On the Approximability of Independent Set Problem on Power Law Graphs

Mathias Hauptmann[*] Marek Karpinski[†]

Abstract

We give the first nonconstant lower bounds for the approximability of the Independent Set Problem on the Power Law Graphs. These bounds are of the form n^ϵ in the case when the power law exponent satisfies $\beta < 1$. In the case when $\beta = 1$, the lower bound is of the form $\log(n)^\epsilon$. The embedding technique used in the proof could also be of independent interest.

1 Introduction

In this paper we prove new inapproximability results for the Maximum Independent Set (MIS) problem on Power Law Graphs.

The Independent Set Problem on general graph instances is equivalent to the Max Clique Problem. In [13] it was shown that the Max Clique Problem is NP-hard to approximate within a factor n^c for some constant $c > 0$. A randomized reduction from PCPs to Max Clique has been constructed in [16]. This implies that if NP $\neq$ ZPP, then Max Clique cannot be approximated within $n^{1-\epsilon}$ for any $\epsilon > 0$. Khot [19] showed that under the Unique Games Conjecture (UGC), there exists some $\gamma > 0$ such that Max Clique cannot be approximated within $n/2^{(\log n)^{1-\gamma}}$. These results have been derandomized in [22]. This implies that unless P = NP, Max Clique cannot be approximated within $n^{1-\epsilon}$ for any $\epsilon > 0$. Furthermore, there exists some $\gamma > 0$ such that, unless $\tilde{\mathrm{P}} = \tilde{\mathrm{NP}}$, no quasi-polynomial time algorithm approximates Max Clique within $n/2^{(\log n)^{1-\gamma}}$.

It has been discovered recently that many real-world large scale networks have a node degree distribution which follows a power law. This has been observed for

---

[*] Dept. of Computer Science, University of Bonn. e-mail: `hauptman@cs.uni-bonn.de`

[†] Dept. of Computer Science, University of Bonn and Hausdorff Center for Mathematics, University of Bonn. e-mail: `marek@cs.uni-bonn.de`

the graphs of the Internet [12] and the World Wide Web (WWW) [4], peer-to-peer networks, gene regulatory networks and protein interaction networks [8] and for social networks ([20], [18]).

A random graph model for power law graphs has been introduced in [1],[2]. Power law graphs have the property that their node degree distribution follows a power law, i.e. the number of nodes of degree i is proportional to $i^{-\beta}$, for some fixed $\beta > 0$. This parameter β is called the power law exponent.

It has been observed experimentally that some optimization problems tend to be much easier to solve on power law graphs than on general graph instances [11]. This rises the question on whether one can show differences in terms of approximability or approximation hardness of several optimization problems between general graph instances and power law graphs [21]. Ferrante et al. [14] showed the NP-hardness of Max Clique in power law graphs for $\beta \geq 1$ and of MIS in power law graphs for all $\beta > 0$. Shen et al. [21] proved the APX-hardness of Max Clique and Maximum Independent Set on power law graphs, also for $\beta > 1$. Their result is based on an efficient embedding of bounded degree graphs into power law graphs.

We consider the Maximum Independent Set (MIS) problem in Power Law Graphs (PLGs) for $\beta \leq 1$. For $\beta \in (0, 1)$ we show that the MIS problem in (α, β)-PLGs is hard to approximate within n^{ϵ}, for some $\epsilon > 0$ being constant. For $\beta = 1$, we give a lower bound of Δ^{ϵ_2} for some $\epsilon_2 > 0$, where $\Delta = e^{\alpha}$ is the maximum degree of a power law graph.

The paper is organized as follows. In Section 2 we give the formal definition of a power law graph according to [1]. Section 3 provides an important tool which we will use in our constructions, namely a method how to complete constructively fragments of a given power law node degree distribution such that the maximum independent set size of the resulting graph is small. In Section 4 we will use this auxiliary construction to obtain our hardness result for the case when $\beta < 1$. The case $\beta = 1$ is considered in Section 5. In Section 6 we give a summary and a number of open questions.

2 Power Law Graphs

In this section we start with giving the formal definition of the (α, β) Power-Law Graphs (cf. [1],[2]). We introduce also the notion of a node degree interval, which will be of particular importance within our constructions. Informally, a *node degree interval* (or just *interval* for short) is a subset of nodes in a graph whose node degrees are all within a given interval $[a, b]$. In the subsequent sections we will be concerned with the construction of embeddings of a given graph G into some node degree interval of a power law graph. In this section we provide estimates for sizes and volumes (i.e. sums of node degrees) of intervals in (α, β) Power-Law Graphs

(PLGs).

Definition 2.1 *[3] An undirected multigraph $G = (V, E)$ with self loops is called an (α, β) Power-Law Graph if the following conditions hold:*

- *The maximum degree is $\Delta = \lfloor e^{\alpha/\beta} \rfloor$.*
- *For $i = 1, \ldots, \Delta$, the number y_i of nodes of degree i in G satisfies*

$$y_i \;=\; \left\lfloor \frac{e^{\alpha}}{i^{\beta}} \right\rfloor$$

The following estimates for the number n of vertices of an (α, β) Power-Law Graph are well known (cf. [3]):

$$n \;\approx\; \begin{cases} \frac{e^{\alpha/\beta}}{1-\beta} & \text{for } 0 < \beta < 1, \\ \alpha \cdot e^{\alpha} & \text{for } \beta = 1, \\ \zeta(\beta) \cdot e^{\alpha} & \text{for } \beta > 1. \end{cases} \qquad m \;\approx\; \begin{cases} \frac{1}{2}\frac{e^{2\alpha/\beta}}{2-\beta} & \text{for } 0 < \beta < 2, \\ \frac{1}{4}\alpha e^{\alpha} & \text{for } \beta = 2, \\ \frac{1}{2}\zeta(\beta - 1)e^{\alpha} & \text{for } \beta > 2. \end{cases}$$

Here $\zeta(\beta) = \sum_{i=1}^{\infty} i^{-\beta}$ is the *Riemann Zeta Function.*

Given an (α, β) Power-Law Graph $G = (V, E)$ with n vertices and maximum degree Δ and two integers $1 \leq a \leq b \leq \Delta$, an *interval* $[a, b]$ is defined as the subset of V

$$[a, b] \;=\; \{v \in V | a \leq \deg_G(v) \leq b\} \tag{1}$$

If $U \subseteq V$ is a subset of vertices, the *volume* $\mathrm{vol}(U)$ of U is defined as the sum of node degrees of nodes in U. We have the following estimates for sizes and volumes of node intervals in (α, β)-PLGs (cf. [15]).

