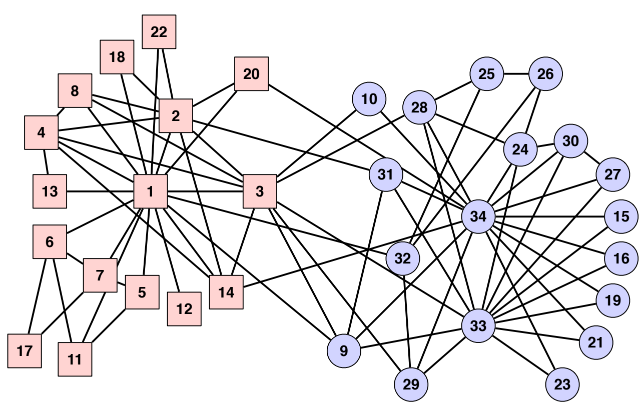
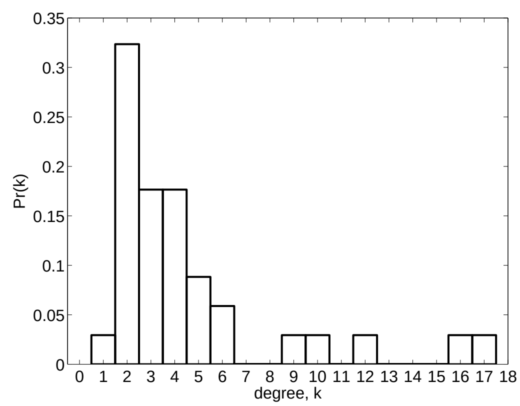
Network Analysis and Modeling, CSCI 5352 **Prof. Aaron Clauset**
Lecture 6 **17 September 2013**

1 Degree distributions

In many ways, the degrees of vertices in a network are a fundamental network property, correlating with or driving many other kinds of network patterns. We already saw some of this in our discussion of centrality scores and degree assortativity. The same is true for many other phenomena, including processes that run on top of the network structure. A network's *degree distribution* captures the pattern of a network's degrees by quantifying the relative frequency or probability of different levels of connectivity.[1]

The degree distribution may be denoted by p_k or $\Pr(k)$ or $p(k)$. For instance, consider again the karate club network. In this network, a few vertices (1, 33 and 34) have very high degree, while most other vertices have relatively low degree. We tabulate this network's degree distribution by counting the number of times each possible degree value occurs, and then normalizing by the number of vertices: $p_k = (\#\text{ vertices with degree } k)/n$, for $k \geq 0$. This probability mass function or distribution (pdf) is a normalized histogram of the observed degree values.

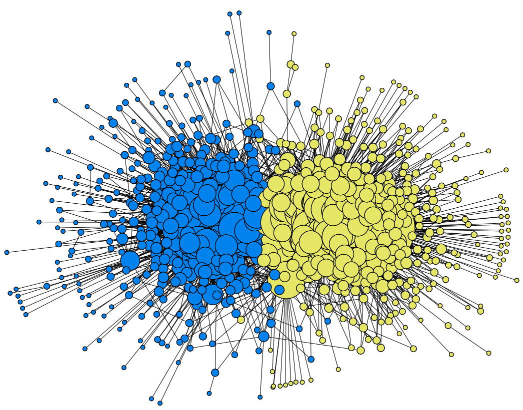


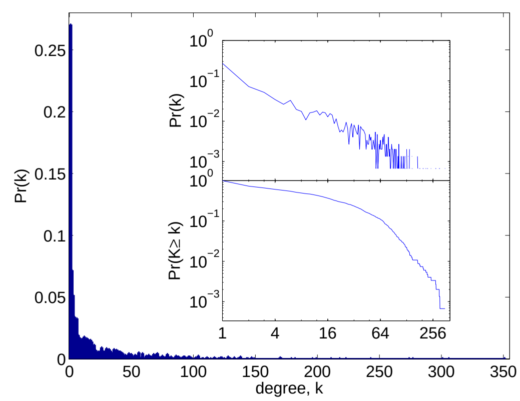


Degree distributions in real networks are often "heavy tailed," meaning that as k increases, the remaining proportion of vertices with degree *at least* k decreases more slowly than it would in a geometric or exponential distribution. In these situations, the shape of the pdf becomes very noisy for large values of k, because there are either zero or one (usually zero) vertices with that value in the network. This makes the pdf difficult to interpret, particularly around the higher-degree vertices, which are often of specific interest.

