}^{j-1} \left(\frac{\sum_{i=1}^{\Delta} \frac{e^{\alpha}}{i^{\beta-1}} - d(u) - d(U) - (k-1)}{\sum_{i=1}^{\Delta} \frac{e^{\alpha}}{i^{\beta-1}} - 2(k-1) - 1} \right).
\end{aligned}$$

Now define $N = \sum_{i=1}^{\Delta} \frac{e^{\alpha}}{i^{\beta-1}}$. We have

$$\begin{aligned}
\eta(u, U) &= \sum_{j=1}^{d(u)} \frac{d(U)}{N - 2j + 1} \prod_{k=1}^{j-1} \frac{N - d(U) - d(u) - (k-1)}{N - 2k + 1} \\
&\geqslant \sum_{j=1}^{d(u)} \frac{d(U)}{N - 2j + 1} \left(\frac{N - d(U) - d(u) - (j-1)}{N - 2j + 1} \right)^{j-1} \\
&\geqslant \sum_{j=1}^{d(u)} \frac{d(U)}{N} \left(\frac{N - d(U) - 2d(u) + 1}{N - 2d(u) + 1} \right)^{j-1} \\
&= \frac{d(U)}{N} \left[\frac{1 - \left(\frac{N - d(U) - 2d(u) + 1}{N - 2d(u) + 1} \right)^{d(u)}}{1 - \frac{N - d(U) - 2d(u) + 1}{N - 2d(u) + 1}} \right] \\
&= \frac{d(U)}{N} \left[1 - \left(\frac{N - d(U) - 2d(u) + 1}{N - 2d(u) + 1} \right)^{d(u)} \right] \cdot \frac{N - 2d(u) + 1}{d(U)} \\
&= \frac{N - 2d(u) + 1}{N} \left[1 - \left(\frac{N - d(U) - 2d(u) + 1}{N - 2d(u) + 1} \right)^{d(u)} \right].
\end{aligned}$$

Since the function $\left(\frac{N-d(U)-2d(u)+1}{N-2d(u)+1}\right)^{d(u)}$ is monotone decreasing in $d(u)$ it follows that we can choose $d(u) = 3$ in order to obtain a lower bound on $\eta(u, U)$, which yields

$$\begin{aligned}\eta(u,U) &\geqslant \frac{N-\Delta+1}{N}\left[1-\left(\frac{N-d(U)-2\cdot 3+1}{N-2\cdot 3+1}\right)^3\right] \\ &= \frac{\sum_{i=1}\frac{e^\alpha}{i^{\beta-1}} - e^{\frac{\alpha}{\beta}}+1}{\sum_{i=1}\frac{e^\alpha}{i^{\beta-1}}}\left[1-\left(\frac{\sum_{i=1}\frac{e^\alpha}{i^{\beta-1}}-d(U)-6+1}{\sum_{i=1}\frac{e^\alpha}{i^{\beta-1}}-6+1}\right)^3\right] .\end{aligned}$$

□

Because of Equation 8.1 we have $\mathbb{E}[x(V^*)] \geqslant 1/2 \cdot \sum_{u:\, d(u)\geqslant 3}\eta(u, U)$ and we obtain the following approximation ratio of the algorithm `Deterministic_Rounding`:

$$\begin{aligned}\rho &\leqslant \mathbb{E}\left[2-\frac{1}{2}\frac{x(V^*)}{x(V)}\right] \\ &\leqslant 2-\frac{1}{2}\frac{\left(\sum_{i=1}^{\Delta}\frac{e^\alpha}{i^\beta}-e^\alpha-\frac{e^\alpha}{2^\beta}\right)\frac{\sum_{i=1}^{\Delta}\frac{e^\alpha}{i^{\beta-1}}-e^{\frac{\alpha}{\beta}}+1}{\sum_{i=1}^{\Delta}\frac{e^\alpha}{i^{\beta-1}}}\left[1-\left(\frac{\sum_{i=1}^{\Delta}\frac{e^\alpha}{i^{\beta-1}}-d(U)-6+1}{\sum_{i=1}^{\Delta}\frac{e^\alpha}{i^{\beta-1}}-6+1}\right)^3\right]}{\frac{1}{2}\sum_{i=1}^{\Delta}\frac{e^\alpha}{i^\beta}} \\ &= 2-\frac{\left(\sum_{i=1}^{\Delta}\frac{1}{i^\beta}-1-\frac{1}{2^\beta}\right)\left(\sum_{i=1}^{\Delta}\frac{1}{i^{\beta-1}}-\frac{\Delta}{e^\alpha}+\frac{1}{e^\alpha}\right)}{\left(\sum_{i=1}^{\Delta}\frac{1}{i^{\beta-1}}\right)\left(\sum_{i=1}^{\Delta}\frac{1}{i^\beta}\right)}\left[1-\left(\frac{\sum_{i=1}^{\Delta}\frac{1}{i^{\beta-1}}-\frac{d(U)}{e^\alpha}-\frac{5}{e^\alpha}}{\sum_{i=1}^{\Delta}\frac{1}{i^{\beta-1}}-\frac{5}{e^\alpha}}\right)^3\right]\end{aligned} \tag{8.6}$$

We observe that the terms $\sum_{i=1}\frac{e^\alpha}{i^\beta}$ and $\sum_{i=1}\frac{e^\alpha}{i^{\beta-1}}$ converge to $\zeta(\beta)$ and $\zeta(\beta-1)$ respectively as $\alpha \to \infty$. Furthermore, $\frac{2\Delta}{e^\alpha}, \frac{1}{e^\alpha} \to 0\,(\alpha \to \infty)$. Hence the upper bound on the expected approximation ratio in inequality (8.6) converges to

$$\rho' \leqslant 2-\frac{\left(\zeta(\beta)-1-\frac{1}{2^\beta}\right)\cdot\zeta(\beta-1)}{\zeta(\beta-1)\cdot\zeta(\beta)}\left[1-\left(\frac{\zeta(\beta-1)-\left(1+\frac{1}{2^{\beta-1}}\right)}{\zeta(\beta-1)}\right)^3\right].$$

Thus we obtain the following theorem:

Theorem.
For all $\beta > 2$ *and* $\varepsilon > 0$ *the* MINIMUM VERTEX COVER *problem in random power law graphs* $G_{\alpha,\beta} \in \mathcal{G}_{\alpha,\beta}$ *can be approximated within expected approximation ratio*

$$\rho' \leqslant 2 - \frac{\left(\zeta(\beta) - 1 - \frac{1}{2^\beta}\right) \cdot \zeta(\beta-1)}{\zeta(\beta-1) \cdot \zeta(\beta)} \left[1 - \left(\frac{\zeta(\beta-1) - \left(1 + \frac{1}{2^{\beta-1}}\right)}{\zeta(\beta-1)}\right)^3\right] + \varepsilon \ .$$

We conclude this last main chapter with a short summary, directions for further research and some bibliographic notes.

