Adjacency labeling schemes for sparse and power law graphs.

Noy Rotbart, Jakob Grue Simonsen and Christian Wulff-Nilsen

February 5, 2015

Abstract

We devise adjacency labeling schemes for power-law graphs, where the number of vertices of degree k is proportional to $k^{-\alpha}$ for some $\alpha > 2$. This sub-family of sparse graphs received a considerable attention in the literature in the last decade. We provide upper and lower bounds for the size of the label for both families. We also compare the label size sufficient for graphs that are constructed from the standard generative model for power law graphs, namely, the BA model. Finally, demonstrate our labeling scheme on a collaboration graph that corresponds to a power law graph.

1 Introduction

Power-law graphs appear in vast number of places. The number of nodes in power-law graphs seen in practice is of magnitude 10^{12}, label of magnitude $\sqrt{10^{12}}$ is reasonable to handle given the current state of hardware.

1.1 Previous work

A graph G is universal, respectively induced universal to the graph family $\mathcal{F}$ if it contains all graphs in $\mathcal{F}$ as subgraphs, respectively induced subgraphs. The size of the smallest induced universal graph was first studied by Moon [20]. Kannan, Naor and Rudich [17] showed that a $f(n) \log n$ adjacency labeling scheme for a family of graphs constructs an induced universal graph for this family of size $2^{f(n)}$. Moon [20] showed a lower bound of $n/2$ on the label size for general graphs, and an upper bound of $n/2 + \log n$. This gap was recently closed by Alstrup et al. [4]. Universal graphs for sparse graphs were investigated first by Babai et al. [7] and improved by Alon and Asodi [2]. Routing schemes (hinging on efficient routing labeling schemes) for power-law graphs were investigated by Chen et. al [12], and by Brady and Cowen [10].

In the literature, a number of generative models that "grow" random graphs whose degree distributions are, with high probability, asymptotically "close" to $\mathrm{ddist}_G(k) \sim Ck^{-\alpha}$ have been proposed for various values of α, most prominently the Barabasi-Albert model [8], and the Aiello-Chung-Lu model [1]. For a survey on the topic see [19]. Common to these are that they work in discrete time-steps with each step involving some randomization (e.g., in the Barabasi-Albert model, each step introduces a fresh node that connects to a fixed number of existing edges with some probability dependent on the degrees of the existing nodes).

2 Preliminaries

In the following we deal with n-vertex undirected finite graphs. A graph $G = (V, E)$ is *sparse* if $|E| = O(n)$, and more precisely a graph with at most cn edges is called a *c-sparse* graph. Let $1 \leq c' \leq 2$ and denote a graph with $O(n^{c'})$ edges as *c'-polysparse*. From hereon, we denote the family of c-sparse graphs with n nodes as $\mathcal{S}_{c,n}$. The collection of graphs of n vertices among a graph family $\mathcal{F}$ is denoted $\mathcal{F}_n$. The degree of a vertex v in a graph $G = (V, E)$ is denoted by $\Delta(v)$. For all $k \geq 0$ the collection of vertices of degree k is denoted V_k. A binary string x is a member of the set $\{0, 1\}^*$, and we denote its length by $|x|$. We denote the concatenation of two binary strings x, y by $x \circ y$. Let $G = (V, E) \in \mathcal{G}(n)$, and let $u, v \in V$. The boolean function *adjacency*(v, u), define over vertices in $G \in \mathcal{G}$, returns **true** if and only if u and v are adjacent in G. A *label assignment* for $G \in \mathcal{G}$ is a mapping of each $v \in V(G)$ to a bit string $\mathcal{L}(v)$, called the *label* of v. An *adjacency labeling scheme* for $\mathcal{G}$ consists of an encoder and decoder. The *encoder* is an algorithm that receives $G \in \mathcal{G}$ as input and computes the label assignment e_G. The *decoder* is an algorithm that receives any two labels $\mathcal{L}(v), \mathcal{L}(u)$ and computes the query $d(\mathcal{L}(v), \mathcal{L}(u))$, such that $d(\mathcal{L}(v), \mathcal{L}(u)) = \mathit{adjacency}(v, u)$. The *size* of the labeling scheme is the maximum label length. Hereafter, we refer to adjacency labeling schemes simply as labeling schemes. For the encoding and decoding algorithms, we assume a $\Omega(\log n)$ word size RAM model (see [3] for additional details).