[1] Another property that captures similar information is the *degree sequence*, which is the (unordered) list of the degrees in the network, i.e., $\{k_1, k_2, k_3, \ldots, k_n\}$. We will encounter the degree sequence again when we cover random graphs.

For instance, consider the political blogs network[2] we saw in a previous lecture, shown here along side the pdf of its degree distribution (where we treat edges as undirected; $n = 1490$, $m = 33430$). The main figure shows a long tail stretching out above $k = 50$, which is made almost invisible by the large fraction of vertices with degree $k < 10$. The upper inset shows the same data on double logarithmic or "log-log" axes. Here, the tail is more visible but is extremely noisy, and the highest-degree vertices appear as a few dots on the far right side of the figure.





The solution is to instead plot the complementary cumulative distribution function (ccdf), which is exactly the fraction of vertices with degree at least k, and is denoted $\Pr(K \geq k)$, where K represents a random variable drawn iid from the distribution.[3] This empirical function always begins at 1, as all vertices have degree at least as large as the small value. As we increase k, the ccdf decreases by a factor of $1/n$ for each vertex with degree k, until it reaches a value of $1/n$ at $k = \max(k_i)$, the largest degree vertex in the network. The ccdf is typically plotted on doubly-logarithmic axes, which allows us to plot both small and very large degrees, and moderate and very small probabilities, on the same figure.

The lower inset above shows the ccdf for political blogs network, which has a much smoother shape, revealing interesting structure: the curvature of the ccdf seems to change around $k = 64$ or so, decreasing slowly before that value and much more quickly after. Furthermore, about 11% of the vertices have degree $k \geq 64$, making the tail a non-trivial fraction of the network.

[2]Network image from Karrer and Newman, *Phys. Rev. E* **83**, 016107 (2011) at `arxiv:1008.3926`. Vertices are colored according to their ideological label (liberal or conservative), and their sizes are proportional to their degree. Data from Adamic and Glance, *WWW* Workshop on the Weblogging Ecosystem (2005).

[3]Mathematically, $\Pr(K \geq k) = 1 - \Pr(K < k)$, where $\Pr(K < k)$ is the cumulative distribution function or cdf.

Commentary on degree distributions.
The shape of the degree distribution is of general interest in network science. It tells us how skewed the distribution of connections is, which has implications for other network summary statistics, inferences about large-scale structural patterns, and the dynamics of processes that run on top of networks. The degree distribution is also often the first target of analysis or modeling: What pattern does the degree distribution exhibit? Can we model that pattern simply? Can we identify a social or biological process model that reproduces the observed pattern?

This latter point is of particular interest, as in network analysis and modeling we are interested not only in the pattern itself but also in understanding the process(es) that produced it. The shape of the degree distribution, and particularly the shape of its upper tail, can help us distinguish between distinct classes of models. For instance, a common claim in the study of empirical networks is that the observed degree distribution follows a *power law* form (see below), which in turn implies certain types of exotic processes. Although many of these claims end up being wrong, the power-law distribution is of sufficient importance that we will spend the rest of this lecture learning about their interesting properties.

2 Power-law distributions

A power-law distribution is a special kind of probability distribution. The simplest form of a power-law distribution is defined on a continuous random variable as

$$p(x) = Cx^{-\alpha} \qquad \text{for } x \geq x_{\min} \ , \tag{1}$$

where the normalization constant $C = (\alpha - 1)x_{\min}^{\alpha-1}$ is derived in the usual way.[4] Note that this expression only makes sense for $\alpha > 1$, which is indeed a requirement for a power law to be a valid probability distribution.[5] A little algebra allows us to rewrite Eq. (1) in a more compact form

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha} \qquad \text{for } x \geq x_{\min} \ . \tag{2}$$

The cumulative distribution function (cdf) also has a very simple form. (Exercise: derive it.) Figure 1a shows examples of three power-law distributions, illustrating their signature pattern: a straight line on log-log axes.

[4] For $p(x)$ to be a pdf, it must satisfy the identity $1 = \int_{-\infty}^{\infty} p(x) \mathrm{d}x$. The power-law distribution is only defined on $x \geq x_{\min}$, and thus we implicity define $p(x) = 0$ for $x < x_{\min}$.

