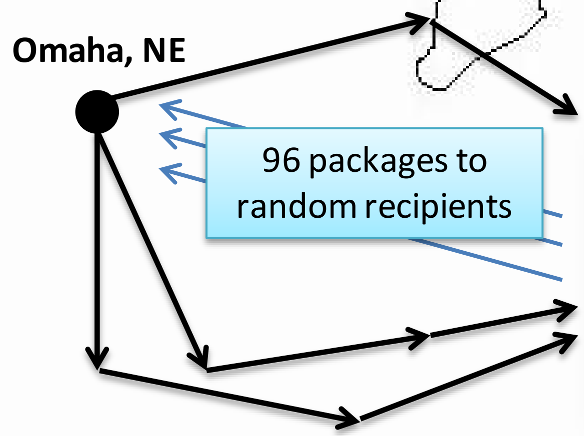
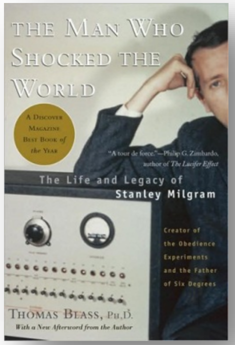
# Small-world Phenomena

- *Name, address, occupation* of the target were known; no sending was allowed
- 18 packages returned back to Boston
- mean path result was just 5.9 steps
- small-world effect was confirmed in many other experiments

Bonus observations in the experiment
- most of the packages were received through 3 target's friends
- people are good in finding short paths (later was shown that it is hard to find shortest path without knowing full information)

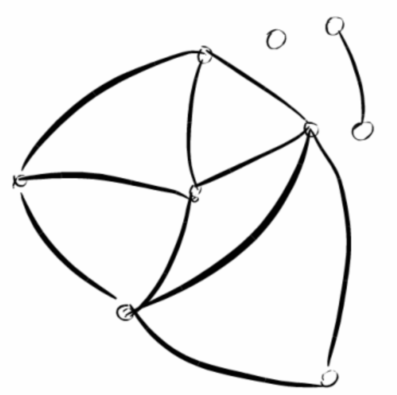



**Stanley Milgram**
**1933-1984**




Similar experiments
- emails: only 384 out of 24K were received/ results confirmed, 4 steps
- Microsoft .NET Messenger Service: 6.6 people

# Degree Distributions



$$p_0 = \frac{1}{9},\ p_1 = \frac{2}{9},\ p_2 = \frac{1}{9},\ p_3 = \frac{2}{9},\ p_4 = \frac{3}{9}$$

The probability that randomly chosen node has degree *k*

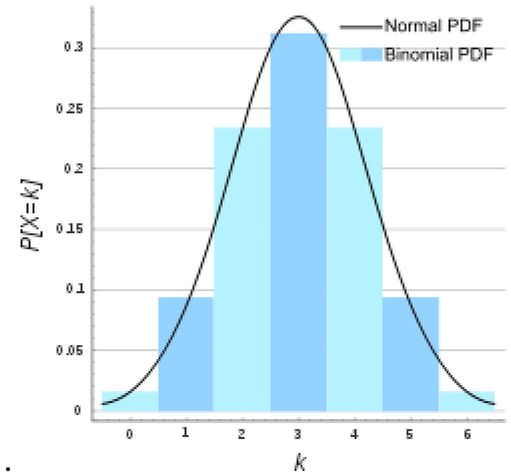
Classical undirected random graph models $G_{n,p}$

choose *k* neigh among *n-1*

prob of being connected to exactly k neigh

$$\binom{n-1}{k} p^k (1-p)^{n-1-k} \quad \text{(Binomial distribution)}$$

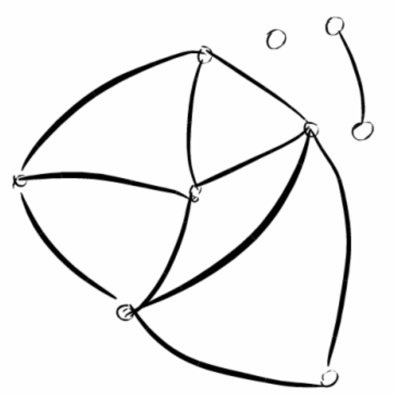
when graphs are small




when graphs are large (*n* is assumed to be large, mean degree is approximately constant as the network grows)