Theorem 2.1 *[15] For $0 < \beta < 1$, we have*

$$\begin{aligned} |[x\Delta, y\Delta]| &\in \left[\tfrac{\Delta}{1-\beta}(y^{1-\beta} - x^{1-\beta}) - \left(\tfrac{1}{x^{\beta}} - \tfrac{1}{y^{\beta}}\right),\ \tfrac{\Delta}{1-\beta}(y^{1-\beta} - x^{1-\beta})\right] \\ |[x\Delta, \Delta]| &\in \left[\tfrac{\Delta}{1-\beta}(1 - x^{1-\beta}) - \left(\tfrac{1}{x^{\beta}} - 1\right),\ \tfrac{\Delta}{1-\beta}(1 - x^{1-\beta})\right] \\ vol([x\Delta, \Delta]) &\geq \Delta^2\left(\tfrac{1-x^{2-\beta}}{2-\beta} - \tfrac{1}{2} + \tfrac{x^2}{2}\right) - \Delta\left(1 - x^{1-\beta} - \tfrac{1}{2} + \tfrac{x}{2}\right) \end{aligned}$$

For $\beta = 1$, we have

$$\begin{aligned} |[x\Delta, y\Delta]| &\in \left[e^{\alpha} \cdot \left(\ln\left(\tfrac{1}{x}\right) - \ln\left(\tfrac{1}{y}\right)\right) - (y - x)e^{\alpha},\ e^{\alpha} \cdot \left(\ln\left(\tfrac{1}{x}\right) - \ln\left(\tfrac{1}{y}\right)\right)\right] \\ vol([x\Delta, y\Delta]) &\in \left[e^{\alpha}(y - x)\Delta - \left(\tfrac{y\Delta(y\Delta+1)}{2} - \tfrac{x\Delta(x\Delta+1)}{2}\right), e^{\alpha}(y - x)\Delta\right] \end{aligned}$$

These estimates will be used in subsequent sections where we construct efficient reductions from the general Maximum Independent Set (MIS) problem to MIS in (α, β)-PLGs for $\beta \leq 1$. In particular we will be concerned with the embedding of the node degree distribution of a given graph G into an interval of the form $[x\Delta, \Delta]$ of a power law node degree distribution with parameters α and β.

3 An Auxiliary Graph Construction

We start with an auxiliary problem.

Our hardness results for the MIS problem in Power Law Graphs rely on the construction and analysis of embeddings of a given class of graphs G, namely a class of instances on which the MIS problem is known to be hard to approximate into (α, β)-PLG. Such an instance G will be mapped to a node degree interval of the form $[x\Delta, \Delta]$, where the parameter α (and hence $\Delta = e^{\alpha/\beta}$) has been chosen appropriately. Then the task will be to construct the remaining part of the power law graph, which consists of a subgraph corresponding to the interval $[1, x\Delta - 1]$ and the rest of the interval $[x\Delta, \Delta]$. This leads us to the following construction problem. We are given a sequence of node degrees $d = (d_1, \ldots, d_n)$. We want to construct a multigraph G_d with n vertices such that d is the degree sequence of G_d and furthermore, the size $IS(G_d)$ of a maximum independent set in G_d is *as small as possible*. In particular we have to consider the case when the degree sequence d corrsponds to an *interval* $[a, b]$ of a power law distribution with $1 \leq a \leq b \leq \Delta = \lfloor e^{\alpha/\beta} \rfloor$. This means that for each $a \leq i \leq b$, the sequence d contains $\lfloor \frac{e^\alpha}{i^\beta} \rfloor$ entries all equal to i. We may assume that the sequence d is always *sorted*, i.e. the entries in d are in increasing order $d_1 \leq \ldots \leq d_n$.

The idea of the construction is as follows. We cover the vertices of the auxiliary graph G_d by a small number n_c of cliques, in increasing order by their node degrees. Then $\lceil n/n_c \rceil$ is an upper bound for $IS(G_d)$. Let us now describe this construction in more detail.

We consider the following situation. We are given a node degree interval $[a, b]$ with $1 \leq a < b \leq \Delta = \left\lfloor e^{\alpha/\beta} \right\rfloor$. We want to construct a graph $G_{a,b}$ such that the degree distribution of this graph is precisely equal to the part of the power law node degree distribution corresponding to this interval. Moreover, we want to achieve that, informally, the size IS(G_{ab} of a maximum independent set in the graph $G_{a,b}$ is sufficiently small.

Let us now describe our construction in detail. From the node degree interval $[a, b]$ we first construct the associated degree sequence $(d_1, \ldots, d_m)$ with $m = \sum_{j=a}^{b} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor$. The set of vertices is

$$V_{a,b} = \{v_1, \ldots, v_m\} \quad \text{for } m = \sum_{j=a}^{b} \left\lfloor \frac{e^\alpha}{j^\beta} \right\rfloor$$

The degree sequence $(d_1, \ldots, d_m)$ has the following form:

$$\begin{aligned} & d_1 = \ldots = d_{\lfloor e^\alpha / a^\beta \rfloor} = a, \\ & d_{\lfloor e^\alpha / a^\beta \rfloor + 1} = \ldots = d_{\lfloor e^\alpha / a^\beta \rfloor + e^\alpha / (a+1)^\beta \rfloor} = a + 1 \end{aligned}$$

and in general, if j is the last index of a node of degree i, then the nodes with indices $j + 1, \ldots, j + \lfloor e^\alpha / (i + 1)^\beta \rfloor$ are of degree $i + 1$.

We generate the set of edges as follows:

- We take the first $d_1 + 1$ nodes $v_1, \ldots, v_{d_1+1}$ and connect them by a clique.
- We compute residual degrees accordingly.
- We take the next d_{d_1+2} nodes, connect them by a clique, keep track of residual degrees and iterate.