[5] Mathematically, the only way to have something that behaves like a power-law distribution but with a heavier tail than $\alpha \gtrsim 1$ is to effectively truncate its upper range, e.g., by adding an exponential cutoff in the upper tail $\Pr(x) \propto x^{-\alpha}\mathrm{e}^{-\lambda x}$, or by having a finite range or hard upper cutoff $x_{\min} \leq x \leq x_{\max}$.

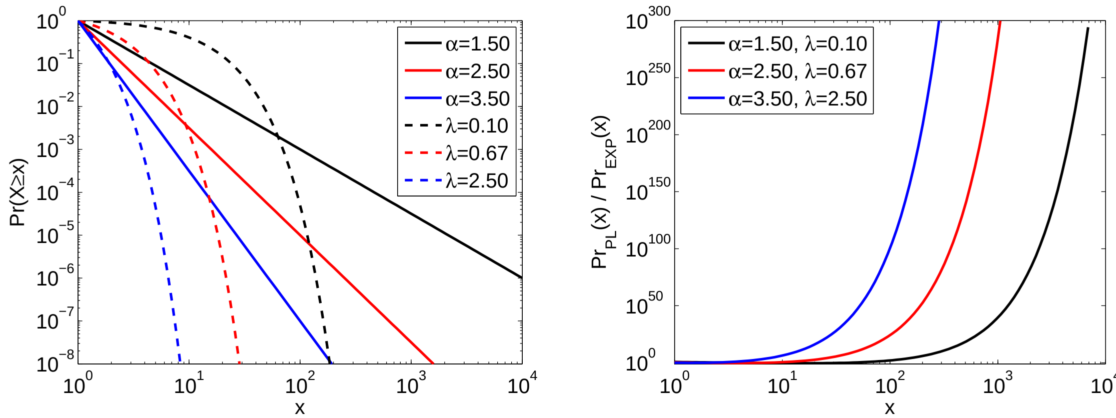


Figure 1: (a) Power-law and exponential distributions, for several choices of parameters, all with $x_{\min} = 1$. (b) The ratio of power-law and exponential distributions, illustrating that events that are effectively "impossible" (negligible probability under an exponential distribution) become practically commonplace under a power-law distribution.

2.1 Power laws have unusual properties

Power-law distributions are interesting in part because they exhibit unusual properties, but also because they can indicate the presence of interesting network mechanisms. They also exhibit several counter-intuitive behaviors.

Many empirical quantities tend to cluster around a typical value. For instance, the speeds of cars on a highway, the weights of apples, air pressure in a room, sea level over a year, the temperature in New York at noon on Midsummer's Day. All of these things vary somewhat, but their distributions place a negligible amount of probability far from the *typical* value, making that typical value representative of nearly all observations. For instance, it is entirely reasonable to say that adult males American are about 170cm (about 5.6 feet) tall, because not one of the 100 million-odd members of this group are more than a factor of 2 above or below it. Even the largest deviations, which are exceptionally rare (one person out of 100 million is a probability of 10^{-8}), are within this range and hence quoting a mean and standard deviation provides an accurate summary. The underlying processes that generate such distributions fall into a general class well-described by the Central Limit Theorem.

However, many quantities in complex social, biological or technological systems, and many properties associated with networks, do not fit this pattern. Heavy-tailed distributions in general, but power laws in particular, can indicate the presence of interesting endogenous processes like feedback loops, self-organization, network effects, etc.

A classic example of a power-law distributed quantity is the population of cities. Data from the 2000 United States Census on the 600 largest cities exemplifies the weirdness of power laws.[6] Among these, the average population is $\langle x\rangle = 165{,}719$, and metropolises like New York City and Los Angles seem to be "outliers" relative to this size. A clue that city sizes are not well explained by a Normal distribution is that the sample standard deviation $\sigma = 410{,}730$ is significantly larger than the sample mean. If we modeled the city data as a Normal, we would expect to see roughly half as many cities as large or larger than Albuquerque (population 448,607) than we actually do. Furthermore, we would never expect to see a city as large as New York City (population 8,008,278), which is more than 12σ above μ, and expected size of the largest city would be Indianapolis (population 781,870).[7]

