Adjacency labeling schemes for sparse and power law graphs.

Jakob Grue Simonsen, Noy Rotbart and Christian Wulff Nielsen

January 12, 2015

Abstract

We discuss adjacency labeling schemes for power-law graphs. These are a sub-family of sparse graphs that received a considerable attention in the literature in the last decade. We first find a lower bound for the size of the induced universal graphs required to each of the families. We then provide a constructive upper bound for these induced universal graphs using adjacency labeling schemes of similar asymptotic size.

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 of 10^{12}. A label of $\sqrt{10^{12}}$ bits 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. Kannan, Naor and Rudich [] showed that an adjacency labeling scheme of $f(n)$ bits for some family of graphs can construct an induced universal graph for this family of $2^{f(n)}$ vertices. The size of the smallest induced universal graph was first studied by Moon [5]. Seen from labelling schemes, he 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. [2]. Universal graphs for sparse graphs were investigated first by Babai et al. [3] and improved by Alon and Asodi [1].

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 an n-vertex graph with at most cn edges is called c-sparse. We denote the family of c-sparse graphs with n nodes as $\mathcal{S}_C$. The degree of a vertex v in a graph $G = (V, E)$ is denoted by $\Delta(v)$. For every $1 \leq k \leq n$ 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* $\text{ddist}_G : \mathbb{N}_0 \longrightarrow \mathbb{Q}$ *defined by* $\text{ddist}_G(k) := \frac{|V_k|}{n}$.

We also define a family of graphs where this fraction exponentially decreases as the degree increases.

Definition 2. *We say that a graph* $G = (V, E)$ *is a* (C, α) power-law graph *if for every* $1 \leq i \leq n$*:* $n \cdot \text{ddist}_G(k) = \lfloor C \frac{n}{i^\alpha} \rfloor$ *for some real numbers* $C > 0, \alpha > 1$.

By this definition:

$$|E| = \sum_{i=1}^{n} i \lfloor C \frac{n}{i^\alpha} \rfloor = \sum_{i=1}^{n} \lfloor C \frac{n}{i^{\alpha-1}} \rfloor.$$

Since $\sum_{i=1}^{n} 1/i^a$ converges when $a > 1$, all power law graphs where $\alpha > 2$ are sparse. Moreover, for $\alpha > 3$, a (C, α) power law graph is $2C$ sparse[1]. We denote $\mathcal{P}_{C,\alpha}$ the family of power law graphs with n nodes, and parameters C and α.

We also ignore rounding and remark the two following equalities.

1. $\text{ddist}_G(k) = \frac{C}{k^\alpha}$
2. $|V_k| = \frac{nC}{k^\alpha}$.

3 The Labeling Schemes

We first handle sparse graphs.

Proposition 1. *There exist a* $\sqrt{2cn} \log n + \log n$ *adjacency labeling scheme for* $\mathcal{S}_c$.

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 $ID(v)$

[1] It is in fact $\frac{\pi^2}{6} C \sim 1.64C$ sparse, which is the result of $\zeta(2)$, the Riemman Zeta function.

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 $\sqrt{2cn}$. Decoding the label is now obvious, and will take $O(\sqrt{n})$ operations. □

Remark 2. It is easy to see that $f(n)$-sparse graphs[2] enjoy a $\sqrt{2f(n)n}\log n$ labeling scheme by setting the degree threshold to $\sqrt{2f(n)n}$.

Recall that $\mathcal{P}_{C,\alpha} \in \mathcal{S}_{2C}$ when $\alpha \geq 2$. This yields a $\sqrt{4Cn}\log n$ labeling scheme for $\mathcal{P}_{C,\alpha}$. We now show that this label can be significantly improved. To do so we first need to account for the number of vertices of degree at least k for any $1 \leq k \leq n$.

Proposition 2. *Let $f : \mathbb{N} \longrightarrow \mathbb{N}$ be a mapping such that $f(n) = o(n)$. Let $C > 0$ and $\alpha > 1$ be real numbers. Then there is $N \in \mathbb{N}$ such that if G is a power-law graph with* $\mathrm{ddist}_G(k) = Ck^{-\alpha}$ *and at least N vertices, then the fraction of vertices in G with degree at least $f(n)+1$ is bounded above by $O(f(n)^{-(\alpha-1)})$.*

Proof. For $1 \leq j \leq n-1$, the fraction of vertices of degree at least j is $C\sum_{i=j}^{n-1} 1/i^\alpha$. Note that $d/dx(Cx^{-\alpha}) = -\alpha Cx^{-(\alpha+1)} < 0$ and thus for all $j \geq 1$ we have $\int_j^{j+1} Cx^{-\alpha}dx > C(j+1)^{-\alpha}$. Hence, the fraction of vertices of degree at least $j+1$ is at most

$$\int_j^{n-1} Cx^{-\alpha}dx = \left[\frac{C}{-(\alpha-1)}x^{-(\alpha-1)}\right]_j^{n-1} = \frac{C}{\alpha-1}\left(j^{-(\alpha-1)} - (n-1)^{-(\alpha-1)}\right)$$