$$\frac{(np)^k}{k!} e^{-np} \quad \text{(Poisson distribution)}$$

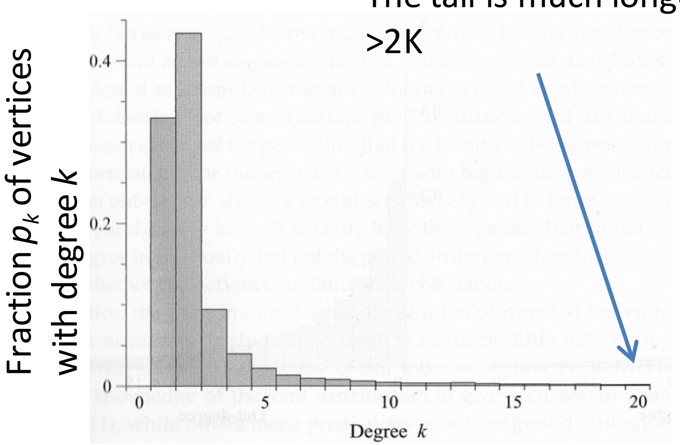
# Degree Distributions



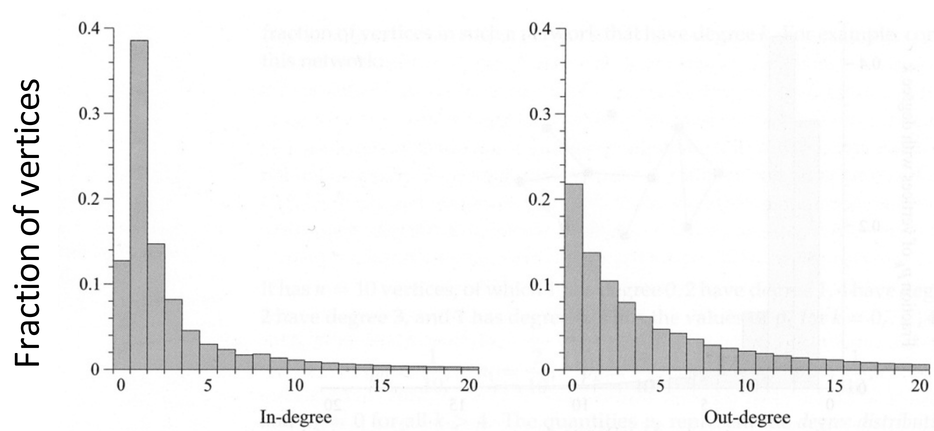
$$p_0 = \frac{1}{9},\ p_1 = \frac{2}{9},\ p_2 = \frac{1}{9},\ p_3 = \frac{2}{9},\ p_4 = \frac{3}{9}$$