We begin by a definition relating to the fraction of vertices of a certain degree.

Definition 1. *The* degree distribution *of a graph $G = (V, E)$ is the mapping* $\mathrm{ddist}_G(k) : \mathbb{N}_0 \longrightarrow \mathbb{Q}$ *defined by* $\mathrm{ddist}_G(k) := \frac{|V_k|}{n}$.

We now define *power-law* graphs similarly to other definitions in the literature [19, 1, 8].

Definition 2. *We call $\mathcal{P}_{\alpha,n}$ the family of graphs of n nodes in which $\mathrm{ddist}_G(k) \sim Ck^{-\alpha}$ for real numbers and $\alpha > 1$ and $C = 1/\sum_{k=1}^{\infty} i^{-\alpha} = 1/\zeta(\alpha)$ where ζ is the Riemann zeta function.*

First, note that "$\sim$" entails some rounding of $nCk^{-\alpha}$ since $n\mathrm{ddist}_G(k)$ is a positive integer, but $nCk^{-\alpha}$ is, in general, a non-integral real. Also note that the choice of C arise since we require that $k \mapsto C\frac{n}{k^\alpha}$ be a probability distribution, and thus $C \sum_{k=1}^{\infty} i^{-\alpha} = 1$.

3 Graph families that contain or are contained in $\mathcal{P}_{\alpha,n}$

In this section we define two families of graphs $\mathcal{P}'_{\alpha,n}$ and $\mathcal{P}''_{\alpha,n}$ such that $\mathcal{P}''_{\alpha,n} \subset \mathcal{P}_{\alpha,n} \subset \mathcal{P}'_{\alpha,n}$ and prove some properties on those families.

Definition 3. *We call $\mathcal{P}'_{\alpha,n}$ the family of graphs of n nodes where $\sum_i^{n-1} V_k \leq c(\frac{n}{i^{\alpha-1}})$ for all $0 \leq i \leq n-1$ and some fixed c.*

The class of *proper* power law graphs contains graphs where the number of vertices of degree k must be $C\frac{n}{k^\alpha}$ rounded either up or down and the number of vertices of degree k is non-increasing with k. Note that the function $k \mapsto C\frac{1}{k^\alpha}$ is strictly decreasing.

Definition 4. *Let $\alpha > 1$ be a real number. We say that a graph $G = (V, E)$ is an α-*proper power-law graph *if (i) for every $1 \leq i \leq n$: $|V_i| = n \cdot \mathrm{ddist}_G(k) \in \{\lfloor\frac{1}{\zeta(\alpha)}\frac{n}{i^\alpha}\rfloor, \lceil\frac{1}{\zeta(\alpha)}\frac{n}{i^\alpha}\rceil\}$, and (ii) for every $1 \leq i \leq n-1$: $|V_i| \geq |V_{i+1}|$. The family of proper power-law graphs of n nodes is denoted $\mathcal{P}''_{\alpha,n}$.*

Proposition 1. $\mathcal{P}''_{\alpha,n} \subset \mathcal{P}_{\alpha,n} \subset \mathcal{P}'_{\alpha,n}$

Proof. □

As a result of this proposition, labeling schemes for $\mathcal{P}'_{\alpha,n}$ hold for $\mathcal{P}_{\alpha,n}$ and lower bounds on the label size for $\mathcal{P}''_{\alpha,n}$ hold for $\mathcal{P}_{\alpha,n}$.

Finally, we show the following properties:

Proposition 2. *The maximum degree of a node in a graph of $\mathcal{P}''_{\alpha,n}$ is at most* $\left(\frac{1}{\zeta(\alpha)(\alpha-1)} + 2\right)\sqrt[\alpha]{n} + 2$.