8.9 SUMMARY AND FURTHER RESEARCH

Let us give a short summary of the findings of this chapter. In Section 8.5 we presented a new approximation algorithm for MIN-VC in (α, β)-PLG with expected approximation ratio of

$$\rho \leqslant 2 - \frac{\zeta(\beta) - 1 - \frac{1}{2^\beta}}{2^\beta \zeta(\beta-1)\zeta(\beta)}$$

in our first analysis of Section 8.7. Moreover, in Section 8.8 we showed an expected asymptotic approximation ratio of

$$\rho' \leqslant 2 - \frac{\left(\zeta(\beta) - 1 - \frac{1}{2^\beta}\right) \zeta(\beta-1)}{\zeta(\beta-1)\zeta(\beta)} \left[1 - \left(\frac{\zeta(\beta-1) - \left(1 + \frac{1}{2^{\beta-1}}\right)}{\zeta(\beta-1)}\right)^3\right] \ .$$

The algorithm itself basically consists of a deterministic rounding procedure on an optimal half-integral solution x that produces a rounded integer solution y. We showed that this rounding procedure yields an approximation ratio of $3/2$ in the subgraph $G_{\alpha,\beta}[V^*]$ induced by the low-degree vertices and a 2-approximation in the residual graph $G_{\alpha,\beta}[V \setminus V^*]$. Furthermore, we proved upper and lower bounds on the total value of the initial half-integral solution x for the vertex set V and V^*, respectively. These bounds for the vertex subsets, together with the acheived approximation ratios

on the corresponding induced subgraphs, yield the above expected approximation ratios.

Further research is aimed on improving the upper and lower bounds presented in the analysis of the algorithm in order to provide further improvements on the overall approximation ratio. Furthermore, an extension of the capabilities of the deterministic rounding procedure to larger vertex subsets and including higher vertex degrees would be feasible to achieve better approximation factors.

8.10 BIBLIOGRAPHIC NOTES

The material and the results presented in this chapter are based on the following publication: Mikael Gast and Mathias Hauptmann. "Approximability of the vertex cover problem in power law graphs." In: *Computing Research Repository (CoRR) preprint arXiv:1204.0982 [cs.DS]; also submitted to Theoretical Computer Science* (2012), pp. 1–16. arXiv: 1204.0982.

The LP-relaxation formulation of MIN-VC and the proof of existence and computability of half-integral solutions (Theorem 8.3 in Section 8.3.1 on page 178) is due to Nemhauser and Trotter [NT75]. Furthermore, the observation that rounding on half-integral solutions yield good approximations for MIN-VC is due to Hochbaum et al. [Hoc+93].

The proofs of our main theorems (Theorem 8.1 on page 175 and Theorem 8.2 on page 175) and the auxiliary theorem (Theorem 8.4 on page 185) also appeared in [GH12].

9 CONCLUSION AND FURTHER RESEARCH

In this thesis we have shown the APX-hardness of the MINIMUM VERTEX COVER (MIN-VC) problem in *connected* (α, β)-PLG multigraphs for $0 < \beta < \beta_{max} \approx 2.48$ and therefore rule out the existence of a PTAS. This partially answered an open question posed by Ferrante, Pandurangan, and Park [FPP08]. We gave explicit approximation lower bounds for the various ranges of the model parameter β. Furthermore, we presented a detailed analysis of the phase transition for $\beta = 1$ by extending the model of (α, β)-PLG to a functional case where the parameter is of the form $\beta_f = 1 \pm 1/f(n)$. For a sufficiently fast growing function $f(n)$, we showed that our new inapproximability results for the case $\beta = 1$ in the original model also hold for this functional model. It remains an important open question to close the gaps between inapproximability and approximability bounds of the underlying problems. We also believe that our results for the two functional cases $\beta = 1 \pm \frac{1}{f(n)}$ can be extended to hold for any $\beta_f = \beta \pm \frac{1}{f(n)}$ with $0 < \beta < \beta_{max} \approx 2.48$.

Then we studied the approximation complexity of the MINIMUM DOMINATING SET (MIN-DS) problem in connected (α, β)-PLG. Particularly, we gave the first logarithmic lower bounds for the approximability of this problem for the parameter range $0 < \beta \leqslant 2$. Our results are based on a reduction from the SET COVER problem combined with the logarithmic lower bound given by Feige [Fei98]. Thereby we also improve over the previously known constant factor lower bounds due to Shen et al. [She+12] for the case of disconnected (α, β)-PLG. For $\beta > 2$ we show that MIN-DS in (α, β)-PLG is in APX and improve on the approximation upper bounds of Shen et al. [She+12] for a greedy algorithm for MIN-DS. Finally, we take a very close look at the phase transition at $\beta = 2$ and considered the case when $\beta_f = 2 + \frac{1}{f(n)}$ is a function of the graph size n which converges to 2 from above. We obtained the surprising result that for every function $f(n)$ with $f(n) = o(\log(n))$, the problem is in APX and for every function $f(n)$ with $f(n) = \omega(\log(n))$ MIN-DS provides a logarithmic approximation lower bound as in the case $0 < \beta \leqslant 2$. The further improvements on both lower and upper approximation bounds are important open questions in the area, especially the upper approximation bounds for $\beta \leqslant 2$.

Finally, we constructed an approximation algorithm for the MIN-VC problem with an expected approximation ratio of $2 - f(\beta)$ for *random* (α, β)*-PLG*, where $f(\beta)$ is a strictly positive function of the model parameter β. In particular $f(\beta)$ does not depend on the size $|V|$ of the graph and thus—for large graph sizes—our approxima-

tion ratio falls below current upper bounds for MIN-VC in general graphs (which is $2 - \Theta(1/\sqrt{\log n})$ and due to Karakostas [Kar09]). The result is based on a deterministic rounding procedure that acts an a given half-integral solution for MIN-VC. Further research is aimed on improving the upper and lower bounds on the size of half-integral solutions in (α, β)-PLG. This would provide further improvements on the overall approximation ratio. Moreover, an extension of the capabilities of the deterministic rounding procedure to higher degree vertices would be helpful for achieving good approximation factors on larger vertex subsets.

BIBLIOGRAPHY

[ACL00] William Aiello, Fan Chung, and Linyuan Lu. "A random graph model for massive graphs." In: *Proceedings of the 32nd annual ACM Symposium on Theory of Computing*. STOC '00. ACM Press, 2000, pp. 171–180. DOI: 10.1145/335305.335326 (cit. on pp. 11, 48, 49, 138).

[ACL01] William Aiello, Fan Chung, and Linyuan Lu. "A random graph model for power law graphs." In: *Experimental Mathematics* 10.1 (2001), pp. 53–66 (cit. on pp. 11, 48–51, 53, 112, 138, 174).

[AB02] Réka Albert and Albert-László Barabási. "Statistical mechanics of complex networks." In: *Reviews of Modern Physics* 74.1 (Jan. 2002), pp. 47–97. DOI: 10.1103/RevModPhys.74.47 (cit. on p. 5).

[AJB00] Réka Albert, Hawoong Jeong, and Albert-László Barabási. "Error and attack tolerance of complex networks." In: *Nature* 406.6794 (July 2000), pp. 378–382. DOI: 10.1038/35019019 (cit. on p. 14).

[AH76] Kenneth Appel and Wolfgang Haken. "Every planar map is four colorable." In: *American Mathematical Society* 82.5 (1976) (cit. on p. 31).

[AB09] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009, pp. I–XXIV, 1–579 (cit. on p. 71).

[Aro+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. "Proof verification and the hardness of approximation problems." In: *Journal of the ACM* 45.3 (May 1998), pp. 501–555. DOI: 10.1145/278298.278306 (cit. on pp. 70, 71).

[AS98] Sanjeev Arora and Shmuel Safra. "Probabilistic checking of proofs: a new characterization of NP." In: *Journal of the ACM* 45.1 (Jan. 1998), pp. 70–122. DOI: 10.1145/273865.273901 (cit. on pp. 70, 71).