The probability that randomly chosen node has degree *k*

The tail is much longer >2K



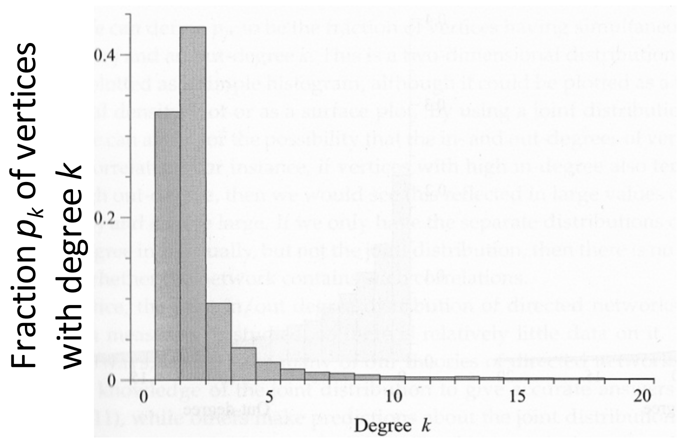

**Internet at the level of autonomous systems**




**World Wide Web**

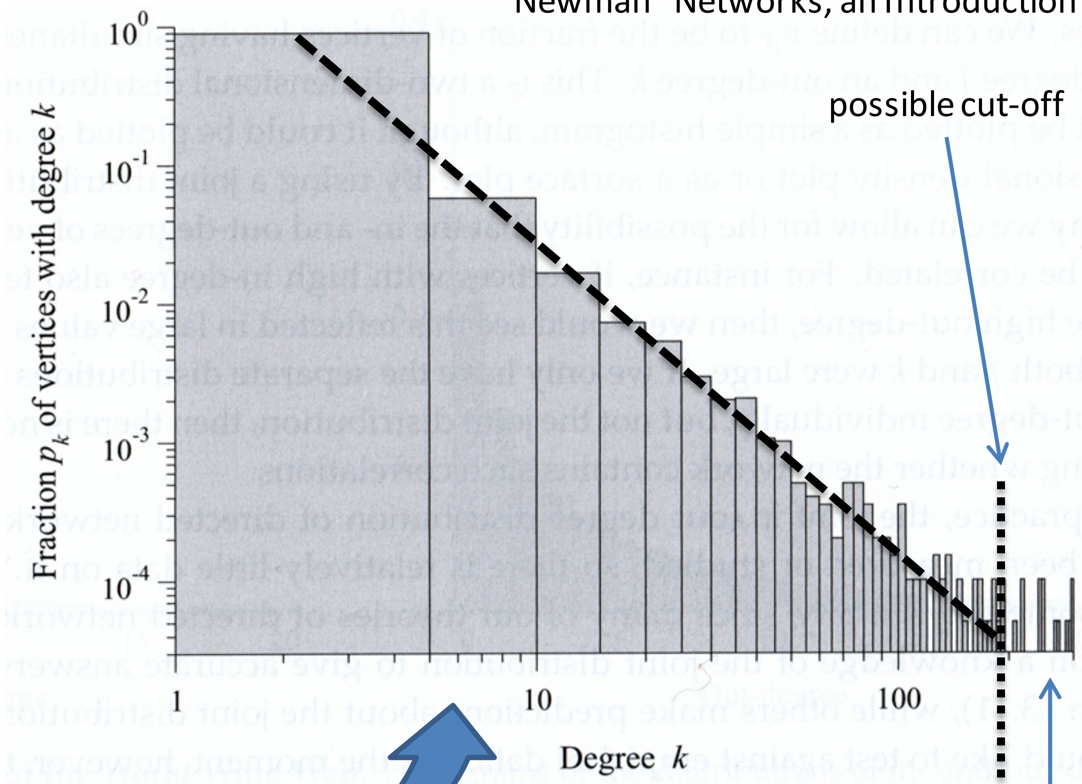
Newman "Networks, an Introduction"

# Power Laws (aka scale-free)




**Internet at the level of autonomous systems**

Newman "Networks, an Introduction"




area of possible fluctuations

logarithmic scales; bigger range of bins

$$\ln p_k = -\alpha \ln k + c \text{ or } p_k = Ck^{-\alpha}, \text{ where } C = e^c$$