A more whimsical and fictional example would be a world where the heights of Americans were distributed as a power law, with approximately the same average as the true distribution (which is convincingly Normal when certain exogenous factors are controlled). In this case, we would expect nearly 60,000 individuals to be as tall as the tallest adult male on record (2.72m tall). However, we would also expect 10,000 individuals as tall as a giraffe, one individual as tall as the Empire State Building (381m), and 180 million diminutive individuals standing 17cm tall. This same analogy was cleverly used in 2006 to describe the counter-intuitive nature of the extreme inequality in the wealth distribution in the United States, whose upper tail is often said to follow a power law.[8]

Moments and fluctuations.
Power laws also exhibit a number of interesting mathematical properties, many of which derive from the distribution's extreme right-skewness and the fact that only the first $\lfloor \alpha - 1 \rfloor$ moments of a power-law distribution exist; all the rest are infinite. In general, the kth moment is defined as

$$\begin{aligned} \langle x^k \rangle &= \int_{x_{\min}}^{\infty} x^k \, p(x) \mathrm{d}x \\ &= (\alpha-1)/x_{\min}^{\alpha-1} \int_{x_{\min}}^{\infty} x^{-\alpha+k} \mathrm{d}x \\ &= x_{\min}^{k} \left(\frac{\alpha-1}{\alpha-1-k} \right) \qquad \text{for } \alpha > k+1 \ . \end{aligned} \tag{3}$$

Thus, when $1 < \alpha < 2$, the first moment (the mean or average) is infinite, along with all the higher moments. When $2 < \alpha < 3$, the first moment is finite, but the second (the variance) and higher

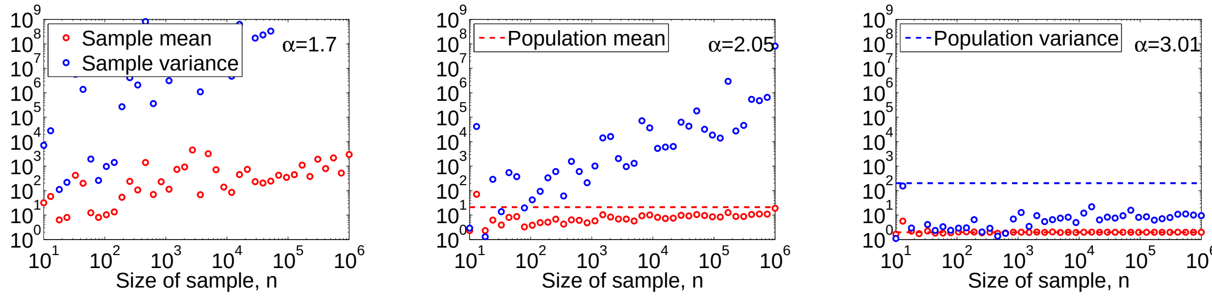
[6] See `http://www.demographia.com/db-uscity98.htm`

[7] The expected maximum size for n iid random variables is given by solving for $x_{\max}$ in $\frac{1}{n} = \int_{x_{\max}}^{\infty} \Pr(x) \mathrm{d}x$.

[8] See `http://www.theatlantic.com/magazine/archive/2006/09/the-height-of-inequality/5089/` .
The upper tail of the wealth distribution does not in fact follow a perfect power law because there are statistically significant deviations in it, which appear because the wealth of individuals are not iid random variables.

moments are infinite. In contrast, all the moments of the vast majority of other pdfs are finite.

Divergent moments have a real impact on the convergence rates of sample statistics like the sample mean or variance. Figure 2 demonstrates this point numerically using synthetic data drawn from power laws with parameters $\alpha = \{1.7, 2.05, 3.01\}$, for different sizes of the sample.



Scale invariance.
Another interesting property of power-law distributions is "scale invariance." Consider the density at $p(x)$ and at some other $p(c\,x)$, where c is some constant. For a power-law distribution, no matter the choice of x, these densities are always proportional, $p(c\,x) \propto p(x)$. Mathematically:

$$\begin{aligned} p(c\,x) &= (\alpha-1)x_{\rm min}^{\alpha-1}(c\,x)^{-\alpha} \\ &= c^{-\alpha}\left[(\alpha-1)x_{\rm min}^{\alpha-1}x^{-\alpha}\right] \\ &\propto p(x) \ . \end{aligned}$$

Thus, under a power law, the relative likelihood between "small" and "large" events is always the same, no matter what scale we choose to make the comparison. This behavior is what we mean mathematically by the term "scale invariant." In fact, the power-law distribution is the only distribution with this property. (Can you prove this?).