[Aue13] Felix Auerbach. "Das Gesetz der Bevölkerungskonzentration." In: *Petermann's Geographische Mitteilungen* 59 (1913), pp. 74–76 (cit. on pp. 9, 41, 42).

[Aus+99] Giorgio Ausiello, Pierluigi Crescenzi, Giorgio Gambosi, Viggo Kann, Aalberto Marchetti-Spaccamela, and Marco Protasi. *Complexity and Approximation: Combinatorial optimization problems and their approximability properties*. Springer-Verlag, Berlin Heidelberg, 1999, p. 524. DOI: 10.1007/978-3-642-58412-1 (cit. on p. 64).

[AKS09] Per Austrin, Subhash Khot, and Muli Safra. "Inapproximability of Vertex Cover and Independent Set in Bounded Degree Graphs." In: *Proceedings of the 24th annual IEEE Conference on Computational Complexity*. CCC '09. IEEE Comput. Soc. Press, July 2009, pp. 74–80. DOI: 10.1109/CCC.2009.38 (cit. on pp. 16, 86).

[BFL91] László Babai, Lance Fortnow, and Carsten Lund. "Non-deterministic exponential time has two-prover interactive protocols." In: *Computational Complexity* 1.1 (Mar. 1991), pp. 3–40. DOI: 10.1007/BF01200056 (cit. on pp. 68, 71).

[Bac94] Paul G. H. Bachmann. *Zahlentheorie: Die analytische Zahlentheorie*. Vol. 2. Leipzig: B. G. Teubner, 1894 (cit. on p. 22).

[BB06] Balabhaskar Balasundaram and Sergiy Butenko. "Graph domination, coloring and cliques in telecommunications." In: *Handbook of Optimization in Telecommunications*. Springer-Verlag, 2006, pp. 865–890. DOI: 10.1007/978-0-387-30165-5_30 (cit. on p. 13).

[Bal+05] Nicole E. Baldwin, Elissa J. Chesler, Stefan Kirov, Michael A. Langston, Jay R. Snoddy, Robert W. Williams, and Bing Zhang. "Computational, integrative, and comparative methods for the elucidation of genetic coexpression networks." In: *Journal of Biomedicine & Biotechnology* 2005.2 (June 2005), pp. 172–80. DOI: 10.1155/JBB.2005.172 (cit. on p. 15).

[BA99] Albert-László Barabási and Réka Albert. "Emergence of Scaling in Random Networks." In: *Science* 286.5439 (Oct. 1999), pp. 509–512. DOI: 10.1126/science.286.5439.509 (cit. on pp. 9–11, 43–45).

[BE85] Reuven Bar-Yehuda and Shimon Even. "A local-ratio theorem for approximating the weighted vertex cover problem." In: *Annals of Discrete Mathematics* 25 (1985), pp. 27–46. DOI: 10.1016/S0304-0208(08)73101-3 (cit. on p. 86).

[BM84] Reuven Bar-Yehuda and Shlomo Moran. "On approximation problems related to the independent set and vertex cover problems." In: *Discrete Applied Mathematics* 9.1 (Sept. 1984), pp. 1–10. DOI: 10.1016/0166-218X(84)90086-6 (cit. on p. 139).

[Ben+88] Michael Ben-Or, Shafi Goldwasser, Joe Kilian, and Avi Widgerson. "Multi-prover interactive proofs: how to remove intractability." In: *Proceedings of the 20th annual ACM Symposium on Theory of Computing*. STOC '88. New York, New York, USA: ACM Press, Jan. 1988, pp. 113–131. DOI: 10.1145/62212.62223 (cit. on p. 68).

[BK99] Piotr Berman and Marek Karpinski. "On some tighter inapproximability results." In: *Proceedings of the 26th international Conference on Automata, Languages and Programming*. ICALP '99. Springer-Verlag, 1999, pp. 200–209. DOI: 10.1007/3-540-48523-6_17 (cit. on pp. 16, 76–82, 87).

[BK01] Piotr Berman and Marek Karpinski. "Efficient amplifiers and bounded degree optimization." In: *ECCC TR01-053* (2001) (cit. on pp. 76, 77, 82).

[BK03] Piotr Berman and Marek Karpinski. "Improved approximation lower bounds on small occurrence optimization." In: *ECCC TR03-008* (2003) (cit. on pp. 76, 77, 80, 82, 87, 89).

[BLW86] Norman L. Biggs, E. Keith Lloyd, and Robin J. Wilson. *Graph Theory, 1736-1936*. Clarendon Press, Oct. 1986, p. 240 (cit. on pp. 31, 178).

[Boc+06] Stefano Boccaletti, Vito Latora, Yamir Moreno, Mario Chavez, and Dong-Uk Hwang. "Complex networks: Structure and dynamics." In: *Physics Reports* 424.4-5 (Feb. 2006), pp. 175–308. DOI: 10.1016/j.physrep.2005.10.009 (cit. on p. 5).

[Bol81] Béla Bollobás. "The diameter of random graphs." In: *Transactions of the American Mathematical Society* 267.1 (Sept. 1981), p. 41. DOI: 10.2307/1998567 (cit. on p. 40).

[Bol98] Béla Bollobás. *Modern graph theory*. Vol. 184. Springer Verlag, 1998 (cit. on pp. 31, 53).

[BF82] Béla Bollobás and Wenceslas Fernandez de la Vega. "The diameter of random regular graphs." In: *Combinatorica* 2.2 (June 1982), pp. 125–134. DOI: 10.1007/BF02579310 (cit. on pp. 7, 40, 53).

[BR04] Béla Bollobás and Oliver M. Riordan. "The diameter of a scale-free random graph." In: *Combinatorica* 24.1 (Jan. 2004), pp. 5–34. DOI: 10.1007/s00493-004-0002-2 (cit. on pp. 46, 47, 53).

[BR05] Béla Bollobás and Oliver M. Riordan. "Mathematical results on scale-free random graphs." In: *Handbook of Graphs and Networks* (2005), pp. 1–34. DOI: 10.1002/3527602755.ch1 (cit. on pp. 11, 39, 44, 45, 53).

[Bor06] Stephen P. Borgatti. "Identifying sets of key players in a social network." In: *Computational and Mathematical Organization Theory* 12.1 (Apr. 2006), pp. 21–34. DOI: 10.1007/s10588-006-7084-x (cit. on p. 13).

[Bor+04] Peer Bork, Lars J. Jensen, Christian von Mering, Arun K. Ramani, Insuk Lee, and Edward M Marcotte. "Protein interaction networks from yeast to human." In: *Current Opinion in Structural Biology* 14.3 (June 2004), pp. 292–299. DOI: 10.1016/j.sbi.2004.05.003 (cit. on p. 7).

[Bro+00] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew S. Tomkins, and Janet Wiener. "Graph structure in the Web." In: *Computer Networks* 33.1-6 (2000), pp. 309–320. DOI: 10.1016/S1389-1286(00)00083-9 (cit. on pp. 9, 41).

[BO04] Pierce G. Buckley and Deryk Osthus. "Popularity based random graph models leading to a scale-free degree sequence." In: *Discrete Mathematics* 282.1-3 (May 2004), pp. 53–68. DOI: 10.1016/j.disc.2003.10.020 (cit. on pp. 11, 47).

[Cay57] Arthur Cayley. "XXVIII. On the theory of the analytical forms called trees." In: *Philosophical Magazine Series 4* 13.85 (1857), pp. 172–176. DOI: 10.1080/14786445708642275 (cit. on p. 31).