$$\text{typical } \alpha \in [2,3] \text{ (see handout Table 8.1)}$$

**Problem of histograms**: statistics is poor at the tail of the distribution

**Solution I**: different sizes of bins

# Power Laws: Logarithmic Binning

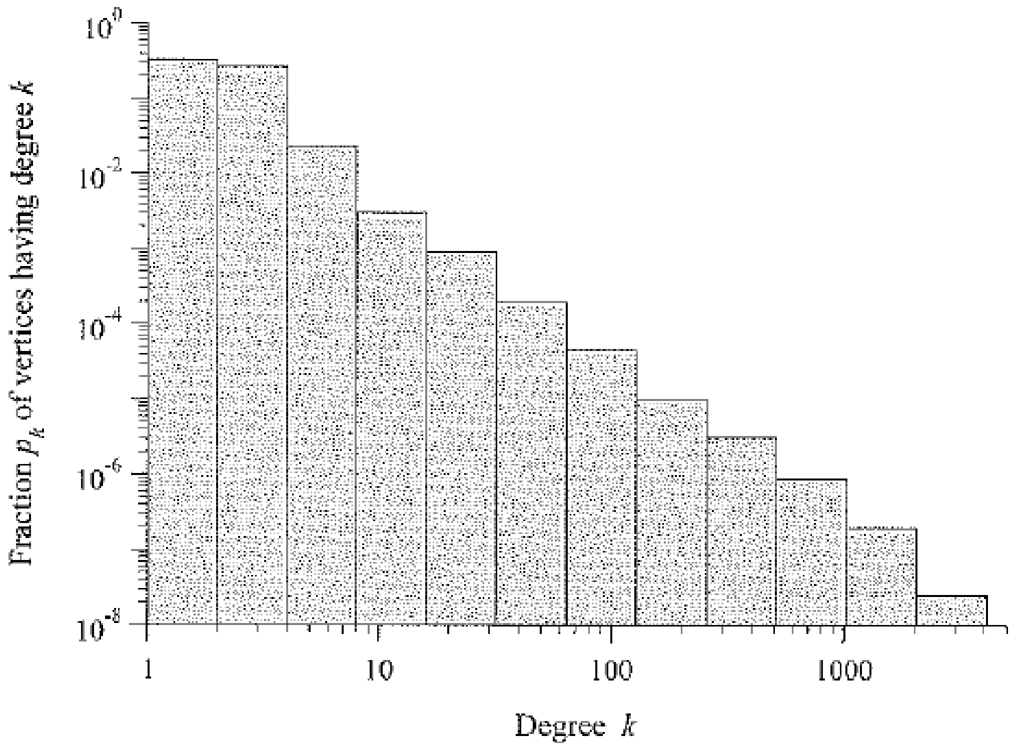



**Figure 8.6: Histogram of the degree distribution if the Internet, created using logarithmic binning.** In this histogram the widths of the bins are constant on a logarithmic scale, meaning that on a linear scale each bin is wider by a constant factor than the one to its left. The counts in the bins are normalized by dividing by bin width to make counts in different bins comparable.

- Bin 1 covers degrees in [1,2)
- Bin 2 covers degrees in [2, 4)
- Bin 3 covers degrees in [4, 8)
- …

Width of bins can vary

# Cumulative Distribution

Probability at a random vertex has degree $k$ or greater

$P_k = \sum_{k'=k}^{\infty} p_{k'}$

Let $p_k$ follows a power law in its tail, i.e., $p_k = Ck^{-\alpha}$ for $k \geq k_{\min}$. Then

$$P_k = C \sum_{k'=k}^{\infty} k'^{-\alpha}$$
$$\approx C \int_k^{\infty} k'^{-\alpha} \, \mathrm{d}k' = \frac{C}{\alpha - 1} k^{-(\alpha-1)}$$