Proof. Let $n > 0$ be an arbitrary integer and let $k' \triangleq \lceil\sqrt[\alpha]{n}\rceil$. Furthermore, let $S_{k'} = \sum_{i=1}^{k'} V_i$, that is $S_{k'}$ is the number of nodes of degree at most k'

Let $S^-_{k'} = \sum_{i=1}^{k'}\lfloor\frac{n}{\zeta(\alpha)}i^{-\alpha}\rfloor$. Then, $S_{k'} \geq S^-_{k'}$. We now bound $S^-_{k'}$ from below. For every i with $1 \leq i \leq k'$ we have $\lfloor\frac{n}{\zeta(\alpha)}i^{-\alpha}\rfloor + 1 \geq \frac{1}{\zeta(\alpha)}i^{-\alpha}$, and hence

$$
\begin{aligned}
S^-_{k'} + k' = {} & \sum_{i=1}^{k'}\left(\left\lfloor\frac{n}{\zeta(\alpha)}i^{-\alpha}\right\rfloor + 1\right) \geq \sum_{i=1}^{k'}\frac{n}{\zeta(\alpha)}i^{-\alpha} = \frac{n}{\zeta(\alpha)}\sum_{i=1}^{k'} i^{-\alpha} \geq \\
& n\left(1 - \frac{1}{\zeta(\alpha)}\sum_{i=k'+1}^{\infty} i^{-\alpha}\right) \geq n\left(1 - \frac{1}{\zeta(\alpha)}\int_{k'}^{\infty} x^{-\alpha}dx\right) = \\
& n\left(1 - \frac{1}{\zeta(\alpha)}\left[\frac{1}{\alpha-1}x^{-\alpha+1}\right]_{\infty}^{k'}\right) = n\left(1 - \frac{1}{\zeta(\alpha)(\alpha-1)}\left(\lceil n^{\frac{1}{\alpha}}\rceil\right)^{-\alpha+1}\right) \geq \\
& n\left(1 - \frac{1}{\zeta(\alpha)(\alpha-1)}\left(n^{\frac{1}{\alpha}}\right)^{-\alpha+1}\right) = n - \frac{n}{\zeta(\alpha)(\alpha-1)}n^{-1+\frac{1}{\alpha}} = \\
& n - \frac{1}{\zeta(\alpha)(\alpha-1)}\sqrt[\alpha]{n}
\end{aligned}
$$

Thus, $S_{k'} \geq S^-_{k'} \geq n - \frac{1}{\zeta(\alpha)(\alpha-1)} \sqrt[\alpha]{n} - \lceil \sqrt[\alpha]{n} \rceil$

As for every $1 \leq i \leq n-1$: $|V_i| \geq |V_{i+1}|$, there are thus at most $1/(\zeta(\alpha)(\alpha - 1)) \sqrt[\alpha]{n} + \lceil \sqrt[\alpha]{n} \rceil$ nodes of degree strictly more than $k' = \lceil \sqrt[\alpha]{n} \rceil$. Hence, the maximum degree of any α-proper power-law graph is at most $\left(\frac{1}{\zeta(\alpha)(\alpha-1)} + 2\right) \sqrt[\alpha]{n} + 2$. □

Proposition 3. *For $\alpha > 2$, all the graphs in $\mathcal{P}''_{\alpha,n}$ are sparse.*

Proof. By Proposition 2, the maximum degree of a node in an α-proper power-law graph is at most $k' \triangleq \left(\frac{1}{\zeta(\alpha)(\alpha-1)} + 2\right) \sqrt[\alpha]{n} + 2$, whence the total number of edges is at most $\frac{1}{2} \sum_{i=1}^{k'} kV_k$. By definition, $V_k \leq \lceil \frac{1}{\zeta(\alpha)} \frac{n}{k^\alpha} \rceil \leq \frac{1}{\zeta(\alpha)} \frac{n}{k^\alpha} + 1$, and thus

$$
\begin{aligned}
\frac{1}{2} \sum_{i=1}^{k'} kV_k &\leq \frac{1}{2} \sum_{i=1}^{k'} k \left(\frac{1}{\zeta(\alpha)} \frac{n}{k^\alpha} + 1 \right) \leq \frac{k'}{2} + \frac{n}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{-\alpha+1} \\
&\leq \left(\frac{1}{2\zeta(\alpha)(\alpha-1)} + 1 \right) \sqrt[\alpha]{n} + 1 + \frac{n\zeta(\alpha-1)}{\zeta(\alpha)}
\end{aligned}
$$

□

Note that 2-proper power-law graphs have $O(n \log n)$ edges.

