Power laws

Definition, properties and how to plot them

Power-law degree distributions

- The degree distribution of many networks follows a *power-law distribution*:
- $$p(x) = C \cdot x^{-k}$$
- The proportion of nodes of degree *x* is (inversely) polynomial in *x*

Power-law distributions

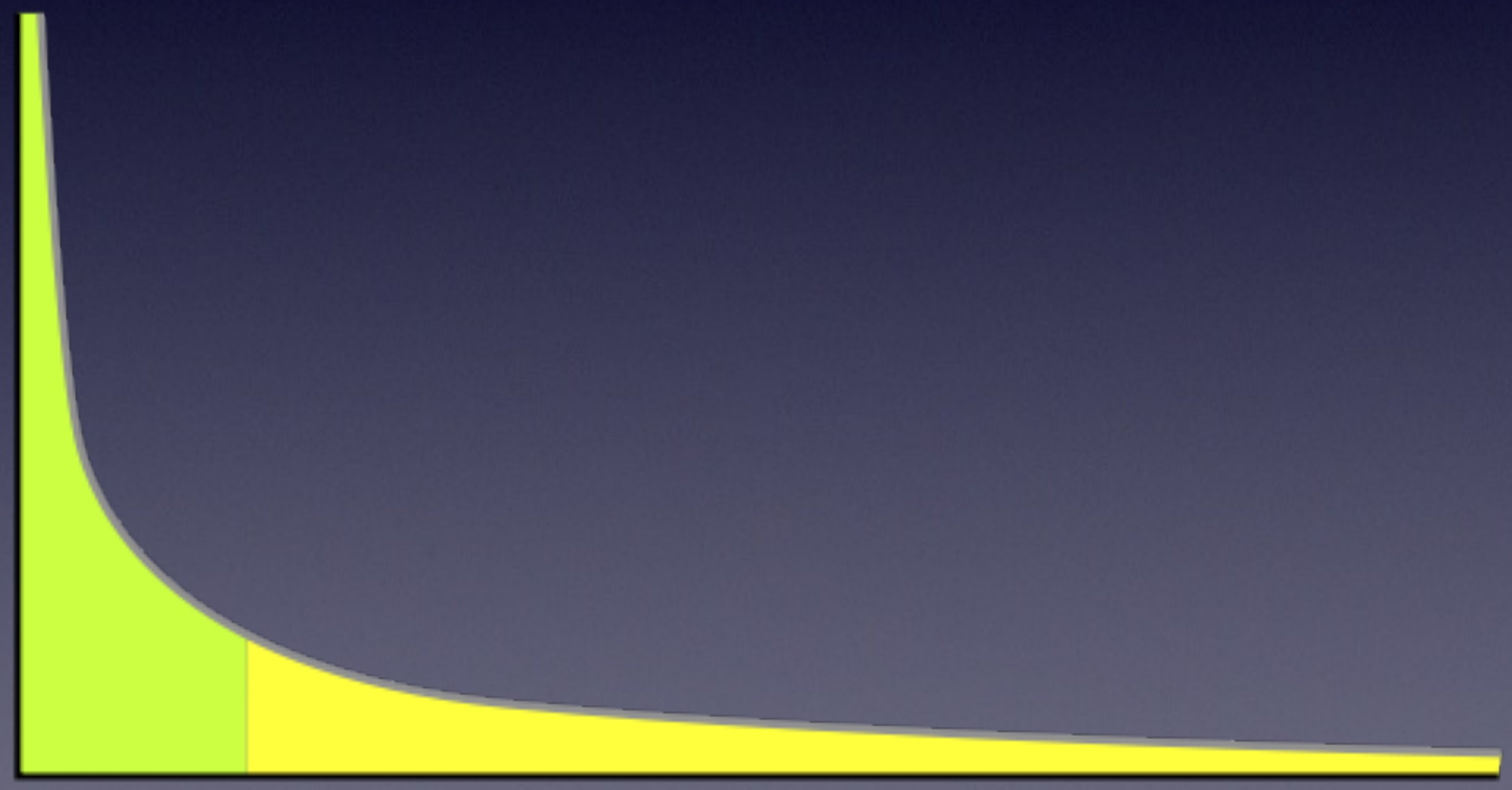
- In general, a discrete power-law random variable X with parameter k is defined by
- $$Pr(X = x) = C \cdot x^{-k} \quad \text{for all } x > x_{min}$$
- A continuous power-law random variable X with parameter k is defined by a probability density function
- $$p(x) = C \cdot x^{-k} \quad \text{for all } x > x_{min}$$