$$\alpha = 1 + N \left( \sum_i \ln \frac{d(i)}{k_{\min} - 1/2} \right)^{-1}$$

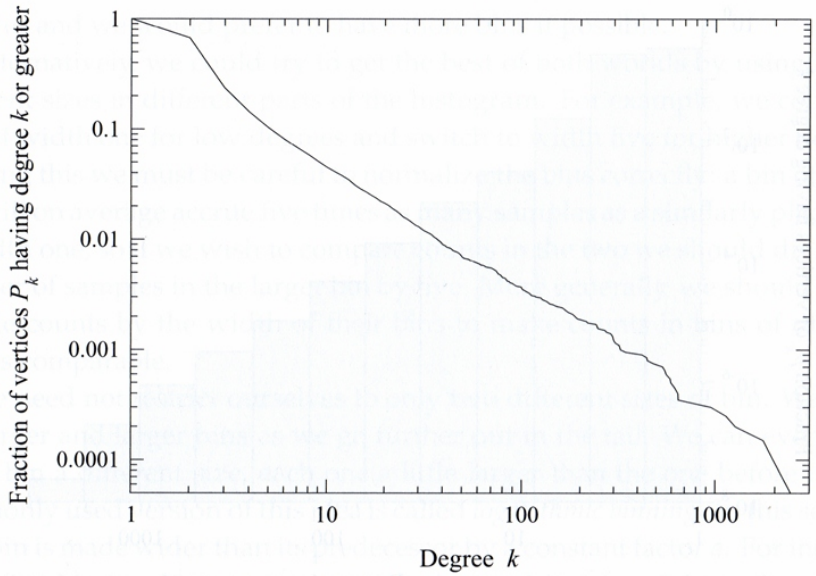


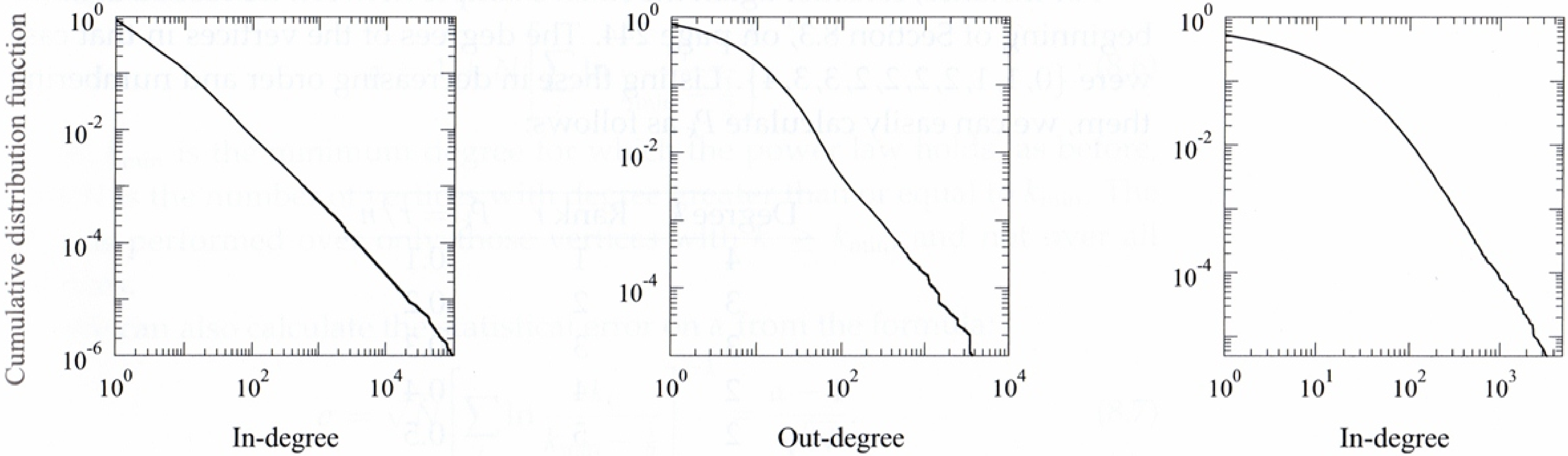

**Figure 8.7: Cumulative distribution function for the degrees of vertices on the Internet.** For a distribution with a power-law tail, as is approximately the case for the degree distribution of the Internet, the cumulative distribution function, Eq. (8.4), also follows a power law, but with a slope 1 less than that of the original distribution.

Newman "Networks, an Introduction"

Advantages:
- no bins
- easy calculation
- can be plotted as normal function at log-log scale
- binning loses the information; cumulative distribution preserves everything

Disadvantages
- less easy to interpret than histograms
- successive points are correlated

# Cumulative Distribution




(a) World Wide Web (b) World Wide Web (c) Citation

**Figure 8.8: Cumulative distribution functions for in- and out-degrees in directed networks.** (a) The in-degree distribution of the World Wide Web, from the data of Broder *et al.* [56]. (b) The out-degree distribution for the same Web data set. (c) The in-degree distribution of a citation network, from the data of Redner [280]. The distributions follow approximate power-law forms in each case.