4 The Labeling Schemes

We first handle c-sparse graphs.

Proposition 4. *There exist a $\sqrt{2cn}\log n + \log n$ labeling scheme for $\mathcal{S}_{c,n}$.*

Proof. Let $G = (V, E)$ be a c-sparse graph. We first assign each vertex $v \in V$ a unique identifier $ID(v)$, using $\log n$ bits. A vertex of degree at least $\sqrt{2cn}$ is called *fat* and *thin* otherwise. From hereon, we use the terminology degree threshold to describe the value separating these two groups. The first bit of $\mathcal{L}(v), v \in V$ is set to zero if v is fat and to one if it is thin. Since there are at most $2cn$ edges, the number of fat vertices is at most $\sqrt{2cn}$. Let (u, v) be an edge in G such that $ID(u) < ID(v)$. If u and v are both either thin or fat $ID(v)$ will appear in $\mathcal{L}(u)$ and vice versa. If u is fat and v is thin, $ID(u)$ will appear in $ID(v)$. Since there are at most $\sqrt{2cn}$ fat vertices, the size of the largest label is bounded by $\sqrt{2cn}\log n + \log n$. Similarly, thin vertices enjoy the same label size as they have at degree at most $\sqrt{2cn}$. Let $\mathcal{L}(v), \mathcal{L}(w)$ be two labels assigned by our suggested decoder to vertices $v, w \in V$. If v and w are both fat or both thin, then there is an edge from v to w if and only if w is listed in the label of v and vice versa. Assume w.l.o.g that v is thin and w is fat, then there is an edge from v to w if and only if w is listed in the label of w. □

Remark 1. It is easy to see that $f(n)$-sparse graphs enjoy a $\sqrt{2f(n)}\log n$ labeling scheme by setting the degree threshold to $\sqrt{2f(n)}$. In addition c-polysparse graphs enjoy a $n^{\frac{c}{2}}\log n$ labeling scheme by setting the threshold to $n^{\frac{c}{2}}$.

Recall that $\mathcal{P}_{\alpha,n} \in \mathcal{S}_{c\alpha}$ when $\alpha \geq 2$. This yields a $(\sqrt{2cn}+1)\log n$ labeling scheme for $\mathcal{P}_{\alpha,n}$. We now show that this label can be significantly improved, by constructing a labeling scheme for $\mathcal{P}'_{\alpha,n}$ which contains $\mathcal{P}_{\alpha,n}$.

Proposition 5. *For all sufficiently large n, The family of graphs $\mathcal{P}'_{\alpha,n}$ enjoys an $O(\sqrt[\alpha]{n}\log n)$ labeling scheme.*

Proof. We set the degree threshold at $f(n)$. By that we mean that a vertex $v \in V$ is *thin* if $\Delta(v) \leq f(n)$ and *fat* otherwise. By definition 3 there are at most $c'n/f(n)^{\alpha-1}$ fat vertices for some c'. We first assign each vertex a unique identifier from $1 \dots n$, using bit strings of size $\log n$. The label of a thin vertex consists of two parts: its unique identifier and a list of the identifiers of all its neighbors. The size of such label is thus at most $(f(n)+1)\log n$. The label of a fat vertex consists also of two parts: a unique identifier and a bit string of length $c' \cdot n \cdot f(n)^{-(\alpha-1)}$ such that position i is 1 in this bit string if the vertex is adjacent to the i'th largest large vertex. The size of a such label is at most $c'n/f(n)^{\alpha-1} + \log n$. Decoding the label is now identical to that of Proposition 4.

From definition 3 it follows that the smallest label size is attained when $f(n)\log n = c'n/f(n)^{\alpha-1}$ for some constant c'. We rearrange the equation and get $f(n)^\alpha = c'n$ and thus the optimum label size occurs when $f(n) = \sqrt[\alpha]{c'n}$, and the resulting label size is $O(\sqrt[\alpha]{n}\log n)$.

