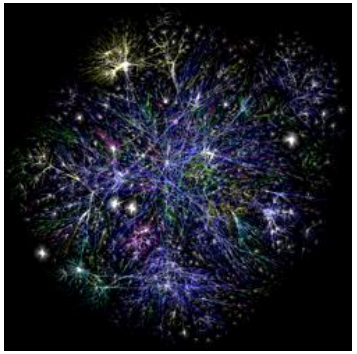
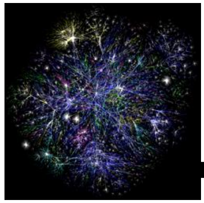
Information Networks

Power Laws and Network Models

Lecture 3





Erdös-Renyi Random Graphs

- For each pair of vertices (i,j), generate the edge (i,j) independently with probability p
  - degree sequence: Binomial

$$p(k) = B(n;k;p) = \binom{n}{k} p^k (1-p)^{n-k}$$

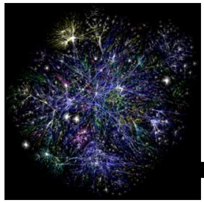
  - in the limit: Poisson

$$p(k) = P(k;z) = \frac{z^k}{k!} e^{-z}$$

  - real life networks: power-law distribution

$$p(k) = Ck^{-\alpha}$$

  - Low clustering coefficient: C = z/n
  - Short paths, but not easy to navigate
  - Too random…



Departing from the ER model

- We need models that better capture the characteristics of real graphs
  - degree sequences
  - clustering coefficient
  - short paths

Graphs with given degree sequences

- Assume that we are given the degree sequence $[d_1,d_2,\ldots,d_n]$

- Create d_i copies of node i
- Take a random matching (pairing) of the copies
  - self-loops and multiple edges are allowed

- Uniform distribution over the graphs with the given degree sequence

The giant component

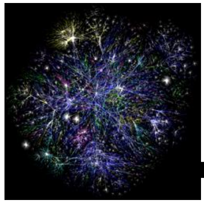
- The phase transition for this model happens when

$$\sum_{k=0}^{\infty} k(k-2)p_k = 0$$

p_k: fraction of nodes with degree k

- The clustering coefficient is given by

$$C = \frac{z}{n}\left(\frac{\langle d^2\rangle - \langle d\rangle}{\langle d\rangle^2}\right)^2$$



Power-law graphs

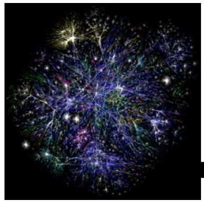
- The critical value for the exponent α is

$$\alpha = 3.4788\ldots$$

- The clustering coefficient is

$$C = n^{-\beta} \qquad \beta = \frac{3\alpha - 7}{\alpha - 1}$$

- When α<7/3 the clustering coefficient increases with n



A different approach

- Given a degree sequence $[d_1,d_2,\ldots,d_n]$, where $d_1 \geq d_2 \geq \ldots \geq d_n$, there exists a simple graph with this sequence (the sequence is realizable) if and only if

$$\sum_{i=1}^{k} d_i \leq k(k-1) + \sum_{i=k+1}^{n} \min\{d_i, k\}$$

- There exists a connected graph with this sequence if and only if

$$\sum_{i=1}^{k} d_i \geq 2(n-1)$$

Getting a random graph

- Perform a random walk on the space of all possible simple connected graphs
  - at each step swap the endpoints of two randomly picked edges

Graphs with given degree sequence

- The problem is that these graphs are too contrived
- It would be more interesting if the network structure emerged as a side product of a stochastic process

Power Laws

- Two quantities y and x are related by a power law if

$$y \approx x^{-\alpha}$$

- A (continuous) random variable X follows a power-law distribution if it has density function

$$f(x) = Cx^{-\alpha}$$

- Cumulative function

$$P[X \geq x] = \frac{C}{\alpha - 1} x^{-(\alpha-1)}$$

Normalization constant

- Assuming a minimum value x_{min}

$$C = (\alpha - 1) x_{min}^{a-1}$$

- The density function becomes

$$f(x) = \frac{(\alpha - 1)}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}$$