So in each iteration i, we construct a clique of size $d_{p(i)} + 1$ on the set of nodes $\{v_{p(i)}, \ldots, v_{p(i)+d_{p(i)}}\}$. The function $p(\cdot)$ satisfies:

- $p(1) = 1$
- $p(i+1) = p(i) + d_{p(i)} + 1$

We give an upper bound on the size of an independent set in this graph $G_{a,b}$. If we would cover each set $V_i = \{v \in V(G_{a,b}) | \deg(v) = i\}$ separately, the number of cliques needed for V_i would be bounded by

$$\left\lceil \frac{\left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor}{i} \right\rceil \leq \frac{\left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor + i}{i} \leq \frac{\frac{e^\alpha}{i^\beta} + i}{i} = \frac{e^\alpha + i^{\beta+1}}{i^{\beta+1}} \tag{2}$$

Thus the size of a maximum independent set in this graph is bounded by

$$\begin{aligned}\sum_{i=a}^{b} \frac{e^\alpha + i^{\beta+1}}{i^{\beta+1}} &\leq \int_a^{b+1} \frac{e^\alpha}{i^{\beta+1}} dx + \frac{e^\alpha}{a^{\beta+1}} - \frac{e^\alpha}{(b+1)^{\beta+1}} + b + 1 - a \\ &= \left[\frac{e^\alpha}{-\beta \cdot x^\beta} \right]_a^{b+1} + \frac{e^\alpha}{a^{\beta+1}} - \frac{e^\alpha}{(b+1)^{\beta+1}} + b + 1 - a \\ &= \frac{e^\alpha}{\beta} \cdot \left(\frac{1}{a^\beta} - \frac{1}{(b+1)^\beta} \right) + \frac{e^\alpha}{a^{\beta+1}} - \frac{e^\alpha}{(b+1)^{\beta+1}} + b + 1 - a\end{aligned}$$

The number of nodes of the graph $G_{a,b}$ is $\sum_{i=a}^{b} \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor$, which is contained in the interval

$$\begin{aligned}&\left[\frac{\Delta}{1-\beta} \left(\left(\frac{b}{\Delta}\right)^{1-\beta} - \left(\frac{a}{\Delta}\right)^{1-\beta} \right) - \left(\frac{\Delta^\beta}{a^\beta} - \frac{\Delta^\beta}{b^\beta} \right), \frac{\Delta}{1-\beta} \left(\left(\frac{b}{\Delta}\right)^{1-\beta} - \left(\frac{a}{\Delta}\right)^{1-\beta} \right) \right] \\ &= \left[\frac{e^\alpha}{1-\beta} \left(b^{1-\beta} - a^{1-\beta} - \frac{1}{a^\beta} + \frac{1}{b^\beta} \right), \frac{e^\alpha}{1-\beta} \left(b^{1-\beta} - a^{1-\beta} \right) \right]\end{aligned}$$

These estimates will be used in Section 4 and Section 5 in the analysis of our embeddings of graphs into (α, β) PLGs.

4 An Embedding of Graphs into PLGs for $\beta < 1$

In this section we show that for each $\beta < 1$, the MIS problem on (α, β)-PLGs is NP-hard to approximate within n^ϵ for some constant $\epsilon \in (0,1)$ which only depends on β. This result is based on the construction of an efficiently computable embedding of arbitrary graphs into (α, β)-PLGs for $\beta < 1$. The global structure of this embedding is as follows. Given a graph G which we want to embed, we map the vertices of G to a node degree interval of the form $[x\Delta, \Delta]$ of some power law distribution with parameters α, β, where $\Delta = \lfloor e^{\alpha/\beta} \rfloor$. Then we make use of our auxiliary graph construction from Section 3 in order to construct the remaining parts of the power law graph $G_{\alpha,\beta}$. By a careful choice of the parameters x and α of this construction, we will be able to bound the size of a maximum independent set in the residual graph $G_{\alpha,\beta} \setminus G$ such that the approximation hardness carries over from general graph instances to (α, β)-PLGs.

Given a graph G with m nodes, we start by replacing G by the graph G' which contains for each node v_i in G a clique of size 2 consisting of nodes $v_{i,1}, v_{i,2}$. Now for all $1 \le i < j \le m$, nodes $v_{i,k}$ and $v_{j,l}$ ($k, l \in \{1,2\}$) are connected by an edge iff G contains an edge between v_i and v_j. The number of nodes of G' is $n = 2m$, and we have $IS(G') = IS(G)$. Furthermore G' contains a perfect matching, consisting of edges $\{v_{i,1}, v_{i,2}\}$ ($1 \le i \le m$). This will enable us to replace these edges by multi-edges in order to fit the graph into some given part of a power law node distribution. We will now construct an embedding of the graph G' into an (α, β)-PLG. In particular, the nodes of G' will be mapped to nodes in the node interval $[x\Delta, \Delta]$, where x is a parameter of the construction. It turns out that we can choose α and x in such a way that $n \ge \frac{1}{2} \cdot |[x\Delta, \Delta]|$.

Let us now give the details of the construction. First we choose the parameters α and $0 < x < 1$ such that

$$n \;\le\; x\Delta \text{ and } n \le |[x\Delta, \Delta]|. \tag{3}$$

The first inequality in (3) will enable us to implement the node degrees within $[x\Delta, \Delta]$ by replacing edges $\{(v,1),(v,2)\}$ in G' by multi-edges. The second condition ensures that the interval $[x\Delta, \Delta]$ is sufficiently large such that G' can be embedded into it. Then the size of the node degree interval $[x\Delta, \Delta]$ can be estimated as follows:

$$|[x\Delta, \Delta]| \in \left[\frac{\Delta}{1-\beta} \cdot \left(1 - x^{1-\beta}\right) - (x^{-\beta} - 1), \frac{\Delta}{1-\beta} \cdot \left(1 - x^{1-\beta}\right)\right] \tag{4}$$

This means that we have to choose x such that $x \le \frac{1-x^{1-\beta}}{1-\beta}$, i.e.

$$(1-\beta)x \;+\; x^{1-\beta} \;-1 \;\le\; 0 \text{ with } x \in (0,1) \tag{5}$$

We observe that (5) holds provided we choose x such that

$$\max\left\{(1-\beta)x,\ x^{1-\beta}\right\} \ \le\ \frac{1}{2} \tag{6}$$

which is equivalent to

$$x \ \le\ \min\left\{\frac{1}{2(1-\beta)}, \left(\frac{1}{2}\right)^{\frac{1}{1-\beta}}\right\} \tag{7}$$