Scale invariance implies the signature pattern of a power law: a straight line on log-log axes. Taking the logarithm of both sides of Eq. (1), we obtain an expression for $\ln p(x)$ that is linear in $\ln x$,

$$\begin{aligned} \ln p(x) &= \ln\left[(\alpha-1)x_{\rm min}^{\alpha-1}(x)^{-\alpha}\right] \\ &= \ln C - \alpha \ln x \ . \end{aligned}$$

Thus, changing scales from $x \to c\,x$ (e.g., changing our measurement unit from meters to millimeters, or from micrograms to kilograms) simply shifts the power law up or down on a logarithmic scale. Non-scaling distributions, like the exponential, instead exhibit curvature on log-log axes, which implies a sensitivity to the choice of scale.

2.2 Top-heavy distributions and the 80–20 "rule"

All right-skewed or heavy-tailed distributions exhibit strong forms of inequality, which we can represent using a Lorenz curve (after Max Otto Lorenz, 1880–1962, an American economist). This curve plots the fraction of total "wealth" W held by the richest fraction P of the population. For a network, the wealth of a vertex is simply its degree. The mathematical form of a power-law distribution produces a Lorenz curve with particularly simply structure, but any distribution, including an empirical one, can be converted into a Lorenz curve.

Under the power law, the fraction P of the population whose wealth is at least x is given by the ccdf:

$$P(x) = \int_{x}^{\infty} C\, y^{-\alpha} \mathrm{d}y = \left(\frac{x}{x_{\min}}\right)^{-\alpha+1} \quad , \tag{4}$$

where C is the normalization constant. The fraction all wealth held by those people is

$$W(x) = \frac{\int_{x}^{\infty} y\, p(y) \mathrm{d}y}{\int_{x_{\min}}^{\infty} y\, p(y) \mathrm{d}y} = \left(\frac{x}{x_{\min}}\right)^{-\alpha+2} \quad , \tag{5}$$

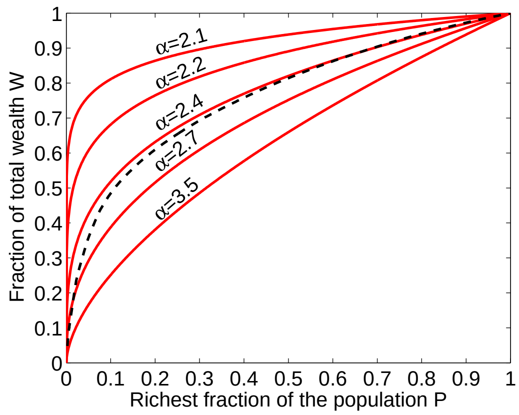
for $\alpha > 2$. Solving Eq. (4) for $x/x_{\min}$, and substituting the result into Eq. (5) produces an expression, the Lorenz curve, equating W to P that does not depend on x

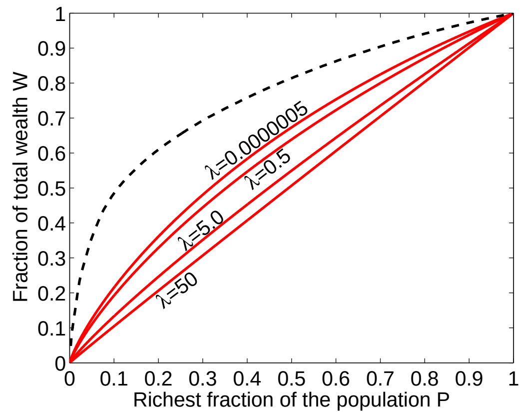
$$W = P^{(\alpha-2)/(\alpha-1)} \quad . \tag{6}$$

To illustrate the Lorenz curve, the figures on the next page show the curves for several different power-law distributions along side several curves for an exponential distribution. The steep increase of the power-law Lorenz curves illustrates the idea of a "top-heavy" distribution, in which a very small fraction of individuals (vertices) hold a very large fraction of the wealth (edges). In contrast, the exponential Lorenz curves exhibit very little variation, indicating that the wealth is spread fairly evenly among the population.

