Power Laws and Preferential Attachment

Lin Chen

1Department of Electrical Engineering
Yale University
lin.chen@yale.edu

Theoretical Challenges in Network Science
Acknowledgement

This presentation contains some material adapted from Professor Daniel Spielman’s lecture note http://cs-www.cs.yale.edu/homes/spielman/462/lect04-13.pdf.
Outline

1. A History of Power Law
2. Preferential Attachment Model
3. Distribution of Low In-degrees
4. Distribution of In-degree
In the late 90’s, a number of papers were written that claimed that the distributions of the degrees of vertices in real-world graphs followed power laws.

Power law: there are constants $a$ and $c$ such that the fraction of vertices of degree $d$ equals

$$ad^{-c}.$$ 

The constant $c$ was usually between 2 and 3.

This was an attempt to quantify the observation that real graphs have an unusually large number of vertices of high degree.
Plots of Real-World Graphs

- The log of the fraction of nodes of degree $d$ would be
  \[ a - c \log d. \]
  This is a straight line in a log-log plot.

- We can also look at the Cumulative Distribution Function (CDF). The fraction of nodes of degree $\geq d$ is expected to be
  \[ \sum_{x \geq d} ax^{-c} \approx \int_{x \geq d} ax^{-c} = ax^{-c+1}/(c+1), \]
  which should also look linear on a log-log plot as well, but with a different slope.
Parameters: some probability $p$ and network size $n$.
Here are the steps (the rich get richer):

1. We begin with one node, which contains a directed edge to itself. This is the only self-loop we will include in this construction.

2. Add vertices one-by-one and create one edge leaving each.

3. When we add vertex $t + 1$:
   
   1. With probability $p$, choose the endpoint of the edge uniformly from the $t$ existing vertices.
   2. With probability $1 - p$, choose a random edge already in the graph and use its endpoint.
Let’s examine the expected number of in-degree 0 nodes. This approach is somewhat informal but can be made rigorous.

- A node only goes from in-degree 0 to in-degree 1 when it is chosen as a uniform random neighbor of some node that comes along later.

- If node \( j \) has in-degree 0 when node \( t + 1 \) is added, then the chance that node \( t + 1 \) links to node \( j \) is \( p/t \). Thus, the probability that node \( j \) has in-degree 0 after \( n \) nodes have been added is

\[
\prod_{t=j+1}^{n} \left(1 - \frac{p}{t-1}\right)
\]
For $j > \sqrt{n}$, this is well-approximated by

$$\prod_{t=j+1}^{n} \left(1 - \frac{p}{t-1}\right) \approx \prod_{t=j+1}^{n} \exp\left(-\frac{p}{t-1}\right) = \exp\left(-p \sum_{t=j}^{n-1} \frac{1}{t}\right)$$

$$\approx \exp\left(-p \log\left(\frac{n}{j}\right)\right) = \left(\frac{j}{n}\right)^{p}.$$ 

Use Euler’s famous result $\sum_{i=1}^{k} \frac{1}{i} \rightarrow \log k + \gamma$, where $\gamma$ is Euler’s constant. We have

$$\sum_{t=j}^{n-1} \frac{1}{t} \approx \log(n - 1) - \log(j - 1) \approx \log(n/j).$$

Thus the number of nodes of in-degree 0 should be approximately

$$\sum_{j=1}^{n} \left(\frac{j}{n}\right)^{p} = n^{-p} \sum_{j=1}^{n} j^{p} \approx n^{-p} \frac{1}{p + 1} n^{1+p} = \frac{n}{p + 1}.$$
Let $X_0(t)$ be the number of vertices of in-degree 0 after $t$ vertices have been added.

When we add node $t+1$, it has in-degree 0. Thus the number of in-degree 0 nodes will increase unless that node points to another of in-degree 0, which happens with probability $pX_0(t)/t$. So,

$$X_0(t+1) = \begin{cases} 
X_0(t) & \text{with probability } pX_0(t)/t \\
X_0(t) + 1 & \text{with probability } 1 - pX_0(t)/t.
\end{cases}$$

Thus

$$\mathbb{E}[X_0(t+1)] = \mathbb{E}[X_0(t)] + 1 - p\mathbb{E}[X_0(t)]/t.$$
Let \( Y_0(t) = \mathbb{E}[X_0(t)] \). We have

\[
Y_0(t+1) = Y_0(t) + 1 - pY_0(t)/t.
\]

We hope that \( Y_0 \) approaches \( c_0 t \) for some constant \( c_0 \), as \( t \) becomes large. In this case, \( c_0 \) should satisfy

\[
c_0 = 1 - pc_0 t/t = 1 - pc_0,
\]

which yields

\[
c_0 = \frac{1}{1+p},
\]

the result we obtained before.
Expected Number of In-degree \( k \) Nodes

- For higher in-degrees, let \( X_k(t) \) be the number of nodes of in-degree \( k \) after \( t \) vertices have been added.
- \( X_k \) decreases if the edge from node \( t + 1 \) hits a node of in-degree \( k \), which happens with probability

\[
\frac{pX_k(t)}{t} + \frac{(1 - p)kX_k(t)}{t}.
\]

- \( X_k \) increases if the edge hits a node of in-degree \( k - 1 \), which happens with probability

\[
\frac{pX_{k-1}(t)}{t} + \frac{(1 - p)(k - 1)X_{k-1}(t)}{t}.
\]
Thus

\[\mathbb{E}[X_k(t+1)] - \mathbb{E}[X_k(t)] = \frac{1}{t}(pX_{k-1}(t) + (1-p)(k-1)X_{k-1}(t) - pX_k(t) - (1-p)kX_k(t)).\]

If we look for a solution to these equations of the form \(Y_k = c_k t\), we get

\[c_k = pc_{k-1} + (1-p)(k-1)c_{k-1} - pc_k - (1-p)kc_k,\]

which gives

\[
\frac{c_k}{c_{k-1}} = \frac{p + (1-p)(k-1)}{1 + p + (1-p)k} = 1 - \frac{2-p}{(k+1) - p(k-1)}.
\]
When $k$ is large,

$$\frac{c_k}{c_{k-1}} = 1 - \frac{2-p}{(k+1) - p(k-1)} \approx 1 - \frac{2-p}{k(1-p)}.$$

So

$$\frac{c_k}{c_j} = \prod_{i=j+1}^{k} \left(1 - \frac{2-p}{1-p} \cdot \frac{1}{i}\right) \approx \left(\frac{j}{k}\right)^{\frac{2-p}{1-p}} = \left(\frac{j}{k}\right)^{1+\frac{1}{1-p}}.$$

This is equivalent to saying that for large $k$,

$$c_k \approx \beta \left(\frac{1}{k}\right)^{1+\frac{1}{1-p}}.$$