Pareto distribution

- Pareto distribution is pretty much the same but we have

$$P[X \geq x] = C' x^{-\beta}$$

- and we usually we require

$$x \geq x_{min}$$

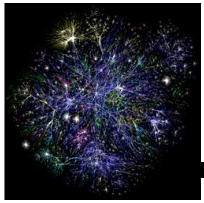
Zipf's Law

- A random variable X follows Zipf's law if the r-th largest value x_r satisfies

$$x_r \approx r^{-\gamma}$$

- Same as requiring a Pareto distribution

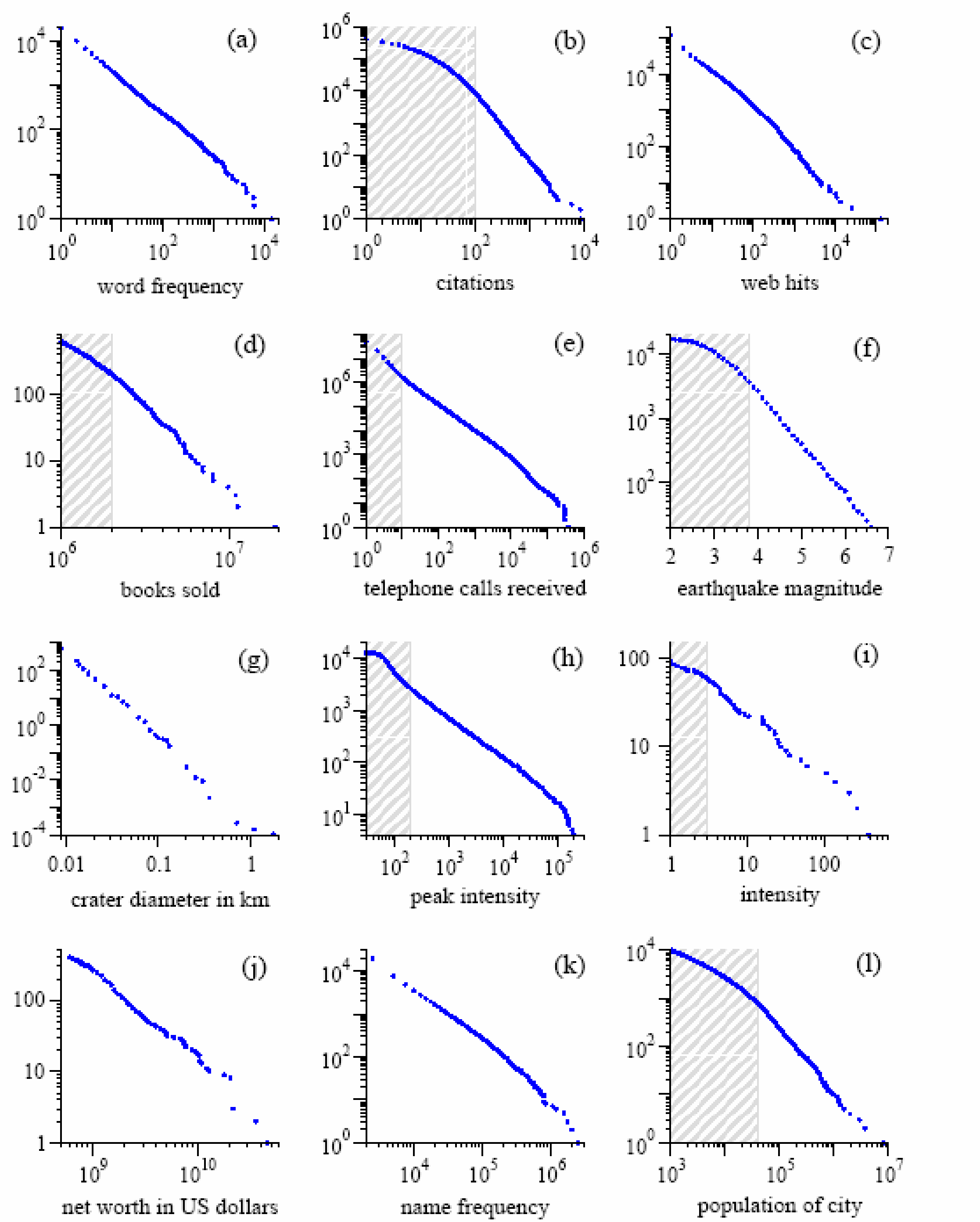
$$P[X \geq x] \approx x^{-1/\gamma}$$



Power laws are ubiquitous

	quantity	minimum $x_{\min}$	exponent α
(a)	frequency of use of words	1	2.20(1)
(b)	number of citations to papers	100	3.04(2)
(c)	number of hits on web sites	1	2.40(1)
(d)	copies of books sold in the US	2 000 000	3.51(16)
(e)	telephone calls received	10	2.22(1)
(f)	magnitude of earthquakes	3.8	3.04(4)
(g)	diameter of moon craters	0.01	3.14(5)
(h)	intensity of solar flares	200	1.83(2)
(i)	intensity of wars	3	1.80(9)
(j)	net worth of Americans	\$600m	2.09(4)
(k)	frequency of family names	10 000	1.94(1)
(l)	population of US cities	40 000	2.30(5)