□

5 Lower Bounds

In this section we show lower bounds for the label size of any labeling schemes for both $\mathcal{S}_{c,n}$ and $\mathcal{P}_{\alpha,n}$. Our proofs rely on Moon's [20] lower bound of $\lfloor n/2 \rfloor$ bits for labeling scheme for general graphs. We first show that the upper bound achieved for sparse graphs are fairly close to the best possible[1]. We first present a more detailed account of the lower bound suggested by Spinard [21].

Proposition 6. *Any labeling scheme for $\mathcal{S}_{c,n}$ requires labels of size strictly larger than $\frac{\sqrt{n}}{2\sqrt{c}}$ bits.*

Proof. Assume for contradiction that there exist a labeling scheme assigning labels of size strictly less than $\frac{\sqrt{n}}{2\sqrt{c}}$. Let G be an n-vertices graph. Let G' be the graph resulting by adding $\frac{n(n-1)}{c}$ isolated vertices to G, and note that now G' is $c - sparse$. The graph G is an induced subgraph of G'. It now follows that the nodes of G have labels of size strictly less than $\frac{\sqrt{n^2/c}}{2\sqrt{c}} = n/2$ bits. As G was an arbitrary graph, we obtain a contradiction. $\square$

In the remainder of the section we assume that $\alpha > 2$ and prove the following:

Proposition 7. *Any labeling scheme for $\mathcal{P}_{\alpha,n}$ is of size $\Omega(\sqrt[\alpha+1]{n})$.*

More precisely, we present a lower bound for $\mathcal{P}''_{\alpha,n}$ which is contained in $\mathcal{P}_{\alpha,n}$.

5.1 Lower bound for power-law graphs

We first prove that constructing $\mathcal{P}''_{\alpha,n}$ graphs can be done easily. A *degree sequence* is a sequence of integers $d_1 \dots d_n$ such that $0 \le d_i \le n-1$. We denote $v_1 \dots v_n$ the vertices the vertices of a graph $G = (V, E)$ in non-increasing order according to their degree. We denote $v_i \in V$ the i'th vertex in this ordering. The *degree sequence* of G is $d_1 \ge d_2 \dots \ge d_n$ where d_i is the degree of v_i. While every graph has a degree sequence, not all degree sequences form graphs. We say that a degree sequence is *graphic* if it has a corresponding simple graph. The *Erdös-Gallai theorem* [15] states the following: A degree sequence[2] $d_1 \dots d_N$ is graphic if and only if:

1. For every $1 \le k \le N-1$: $\sum_{i=1}^{k} d_i \le k(k-1) + \sum_{i=k+1}^{N} \min\{k, d_i\}$.
2. $\sum_{i=1}^{n} d_i$ is even.

We now show the following:

Lemma 1. *Any degree sequence $d_1 \ge d_2 \dots \ge d_n$ that abides to a proper-power-law distribution is graphic.*

[1] For brevity, we assume now that n is an even number.

[2] In this context, the *configuration model* [9] is a random multigraph constructed from any given degree sequence, tough it is claimed that the number of self-loops and multi edges is negligible [13].

Proof. First, since graphs in $\mathcal{P}_{\alpha,n}$ are sparse, for every $k > O(\sqrt[\alpha]{n})$ the condition holds trivially by observing that $k(k+1) > O(n)$. The maximum degree of a vertex in a power law graph is $O(\sqrt[\alpha]{n})$. When $k \leq O(\sqrt[\alpha]{n})$ we have that:

$$\sum_{i=1}^{k} d_i < O(\sqrt[\alpha]{n}\sqrt[\alpha]{n}) = O(n^{2/\alpha}) = o(n)$$

since $\alpha > 2$. It is easy to see that in this case $\sum_{i=k+1}^{n} \min\{k, d_i\} = O(n)$, and thus the inequality holds.