We have

$$\left(\frac{1}{2}\right)^{\frac{1}{1-\beta}} \le \frac{1}{2(1-\beta)} \text{ or equivalently } 2 \ge 2^{1-\beta}(1-\beta)^{1-\beta},$$

and thus if we choose $x = \left(\frac{1}{2}\right)^{\frac{1}{1-\beta}}$, then (5) holds. This yields $|[x\Delta,\Delta]| = (1-o(1))\frac{1}{2}\cdot\frac{\Delta}{1-\beta}$, and furthermore $n = x\Delta = (\frac{1}{2})^{\frac{1}{1-\beta}}\cdot\Delta$. Now we have to construct the residual graph $G_{\alpha,\beta}\setminus G'$. A straight forward approach is to construct two auxiliary graphs $G_{[1,x\Delta]}$ and $G'_{x\Delta,\Delta]}$, where $G_{[1,x\Delta]} = G_d$ for d being the degree sequence of the interval $[1,x\Delta]$ and $G'_{[x\Delta,\Delta]} = G_{d'}$, where the sequence d' is constructed as follows: d' is a degree sequence for the remaining $|[x\Delta,\Delta]|-n$ nodes in $[x\Delta,\Delta]$. It turns out that in this case, the upper bound for the size of an independent set in $G'_{[x\Delta,\Delta]} = G_{d'}$ would be too large compared to n. Therefore we construct the residual graph $G_{\alpha,\beta}\setminus G'$ as follows. We split the interval $[1,x\Delta]$ into two parts and construct one auxiliary subgraph G_1 for the node degree interval $[1,e^{\alpha/(\beta+1)})$ and one subgraph G_2 for the degree sequence d consisting of the full interval $[e^{\alpha/(\beta+1)},x\Delta-1]$ and the degree sequence for the remaining $|[x\Delta,\Delta]|-n$ nodes within the interval $[x\Delta,\Delta]$. This construction is also shown in Figure 1. An upper bound for $IS(G_2)$ is given by the bound for the size of a maximum independent set of the graph $G_{[e^{\alpha/(\beta+1)},\Delta]} =: G_3$. We have

$$\begin{aligned} IS(G_3) \ &\le\ \left\lceil \frac{\sum_{i=e^{\alpha/(\beta+1)}}^{e^{\alpha/\beta}} \left\lfloor \frac{e^{\alpha}}{i^{\beta}} \right\rfloor}{e^{\alpha/(\beta+1)}} \right\rceil \ \le\ \sum_{i=e^{\alpha/(\beta+1)}}^{e^{\alpha/\beta}} \frac{e^{\alpha\cdot(1-\frac{1}{\beta+1})}}{i^{\beta}} \\ &\le\ e^{\alpha\cdot\frac{\beta}{\beta+1}}\cdot\left(\int_{e^{\alpha/(\beta+1)}}^{e^{\alpha/\beta}+1} x^{-\beta}dx \ +\frac{1}{e^{\alpha}} - \frac{1}{e^{\alpha\cdot\frac{\beta}{\beta+1}}}\right) \ \le\ \frac{e^{\frac{\alpha}{\beta+1}}}{1-\beta} \end{aligned}$$

Now we consider the graph G_1. We obtain the following bound for the number of nodes of G_1:

$$|[1,x\Delta]| \ \le\ \frac{\Delta}{1-\beta}\cdot\left(x^{1-\beta} - \left(\frac{1}{\Delta}\right)^{1-\beta}\right) \ =\ \frac{x^{1-\beta}}{1-\beta}\cdot\Delta \ -\ e^{\alpha}$$

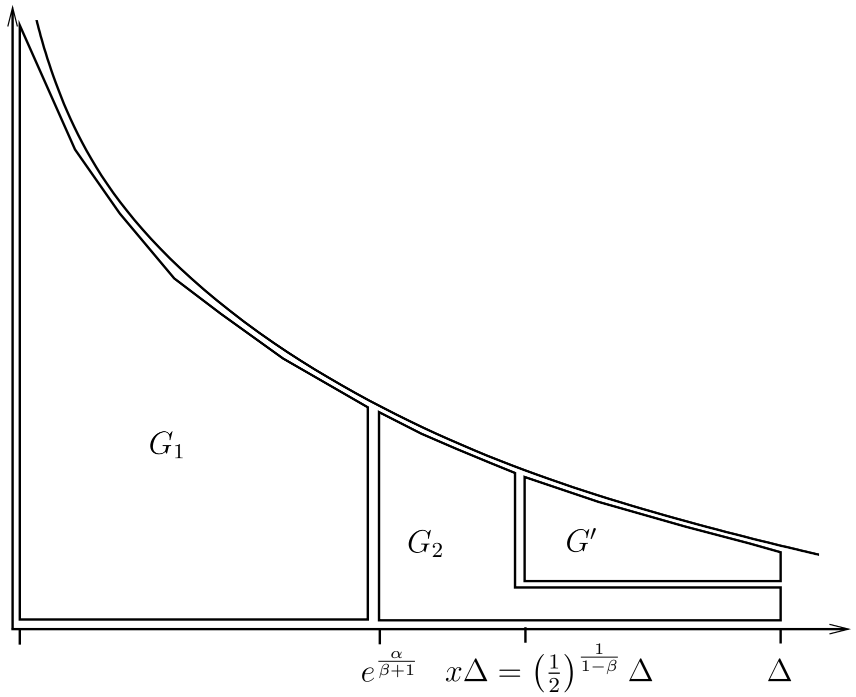
G_1
G_2
G'
$e^{\frac{\alpha}{\beta+1}}$
$x\Delta = \left(\frac{1}{2}\right)^{\frac{1}{1-\beta}} \Delta$
Δ

Figure 1: Construction of the embedding for the case of $\beta < 1$

The size of a maximum independent set in G_1 can be bounded as follows. We split the interval $[1, x\Delta]$ into two subintervals $I_{y,1} = [1, y\Delta)$ and $I_{y,2} = [y\Delta, x\Delta]$, where y has to be chosen appropriately within the interval $(0, x)$. The size of an independent set in the subgraph $G_1[I_{y,1}]$ of G_1 induced by the node interval $I_{y,1} = [1, y\Delta)$ is estimated as follows:

$$\begin{aligned} IS(G_1[I_{y,1}]) &\leq \sum_{i=1}^{y\Delta} \left\lceil \frac{\left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor}{i} \right\rceil \leq \sum_{i=1} y\Delta \frac{e^\alpha}{i^{\beta+1}} + y\Delta \\ &\leq \left[\frac{e^\alpha}{-\beta x^\beta} \right]_1^{y\Delta+1} + e^\alpha \left(1 - \frac{1}{(y\Delta+1)^{\beta+1}} \right) + y\Delta \\ &= \frac{e^\alpha}{\beta} \left(1 - \frac{1}{(y\Delta+1)^\beta} \right) + e^\alpha \left(1 - \frac{1}{(y\Delta+1)^{\beta+1}} \right) + y\Delta \end{aligned}$$

