Adjacency labeling schemes for sparse and power law graphs.

Noy Rotbart, Jakob Grue Simonsen and Christian Wulff Nielsen

January 23, 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 [14]. Kannan, Naor and Rudich [11] 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 [14] 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. [3]. Universal graphs for sparse graphs were investigated first by Babai et al. [6] and improved by Alon and Asodi [2].

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 [7], and the the Aiello-Chung-Lu model [1]. For a survey on the topic see [13]. 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).

Remark 1. Let N be the smallest number of nodes in the an induced universal graph $G = (V, E)$ for a family $\mathcal{F}$ of graphs with n vertices . If this family contains no isomorphic graphs, then $\binom{N}{n}$ is clearly an upper bound on $|\mathcal{F}|$.

2 Preliminaries

In the following we deal with n-vertex undirected finite and connected 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 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. 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* $\mathrm{ddist}_G(k) := \frac{|V_k|}{n}$.

In his survey, Mitzenmacher [13] defines *power-law graphs* in which $\mathrm{ddist}_G(k) \sim Ck^{-\alpha}$ for real numbers $C > 0$ and $\alpha > 1$. As $n\mathrm{ddist}_G(k)$ is a positive integer, but $nCk^{-\alpha}$ is, in general, a non-integral real, the "$\sim$" entails some rounding of $nCk^{-\alpha}$.

We require that $k \mapsto C\frac{n}{k^\alpha}$ be a probability distribution. Hence, in particular $C\sum_{k=1}^{\infty} i^{-\alpha} = 1$, and hence $C = 1/\sum_{k=1}^{\infty} i^{-\alpha} = 1/\zeta(\alpha)$ where ζ is the Riemann zeta function. Thus, C is dictated by the choice of α. We henceforth denote the family of n vertex power-law graphs as $\mathcal{P}_{\alpha,n}$.

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 $\mathcal{P}''_{\alpha,n}$ and prove that $\mathcal{P}''_{\alpha,n} \subset \mathcal{P}_{\alpha,n} \subset \mathcal{P}'_{\alpha,n}$.

Definition 2. *We call* $\mathcal{P}'_{\alpha,n}$ *the family of graphs of* n *nodes where* $\sum_{i}^{n-1} V_k = O(\frac{n}{i^{\alpha-1}})$ *for all* $0 \leq i \leq n-1$.

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 3. *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

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

□

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 $\sqrt{2cn}\log n + \log n$ adjacency 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 2. 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 $O(\sqrt[\alpha]{n}\log n)$ adjacency 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 2 there are at most $c'n/f(n)^{\alpha-1}$ fat vertices for some c'. We first assign each vertex a unique identifier from $1 \ldots n$, 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 2 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 adjacency labeling schemes for both $\mathcal{S}_{c,n}$ and $\mathcal{P}_{\alpha,n}$. Our proofs rely on Moon's [14] lower bound of $\lfloor n/2 \rfloor$ bits for adjacency 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 [15].

Proposition 6. *Any adjacency 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 for adjacency 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 adjacency 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. □

In the remainder of the section we prove the following:

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

More precisely, we show that a subset of $\mathcal{P}_{\alpha,n}$ which we denote *proper* power law graphs enjoys a similar lower bound.

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

5.1 Lower bound for power-law graphs

We show a lower bound for any adjacency labeling scheme for $\mathcal{P}''_{\alpha,n}$ where $\alpha > 2$. To that end, we first prove that constructing $\mathcal{P}''_{\alpha,n}$ graphs can be done easily.

A *degree sequence* is a sequence of integers $d_1 \ldots d_n$ such that $0 \leq d_i \leq n-1$. We denote $v_1 \ldots v_n$ 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 $d_1 \geq d_2 \cdots \geq d_n$ where d_i is the degree of v_i. While every graph has a degree sequence, not all degree sequences have graphs. We say that a degree sequence is *realizable* if it has a corresponding graph. The *Erdös-Gallai theorem* [10] states the following: A degree sequence[2] $d_1 \ldots d_N$ is realizable if and only if:

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

We now show the following:

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

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

[2]In this context, the *configuration model* [8] 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 [9].

Proof. Assume for contradiction that there exist a labeling scheme for adjacency 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 an adjacency labeling scheme for adjacency 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. □

Since proper-power-law graphs are a subset of $\mathcal{P}_{\alpha,n}$this concludes the proof of Proposition 7.

6 Scale free graphs from constructive model

6.1 Constructive models and the implication of the lower bound.

The two most used power-law constructors are Waxman [] and N-level Hierarchical []. 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. We now prove that the constructions by Waxman and Barabasi-Albert can not possibly construct all power-law graphs. It is easy to see that graphs constructed by the Barabasi-Albert model has an $m \log n$ adjacency labeling scheme. Upon vertex insertion, simply store the identifiers of all vertices attached. This along with proposition 8 suggests that there are a lot more power-law graphs than ones that can be created by preferential attachment, as in the Barabasi-Albert model.

Suppose now that we are given a graph $G = (V, E)$ that was constructed by the Barabasi-Albert model with some parameter m. Clearly, G has density m and accordingly, arboricity[3] m. While it is not known how to compute the arboricity of a graph efficiently, it is possible in near-linear time to compute this partition, with at most twice [4] the number of forests in comparison to the optimal [5]. We can thus decompose the graph to $2m$ forests in near linear time and label each forest using Alstrup and Rauhe's [4] $\log n + O(\log^* n)$ labeling scheme for trees, and achieve a $2m(\log n + O(\log^* n))$ labeling scheme for G, in near linear time.

7 Alternative definitions

Here we define two variants of power law graphs which are closely related to the original definition. We also prove some properties guaranteed by these definitions.

[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 [12] 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.

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

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

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

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

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

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

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

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

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