TABLE I Parameters for the distributions shown in Fig. 4. The labels on the left refer to the panels in the figure. Exponent values were calculated using the maximum likelihood method of Eq. (5) and Appendix B, except for the moon craters (g), for which only cumulative data were available. For this case the exponent quoted is from a simple least-squares fit and should be treated with caution. Numbers in parentheses give the standard error on the trailing figures.



But not everything is power law

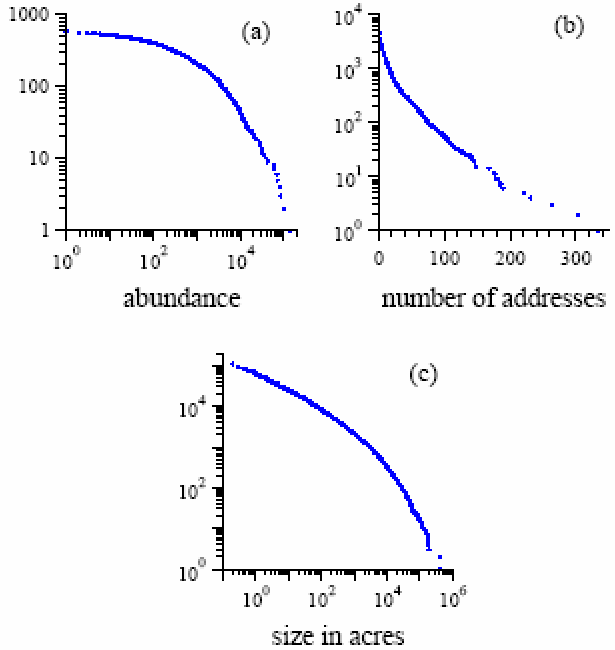
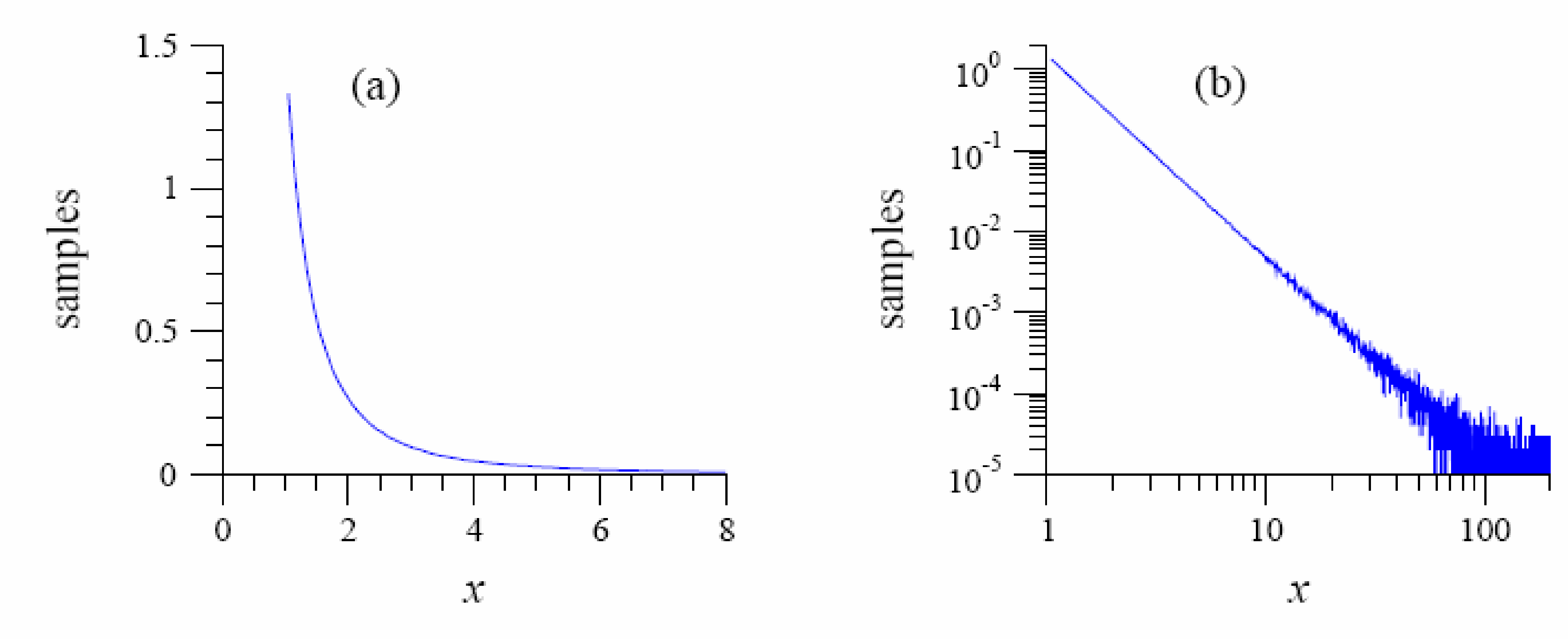