Moreover, the size of a maximum independent set in the subgraph $G_1[I_{y,2}]$ induced by the node degree interval $I_{y,2}$ can be bounded as follows:

$$\begin{aligned}
IS(G_1[I_{y,2}]) &\leq \left\lceil \frac{\sum_{i=y\Delta}^{x\Delta} \left\lfloor \frac{e^\alpha}{i^\beta} \right\rfloor}{y\Delta} \right\rceil \leq \sum_{i=y\Delta}^{x\Delta} \frac{e^\alpha}{y \cdot i^\beta \cdot e^{\alpha/\beta}} \; + 1 \\
&\leq \frac{e^{\alpha\cdot(1-1/\beta)}}{y} \cdot \left(\int_{y\Delta}^{x\Delta+1} \frac{1}{z^\beta} dz + \frac{1}{(y\Delta)^\beta} - \frac{1}{(x\Delta)^\beta} \right) + 1 \\
&\leq \frac{e^{\alpha\cdot(1-1/\beta)}}{y} \cdot \left(\frac{(x\Delta)^{1-\beta}}{1-\beta} - \frac{(y\Delta)^{1-\beta}}{1-\beta} + \frac{1}{(y\Delta)^\beta} - \frac{1}{(x\Delta)^\beta} \right) + 1 \\
&= (1 \pm o(1)) \cdot \frac{e^\alpha}{y\Delta} \cdot \frac{\Delta^{1-\beta}}{1-\beta} \left(x^{1-\beta} - y^{1-\beta} \right) \; = \; \Theta(y^{-1})
\end{aligned}$$

Thus we obtain the following estimate for the size of a maximum independent set in the graph G_1:

$$IS(G_1) \; = \; \Theta\left(y \cdot \Delta \; + \; \frac{1}{y} \right) \tag{8}$$

In order to obtain an upper bound for the right hand side in (8), we observe that

$$\frac{d}{dy} \left(\frac{1}{y} + y \cdot \Delta \right) \; = \; \Delta - \frac{1}{y^2},$$

Thus we choose $y = \Theta\left(\frac{1}{\sqrt{\Delta}}\right)$, say $y = \Delta^{-1/2}$. This yields $IS(G_1) = \Theta(\sqrt{\Delta})$.

Now we start from a graph $G \in \mathcal{G}_{\frac{1}{n^{1-\epsilon}}} \cup \mathcal{G}_{\frac{1}{n^\epsilon}}$. We construct the graph G'. Let n be the number of nodes of G'. Thus we have

$$G' \in \mathcal{G}_{\frac{2^\epsilon}{n^{1-\epsilon}}} \cup \mathcal{G}_{\frac{1}{2^{1-\epsilon} n^\epsilon}} \tag{9}$$

Then we choose the parameters x and α:

$$x = \left(\frac{1}{2}\right)^{\frac{1}{1-\beta}}, \quad \alpha = \beta \cdot \ln\left(\frac{n}{x}\right) \tag{10}$$

We construct the graph $G_{\alpha,\beta}$ as described above. In particular, G' will be embedded into the node degree interval $[x\Delta, \Delta]$, which is of size at least $(1 - o(1)) \cdot \frac{\Delta}{1-\beta}(1 - x^{1-\beta})$.

Altogether we obtain the following result.

Theorem 4.1 *For every $\beta \in (0,1)$, for every $\epsilon > 0$ the Maximum Independent Set Problem on (α,β)-PLGs is NP-hard to approximate within $n^{1-\epsilon}$.*

5 The Case $\beta = 1$

Now we consider the case when the power law exponent is equal to 1. First, we observe that in this case, there is a simple $O(\ln(n))$-approximation algorithm for the Independent Set Problem in $(\alpha, 1)$-PLGs. Namely, the number of nodes is αe^{α}, and we have e^{α} nodes of degree 1. Thus taking half of them yields a $O(\ln(n))$-approximate independent set.

We will now give a lower bound of the form Δ^{ϵ} for maximum degree Δ and some constant $\epsilon > 0$ for the Independent Set Problem in $(\alpha, 1)$-PLGs. As in the previous case $\beta < 1$, we want to proceed as follows: We start from a class $\mathcal{G}$ of graphs for which the Independent Set Problem is hard to approximate. Then we construct a polynomial time reduction which embeds every $G \in \mathcal{G}$ into an $(\alpha, 1)$-PLG $G_{\alpha,1}$. Since the upper bound is $O(\log n)$, we should choose $\mathcal{G}$ appropriately such that the approximation lower bound for Independent Set restricted to graphs from $\mathcal{G}$ is at most logarithmic.

In [5] it was shown that the Independent Set Problem on graphs of degree bounded by Δ is hard to approximate within Δ^{ϵ} for some fixed $\epsilon > 0$. This result also holds for $\Delta = \Theta(\log n)$. We will embed those graphs into $(\alpha, 1)$-PLGs $G_{\alpha,1}$. The construction in [5] starts from a class of graphs $\mathcal{G} = \mathcal{G}_a \cup \mathcal{G}_b$, where $\alpha(G) \leq a \cdot n$ for $G \in \mathcal{G}_a$ and $\alpha(G) \geq b \cdot n$ for all $G \in \mathcal{G}_b$ and $0 < a < b < 1$ are constant. Given a graph $G \in \mathcal{G}$ with n vertices, a degree d-bounded Ramanujan graph H on n vertices is chosen, and the graph $\tilde{DG}^k$ is constructed as follows. The vertices of $\tilde{DG}^k$ are walks $(v_1, \ldots, v_k)$ of length k in H. Two such vertices $v = (v_1, \ldots, v_k)$ and $u = (u_1, \ldots, u_k)$ are connected by an edge iff the vertex subset $\{v_1, \ldots, v_k, u_1, \ldots, u_k\}$ is not an independent set in G. In [5], Theorem 2.1 it is shown that $\alpha(\tilde{DG}^k)$ is within the interval

$$\left[\alpha(G)d^{k-1} \left(\tfrac{\alpha(G)}{n} + \lambda_{n-1} \left(1 - \tfrac{\alpha(G)}{n}\right)\right)^{k-1}, \; \alpha(G)d^{k-1} \left(\tfrac{\alpha(G)}{n} + \lambda_1 \left(1 - \tfrac{\alpha(G)}{n}\right)\right)^{k-1} \right]$$

where A is the transition matrix of the random walk on the Ramanujan graph H, $\lambda_0 \geq \ldots \geq \lambda_{n-1}$ are the eigenvalues of A (note that A is symmetric and has only real eigenvalues) and we have $\lambda_0 = 1, \lambda := \max\{\lambda_1, |\lambda_{n-1}|\} \leq 2\sqrt{d-1}/d$.