"Scale-free" networks

- A power-law distribution is also called *scale-free*, since scaling the input does not change the form of the function:
- $$p(a \cdot x) = C \cdot a^{-k} \cdot x^{-k} = C' \cdot x^{-k}$$

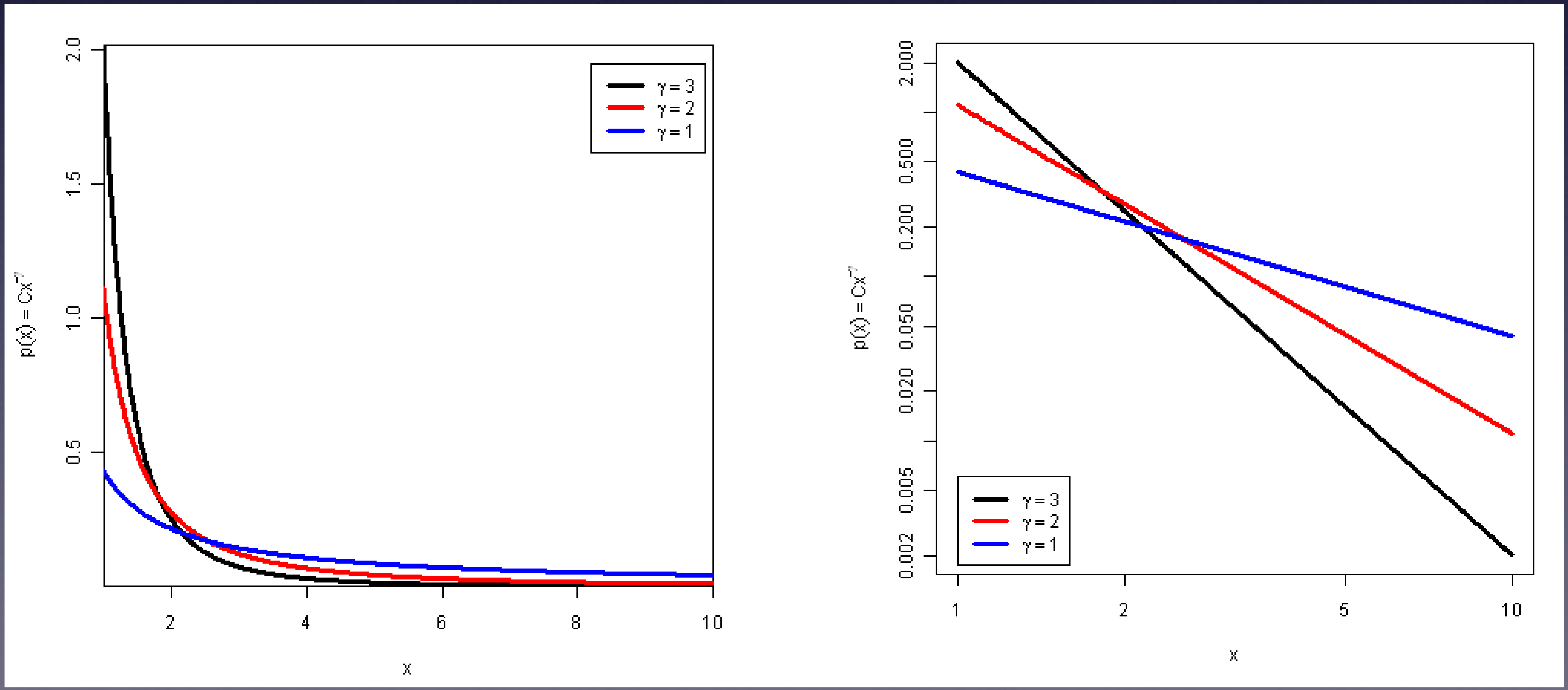
The long tail

- A power-law distribution has a *long tail*: there is a relatively large number of samples of the population for which the variable has a large value