[CLL04] Yuanzhu Peter Chen, Arthur L. Liestman, and Jiangchuan Liu. "Clustering algorithms for ad hoc wireless networks." In: *Ad Hoc and Sensor Networks* (2004) (cit. on p. 13).

[Che+05] Elissa J. Chesler, Lu Lu, Siming Shou, Yanhua Qu, Jing Gu, Jintao Wang, Hui Chen Hsu, John D. Mountz, Nicole E. Baldwin, Michael A. Langston, David W. Threadgill, Kenneth F. Manly, and Robert W. Williams. "Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function." In: *Nature genetics* 37.3 (Mar. 2005), pp. 233–242. DOI: 10.1038/ng1518 (cit. on p. 15).

[CL01] Fan Chung and Linyuan Lu. "The diameter of sparse random graphs." In: *Advances in Applied Mathematics* 26.4 (May 2001), pp. 257–279. DOI: 10.1006/aama.2001.0720 (cit. on p. 40).

[CL06] Fan Chung and Linyuan Lu. *Complex graphs and networks*. American Mathematical Society, 2006, p. 264 (cit. on p. 53).

[Chv79] Vašek Chvátal. "A Greedy Heuristic for the Set-Covering Problem." In: *Mathematics of Operations Research* 4.3 (Aug. 1979), pp. 233–235. DOI: 10.1287/moor.4.3.233 (cit. on p. 134).

[Coh+00] Reuven Cohen, Keren Erez, Daniel Ben-Avraham, and Shlomo Havlin. "Resilience of the Internet to Random Breakdowns." In: *Physical Review Letters* 85.21 (Nov. 2000), pp. 4626–4628. DOI: 10.1103/PhysRevLett.85.4626 (cit. on p. 14).

[CH03] Reuven Cohen and Shlomo Havlin. "Scale-free networks are ultrasmall." In: *Physical Review Letters* 90.5 (Feb. 2003), p. 4. DOI: 10.1103/PhysRevLett.90.058701 (cit. on pp. 46, 47).

[Cor+09] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT Press, 2009, pp. I–XIX, 1–1292 (cit. on pp. 53, 64).

[Cre97] Pierluigi Crescenzi. "A short guide to approximation preserving reductions." In: *Proceedings of the 12th annual IEEE Conference on Computational Complexity*. CCC '97. IEEE Comput. Soc. Press, 1997, pp. 262–273. DOI: 10.1109/CCC.1997.612321 (cit. on pp. 59, 60, 64).

[Cre+99] Pierluigi Crescenzi, Viggo Kann, Riccardo Silvestri, and Luca Trevisan. "Structure in Approximation Classes." In: *SIAM Journal on Computing* 28.5 (Jan. 1999), pp. 1759–1782. DOI: 10.1137/S0097539796304220 (cit. on p. 64).

[DB97] Bevan Das and Vaduvur Bharghavan. "Routing in ad-hoc networks using minimum connected dominating sets." In: *Proceedings of the International Conference on Communications*. Vol. 1. ICC '97. IEEE, 1997, pp. 376–380. DOI: 10.1109/ICC.1997.605303 (cit. on p. 14).

[Die10] Reinhard Diestel. *Graph Theory*. 4th Ed. Vol. 3. Springer-Verlag, Heidelberg, 2010, p. 451 (cit. on pp. 31, 53).

[Din07] Irit Dinur. "The PCP theorem by gap amplification." In: *Journal of the ACM* 54.3 (June 2007), pp. 1–44. DOI: 10.1145/1236457.1236459 (cit. on p. 71).

[DS05] Irit Dinur and Samuel Safra. "On the hardness of approximating vertex cover." In: *Annals of Mathematics* 162.1 (July 2005), pp. 439–485. DOI: 10.4007/annals.2005.162.439 (cit. on pp. 12, 86, 174).

[DM03] Sergey N. Dorogovtsev and José F. F. Mendes. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, 2003, p. 280 (cit. on p. 5).

[DM04] Sergey N. Dorogovtsev and José F. F. Mendes. "The shortest path to complex networks." In: *Arxiv preprint cond-mat/0404593* (2004), pp. 1–25. arXiv: 0404593v4 [arXiv:cond-mat] (cit. on p. 5).

[DMS00] Sergey N. Dorogovtsev, José F. F. Mendes, and Alexander N. Samukhin. "Structure of Growing Networks with Preferential Linking." In: *Physical Review Letters* 85.21 (2000), pp. 4633–4636. DOI: 10.1103/PhysRevLett.85.4633 (cit. on pp. 11, 47).

[DEM01] Eleni Drinea, Mihaela Enachescu, and Michael D. Mitzenmacher. *Variations on random graph models for the web*. Tech. rep. Cambridge, Massachusetts: Harvard University, 2001, pp. 1–3 (cit. on pp. 11, 47).

[EMB02] Holger Ebel, Lutz-Ingo Mielsch, and Stefan Bornholdt. "Scale-free topology of e-mail networks." In: *Physical Review E* 66.3 (Sept. 2002), p. 035103. DOI: 10.1103/PhysRevE.66.035103. arXiv: 0201476 [cond-mat] (cit. on p. 41).

[ER59] Paul Erdős and Alfréd Rényi. "On random graphs, I." In: *Publicationes Mathematicae (Debrecen)* 6 (1959), pp. 290–297 (cit. on pp. 37, 38, 53).

[ER60] Paul Erdős and Alfréd Rényi. "On the evolution of random graphs." In: *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5 (1960), pp. 17–61 (cit. on pp. 4, 7, 37–39, 41, 53).

[Est16] Jean-Baptiste Estoup. "Les Gammes Sténographiques." In: *Institut Sténographique de France* (1916) (cit. on pp. 9, 41).

[Eub+04a] Stephen Eubank, Hasan Guclu, V. S. Anil Kumar, Madhav V. Marathe, Aravind Srinivasan, Zoltán Toroczkai, and Nan Wang. "Modelling disease outbreaks in realistic urban social networks." In: *Nature* 429.6988 (May 2004), pp. 180–184. DOI: 10.1038/nature02541 (cit. on pp. 13, 132, 135).

[Eub+04b] Stephen Eubank, V. S. Anil Kumar, Madhav V. Marathe, Aravind Srinivasan, and Nan Wang. "Structural and algorithmic aspects of massive social networks." In: *Proceedings of the 15th annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '04. SIAM, 2004, pp. 718–727. DOI: 10.1145/982792.982902 (cit. on pp. 9, 11, 12, 41, 132, 174).

[Eul41] Leonard Euler. "Solutio problematis ad geometriam situs pertinentis." In: *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 8 (1741), pp. 128–140 (cit. on p. 30).

[FFF99] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. "On power-law relationships of the Internet topology." In: *ACM SIGCOMM Computer Communication Review* 29.4 (Oct. 1999), pp. 251–262. DOI: 10.1145/316194.316229 (cit. on pp. 9, 41).

[Fei98] Uriel Feige. "A threshold of ln n for approximating set cover." In: *Journal of the ACM* 45.4 (July 1998), pp. 634–652. DOI: 10.1145/285055.285059 (cit. on pp. 16, 17, 69, 132, 134–136, 140, 143, 144, 171, 196).

[Fel71] William Feller. *An Introduction to Probability Theory and Its Applications*. 3rd Ed. Vol. 2. Wiley, 1971, p. 669 (cit. on p. 53).

[FPP08] Alessandro Ferrante, Gopal Pandurangan, and Kihong Park. "On the hardness of optimization in power-law graphs." In: *Theoretical Computer Science* 393.1-3 (Mar. 2008), pp. 220–230. DOI: 10.1016/j.tcs.2007.12.007 (cit. on pp. 17, 87, 91, 92, 135, 196).