Now we want to choose k appropriately such that the maximum degree of $\tilde{DG}^k$ is logarithmic in the number of its nodes. The number of nodes of $\tilde{DG}^k$ is $n \cdot d^{k-1}$. The maximum degree is $d^{k-1} \cdot 3k^2$. Thus we have to choose k such that for some constant $c > 0$,

$$\begin{aligned} & d^{k-1} \cdot 3k^2 \leq c \cdot \log\left(n \cdot d^{k-1}\right) \\ \Leftrightarrow \quad & d^{k-1} \cdot 3k^2 \leq \log(n) + (k-1)\log(d) + \log(c) \\ \Leftrightarrow \quad & (k-1)\log(d) + \log(3k^2) \leq \log\log(n) + \log\left(1 + \tfrac{(k-1)\log(d) + \log(c)}{\log(n)}\right) \end{aligned}$$

We observe that for $k_l = \frac{\log\log(n)}{3\ln(d)}$ we obtain

$$\Delta_l = d^{k-1} \cdot 3k^2 \approx (\log n)^{1/3} \cdot \frac{(\log\log n)^2}{(\ln d)^2} \tag{11}$$

For $k_u = \frac{\log\log(n)}{\ln d}$ we obtain

$$\Delta_u = d^{k-1} \cdot 3k^2 \approx e^{\log\log(n)} \cdot 3 \cdot \frac{(\log\log(n))^2}{(\ln d)^2} = \log(n) \cdot \frac{3(\log\log n)^2}{(\ln d)^2} \tag{12}$$

Thus we choose $k \in [k_l, k_u] = \left[\frac{\log\log n}{3\ln d}, \frac{\log\log n}{\ln d}\right]$ appropriately such that $\Delta_k = d^{k-1} \cdot 3k^2$ satisfies $\Delta_k = \log n$. We let D denote the product graph $\tilde{DG}^k$. Now we have to choose parameters α and x such as to embed the graph D into the node degree interval $[x\Delta, \Delta]$, where $\Delta = e^\alpha$ is the maximum degree of the $(\alpha, 1)$-PLG $G_{\alpha,1}$. Let $n_d = n \cdot d^{k-1}$ be the number of nodes of D. Thus we have to meet the following conditions:

(I) $n_d \;\leq\; |[x\Delta, \Delta]| \;\approx\; e^\alpha \ln\left(\frac{1}{x}\right)$

(II) $\log(n_d) \;\leq\; x\Delta$

Thus we choose $x = \frac{\log(n_d)}{e^\alpha}$. In order to satisfy the requirement (II), we choose $\alpha = \log(n_d)$. Then (I) holds as well. We obtain

$$|[x\Delta, \Delta]| \approx e^\alpha \ln\left(\frac{1}{x}\right) = e^\alpha \cdot \alpha - e^\alpha \cdot \ln(\alpha) = (1 - o(1))\alpha e^\alpha \tag{13}$$

Now we have to give an estimate for the size of a maximum independent set in the subgraph G_1 induced by the residual set of nodes $[x\Delta, \Delta] \backslash V(G)$. This can be upper-bounded by $IS([x\Delta, \Delta])$. Thus it remains to give an estimate for $IS([x\Delta, \Delta])$, the size of a maximum independent set of a graph implementing the complete interval $[x\Delta, \Delta]$ of the power law distribution with parameters α and 1. If we would use the inequality

$$IS([x\Delta, \Delta]) \;\leq\; \left\lceil \frac{\sum_{i=x\Delta}^{\Delta} \left\lfloor \frac{e^\alpha}{i} \right\rfloor}{x\Delta} \right\rceil, \tag{14}$$

the resulting bound would be e^α plus some lower order terms, which is not sufficient for our purpose. We proceed as follows: We split the interval $[x\Delta, \Delta]$ into L subintervals of the form $[x\Delta \cdot h^j, x\Delta \cdot h^{j+1}], j = 0, \ldots L-1$ and use the estimate

$$IS([x\Delta, \Delta]) \;\leq\; \sum_{j=0}^{L-1} \left\lceil \frac{\sum\limits_{i=x\Delta\cdot h^j}^{x\Delta\cdot h^{j+1}} \left\lfloor \frac{e^\alpha}{i} \right\rfloor}{x\Delta \cdot h^j} \right\rceil \tag{15}$$

where h and L are parameters of this estimate. This means that we have to choose $h^L = \frac{1}{x} = \frac{e^\alpha}{\alpha}$, i.e. $h = \sqrt{e^\alpha/\alpha}$. The whole construction for the case $\beta = 1$ is shown in Figure 2. Now the right hand side in (15) is

$$\leq \sum_{j=0}^{L-1} \sum_{i=x\Delta\cdot h^j}^{x\Delta\cdot h^{j+1}} \frac{e^\alpha}{i \cdot x\Delta \cdot h^j} \quad + L = \sum_{j=0}^{L-1} \sum_{i=x\Delta\cdot h^j}^{x\Delta\cdot h^{j+1}} \frac{1}{x \cdot h^j \cdot i} \quad + L$$