FIG. 5 Cumulative distributions of some quantities whose distributions span several orders of magnitude but that nonetheless do not follow power laws. (a) The number of sightings of 591 species of birds in the North American Breeding Bird Survey 2003. (b) The number of addresses in the email address books of 16 881 users of a large university computer system [34]. (c) The size in acres of all wildfires occurring on US federal land between 1986 and 1996 (National Fire Occurrence Database, USDA Forest Service and Department of the Interior). Note that the horizontal axis is logarithmic in frames (a) and (c) but linear in frame (b).

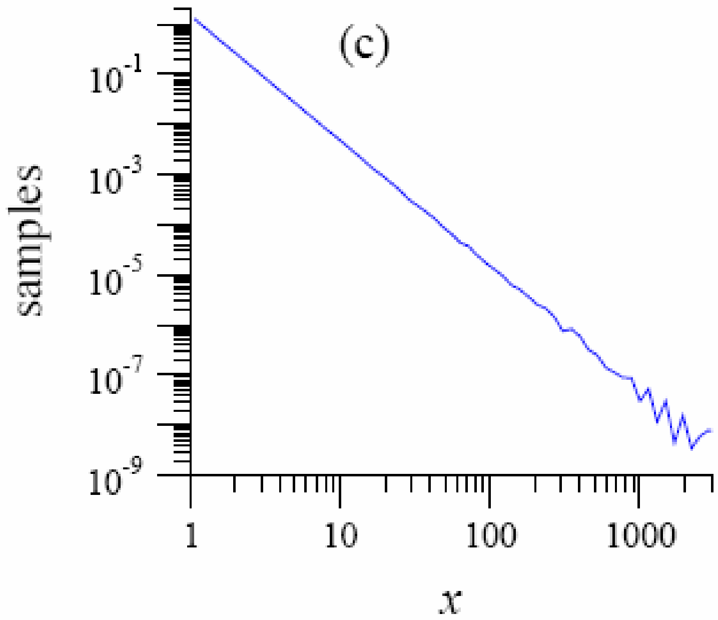
Measuring power laws

Simple log-log plot gives poor estimate

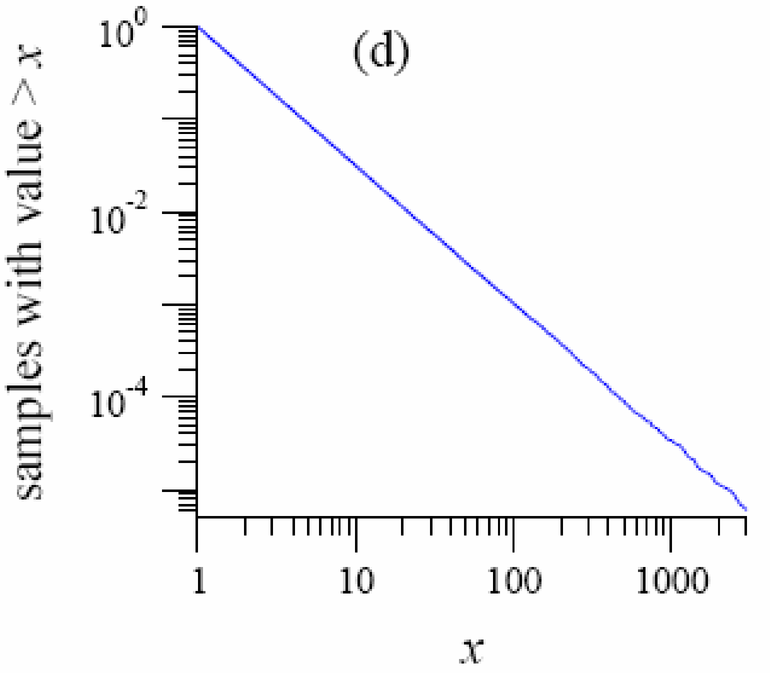


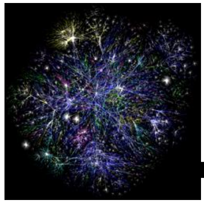
Logarithmic binning

- Bin the observations in bins of exponential size



Cumulative distribution





Computing the exponent

- Least squares fit of a line

- Maximum likelihood estimate

$$\alpha = 1 + n\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}}\right]^{-1}$$

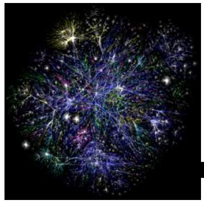
Power-law properties