[For10] Santo Fortunato. "Community detection in graphs." In: *Physics Reports* 486.3-5 (Feb. 2010), pp. 75–174. DOI: 10.1016/j.physrep.2009.11.002 (cit. on p. 13).

[GH12] Mikael Gast and Mathias Hauptmann. "Approximability of the vertex cover problem in power law graphs." In: *Computing Research Repository (CoRR) preprint arXiv:1204.0982 [cs.DS]; also submitted to Theoretical Computer Science* (2012), pp. 1–16. arXiv: 1204.0982 (cit. on p. 194).

[GHK12a] Mikael Gast, Mathias Hauptmann, and Marek Karpinski. "Improved approximation lower bounds for vertex cover on power law graphs and some generalizations." In: *Computing Research Repository (CoRR) preprint arXiv:1210.2698 [cs.CC]* (Oct. 2012), pp. 1–26. arXiv: 1210.2698 (cit. on p. 130).

[GHK12b] Mikael Gast, Mathias Hauptmann, and Marek Karpinski. "Inapproximability of dominating set in power law graphs." In: *Computing Research Repository (CoRR) preprint arXiv:1212.3517 [cs.CC]; also submitted to Theoretical Computer Science* (Dec. 2012), pp. 1–23. arXiv: 1212.3517 (cit. on pp. 171, 172).

[Gil59] Edgar N. Gilbert. "Random Graphs." In: *The Annals of Mathematical Statistics* 30.4 (Dec. 1959), pp. 1141–1144. DOI: 10.1214/aoms/1177706098 (cit. on pp. 4, 7, 37, 41).

[GN02] Michelle Girvan and Mark E. J. Newman. "Community structure in social and biological networks." In: *Proceedings of the National Academy of Sciences of the United States of America* 99.12 (June 2002), pp. 7821–6. DOI: 10.1073/pnas.122653799 (cit. on p. 13).

[GMS03] Christos Gkantsidis, Milena Mihail, and Amin Saberi. "Conductance and congestion in power law graphs." In: *ACM SIGMETRICS Performance Evaluation Review* 31.1 (June 2003), pp. 148–159. DOI: 10.1145/885651.781046 (cit. on p. 174).

[Gol06] Oded Goldreich. "On promise problems: a survey." In: *Theoretical Computer Science* 3895 (Jan. 2006), pp. 254–290. DOI: 10.1007/11685654_12 (cit. on p. 64).

[Gol08] Oded Goldreich. *Computational complexity - a conceptual perspective*. Cambridge University Press, 2008, pp. I–XXIV, 1–606 (cit. on pp. 64, 66, 71).

[GS98] Charles M. Grinstead and J. Laurie Snell. *Introduction to probability*. 2nd Ed. American Mathematical Society, 1998 (cit. on p. 53).

[Gue+02] Nabil Guelzim, Samuele Bottani, Paul Bourgine, and François Képès. "Topological and causal structure of the yeast transcriptional regulatory network." In: *Nature Genetics* 31.1 (May 2002), pp. 60–63. DOI: 10.1038/ng873 (cit. on pp. 9, 41).

[Gut54] Francis Guthrie. "Tinting Maps." In: *The Athenæum* 10 (1854), p. 726 (cit. on p. 31).

[Gut80] Frederick Guthrie. "Note on the colouring of maps." In: *Proceedings of the Royal Society of Edinburgh* 10 (1880), pp. 727–728 (cit. on p. 31).

[Hac67] Alice P. Hackett. *70 years of best sellers, 1895-1965*. New York, NY: R. R. Bowker Co., 1967, p. 280 (cit. on p. 9).

[HR94] Magnús Halldórsson and Jaikumar Radhakrishnan. "Greed is good: approximating independent sets in sparse and bounded-degree graphs." In: *Proceedings of the 26th annual ACM Symposium on Theory of Computing*. STOC '94. ACM Press, May 1994, pp. 439–448. DOI: 10.1145/195058.195221 (cit. on p. 86).

[Hal02] Eran Halperin. "Improved approximation algorithms for the Vertex Cover Problem in graphs and hypergraphs." In: *SIAM Journal on Computing* 31.5 (Jan. 2002), pp. 1608–1623. DOI: 10.1137/S0097539700381097 (cit. on p. 86).

[Hås99] Johan Håstad. "Clique is hard to approximate within n to the power 1-ϵ." In: *Acta Mathematica* 182.1 (Mar. 1999), pp. 105–142. DOI: 10.1007/BF02392825 (cit. on p. 86).

[Hås01] Johan Håstad. "Some optimal inapproximability results." In: *Journal of the ACM* 48.4 (July 2001), pp. 798–859. DOI: 10.1145/502090.502098 (cit. on pp. 75, 86).

[Hoc83] Dorit S. Hochbaum. "Efficient bounds for the stable set, vertex cover and set packing problems." In: *Discrete Applied Mathematics* 6.3 (Sept. 1983), pp. 243–254. DOI: 10.1016/0166-218X(83)90080-X (cit. on p. 86).

[Hoc+93] Dorit S. Hochbaum, Nimrod Megiddo, Joseph Naor, and Arie Tamir. "Tight bounds and 2-approximation algorithms for integer programs with two variables per inequality." In: *Mathematical Programming* 62.1-3 (Feb. 1993), pp. 69–83. DOI: 10.1007/BF01585160 (cit. on pp. 178, 194).

[Jeo+01] Hawoong Jeong, Sean P. Mason, Albert-László Barabási, and Zoltán N. Oltvai. "Lethality and centrality in protein networks." In: *Nature* 411.6833 (May 2001), pp. 41–2. DOI: 10.1038/35075138 (cit. on p. 9).

[Joh74] David S. Johnson. "Approximation algorithms for combinatorial problems." In: *Journal of Computer and System Sciences* 9.3 (Dec. 1974), pp. 256–278. DOI: 10.1016/S0022-0000(74)80044-9 (cit. on p. 133).

[JAB01] Mihajlo Jovanović, Fred S. Annexstein, and Kenneth A. Berman. "Modeling peer-to-peer network topologies through "small-world" models and power laws." In: *IX Telecommunications Forum, TELFOR*. 2001, pp. 1–4 (cit. on pp. 9, 41).

[Kan92] Viggo Kann. "On the approximability of NP-complete optimization problems." PhD thesis. Royal Institute of Technology, Stockholm, 1992 (cit. on pp. 133, 139).

[Kar09] George Karakostas. "A better approximation ratio for the vertex cover problem." In: *ACM Transactions on Algorithms* 5.4 (Oct. 2009), pp. 1–8. DOI: 10.1145/1597036.1597045 (cit. on pp. 18, 85, 86, 197).

[Kar72] Richard M. Karp. "Reducibility among combinatorial problems." In: *Complexity of Computer Computations* (Dec. 1972). Ed. by R E Miller and J W Thatcher, p. 85. DOI: 10.2307/2271828 (cit. on pp. 12, 84, 85, 133, 174, 177).

[Kho02] Subhash Khot. "On the power of unique 2-prover 1-round games." In: *Proceedings of the 34th annual ACM Symposium on Theory of Computing*. STOC '02. New York, New York, USA: ACM Press, May 2002, pp. 767–775. DOI: 10.1145/509907.510017 (cit. on p. 86).

