5

Scale-Free Networks

5.1 Power Laws

In the previous chapter we studied models for graphs with fixed degree sequences. We will continue this topic by studying a specific type of degree distribution that we often observe empirically: power law degree distributions. We will then introduce an important probabilistic model that has been proposed to explain this effect.

A network is said to have a power law degree distribution if the fraction $P(k)$ of nodes in the network that have degree k is (asymptotically)

$$P(k) \sim k^{-\gamma}$$

where γ is a parameter of the network.[1] Such networks are said to be *scale-free*. A power law has a *heavy tail*, which means that values far beyond the mean are likely to occur; e.g., this is in contrast to light-tailed distributions such as exponentials, although many other heavy tailed distributions, such as Lognormal distributions, exist.

[1] Observed values for γ are often in the range $2 < \gamma < 3$, although this is not always the case.

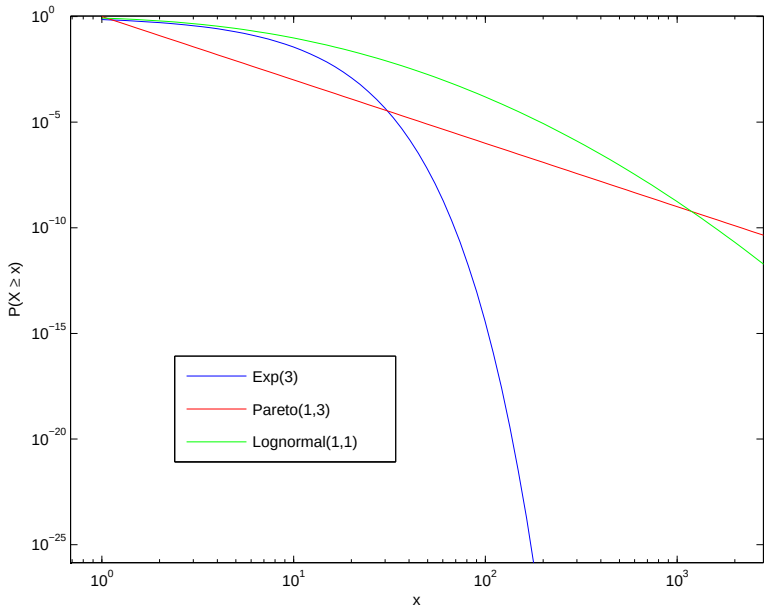

Figure 5.1: The c.c.d.f. of the exponential, Pareto (i.e. power-law), and lognormal distributions.

History

Power laws have a very long and controversial history. Power laws have been observed in many contexts, such as income distributions, file sizes on computers, object sizes on the web, lengths of phone calls, the sizes of cities, and word frequencies in prose. Whenever a phenomenon occurs in such a wide variety of seemingly unrelated scenarios, it is tempting to suspect that some very simple underlying mechanism can be discovered that explains this phenomenon. As an example, another very prominent probability distribution, the normal law, occurs very naturally whenever independent random variables are added, which is captured in the central limit theorem (CLT).

However, for power laws, the quest for such a hidden mechanism is not quite as settled as for the normal law. Most if not all of the above examples of power law distributions "in the wild" have also been challenged at some point or another. Moreover, the mechanism though which they arise is not universally agreed upon. One class of mechanisms that could lead to power law distributions could be termed "the rich get richer"; its study was pioneered by Yule in the 1920s, and by the economist Herbert Simon in the 1950s. In the context of network degree distributions, this mechanism was rediscovered under the name of *preferential attachment* by Albert and Barabási, and is of particular interest to computer scientists because it is a *dynamic* model. Nodes arrive sequentially into a pre-exisitng system, and form connections that depend on the current structure of the network. Such dynamic models are useful for capturing the evolution of networks that grow (or shrink) in size over time, such as the internet (both physical and virtual) and social networks.