- First moment

$$\langle x \rangle = \frac{\alpha - 1}{\alpha - 2} x_{min}$$

- Second moment

$$\langle x^2 \rangle = \frac{\alpha - 1}{\alpha - 3} x^2_{min}$$

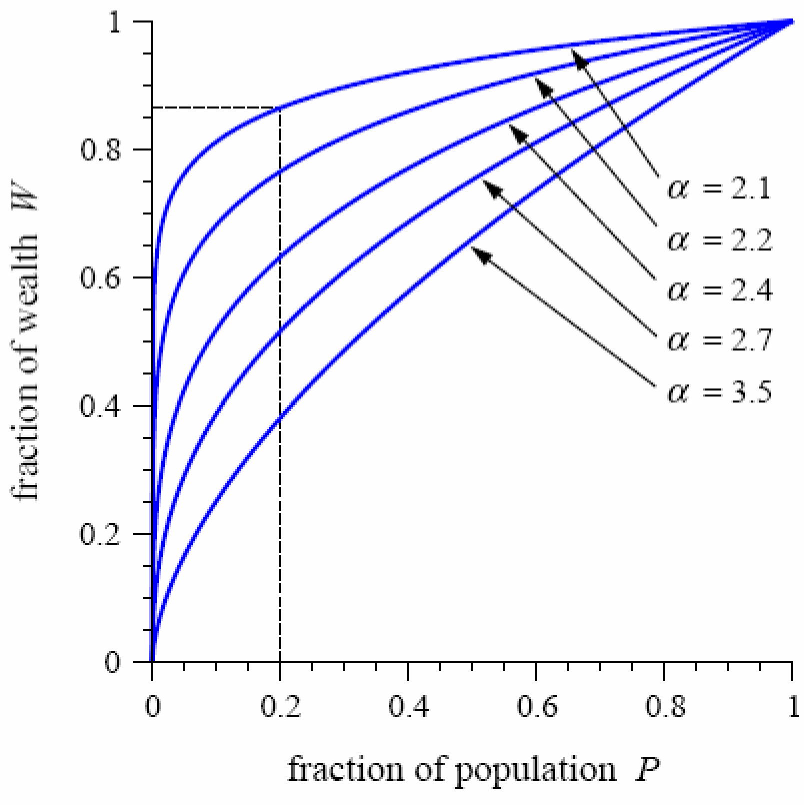


Scale Free property

- A power law holds in all scale
  - f(bx) = g(b)p(x)

The 80/20 rule

- Cumulative distribution is top-heavy



The discrete case

$$p_k = Ck^{-\alpha} = \frac{k^{-\alpha}}{\zeta(\alpha)}$$

$$p_k = CB(\alpha,k) = (\alpha-1)B(\alpha,k) \approx (\alpha-1)k^{-\alpha}$$

$$\langle k \rangle = \frac{\alpha-1}{\alpha-2} \qquad \langle k^2 \rangle = \frac{(\alpha-1)^2}{(\alpha-2)(\alpha-3)}$$

References

- M. E. J. Newman, The structure and function of complex networks, SIAM Reviews, 45(2): 167-256, 2003
- M. E. J. Newman, Power laws, Pareto distributions and Zipf's law, *Contemporary Physics*
- M. Mitzenmacher, A Brief History of Generative Models for Power Law and Lognormal Distributions**,** Internet Mathematics
- M. Mihail, N. Vishnoi ,On Generating Graphs with Prescribed Degree Sequences for Complex Network Modeling Applications, Position Paper, ARACNE (Approx. and Randomized Algorithms for Communication Networks) 2002, Rome, IT, 2002.