Power-law plotting: scales

- If we use a log-log scale, a power-law will show up as a *line*, because
- $$\log p(x) = \log C - k \cdot \log x$$



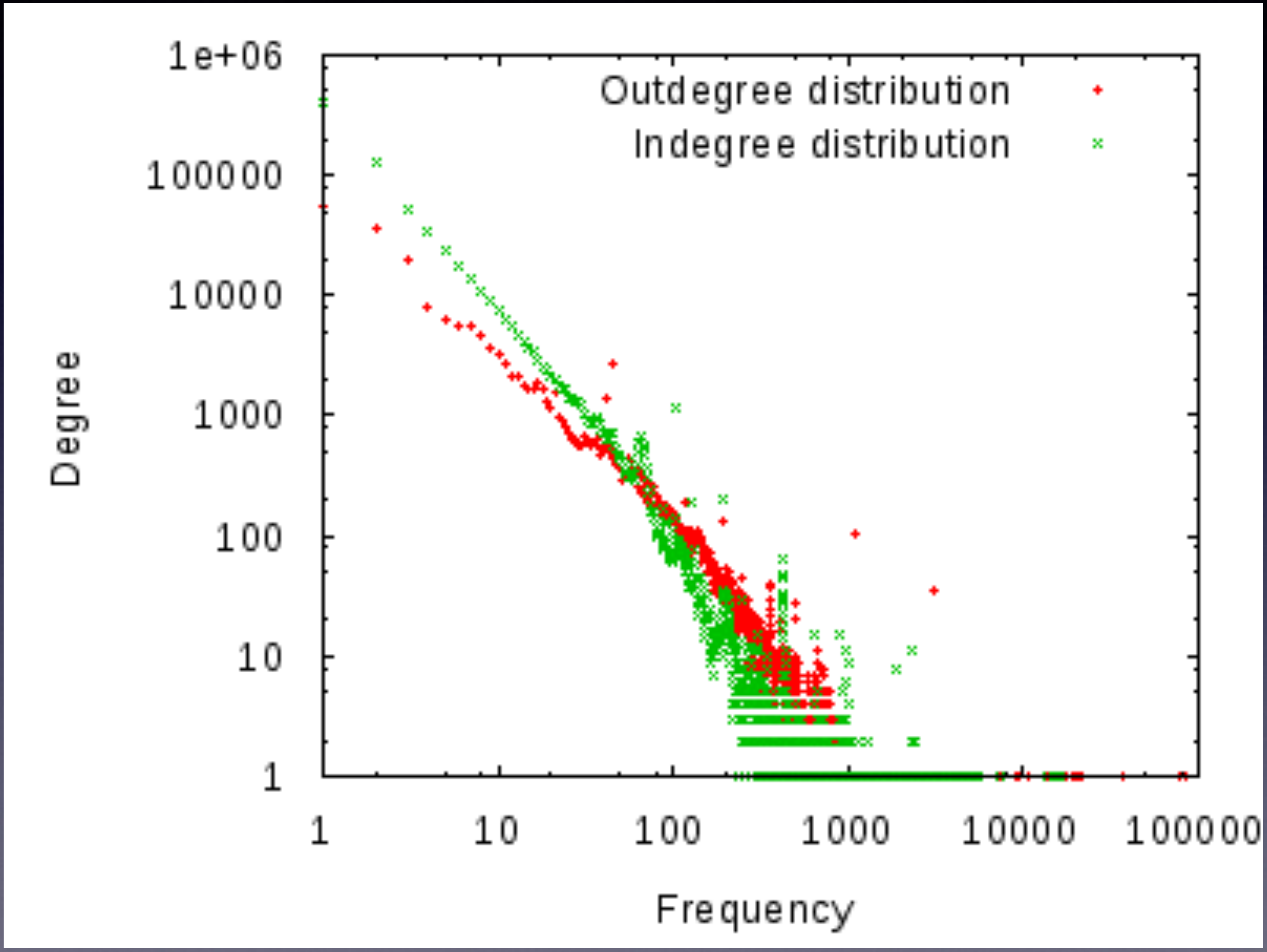
The P.D.F. and the C.D.F.