Finally, if $\sum_{i=1}^{n} d_i$ is not even we add a single edge to the graph. $\square$

Lemma 2. *Any n vertices graph $G = (V, E)$ can be extended to an $N = cn^{\alpha+1}$ vertices graph $G' = (V', E')$ where $G' \in \mathcal{P}''_{\alpha,n}$ and G is an induced subgraph of G'.*

Proof. The number of vertices with degree at least n in G' is $O(\frac{N}{n^\alpha}) = O(n)$. By Lemma 1, we can create G' by first inserting the degree sequence corresponding the vertices of G, and then complete the power lower distribution naively. $\square$

We can now conclude with our lower bound:

Proposition 8. *Any labeling scheme for $\mathcal{P}''_{\alpha,n}$ requires $\Omega(\sqrt[\alpha+1]{n})$ bits.*

Proof. Assume for contradiction that there exist a labeling scheme for assigning labels of size at most $o(\sqrt[\alpha+1]{n})$ bits. Let G be a graph of n-vertices. We now construct the graph G' as described in Lemma 2. The graph G can be reconstructed by a labeling scheme for G' using only the labels of vertices belonging to G inside G'. By the assumption, there is thus a labeling of G' using $o(\sqrt[\alpha+1]{n^{\alpha+1}}) = o(n)$ bits. As G is an arbitrary graph, we obtain a contradiction. $\square$

Since $\mathcal{P}''_{\alpha,n} \subset \mathcal{P}_{\alpha,n}$, this concludes the proof of Proposition 7.

6 Yakob's formulation

We start with a graph of $\sqrt[1+\alpha]{n}$ nodes and construct a graph with n nodes

Observe that Definition 4 and Proposition 2 may be formulated for degree sequences instead of graphs since we do not use any properties of graphs, but rather the distribution of nodes of some degree.

Lemma 3. *Set $n' = n - \sqrt[1+\alpha]{n}$. Let $d_1, ..., d_{n'}$ be any non-increasing sequence of positive integers such that all $d_i \leq \sqrt[1+\alpha]{n} - 1$ (i.e., we have a degree sequence of length n' where every "node" has "degree" bounded above by $\sqrt[1+\alpha]{n} - 1$). Then, the sequence is graphic.*

Proof. Let k be any integer with $1 \leq k \leq n'$; we have:

$$\sum_{i=1}^{k} d_i \leq k \cdot (\sqrt[1+\alpha]{n} - 1)$$

and

$$k(k-1) + \sum_{i=k+1}^{n'} \min(d_i, k) \geq k(k-1) + n' - k$$

Hence, the conditions of the Erdös-Gallai Theorem are fulfilled if

$$\sqrt[1+\alpha]{n} \leq k - 1 + n'/k \tag{1}$$

There are two cases possible:

1. If $k < \sqrt[1+\alpha]{n}$, then

$$n'/k \geq n^{\alpha/(1+\alpha)} - \sqrt[\alpha]{n} = \omega(\sqrt[1+\alpha]{n}).$$

   Hence, for all sufficiently large n, we have $n'/k = \omega(\sqrt[1+\alpha]{n})$, and Eq. 1 is satisfied.

2. If $k \geq \sqrt[1+\alpha]{n}$, we have $k - 1 + n'/k \geq k$, and thus Eq. 1 is satisfied.

Hence, for all k, the sequence is graphic. □

The proof works as follows. We begin with an arbitrary graph G, and construct a degree sequence $d_1, ..., d_n$ which satisfies the requirements of Definition 4. After removing the subsequence of degrees from $d_1, \ldots, d_n$, we will show, using Lemma 3 that the resulting sequence is graphic and denote its resulting graph G'. The graph $G \cup G'$ then has the desired degree sequence $d_1, ..., d_n$ which satisfies Definition 4.

Lemma 4. *Let G be any graph with $q = \sqrt[1+\alpha]{n}$ nodes, and corresponding degree sequence $D_G = d'_1, ..., d'_q$. Assume that we have a degree sequence $d_1, ..., d_n$ that (i) contains D_G as a subsequence; and such that (ii) $d_1, ..., d_n \leq \sqrt[1+\alpha]{n} - 1$. Then, the degree sequence obtained from $d_1, ..., d_n$ by removing $d'_1, ..., d'_q$ satisfy the requirements of Lemma 3.*

(D) For any n, there is a degree sequence satisfying the requirements of Definition 4 (i.e., there is a degree sequence of a proper power-law graph). To see this, let the number of "nodes" of degree j for $1 \leq j \leq \lfloor \sqrt[\alpha]{n/\zeta(n)} \rfloor$ be $\lfloor n/\zeta(\alpha) \cdot j^{-\alpha} \rfloor$.