[KP06] Subhash Khot and Ashok Kumar Ponnuswami. "Better inapproximability results for maxclique, chromatic number and min-3lin-deletion." In: *Proceedings of the 33rd International Conference on Automata, Languages and Programming*. ICALP '06. Springer-Verlag, July 2006, pp. 226–237. DOI: 10.1007/11786986 (cit. on p. 86).

[KR02] Subhash Khot and Venkatesh Raman. "Parameterized complexity of finding subgraphs with hereditary properties." In: *Theoretical Computer Science* 289.2 (Oct. 2002), pp. 997–1008. DOI: 10.1016/S0304-3975(01)00414-5 (cit. on p. 18).

[KR08] Subhash Khot and Oded Regev. "Vertex cover might be hard to approximate to within 2-ϵ." In: *Journal of Computer and System Sciences* 74.3 (May 2008), pp. 335–349. DOI: 10.1016/j.jcss.2007.06.019 (cit. on pp. 13, 18, 86, 174).

[Kir47] Gustav R. Kirchhoff. "Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird." In: *Annalen der Physik und Chemie* 148.12 (1847), pp. 497–508. DOI: 10.1002/andp.18471481202 (cit. on p. 31).

[Kit+10] Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse. "Identification of influential spreaders in complex networks." In: *Nature Physics* 6.11 (Aug. 2010), pp. 888–893. DOI: 10.1038/nphys1746 (cit. on p. 13).

[Kle+99] Jon M. Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S. Tomkins. "The Web as a graph: measurements, models and methods." In: *Proceedings of the 5th annual international Conference on Computing and Combinatorics*. COCOON '99. Springer-Verlag, 1999, pp. 1–17. DOI: 10.1007/3-540-48686-0_1 (cit. on pp. 9, 41).

[KL01] Jon M. Kleinberg and Steve Lawrence. "The structure of the Web." In: *Science* 294.5548 (Nov. 2001), pp. 1849–1850. DOI: 10.1126/science.1067014 (cit. on pp. 9, 41).

[Knu97] Donald E. Knuth. *The Art of Computer Programming, Volume 1: Fundamental Algorithms*. 3rd Ed. Addison-Wesley, 1997, pp. I–XX, 1–650 (cit. on p. 53).

[Kőn16] Dénes Kőnig. "Gráfok és alkalmazásuk a determinánsok és a halmazok elméletére." In: *Matematikai és Természettudományi Értesítő* 34 (1916), pp. 104–119 (cit. on p. 178).

[Kőn31] Dénes Kőnig. "Gráfok és mátrixok." In: *Matematikai és Fizikai Lapok* 38 (1931), pp. 116–119 (cit. on p. 178).

[KSG06] Mehmet Koyutürk, Wojciech Szpankowski, and Ananth Grama. "Assessing significance of connectivity and conservation in protein interaction networks." In: *Journal of Computational Biology* 14.6 (2006), pp. 747–764. DOI: 10.1089/cmb.2007.R014 (cit. on p. 174).

[Kum+00] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, D. Sivakumar, Andrew Tomkins, and Eli Upfal. "Stochastic models for the Web graph." In: *Proceedings of the 41st annual Symposium on Foundations of Computer Science*. FOCS '00. IEEE Comput. Soc. Press, 2000, pp. 57–65. DOI: 10.1109/SFCS.2000.892065 (cit. on pp. 9, 11, 41, 47).

[LF12] Andrea Lancichinetti and Santo Fortunato. "Consensus clustering in complex networks." en. In: *Scientific reports* 2 (Jan. 2012), p. 336. DOI: 10.1038/srep00336 (cit. on p. 13).

[Lan09] Edmund Landau. *Handbuch der Lehre von der Verteilung der Primzahlen*. Vol. 2. Leipzig: B. G. Teubner, 1909, pp. 565–961 (cit. on p. 22).

[Li+05] Lun Li, David L. Alderson, John C. Doyle, and Walter Willinger. "Towards a theory of scale-free graphs: Definition, properties, and implications." In: *Internet Mathematics* 2.4 (2005), pp. 431–523. DOI: 10.1080/15427951.2005.10129111 (cit. on p. 53).

[Li+12] Rong-Hua Li, Jeffrey Xu Yu, Xin Huang, Hong Cheng, and Zechao Shang. "Measuring robustness of complex networks under MVC attack." In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. CIKM '12. ACM Press, Oct. 2012, pp. 1512–1516. DOI: 10.1145/2396761.2398463 (cit. on p. 15).

[Lil+01] Fredrik Liljeros, Christofer R. Edling, Luís A. Nunes Amaral, H. Eugene Stanley, and Yvonne Aberg. "The web of human sexual contacts." In: *Nature* 411.6840 (July 2001), pp. 907–908. DOI: 10.1038/35082140 (cit. on p. 9).

[Lot26] Alfred J. Lotka. "The frequency distribution of scientific productivity." In: *Journal of Washington Academy Sciences* (1926) (cit. on pp. 9, 41).

[Lov75] László Lovász. "On the ratio of optimal integral and fractional covers." In: *Discrete Mathematics* 13.4 (Jan. 1975), pp. 383–390. DOI: 10.1016/0012-365X(75)90058-8 (cit. on p. 134).

[Lov09] László Lovász. "Very large graphs." In: *Current Developments in Mathematics* 2008 (2009), pp. 67–128 (cit. on p. 5).

[Lu01] Linyuan Lu. "The diameter of random massive graphs." In: *Proceedings of the 12th annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '01. SIAM, Jan. 2001, pp. 912–921 (cit. on pp. 52, 53).

[LY94] Carsten Lund and Mihalis Yannakakis. "On the hardness of approximating minimization problems." In: *Journal of the ACM* 41.5 (Sept. 1994), pp. 960–981. DOI: 10.1145/185675.306789 (cit. on p. 134).

[MPS06] Milena Mihail, Christos H. Papadimitriou, and Amin Saberi. "On certain connectivity properties of the internet topology." In: *Journal of Computer and System Sciences* 72.2 (Mar. 2006), pp. 239–251. DOI: 10.1016/j.jcss.2005.06.009 (cit. on p. 11).

[Mil+11] Tijana Milenković, Vesna Memišević, Anthony Bonato, and Nataša Pržulj. "Dominating biological networks." In: *PloS one* 6.8 (Jan. 2011), e23016. DOI: 10.1371/journal.pone.0023016 (cit. on p. 14).

[Mil67] Stanley Milgram. "The small world problem." In: *Psychology today* 2.1 (1967), pp. 60–67 (cit. on p. 6).

[Mit03] Michael Mitzenmacher. "A brief history of generative models for power law and lognormal distributions." In: *Internet Mathematics* 1.2 (Apr. 2003), pp. 226–251 (cit. on p. 53).

[MR95] Michael Molloy and Bruce Reed. "A critical point for random graphs with a given degree sequence." In: *Random Structures and Algorithms* 6.2-3 (Mar. 1995), pp. 161–180. DOI: 10.1002/rsa.3240060204 (cit. on p. 51).

[MR98] Michael Molloy and Bruce Reed. "The Size of the Giant Component of a Random Graph with a Given Degree Sequence." In: *Combinatorics, Probability and Computing* 7.3 (Sept. 1998), pp. 295–305. DOI: 10.1017/S0963548398003526 (cit. on p. 51).

[MS85] Burkhard Monien and Ewald Speckenmeyer. "Ramsey numbers and an approximation algorithm for the vertex cover problem." In: *Acta Informatica* 22.1 (Apr. 1985), pp. 115–123. DOI: 10.1007/BF00290149 (cit. on p. 86).