- Probability distribution function (P.D.F.): p(x)
- Cumulative distribution function (C.D.F.):

$$P(x) = \int_x^\infty p(x')dx'$$

- If $p(x) = C \cdot x^{-k}$, then $P(x) = C' \cdot x^{-(k-1)}$, and the C.D.F. is also a power-law
- Plotting P(x) should be preferred

Example from web-graph.org

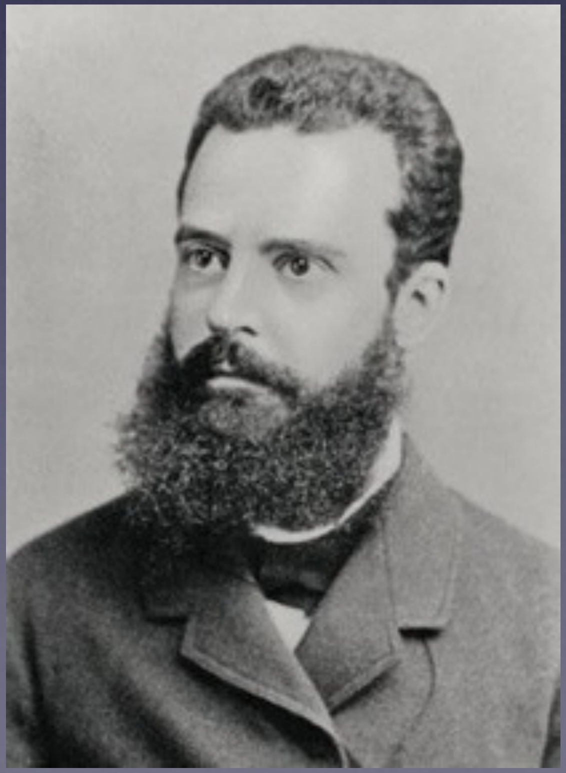


Power-laws everywhere?

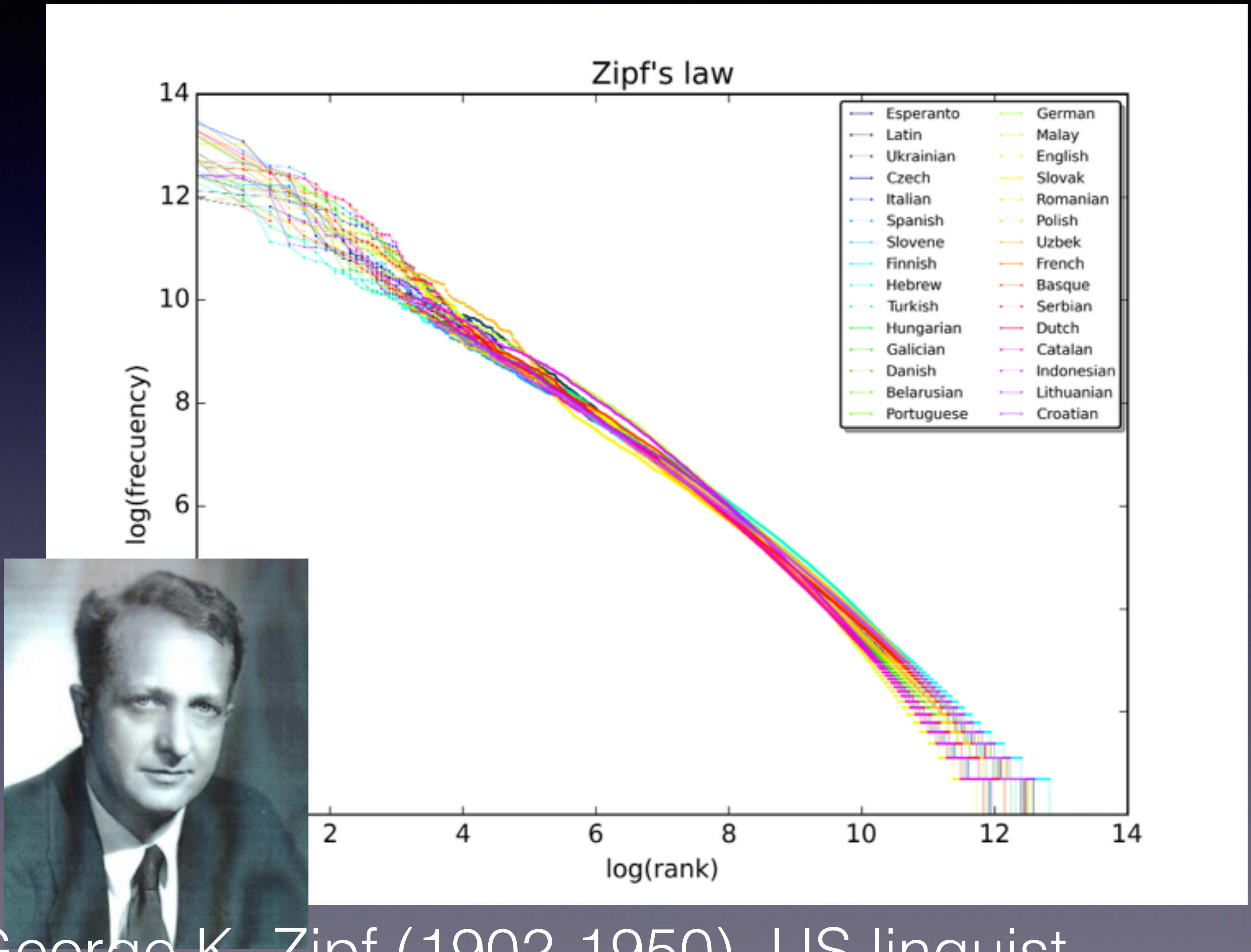
- Power-laws don't appear just in degree distributions
- They characterize many very different phenomena
- Also known as *Zipf laws* or *Pareto distributions*