If the sum of degrees of these nodes is $p < n$, let the number of nodes of degree i for $\lfloor \sqrt[\alpha]{n/\zeta(n)} \rfloor < i \leq n - p$ be 1.

(And there are no "nodes" of higher degree).

Note that the degree sequence above satisfies the requirements of Definition 4, in particular that $|V_i| \geq |V_{i+1}|$.

(E) Let r be a real number and G be any (simple) graph with $q = r \sqrt[1+\alpha]{n}$ nodes. Then the highest degree of any node in G is $q - 1$ and there are at most q such nodes In a proper power-law graph with n nodes, there are at least $\lfloor cn(q-1)^{-\alpha} \rfloor \geq \lfloor r' \sqrt[1+\alpha]{n} \rfloor$ nodes of degree $q - 1$, where r' is a real number depending on c,r, and α. By straightforward computations, we see that we may obtain r' arbitrarily large by choosing r sufficiently small. Hence, the number of nodes of degree $q-1$ is at least $\sqrt[1+\alpha]{n}$ (without rounding).

As the number of nodes of degree j in a proper power-law graph decreases with j, and the maximal number of nodes of *any* degree $d \leq q-1$ in G is q, the number of nodes of degree d in $G \leq$ than the number of nodes of degree d in a proper law graph with n nodes.

Thus, and per (D), we can simply add "nodes" to the degree sequence of G to obtain a degree sequence satisfying the requirements of Definition 4, also satisfying the requirements of Lemma 4 above. This implies that we can satisfy Lemma 3, concluding the proof.

7 Scale free graphs from constructive model

In this section we treat some of the vast literature relating to the construction of power-law graphs in practice. In particular, we are interested in a quantitive comparison. How many graphs can these models produce? How does this number compares to the number of power-law graphs?

The Barabasi-Albert model generates power-law graphs, such that given a parameter m, vertices are inserted to an initially empty graph and attached to at most m existing vertices according to a power-law distribution. First, note that graphs created by this model have low arboricity[3] [16]. We now prove that graphs constructed using Barabasi-Albert model can not possibly construct all power-law graphs.

Graphs constructed by the Barabasi-Albert model have an $O(m \log n)$ adjacency labeling scheme. Let $G = (V, E)$ be an n-node graph resulting by the construction by the Barabasi-Albert model with some parameter m. While it is not known how to compute the arboricity of a graph efficiently, it is possible in near-linear time to compute a partition of G with at most twice [4] the number of forests in comparison to the optimal [6]. We can thus decompose the graph to $2m$ forests in near linear time and label each forest using Alstrup and Rauhe's [5] $\log n + O(\log^* n)$ labeling scheme for trees, and achieve a $2m(\log n + O(\log^* n))$ labeling scheme for G, in near linear time. If the encoder operates at the same time as the creation of the graph an $m \log n$ labeling scheme is possible by simply storing the identifiers of all vertices attached with every vertex insertion.

We now show a connection between the number of distinct (non-isomorphic) graphs in a family and the smallest label size possible.

Lemma 5. *Let $|\mathcal{F}_n|$ denote the number of non-isomorphic graphs in $\mathcal{F}_n$ and let N be the smallest number of nodes of any induced universal graph for $\mathcal{F}_n$ (In other words, any labeling scheme for $\mathcal{F}$ is of size at least $\log N$). Then $\log N - \log n \le \log |\mathcal{F}_n|$.*

Proof. We first claim that an optimal labeling scheme for the family $\mathcal{L}$ will give the same labeling assignment to any two isomorphic graphs G_1 and G_2. From this it follows that $\binom{N}{n} \ge |\mathcal{F}_n|$.

We now show the existence of a a $\log |\mathcal{F}_n| + \log n$ labeling scheme for $\mathcal{F}_n$ in the following way. We first assign each graph $G = (V, E)$ a unique identifier using at most $\log |\mathcal{F}_n|$ bits. A node $v \in V$ receives a label that is a concatenation of a unique ID for v and the unique identifier of G. □

This Lemma with Proposition 8 suggest that there are a lot more power-law graphs than ones that can be created by preferential attachment, as in the Barabasi-Albert model.