We approximate the inner sum by an integral. This yield that the right hand side in (15) is bounded by

$$\begin{aligned}
& \sum_{j=0}^{L-1} \frac{1}{xh^j} \cdot \left(\int_{x\Delta h^j}^{x\Delta h^{j+1}} \frac{1}{y} dy \; + \frac{1}{x\Delta h^{j+1}} - \frac{1}{x\Delta h^j} \right) \; + L \\
= \; & \sum_{j=0}^{L-1} (1+o(1)) \cdot \frac{1}{xh^j} \cdot \left(\ln(x\Delta h^{j+1}) - \ln(x\Delta h^j) + \frac{1}{x\Delta h^{j+1}} - \frac{1}{x\Delta h^j} \right) \; + L \\
= \; & \sum_{j=0}^{L-1} (1+o(1)) \cdot \frac{e^{\alpha(1-\frac{j}{L})}}{\alpha^{1-\frac{j}{L}}} \cdot \left(\frac{j+1}{L} \cdot \alpha - \frac{j}{L} \cdot \alpha \right) \\
= \; & \sum_{j=0}^{L-1} (1+o(1)) \frac{e^{\alpha(1-\frac{j}{L})} \cdot \alpha^{\frac{j}{L}}}{L} \; = \; (1+o(1)) \frac{e^\alpha}{L} \sum_{j=0}^{L-1} \left(\frac{\alpha}{e^\alpha} \right)^{\frac{j}{L}} \\
= \; & (1+o(1)) \frac{e^\alpha}{L} \cdot \frac{1 - \left(\frac{\alpha}{e^\alpha}\right)^{\frac{L}{L}}}{1 - \left(\frac{\alpha}{e^\alpha}\right)^{\frac{1}{L}}} = (1+o(1)) \frac{e^\alpha}{L} \cdot \frac{1 - \frac{\alpha}{e^\alpha}}{1 - \left(\frac{\alpha}{e^\alpha}\right)^{\frac{1}{L}}}
\end{aligned}$$

Now we choose $L = \alpha$. First we observe that

$$\log\left((\frac{\alpha}{e^\alpha})^{1/\alpha} \right) = \frac{1}{\alpha}(\log(\alpha) - \alpha) = \frac{\log \alpha}{\alpha} - 1 \quad \longrightarrow \quad -1 \;\; (\alpha \to \infty) \tag{16}$$

and therefore $\left(\frac{\alpha}{e^\alpha}\right)^{\frac{1}{\alpha}}$ converges to some constant $c' \in (0,1)$, as α tends to infinity. Thus we obtain $|IS([x\Delta, \Delta])| \leq c \cdot \frac{e^\alpha}{\alpha}$ for some constant c.

Now we give an upper bound for the size of an independent set within $[1, x\Delta] = [1, \alpha]$:

$$\begin{aligned}
IS([1, x\Delta]) \;\; & \leq \;\; \sum_{i=1}^{\alpha} \left\lceil \frac{\lfloor \frac{e^\alpha}{i} \rfloor}{i} \right\rceil \; \leq \; \sum_{i=1}^{\alpha} \frac{e^\alpha}{i^2} \; + \alpha \\
& \leq \;\; \alpha \; + \; e^\alpha \left(\int_1^\alpha \frac{1}{y^2} dy \; + \; 1 - \frac{1}{\alpha^2} \right) \\
& \leq \;\; \alpha \; + \; e^\alpha \left(1 - \frac{1}{\alpha} + 1 - \frac{1}{\alpha^2} \right) \;\; = \;\; (1 - o(1)) \cdot 2e^\alpha
\end{aligned}$$

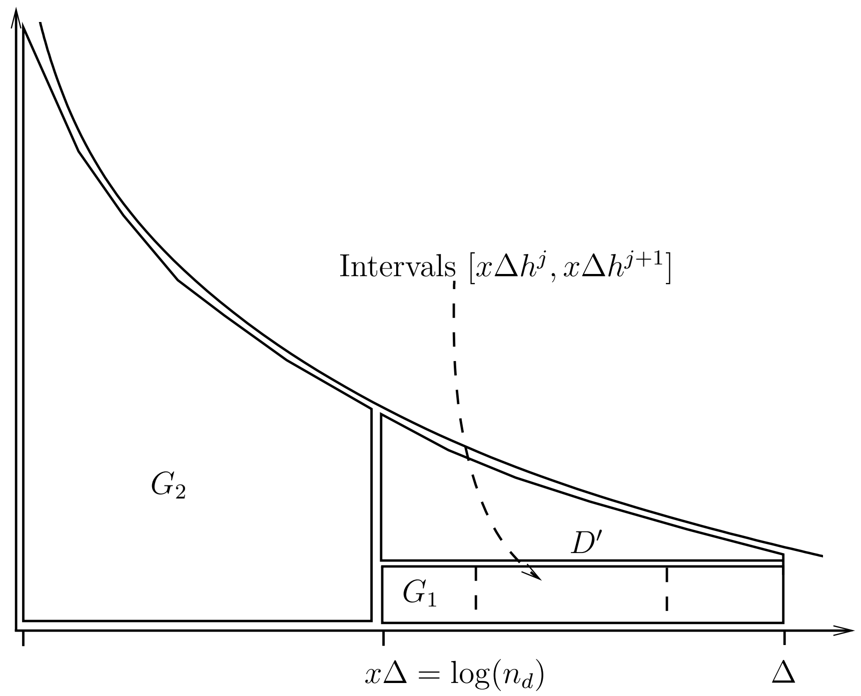

Figure 2: Construction for the case $\beta = 1$

Now we start from a graph $G \in \mathcal{G}_a \cup \mathcal{G}_b$ with n nodes. We construct the graph $D = \widetilde{DG}^k$ for $k = \log\log n$. Thus the maximum degree of D is $\Delta_k = d^{k-1} \cdot 3k^2$, where d is the degree of the Ramanujan graph used in the construction of D. The resulting lower bound for the approximability of Maximum Independent Set in graphs of the form $D = \widetilde{DG}^k, G \in \mathcal{G}_a \cup \mathcal{G}_b$ is

$$bnd^{k-1} \cdot \left(b + \frac{2\sqrt{d-1}}{d}\right)^{k-1} \Big/ \left(and^{k-1} \cdot \left(a - \frac{2\sqrt{d-1}}{d}\right)^{k-1}\right) = \frac{b}{a} \cdot \left(\frac{b - \epsilon_2}{a + \epsilon_2}\right)^{k-1}$$

where we can choose ϵ_2 arbitrary small by choosing d sufficiently large. Now D is embedded into an $(\alpha, 1)$-PLG $G_{\alpha,1}$ (cf. Figure 2). The number of nodes of $G_{\alpha,1}$ is equal to αe^{α}, and $G_{\alpha,1}$ consists of the subgraphs D', G_1 and G_2. We have shown that the size of a maximum independent set in the subgraph G_1 is of order $\frac{1}{\alpha}e^{\alpha}$, while the size of a maximum independent set in the subgraph G_2 is of order e^{α}. Thus it remains to show that if $G \in \mathcal{G}_b$, then this implies that $IS(D') \geq \alpha^{\epsilon} \cdot e^{\alpha}$ for some constant ϵ. It suffices to show that

$$(b + \epsilon_2)^{k-1} > \frac{1}{(\log(nd^{k-1}))^{1/\epsilon}} \tag{17}$$

We let $B' = \frac{1}{b+\epsilon_2}$. Now (17) is equivalent to

$$(\log n)^{1/\epsilon} \left(1 + \frac{(k-1)\log d}{\log(n)}\right)^{1/\epsilon} \; > \; B'^{k-1}, \tag{18}$$

and (by taking logarithms) to $\frac{1}{\epsilon} \cdot \log\log(n) \cdot (1 + o(1)) > (k-1) \cdot \log(B')$. We observe that for $k = \log\log n$ there exists some $\epsilon > 0$ such that this last inequality holds. Thus we obtain the following result.