[MS00] Jose M. Montoya and Ricard V. Sole. "Small world patterns in food webs." In: *Preprint arXiv:0011195 [cond-mat]* (Nov. 2000), p. 6. arXiv: 0011195 [cond-mat] (cit. on p. 9).

[NSS95] Moni Naor, Leonard J. Schulman, and Aravind Srinivasan. "Splitters and near-optimal derandomization." In: *Proceedings of the 36th annual IEEE Symposium on Foundations of Computer Science*. FOCS '95. IEEE Comput. Soc. Press, Oct. 1995, pp. 182–191. DOI: 10.1109/SFCS.1995.492475 (cit. on p. 134).

[NT74] George L. Nemhauser and Leslie E. Trotter. "Properties of vertex packing and independence system polyhedra." In: *Mathematical Programming* 6.1 (Dec. 1974), pp. 48–61. DOI: 10.1007/BF01580222 (cit. on pp. 176, 177).

[NT75] George L. Nemhauser and Leslie E. Trotter. "Vertex packings: Structural properties and algorithms." In: *Mathematical Programming* 8.1 (Dec. 1975), pp. 232–248. DOI: 10.1007/BF01580444 (cit. on pp. 18, 177, 178, 194).

[New01a] Mark E. J. Newman. "Scientific collaboration networks. I. Network construction and fundamental results." In: *Physical Review E* 64.1 (June 2001), p. 8. DOI: 10.1103/PhysRevE.64.016131 (cit. on p. 7).

[New01b] Mark E. J. Newman. "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality." In: *Physical Review E* 64.1 (June 2001), p. 7. DOI: 10.1103/PhysRevE.64.016132 (cit. on p. 7).

[New03] Mark E. J. Newman. "The Structure and Function of Complex Networks." In: *SIAM Review* 45.2 (Jan. 2003), pp. 167–256 (cit. on p. 5).

[New05] Mark E. J. Newman. "Power laws, Pareto distributions and Zipf's law." In: *Contemporary Physics* 46.5 (Sept. 2005), pp. 323–351. DOI: 10.1080/00107510500052444 (cit. on p. 53).

[NW06] Mark E. J. Newman and Duncan J. Watts. *The structure and dynamics of networks*. Princeton University Press, 2006 (cit. on pp. 5, 9).

[NSH04] Vera van Noort, Berend Snel, and Martijn A. Huynen. "The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model." In: *EMBO reports* 5.3 (Mar. 2004), pp. 280–284. DOI: 10.1038/sj.embor.7400090 (cit. on p. 7).

[Pap94] Christos H. Papadimitriou. *Computational complexity*. 1st Editio. Addison-Wesley, 1994 (cit. on p. 76).

[PY91] Christos H. Papadimitriou and Mihalis Yannakakis. "Optimization, approximation, and complexity classes." In: *Journal of Computer and System Sciences* 43.3 (Dec. 1991), pp. 425–440. DOI: 10.1016/0022-0000(91)90023-X (cit. on pp. 12, 60, 64, 76, 82, 84, 85, 174).

[PY88] Christos Papadimitriou and Mihalis Yannakakis. "Optimization, approximation, and complexity classes." In: *Proceedings of the 20th annual ACM Symposium on Theory of Computing*. STOC '88. New York, New York, USA: ACM Press, Jan. 1988, pp. 229–234. DOI: 10.1145/62212.62233 (cit. on p. 64).

[Par96] Vilfredo F. Pareto. *Cours d'economie politique*. Librairie Droz, 1896 (cit. on pp. 9, 41, 43).

[PL01] Kihong Park and Heejo Lee. "On the effectiveness of route-based packet filtering for distributed DoS attack prevention in power-law internets." In: *ACM SIGCOMM Computer Communication Review* 31.4 (Oct. 2001), pp. 15–26. DOI: 10.1145/964723.383061 (cit. on pp. 12, 84, 174).

[PV01] Romualdo Pastor-Satorras and Alessandro Vespignani. "Epidemic Spreading in Scale-Free Networks." In: *Physical Review Letters* 86.14 (Apr. 2001), pp. 3200–3203. DOI: 10.1103/PhysRevLett.86.3200 (cit. on pp. 13, 132).

[PV07] Romualdo Pastor-Satorras and Alessandro Vespignani. *Evolution and Structure of the Internet: A Statistical Physics Approach*. New York: Cambridge University Press, 2007, p. 284 (cit. on p. 5).

[PM81] Azaria Paz and Shlomo Moran. "Non deterministic polynomial optimization problems and their approximations." In: *Theoretical Computer Science* 15.3 (Jan. 1981), pp. 251–277. DOI: 10.1016/0304-3975(81)90081-5 (cit. on p. 139).

[Pri65] Derek John de Solla Price. "Networks of scientific papers." In: *Science* 149.3683 (July 1965), pp. 510–515. DOI: 10.1126/science.149.3683.510 (cit. on pp. 11, 43).

[Pri76] Derek John de Solla Price. "A general theory of bibliometric and other cumulative advantage processes." In: *Journal of the American Society for Information Science* 27.5 (Sept. 1976), pp. 292–306. DOI: 10.1002/asi.4630270505 (cit. on pp. 11, 43).

[RS07] Jaikumar Radhakrishnan and Madhu Sudan. "On Dinur's proof of the PCP theorem." In: *Bulletin of the American Mathematical Society* 44.01 (Jan. 2007), pp. 19–62. DOI: 10.1090/S0273-0979-06-01143-8 (cit. on p. 71).

[RTP04] Emad Ramadan, Arijit Tarafdar, and Alex Pothen. "A hypergraph model for the yeast protein complex network." In: *Proceedings of the 18th International Parallel and Distributed Processing Symposium*. IPDPS '04. IEEE, 2004, pp. 189–196. DOI: 10.1109/IPDPS.2004.1303205 (cit. on p. 14).

[Raz95] Ran Raz. "A parallel repetition theorem." In: *Proceedings of the 27th annual ACM Symposium on Theory of Computing*. STOC '95. New York, New York, USA: ACM Press, May 1995, pp. 447–456. DOI: 10.1145/225058.225181 (cit. on p. 134).

[Raz98] Ran Raz. "A Parallel Repetition Theorem." en. In: *SIAM Journal on Computing* 27.3 (June 1998), pp. 763–803. DOI: 10.1137/S0097539795280895 (cit. on p. 134).

[RS97] Ran Raz and Shmuel Safra. "A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP." In: *Proceedings of the 29th annual ACM Symposium on Theory of Computing*. STOC '97. ACM Press, May 1997, pp. 475–484. DOI: 10.1145/258533.258641 (cit. on pp. 13, 135).

[RSG99] Chantal Roth-Korostensky, Ulrike Stege, and Gaston H. Gonnet. *Algorithms for improving multiple sequence alignments and building evolutionary trees using insertions and deletion events*. Tech. rep. Swiss Federal Institute of Technology, Zürich, 1999, p. 14 (cit. on p. 14).

[Ses+08] Mukund Seshadri, Sridhar Machiraju, Ashwin Sridharan, Jean Bolot, Christos Faloutsos, and Jure Leskovec. "Mobile call graphs." In: *Proceedings of the 14th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining*. KDD '08. ACM Press, Aug. 2008, pp. 596–604. DOI: 10.1145/1401890.1401963 (cit. on pp. 9, 41).

[Sha92] Adi Shamir. "IP = PSPACE." In: *Journal of the ACM* 39.4 (Oct. 1992), pp. 869–877. DOI: 10.1145/146585.146609 (cit. on pp. 68, 71).