Pareto Distributions

A prominent and simple example of a power law distribution is the Pareto distribution. Let X be a random variable denoting the degree of a vertex that follows the Pareto law. Then, the c.c.d.f. (of the Pareto distribution) is

$$\bar{F}(x) = \mathbb{P}r(X > x) = \begin{cases} \left(\frac{x}{\beta}\right)^{-\gamma} & \text{if } X \geq \beta, \\ 1 & \text{otherwise.} \end{cases}$$

The exponent γ is sometimes called the Pareto index. The moments of the Pareto distribution only exist up to $k < \gamma$; higher moments are always infinite. Moreover, all moments are captured by a simple expression:

$$\mathbb{E}\left[X^k\right] = \begin{cases} \left(\frac{\beta^k \gamma}{\gamma - k}\right) & \text{if } k < \gamma, \\ \infty & \text{otherwise.} \end{cases}$$

Thus, if $\gamma \leq 1$, then X is said to have ∞-mean, and for $1 < \gamma \leq 2$, then X is said to have ∞-variance.

5.2 The Rich-Get-Richer Phenomena

Herbert Simon, a famous economist, introduced a model in a 1955 paper to explain the heavy-tailed distribution in incomes first observed by Vilfredo Paret[2]. His model could be termed "the rich get richer"; the intuition is that someone who is rich has more opportunities to earn their next dollar than someone who is poor.

Simon explained his model in terms of word frequencies in prose in the following manner: Assume that $X_i(t)$ is the number of different words that have occurred i times among the first t words in the text. Then, assume that with some probability α, the next word typed is a new word (i.e., a word that has not occurred among the first t words), and with probability $1 - \alpha$, the next word is sampled uniformly amongst the t words in the text so far. Thus, the probability to select a word of class i is proportional to $iX_i(t)$. In other words, there is a bias for the next word to be one that has already occurred quite often.

Another intuitive example arises when we consider the sizes of cities. Suppose you move to a new country and try to decide what city or town to settle in. It is unlikely that you will sample from the set of cities and towns uniformly to decide where to move; rather, your behavior is probably representative of the behavior of people before you, i.e., you are much more likely to settle in a large city than in a small town. In other words, you sample from the *population* rather from the set of cities and towns. This may go same way towards explaining a power law in city sizes. What is remarkable that even if the sizes of cities are initially equal, the cumulative effect of random decisions by new arrivals would give rise to the power law over time; we will see this effect in the preferential attachment model that we study.

Preferential Attachment

The Simon model is an example of a growth model, where objects arrive sequentially in a system, i.e., the system is never in equilibrium. The Barabási-Albert (BA) model [1] proceeds along the same lines: at each time step, a node node arrives to the system, and establishes one or several edges to existing nodes. We will consider one variation of this model, which proceeds as follows: At each time step, a new node arrives and with probability α connects to one of the existing nodes uniformly at random, and with probability $1 - \alpha$ connects to a node v with probability proportional to $d_{in}(v)$ (where $d_{in}(v)$ is the node in-degree of v).

Let us now consider the degree distribution of this network after t time steps. Let $X_j(t)$ denote the number of nodes v with $d_{in}(v) = j$ at time t (with some abuse of notation, we sometimes write X_j for $X_j(t)$ when the t is clear from context). Note that the total number of nodes and the total number of edges at time t is always t. However, the degree distribution is constantly changing. For any $j \geq 1$, we can explicitly write down these dynamics as follows:

$$\mathbb{P}\left\{X_j(t+1) = X_j(t) + 1\right\} = \alpha \frac{X_{j-1}}{t} + (1 - \alpha)(j-1)\frac{X_{j-1}}{t},$$

[2]Who taught at the University of Lausanne.

and

$$\mathbb{P}\{X_j(t+1) = X_j(t) - 1\} = \alpha\frac{X_j}{t} + (1-\alpha)j\frac{X_j}{t}.$$

Thus, the expected change in X_j is

$$\mathbb{E}[X_j(t+1) - X_j(t)] = \frac{\alpha(X_{j-1} - X_j) + (1-\alpha)[(j-1)X_{j-1} - jX_j]}{t}/$$