A particular form of top-heaviness, in which 80% of the wealth is held by the richest 20% of people, is sometimes called the "80–20 rule." This level of inequality is not the most extreme, however. As $\alpha \to 2$ the inequality becomes progressively more extreme, with a smaller fraction of the population holding a greater proportion of the total wealth. When $\alpha < 2$, the integrals in our calculation above diverge and the total wealth is effectively held by a single person, i.e., the sum of all the wealth is largely equal to the largest value in the sum.

A compact summary of the curve, and the degree of inequality it represents, is the *Gini coefficient* G. This coefficient is defined as twice the area between the observed $W(P)$ function and the "perfect equality" function $W = P$. Since the maximum area between the two curves is 1/2 (when





one individual holds all the wealth), G is a real value on the unit interval $[0, 1]$ with larger values indicating more skewed distributions. As a point of reference, the Lorenz curve figures here show the curve (dashed line) for the distribution of wealth among the wealthiest 400 individuals (according to Forbes in 2003), which yields $G = 0.527$. For more information, the Wikipedia page for Gini coefficients[9] has a nice map, derived from the CIA *World Fact Book 2009*, showing Gini coefficients for most countries worldwide.

2.3 Power-law tails

Very few empirical degree distributions, or degree distributions produced by mechanistic models, create perfect power laws. Instead, they often exhibit a "body" or "shoulder" in the ccdf, in which some non-power-law part holds smaller values of x and the power-law part only holds above some value. In this case, we say the distribution has a power-law *tail*. These distributions can generally be expressed in the form $\Pr(x) = L(x)\, x^{-\alpha}$, where $L(x)$ represents a "slowly varying function," i.e., as $x \to \infty$, $L(x) \to c$, where c is some constant, and $p(x) \to x^{-\alpha}$.

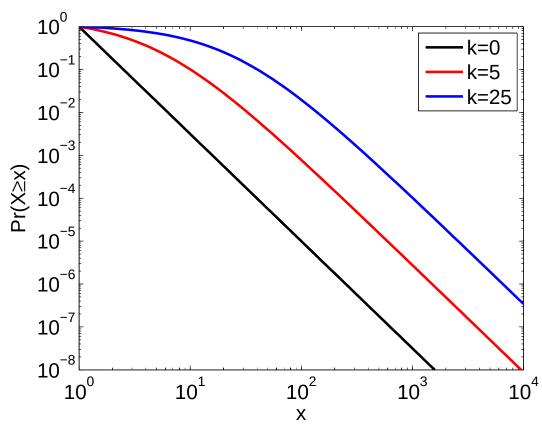
There are many possible forms we might choose for $L(x)$. One that crops up occasionally in network science is the *shifted power-law* distribution, which has the form $\Pr(x) = C(k + x)^{-\alpha}$ for $x \geq x_{\min}$, and constant $k \geq 0$. When $k = 0$, we recover the pure power law exactly. This distribution has a power-law tail:

$$\begin{aligned} \Pr(x) &= C(x+k)^{-\alpha} \qquad\qquad \text{for } x \geq x_{\min} \\ &= C(x+k)^{-\alpha} \left(\frac{x^{-\alpha}}{x^{-\alpha}} \right) \end{aligned}$$

[9] See `http://en.wikipedia.org/wiki/Gini_coefficient`

$$= C\left(1+\frac{k}{x}\right)^{-\alpha} x^{-\alpha}$$
$$= L(x)\, x^{-\alpha} \ ,$$

where $L(x) = C\left(1+\frac{k}{x}\right)^{-\alpha}$. In the limit of $x \to \infty$, $L(x) \to 1$, which satisfies the requirements for a power-law tail. The function $L(x)$ describes exactly how the deviations from the power-law form decay as we move further out into the tail. When $x \lesssim k$, the "body" term $L(x)$ is large compared to the tail term $x^{-\alpha}$, leading to curvature on the log-log plot. To illustrate this point, here are a few shifted power-law distributions, with progressively greater choices of k.



3 At home

1. Read Chapter 8.1–8.4 (pages 235–260) in *Networks*
2. Next time: more degree distributions