[SNT10] Yilin Shen, Dung T. Nguyen, and My T. Thai. "On the hardness and inapproximability of optimization problems on power law graphs." In: *Proceedings of the 4th international Conference on Combinatorial Optimization and Applications*. COCOA '10. Springer-Verlag, 2010, pp. 197–211. DOI: 10.1007/978-3-642-17458-2_17 (cit. on pp. 87, 92, 135).

[She+12] Yilin Shen, Dung T. Nguyen, Ying Xuan, and My T. Thai. "New techniques for approximating optimal substructure problems in power-law graphs." In: *Theoretical Computer Science* 447 (Aug. 2012), pp. 107–119. DOI:

10.1016/j.tcs.2011.10.023 (cit. on pp. 17, 87, 88, 92–96, 132, 135, 136, 139, 157, 161–164, 171, 196).

[Sig+03] Georgos Siganos, Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. "Power laws and the AS-level Internet topology." In: *IEEE/ACM Transactions on Networking* 11.4 (Aug. 2003), pp. 514–524. DOI: 10.1109/TNET.2003.815300 (cit. on pp. 9, 41).

[SR11] Mikhail V. Simkin and Vwani P. Roychowdhury. "Re-inventing Willis." In: *Physics Reports* 502.1 (May 2011), pp. 1–35. DOI: 16/j.physrep.2010.12.004 (cit. on p. 53).

[Sim55] Herbert A. Simon. "On a class of skew distribution functions." In: *Biometrika* 42.3/4 (Dec. 1955), pp. 425–440. DOI: 10.2307/2333389 (cit. on pp. 11, 43).

[Sip97] Michael Sipser. *Introduction to the theory of computation*. PWS Publishing Company, 1997, pp. I–XV, 1–396 (cit. on p. 53).

[Sip12] Michael Sipser. *Introduction to the Theory of Computation*. Cengage Learning, 2012, p. 480 (cit. on p. 64).

[Sla97] Petr Slavík. "A Tight Analysis of the Greedy Algorithm for Set Cover." In: *Journal of Algorithms* 25.2 (Nov. 1997), pp. 237–254. DOI: 10.1006/jagm.1997.0887 (cit. on p. 134).

[Sri95] Aravind Srinivasan. "Improved approximations of packing and covering problems." In: *Proceedings of the 27th annual ACM Symposium on Theory of Computing*. STOC '95. New York, New York, USA: ACM Press, May 1995, pp. 268–276. DOI: 10.1145/225058.225138 (cit. on p. 134).

[SSŽ02] Ivan Stojmenovic, Mahtab Seddigh, and Joviša Žunic. "Dominating sets and neighbor elimination-based broadcasting algorithms in wireless networks." In: *IEEE Transactions on Parallel and Distributed Systems* 13.1 (2002), pp. 14–25. DOI: 10.1109/71.980024 (cit. on pp. 14, 132).

[TM69] Jeffrey Travers and Stanley Milgram. "An Experimental Study of the Small World Problem." In: *Sociometry* 32.4 (Dec. 1969), p. 425. DOI: 10.2307/2786545 (cit. on p. 6).

[Wan+11] Dashun Wang, Zhen Wen, Hanghang Tong, Ching-Yung Lin, Chaoming Song, and Albert-László Barabási. "Information spreading in context." In: *Proceedings of the 20th international Conference on World Wide Web*. WWW '11. New York, New York, USA: ACM Press, Mar. 2011, pp. 735–744. DOI: 10.1145/1963405.1963508 (cit. on p. 13).

[Wat03] Duncan J. Watts. *Small worlds: the dynamics of networks between order and randomness*. Princeton University Press, 2003 (cit. on p. 7).

[WS98] Duncan J. Watts and Steven H. Strogatz. "Collective dynamics of 'small-world' networks." In: *Nature* 393.6684 (June 1998), pp. 440–442. DOI: 10.1038/30918 (cit. on pp. 7, 34, 35).

[WGS01] Jie Wu, Ming Gao, and Ivan Stojmenovic. "On calculating power-aware connected dominating sets for efficient routing in ad hoc wireless networks." In: *Proceedings of the International Conference on Parallel Processing*. ICPP '01. IEEE, 2001, pp. 346–354. DOI: 10.1109/ICPP.2001.952080 (cit. on p. 14).

[WL99] Jie Wu and Hailan Li. "On calculating connected dominating set for efficient routing in ad hoc wireless networks." In: *Proceedings of the 3rd international Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*. DIALM '99. ACM Press, Aug. 1999, pp. 7–14. DOI: 10.1145/313239.313261 (cit. on p. 14).

[Yul25] G. Udny Yule. "A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S." In: *Philosophical Transactions of the Royal Society of London* 213.Series B (1925), pp. 21–87 (cit. on pp. 11, 43).

[Zip35] George K. Zipf. *The psycho-biology of language*. Houghton Mifflin Company, 1935, p. 336 (cit. on pp. 9, 42).

[Zip49] George K. Zipf. *Human behavior and the principle of least effort: an introduction to human ecology*. Addison-Wesley Press, 1949, p. 573 (cit. on pp. 9, 42).

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

LIST OF ACRONYMS

a.s.	almost surely
i.i.d.	independent and identically distributed
Min-DS	Minimum Dominating Set
Max-IS	Maximum Independent Set
Min-VC	Minimum Vertex Cover
PLG	power law graph
w.h.p.	with high probability
CCDF	complementary cumulative distribution function
CDF	cumulative distribution function
LCD	linearized chord diagram
PTM	probabilistic Turing Machine
TM	Turing Machine

GLOSSARY

APX	the class of all NP optimization problems that admit a polynomial time approximation algorithm with constant approximation ratio
$C(G)$	the global clustering coefficient of a graph G
$C_v(G)$	the local clustering coefficient of a vertex v in G
$d(G)$	the average degree of a graph G
$d(v)$	the degree of a vertex v
$\Delta(G)$	the maximum degree of a graph G
$\delta(G)$	the minimum degree of a graph G
DSPACE	the class of deterministic space complexity
DTIME	the class of deterministic time complexity
$E(G)$	the set of edges of a graph G
$G = (V, E)$	a graph consisting of vertices and edges
$\mathbb{N}$	the set of natural numbers $\mathbb{N}_0 \setminus \{0\}$
$\mathbb{N}_0$	the set of natural numbers $\{0, 1, 2, \ldots\}$
$N(v)$	the set of neighboring vertices of a vertex v

NP	the class of decision problems Π for which a non-deterministic polynomial time Turing Machine exists
NPO	the class of all NP optimization problems
NSPACE	the class of non-deterministic space complexity
NTIME	the class of non-deterministic time complexity
P	the class of decision problems Π that are decidable in deterministic polynomial time
Π	a decision or optimization problem
$\overline{\Pi}$	the set of no-instances of a decision problem Π
PO	the class of all P optimization problems
PTAS	the class of all NP optimization problems that admit a polynomial time approximation scheme
$\mathbb{Q}$	the set of rational numbers
$\mathbb{Q}_+$	the set of positive rational numbers
$\mathbb{R}$	the set of real numbers
$\mathbb{R}_+$	the set of positive real numbers
Σ	a finite alphabet
$V(G)$	the set of vertices of a graph G
$\mathbb{Z}$	the set of integers $\{0, 1, -1, 2, -2, \ldots\}$
ZPP	the class of decision problems Π for which a non-deterministic probabilistic Turing Machine exists

ZTIME	the class of non-deterministic probabilistic time complexity

INDEX