We now use a standard trick: a continuous time mean-field approximation. In effect, we will approximate the above discrete process with a *continuous* process which we can then evaluate using basic calculus. In this case, we obtain

$$\frac{dX_j}{dt} = \frac{\alpha(X_{j-1} - X_j) + (1-\alpha)[(j-1)X_{j-1} - jX_j]}{t} \tag{5.1}$$

for $j \geq 1$. Similarly, for $j = 0$, we get that

$$\frac{dX_0}{dt} = 1 - \frac{\alpha X_0}{t}.$$

Let us assume that the fraction of nodes of degree j converges, i.e., $X_j/t \to c_j$. We then solve for c_j:

$$c_0 = \frac{1}{1+\alpha} \tag{5.2}$$

$$\frac{c_j}{c_{j-1}} = \frac{\alpha + (j-1)(1-\alpha)}{1+\alpha+j(1-\alpha)} = 1 - \frac{2-\alpha}{1+\alpha+j(1-\alpha)} \sim 1 - \frac{2-\alpha}{1-\alpha}j^{-1}. \tag{5.3}$$

Note that

$$\left(\frac{j}{j-1}\right)^{-\beta} = (1 - \frac{1}{j})^{\beta} \sim 1 - \frac{\beta}{j}. \tag{5.4}$$

Therefore, the tail behavior of c_j is

$$c_j \sim c \cdot j^{-\frac{2-\alpha}{1-\alpha}} \tag{5.5}$$

Therefore, the exponent of the c.d.f. is $\gamma = 1/(1-\alpha)$.

Obviously, the above argument is not rigorous; in (5.1), we have replaced a discrete random process with a continuous deterministic function. Also, we have assumed that X_j/t converges, which requires a proof. This type of approximation is sometimes referred to as "continuum theory" [1][3]. A mathematically rigorous argument proceeds by first showing that $X_j(t)$, appropriately rescaled, is a martingale, and then applying a concentration result known as Azuma's inequality (which can be roughly viewed as equivalent to the CLT for martingales). This allows us to show that the process $X_j(t)$ is unlikely to depart too much from its expectation. An example of a rigorous treatment of the subject is given in [2].

We now turn to the BA model. Start with m_0 nodes. At each time step t, add $m \leq m_0$ edges from the new node to the set of existing nodes, such that

$$\mathbb{P}\{\text{connect to } v\} = \frac{d_v}{\sum_{u=1}^{t+m_0} d_u} \tag{5.6}$$

After t steps, there are $m_0 + t$ nodes and $mt + e_0$ edges in the graph, where e_0 is the number of initial edges. The analysis of the BA model proceeds analogously to the one for the version with directed edges above, and with $\alpha = 0$. However, in their model, $\gamma = 3$.

[3] although "continuum approximation" would be a more appropriate label.

5.3 Power Law vs. Lognormal Law

There is another probability law called the *lognormal* law, which is easily confused with a power law. A random variable Y has lognormal distribution if it can be written as $X = \log Y$ and X has normal distribution.

The p.d.f. of a lognormal random variable with parameters μ and σ^2 is given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2/2\sigma^2}. \tag{5.7}$$

To see why the lognormal distribution can be hard to distinguish from a power law, write its p.d.f. as

$$\begin{aligned} \log p(x) &= -\log x - \frac{1}{2}\log(2\pi\sigma^2) - \frac{(\log x - \mu)^2}{2\sigma^2} \\ &= -\frac{(\log x)^2}{2\sigma^2} + \big(\mu/\sigma^2 - 1\big)\log x - \left(\frac{1}{2}\log(2\pi\sigma^2) + \frac{\mu^2}{2\sigma^2}\right) \end{aligned} \tag{5.8}$$

Note however that the tail of the lognormal distribution is light enough for all moments to exist. This is because

$$\exp(-(\log x)^2) = \frac{1}{x^{\log x}}, \tag{5.9}$$

which decreases faster than any polynomial.

Bibliography

[1] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, vol. 286:pp. 509–512, 1999.

[2] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the Web graph. In *Proc. Foundations of Computer Science (FOCS)*, pages 57–65, 2000.