From Newman "Networks, an Introduction"

Homework:

- Download network “as-22july06” from UFL matrix collection
- Plot degree distribution histogram
- Plot cumulative degree distribution function
- Compute power law parameters *C*, and $\alpha$

(submit by 2/20/2014)

# Power Laws

More examples: city populations, moon craters, solar flares, computer files, words frequencies in human languages, hits on web pages, publications per scientist, book sales, …

**Normalization**: we have to find $C$ such that $\sum_{k=0}^{\infty} p_k = 1$
After eliminating $k = 0$

$$C = \frac{1}{\sum_{k=1}^{\infty} k^{-\alpha}} = \frac{1}{\zeta(\alpha)}, \text{ i.e., } p_k = \frac{k^{-\alpha}}{\zeta(\alpha)}, \text{ where } p_0 = 0$$

Riemann zeta function

However, pure power-law behavior is not perfect for real-world networks
**Normalization over the tail:**

incomplete Riemann zeta function

$$p_k = \frac{k^{\alpha}}{\sum_{k=k_{\min}}^{\infty} k^{-\alpha}} = \frac{k^{-\alpha}}{\zeta(\alpha, k_{\min})}$$

or if we approximate it then $C \approx 1/\left(\int_{k_{\min}}^{\infty} k^{-\alpha} dk\right) = (\alpha - 1)k_{\min}^{\alpha-1}$

**Moments**: The $m$th moment of the distribution is defined as

$$\langle k^m \rangle = \sum_{k=0}^{\infty} k^m p_k = \sum_{k=0}^{k_{\min}-1} k^m p_k + C \sum_{k=k_{\min}}^{\infty} k^m k^{-\alpha}$$

if power law begins with some $k_{min}$

$m$th moment exists (finite) when $\alpha > m + 1$ (integrate the second term)

Remark: This estimate works for arbitrarily large network with the same power law distribution. For finite network $\langle k^m \rangle = \frac{1}{n} \sum_{i \in V} d(i)^m$

**Top-heavy distributions or 80/20 rule**: how many edges are connected to the highest degree vertices?

$$\int_{x_{1/2}}^{\infty} p(x)\,\mathrm{d}x = \frac{1}{2} \int_{x_{\min}}^{\infty} p(x)\,\mathrm{d}x,$$

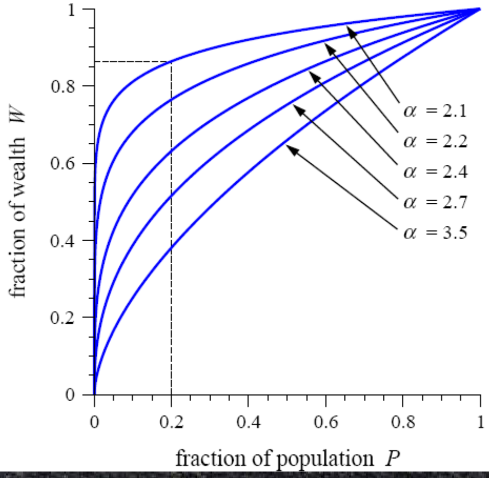
Point that divides distribution in two halves