In particular, the fraction of vertices of degree at least $f(n)+1$ is at most $\frac{C}{\alpha-1}\left(f(n)^{-(\alpha-1)} - (n-1)^{-(\alpha-1)}\right) \leq \frac{C}{\alpha-1}f(n)^{-(\alpha-1)} = O(f(n)^{-(\alpha-1)})$. □

Lemma 1. *Let $f : \mathbb{N} \longrightarrow \mathbb{N}$ be a computable mapping such that $f(n) = o(n)$ and let $C > 0, \alpha > 1$ be real numbers. Then the family of power-law graphs with* $\mathrm{ddist}_G(k) = Ck^{-\alpha}$ *has a adjacency labeling scheme such that for all sufficiently large n, the maximum size of a label is bounded above by:*

$$\log n + \max(f(n)\log n, O(n/f(n)^{\alpha-1})).$$

Proof. We set the degree threshold at $f(n)$. By that we mean that a vertex $v \in V$ is *small* if $\Delta(v) \leq f(n)$ and *large* otherwise. By Proposition 2 there are at most $c'n/f(n)^{\alpha-1}$ large vertices for some c'. We assign each vertex a unique identifier from $1 \ldots n$, such that the large vertices are assigned the last $c'n/f(n)^{\alpha-1}$ identifiers according to their degree in non-decreasing order.

The label of a small 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 large 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

[2]Graphs with n vertices and $f(n)$ edges

this bit string if the vertex is adjacent to the i'th largest large vertex. The size of a such label is at most $c' \cdot n/f(n)^{\alpha+1} + \log n$.

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 small or both large, 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 small and w is large, then there is an edge from v to w if and only if w is listed in the label of w □

From Lemma 1 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 $O(\sqrt[\alpha]{n}\log n)$.

This also means that there are $O(\sqrt[\alpha]{n})$ large vertices, which correlates nicely with the following fact. The largest degree in an (α, C)-power law graph is bounded by $O(\sqrt[\alpha]{n})$. To see this observe that the number of vertices of degree k must to be at least 1. Thus $nC\frac{1}{k^{\alpha}} \geq 1$, which implies that $k \leq \sqrt[\alpha]{nC}$ for some constant C.

Conjecture 1. *Any family of graphs such that* ddist(k) *has "high" positive skewness will have labeling schemes for adjacency with sublinar maximum labeling size. A reasonable way forward would be to consider the third moment of some standard distributions and see what happens.*

4 Lower Bounds

We begin this section by showing that the upper bounds achieved for sparse graphs are fairly close to the best possible. By Moon [5] it follows that any adjacency labeling scheme for general graphs requires at least $\lfloor n/2 \rfloor$ bits. For brevity, we assume now that n is an even number. We present the following extension, due to Spinard [6].

Proposition 3. *Any adjacency labeling scheme for c-sparse graphs 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 less than $\frac{\sqrt{n^2/c}}{2\sqrt{c}} = n/2$ bits. As G was an arbitrary graph, we obtain a contradiction. □

4.1 Lower bounds for $\mathcal{P}_{C,\alpha}$.

We now show that a similar lower bound can be attained for power-law graphs where $\alpha > 2$. To do so, we first must argue that constructing power law graphs in this fashion is possible.

A *degree sequence* is a sequence of integers $d_1 \dots d_n$ such that $0 \leq d_i \leq n-1$. 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 $d_1 \geq d_2 \dots \geq d_n$ 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* [4] states the following: A degree sequence $d_1 \dots d_N$ 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 2. *Any degree sequence $d_1 \geq d_2 \dots \geq d_n$ that abides to an (α, C) power-law distribution is realizable.*

Proof. First, since graphs in $\mathcal{P}_{C,\alpha}$ 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 $k \leq O(\sqrt[\alpha]{n})$, and since $\alpha > 2$ we have that:

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

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. □

Lemma 3. *Any n vertices graph $G = (V, E)$ can be extended to an $N = cn^{\alpha+1}$ vertices graph $G' = (V', E')$ where $G' \in \mathcal{P}_{C,\alpha}$, such that 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 2, we can create G' by first inserting the degree sequence corresponding the vertices of G, and then complete the power lower distribution naively.

□

We can now conclude with our lower bound:

Proposition 4. *Any adjacency labeling scheme for $\mathcal{P}_{C,\alpha}$ requires least $O(\sqrt[\alpha+1]{n})$ bits.*

Proof. Assume for contradiction that there exist a labeling scheme for adjacency assigning labels of size at most $o(\sqrt[\alpha+1]{n}/c)$ bits. Let G be a graph of n-vertices. We now construct the graph G' as described in Lemma 3. 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.

4.2 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 4 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.

□

References

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

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

[3] 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.

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

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

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