Pareto's distribution of income

- Pareto used available data on income distribution for several countries
- He found that income was well-approximated by a power-law with exponent between 2.32 and 2.72
- "30% of the population owns 70% of the land"
- Vilfredo Pareto (1848-1923), Italian economist and engineer



Frequency of words - "Zipf's Law"



George K. Zipf (1902-1950), US linguist

Another example: size of cities

- Let’s re-discover by ourselves another power-law: size of Italian towns
- Get the data publicly available from ISTAT’s website
- http://www.istat.it/storage/codici-unita-amministrative/elenco-comuni-italiani.xls

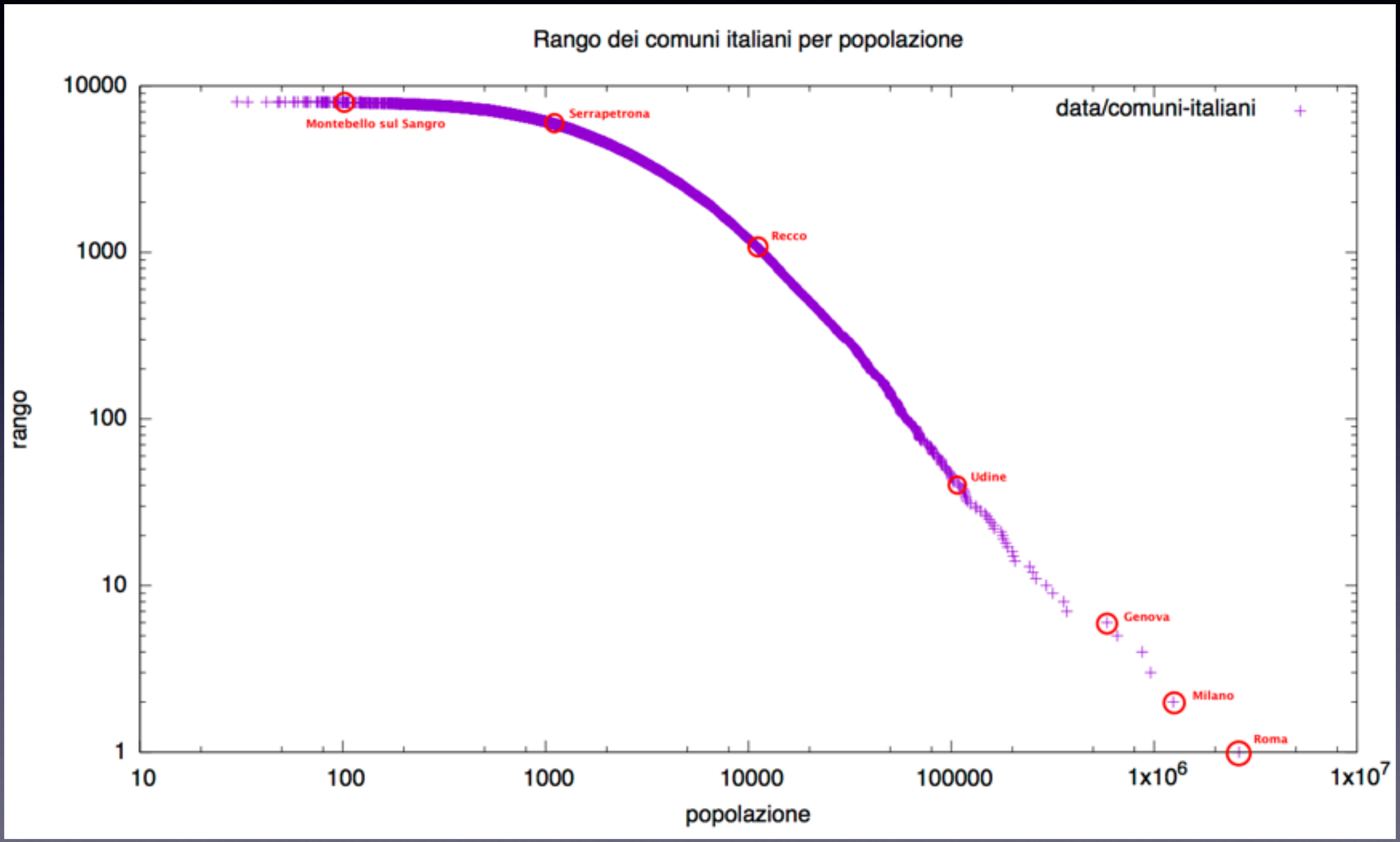
Another example: size of cities

- After processing ISTAT’s .xls file, we arrive to a list of Italian city sizes, ordered by size:
    - 2617175 (Rome)
    - 1242123 (Milan)
    - etc

Another example: size of cities

- The *rank* of each item (city) is its *position* in the ordered list
- We plot the rank of each item over the quantity x of interest; in this case, x = population size
- This is equivalent to plotting P(x)
- We use a log-log scale
- *Gnuplot* is recommended for best results