Theorem 5.1 *For $\beta = 1$, there exists an $\epsilon \in (0,1)$ such that the Maximum Independent Set Problem on (α,β)-PLGs is NP-hard to approximate within $(\log n)^\epsilon$.*

Remark: In this section we have constructed an embedding of graphs $G \in \mathcal{G} = \mathcal{G}_a \cup \mathcal{G}_b$ into $(\alpha,1)$-PLGs, for $0 < a < b < 1$ being constant. It combines the construction from [5] with an embedding of the resulting MIS instances $D = \tilde{DG}^k$ into power law graphs for $\beta = 1$. This construction does not work for the case of a, b being non-constant, e.g. if we try to start from the class of graphs $\mathcal{G} = \mathcal{G}_{1/n^{1-\epsilon}} \cup \mathcal{G}_{1/n^\epsilon}$ from [22]. The reason is that in the construction from [5], a d-regular Ramanujan graph H is used, where $d > \frac{16}{(b-a)^2}$. For $a = 1/n^{1-\epsilon}, b = 1/n^\epsilon$ this yields a non-constant lower bound for d, which implies that the graph D would be super-polynomially larger than G.

6 Further Research

We have given new approximation lower bounds for the MIS problem in (α,β)-PLGs for $\beta \leq 1$. For $\beta < 1$ being constant, the lower bound is $n^{1-\epsilon}$ for every ϵ, while for $\beta = 1$ the lower bound is $(\log n)^\epsilon$ for some constant $\epsilon \in (0,1)$. The further improvements on these lower bounds are important open questions in this area. Another question is the status of the functional cases around value $\beta = 1$, i.e. when β is of the form $\beta_f = 1 \pm \frac{1}{f(n)}$ for some function $f(n)$ depending on the size n of the graph. Another question is the approximability status of the MIS on random PLGs in the preferential attachment model [7].

References

[1] W. Aiello and F. Chung and L. Lu, *A random graph model for massive graphs*, Proc. 32nd ACM STOC, pp. 171–180, 2000.

[2] W. Aiello and F. Chung and L. Lu, *A random graph model for power law graphs*, Experimental Mathematics 10, pp. 53–66, 2000.

[3] W. Aiello and F. Chung and L. Lu, *Random evolution in massive graphs*, Proc. 42th IEEE FOCS, pp. 510–519, 2001.

[4] R. Albert and H. Jeong and A.L. Barabasi, *The diameter of the world wide web*, Nature 401, pp. 130–131, 1999.

[5] N. Alon and U. Feige and A. Wigderson and D. Zuckerman, *Derandomized graph products*, Computational Complexity 5(1), pp. 60–75, 1995.

[6] P. Austrin and S. Khot and M. Safra, *Inapproximability of Vertex Cover and Independent Set in Bounded Degree Graphs*, Theory of Computing 7, pp. 27–43, 2011.

[7] A.-L. Barabasi, *Emergence of Scaling in Random Networks*, Science 10, pp. 509–512, 1999.

[8] S. Bornholdt and H.G. Schuster (Eds.), *Handbook of Graphs and Networks: From the Genome to the Internet*, John Wiley and Sons, Inc., New York 2003.

[9] F. Chung and L. Lu, *Connected Components in Random Graphs with Given Expected Degree Sequences*, Annals of Combinatorics 6, No.2, 2002.

[10] F. Chung and L. Lu, *Complex graphs and networks*, AMS, 2006.

[11] S. Eubank and V.S.A. Kumar and M.V. Marathe and A. Srinivasan and N. Wang, *Structural and algorithmic aspects of massive social networks*, Proc. 15th ACM-SIAM SODA, pp. 718–727, 2004.

[12] M. Faloutsos and P. Faloutsos and C. Faloutsos, *On power-law relationship of the internet topology*, Proc. SIGCOMM'99, pp. 251–261, 1999.

[13] U. Feige and S. Goldwasser and L. Lovasz and S. Safrai and M. Szegedy, *Interactive Proofs and the Hardness of Approximating Cliques*, Journal of the ACM 43, pp. 268–292, 1996.

[14] A. Ferrante, G. Pandurangan, K. Park, *On the Hardness of Optimization in Power-Law Graphs*, Theoretical Computer Science 393, pp. 220–230, 2008.

[15] M. Gast, M. Hauptmann, M. Karpinski, *Inapproximability of Dominating Set on Power Law Graphs*, Theoretical Computer Science 562, pp. 436–452, 2015.

[16] J. Hastad, *Clique is hard to approximate within* $n^{1-\epsilon}$, Acta Mathematica 182, pp. 105–142, 1999.

[17] D. Kempe and J. Kleinberg and E. Tardos, *Maximizing the spread of information through a social network*, KDD, ACM Press, pp. 137–146, 2003.

[18] D. Kempe and J. Kleinberg and E. Tardos, *Influential nodes in a diffusion model for social networks*, Proc. 32nd ICALP, pp. 1127–1138, 2005.

[19] S. Khot, *Improved Inapproximability Results for MaxClique, Chromatic Number and Approximating Graph Coloring*, Proc. 42nd IEEE FOCS, pp. 600–609, 2001.

[20] S. Redner, *How popular is your paper? An empirical study of the citation distribution*, The European Physical Journal B - Condensed Matter and Complex Systems 4, pp. 131–134, 1998.

[21] Y. Shen and D.T. Nguyen and Y. Xuan and M.T. Thai, *New techniques for approximating optimal substructure problems in power-law graphs*, Theoretical Computer Science 447, pp. 107–119, 2012.

[22] D. Zuckerman, *Linear Degree Extractors and the Inapproximability of Max Clique and Chromatic Number*, Theory of Computing 3, pp. 103–128, 2007.