Quick-and-dirty note on adjacency labelling for power law graphs. Caveat lector: there may be a plethora of errors.

Jakob Grue Simonsen

November 12, 2014

Definition 1. *The* degree distribution *of a finite undirected graph* $G = (V, E)$ *is the map* $\text{ddist}_G : \mathbb{N}_0 \longrightarrow \mathbb{Q}$ *defined by* $\text{ddist}_G(k) \triangleq |\{v \in V : \text{deg(v)} = \text{n}\}|/|V|$.

Intuition: $\text{ddist}_G(k)$ is the fraction of nodes in G having k edges incident to it.

Definition 2. *A* power-law graph *is a finite undirected graph* $G = (V, E)$ *with no nodes of degree zero such that, for* $k \geq$, *we have* $|V| \cdot \text{ddist}_G(k) = [|V| \cdot Ck^{-\alpha}]$ *for some real numbers* $C > 0, \alpha > 1$ *(where* $[\cdot]$ *is the nearest integer function).*

Intuition: The fraction of nodes in G having k edges is proportional to $k^{-\alpha}$.

The literature is not very clear on when to use $[\cdot]$, $\lfloor \cdot \rfloor$ or other rounding — most papers assume a little slack, and it is common just to see "$\text{ddist}_G(k) \sim k^{-\alpha}$"[1]. Furthermore, the degree distribution is sometimes only assumed to "kick in" for sufficiently large values of k. Power-law graphs are also known as "scale-free networks", or "graphs with a fat-tailed degree distribution".

An unbelievable amount of literature has been written about power law graphs, almost all of it bad. A very large set of phenomena that "naturally" involve graphs (protein networks, internet AS-level graphs, Facebook friends, ...) have been modelled more-or-less accurately by power-law graphs[2]. The typical fit of $Ck^{-\alpha}$ results in $1 < \alpha < 2$.

In the following, for ease of notation, ignore rounding etc., set $|V| = n$, and thus assume that $\text{ddist}_G(k) = Ck^{-\alpha}$ and that the number of nodes of degree k is $nCk^{-\alpha}$.

Proposition 1. *Let* $f : \mathbb{N} \longrightarrow \mathbb{N}$ *be a map 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* $\text{ddist}_G(k) = Ck^{-\alpha}$ *and at least* N *nodes, then the fraction of nodes in* G *with degree at least* $f(n) + 1$ *is bounded above by* $O(f(n)^{-(\alpha-1)})$.

[1] Also: when truncating the probability mass of a distribution with infinite support (e.g., power-law distributions), the excess probability mass needs to be accounted for. If we ever write a paper on this, we need to do this more formally than almost all existing papers.

[2] I think most of it is crap — ask Casper.

Proof. For $1 \leq j \leq n-1$, the fraction of nodes of degree at least j is $C(j^{-\alpha} + (j+1)^{-\alpha} + \cdots + (n-1)^{-\alpha})$.

Note that $d/dx(Cx^{-\alpha}) = -\alpha C x^{-(\alpha+1)} < 0$ and thus that, for all $j \geq 1$ we have $\int_j^{j+1} Cx^{-\alpha}dx > C(j+1)^{-\alpha}$. Hence, the fraction of nodes of degree at least $j+1$ is at most[3]

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

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

Lemma 1. *Let $f : \mathbb{N} \longrightarrow \mathbb{N}$ be a computable map 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 labelling scheme for adjacency such that for all sufficiently large n, the maximum size of a label is bounded above by $O(\log n(1 + f(n) + n/(f(n)^{(\alpha-1)})))$*

Proof. Say that a node is "small" if it has at most $f(n)$ neighbors and "large" if it has at least $f(n)+1$ neighbors.

The label of each node v consists of (i) an identifier (space cost $\log(n)$), and:

- If v is small, the list of identifers of all its small neighbors (space cost $O(f(n)\log(n))$).
- The list of identifiers of all large neighbors of v (by Proposition 1 there are at most $O(n \cdot f(n)^{-(\alpha-1)})$ such neighbors, for a total space cost of $O(\log(n) \cdot n/(f(n)^{\alpha-1}))$.

Thus, if v is large, the total label size is $O(\log(n)(1 + n/(f(n)^{\alpha-1})))$, and if v is small the total label size is $O(\log(n)(1 + f(n) + n/(f(n)^{\alpha-1})))$.

Observe that if f is computable, the above encoding scheme is computable.

Given two nodes (v, w), the decoder inspects the labels of v and w. If v and w are both small or both large, then there is an edge from v to w iff w is listed in the label of v and vice versa. If v is small and w is large, then there is an edge from v to w iff w is listed in the label of w (the case where w is small and v is large is symmetric). □

Proposition 2. *Let $C > 0, \alpha > 1$ be real numbers. Then the family of power law graphs with $\mathrm{ddist}_G(k) = Ck^{-\alpha}$ has a labelling scheme for adjacency such that for all sufficiently large n, the maximum size of each label is bounded above by $O(n^{1/(\alpha-1)}\log(n))$*

[3] I doubt that a better upper bound can be *computed* easily. It is not hard to see that, for each sufficiently small $\epsilon > 0$, the fraction can be bounded more tightly by $\zeta(\alpha, j+1) - \epsilon$ where ζ is the Hurwitz zeta function; but exact computation of this can be extremely difficult.

Proof. By Lemma 1, setting $f(n) = n^{1/(\alpha-1)}$, we obtain a maximum label size of $O(\log(n)(1 + n^{1/(\alpha-1)} + n/(n^{1/(\alpha-1)})^{\alpha-1})) = O(n^{1/(\alpha-1)} \log(n))$ □

For $\alpha > 2$, that is, for most power law graphs occurring in fits to real-world data, the above proposition yields that the maximum label size is $o(n)$. For larger values of α, the proposition yields that the maximum label size has even better asymptotic bounds, e.g. $\alpha \geq 3$, the maximum label size is $o(\sqrt{n}\log(n))$.

For $1 < \alpha \leq 2$, Proposition 2 yields an unusable bound (because $n^{1/(\alpha-1)} = \Omega(n)$). However, Lemma 1 can be used to show that, for all $\alpha > 1$, there exists a labelling scheme with maximum label size $O(\sqrt{n}\log(n))$ (note, though, that this is actually a *worse* result for $\alpha > 2$ than the bound in Proposition 2).

Proposition 3. *Let $C > 0$ and $\alpha > 1$ be real numbers. Then the family of power law graphs with $\mathrm{ddist}_G(k) = Ck^{-\alpha}$ has a labelling scheme for adjacency such that for all sufficiently large n, the maximum size of each label is bounded above by $O(\sqrt{n}\log(n))$.*

Proof. Set $f(n) = \sqrt{(}n)$. By Lemma 1, there is a labelling scheme with maximum label size $O(log(n)(1+\sqrt{(}n)+n/\sqrt{n}^{\alpha-1})) = O(\log(n)(\sqrt{(}n)+n^{1-(\alpha-1)/2})) = O(\log(n)(\sqrt{n}+n^{(3-\alpha)/2}))$. As $\alpha > 1$, the term $n^{(3-\alpha)/2})$ is asymptotically dominated by $\sqrt{n}$, and we thus obtain that the maximum size of each label is bounded above by $O(\sqrt{n}\log(n))$, as desired. □

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