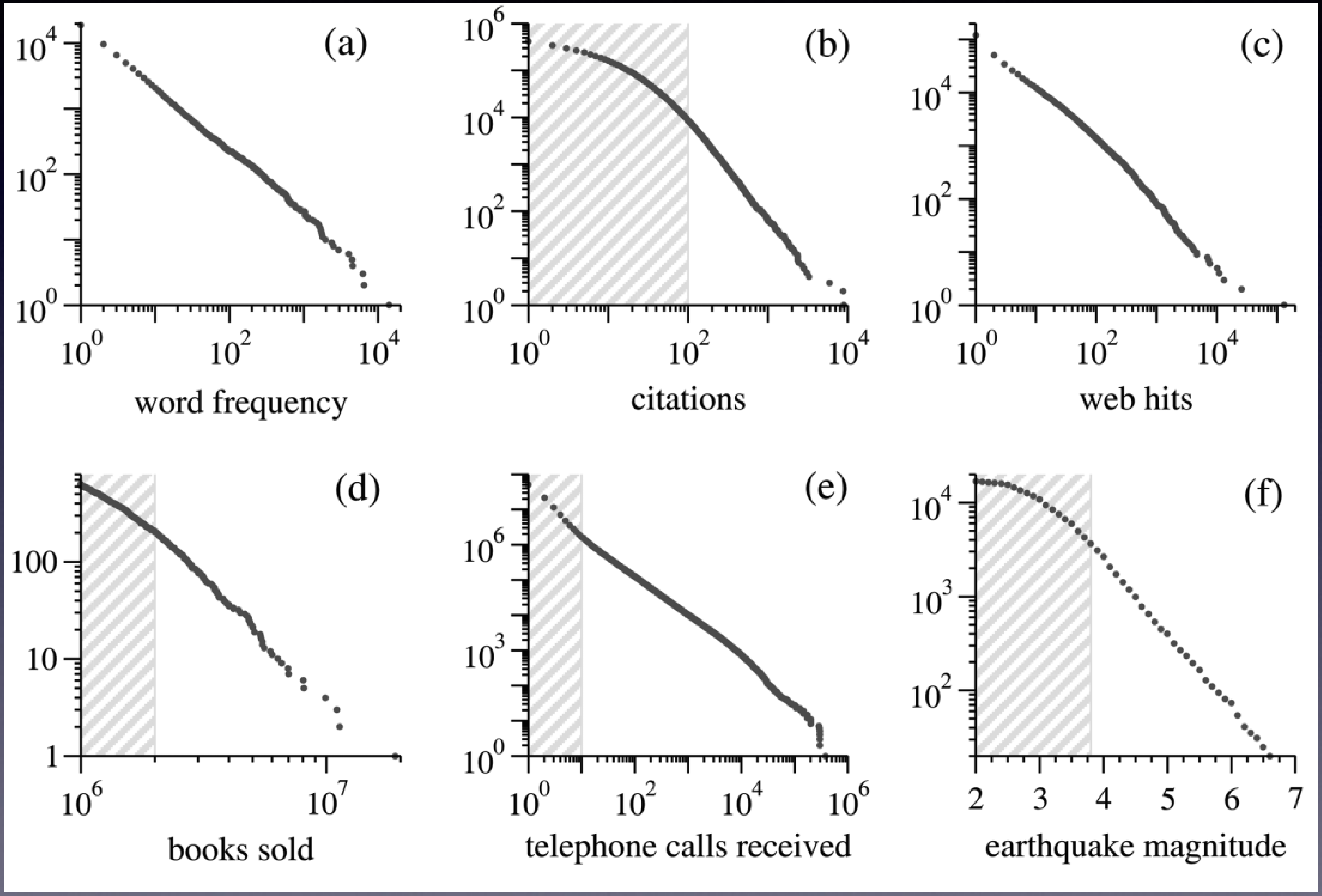
Another example: size of cities



Another example: size of cities

- We find out that the distribution *P(x)* follows a power-law *only when x is large enough*
- This is a common phenomenon; the power-law relation holds for $x > x_{min}$
- Thus, a power-law has two main parameters: the power-law exponent (k) and the threshold (x_{min})

More examples



More examples

Quantity	Threshold (x	Exponent (k)
(a) Frequency of use of words	1	2.20
(b) Number of citations to papers	100	3.04
(c) Number of hits on web sites	1	2.40
(d) Copies of books sold in the US	2000000	3.51
(e) Telephone calls received	10	2.22
(f) Magnitude of earthquakes	3.8	3.04

Estimating the power-law exponent

- Let $x_1, x_2, \ldots, x_n$ be the measured quantities (above x_{min})
- A reliable estimator (for the discrete case) is

$$k = 1 + n \cdot \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min} - 1/2} \right]^{-1}$$

Estimating x_{min}

- If we fix x_{min}, we can estimate k, so do the following:
- Choose x_{min} to minimize "distance" between the data and the fitted distribution (for that x_{min})
- To measure the distance between two distributions, we can use the "Kolmogorov-Smirnov" distance:

$$D = \max_{x \geq x_{\min}} |S(x) - P(x)|$$

- S(x) = cumulative distribution function of the data
- P(x) = cumulative distribution function of the fitted power-law

Finally, an obvious remark:

- Not all right-skewed distributions are power-laws:
  - Lengths of relationships between couples (exponentially distributed)
  - Abundance of North-American bird species (log-normally distributed)
  - etcetera; see Newman (2007)