$$x_{1/2} = 2^{1/(\alpha-1)} x_{\min}.$$

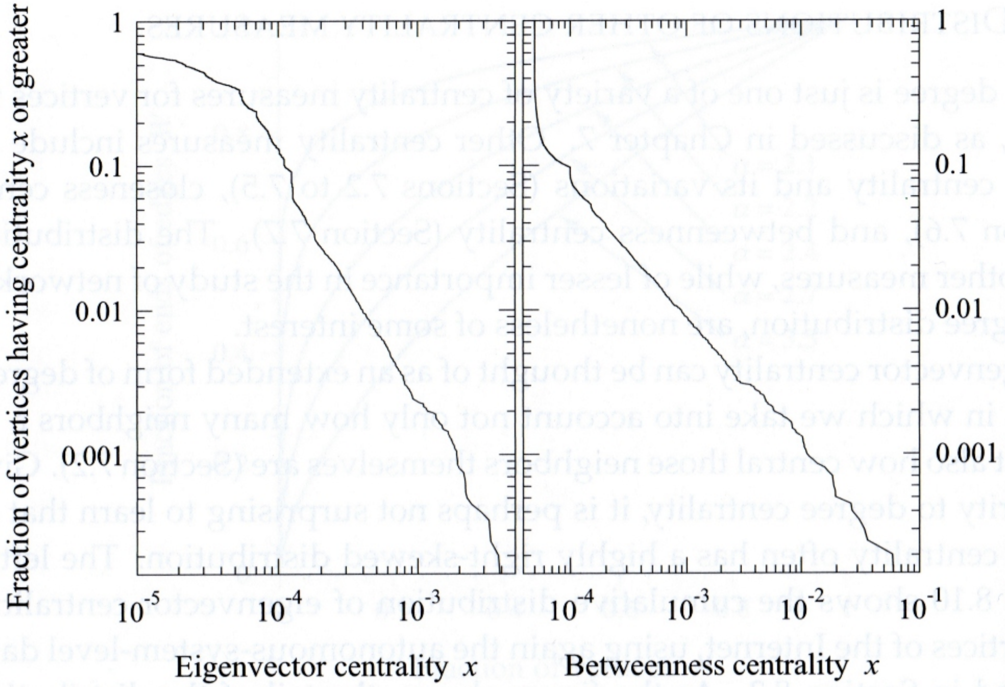
A fraction of edges attached to the highest degree vertices

$$W = P^{(\alpha-2)/(\alpha-1)}$$

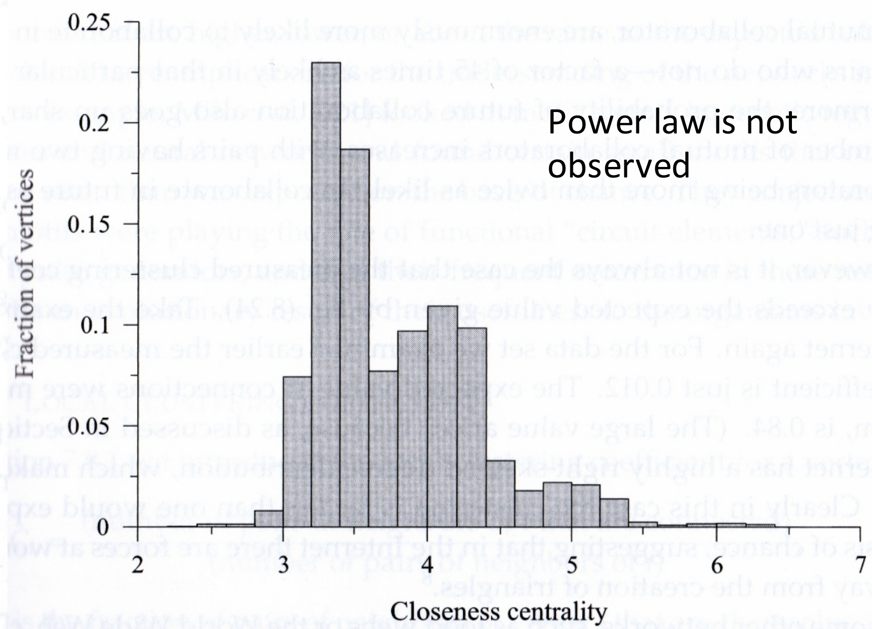
A fraction of highest degree vertices




Further reading: Newman "Power laws, Pareto distributions and Zipf's law"




Cumulative distributions for Internet nodes




Power law is not observed

Noncumulative histogram for Internet nodes

Homework: review of Newman “Power laws, Pareto distributions and Zipf's law”
(submit by 2/20/2014)