[3]The arboricity of a graph is the minimum number of forests into which its edges can be partitioned.

[4]In fact for any $\epsilon \in (0, 1)$ Kowalik [18] showed that there exist an $O(|E(G)|/\epsilon)$ algorithm that computes such partition using at most $(1 + \epsilon)$ times more forests than the optimal.

The two most used power-law constructors for generating network topologies that obey power law distributions are Waxman [22] and N-level Hierarchical [11]. Finally, we mention the model proposed by Chung [14] (Chapter 3) which produces graphs with self loops and multi-edges.

8 Experimental Study

9 Conclusion

References

[1] W. Aiello, F. Chung, and L. Lu. A random graph model for power law graphs. *Experimental Mathematics*, 10(1):53–66, 2001.

[2] N. Alon and V. Asodi. Sparse universal graphs. *J. Comput. Appl. Math.*, 142(1):1–11, May 2002.

[3] S. Alstrup, P. Bille, and T. Rauhe. Labeling schemes for small distances in trees. *SIAM J. Disc. Math.*, 19(2):448–462, 2005.

[4] S. Alstrup, H. Kaplan, M. Thorup, and U. Zwick. Adjacency labeling schemes and induced-universal graphs. *arXiv preprint arXiv:1404.3391*, 2014.

[5] S. Alstrup and T. Rauhe. Small induced-universal graphs and compact implicit graph representations. In *Proceedings of the 43rd Symposium on Foundations of Computer Science*, FOCS '02, pages 53–62, Washington, DC, USA, 2002. IEEE Computer Society.

[6] S. R. Arikati, A. Maheshwari, and C. D. Zaroliagis. Efficient computation of implicit representations of sparse graphs. *Discrete Applied Mathematics*, 78(1):1–16, 1997.

[7] L. Babai, F. R. Chung, P. Erdos, R. L. Graham, and J. Spencer. On graphs which contain all sparse graphs. *Ann. Discrete Math*, 12:21–26, 1982.

[8] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[9] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316, 1980.

[10] A. Brady and L. J. Cowen. Compact routing on power law graphs with additive stretch. In *ALENEX*, volume 6, pages 119–128. SIAM, 2006.

[11] K. L. Calvert, M. B. Doar, and E. W. Zegura. Modeling internet topology. *Communications Magazine, IEEE*, 35(6):160–163, 1997.

[12] W. Chen, C. Sommer, S.-H. Teng, and Y. Wang. A compact routing scheme and approximate distance oracle for power-law graphs. *ACM Transactions on Algorithms (TALG)*, 9(1):4, 2012.

[13] F. Chung, L. Lu, and V. Vu. Eigenvalues of random power law graphs. *Annals of Combinatorics*, 7(1):21–33, 2003.

[14] F. R. Chung and L. Lu. *Complex graphs and networks*, volume 107. American mathematical society Providence, 2006.

[15] P. Erdos and T. Gallai. Graphs with prescribed degrees of vertices (hungarian). *Mat. Lapok*, 11:264–274, 1960.

[16] G. Goel and J. Gustedt. Bounded arboricity to determine the local structure of sparse graphs. In *Graph-Theoretic Concepts in Computer Science*, pages 159–167. Springer, 2006.

[17] S. Kannan, M. Naor, and S. Rudich. Implicit representation of graphs. In *SIAM Journal On Discrete Mathematics*, pages 334–343, 1992.

[18] Ł. Kowalik. Approximation scheme for lowest outdegree orientation and graph density measures. In *Algorithms and computation*, pages 557–566. Springer, 2006.

[19] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet mathematics*, 1(2):226–251, 2004.

[20] J. Moon. On minimal n-universal graphs. In *Proceedings of the Glasgow Mathematical Association*, volume 7, pages 32–33. Cambridge University Press, 1965.

[21] J. P. Spinrad. *Efficient graph representations*. American mathematical society, 2003.

[22] B. M. Waxman. Routing of multipoint connections. *Selected Areas in Communications, IEEE Journal on*, 6(9):1617–1622, 1988.