Power Law Networks

Rik Sarkar

Degree Distribution

- A more sophisticated way of characterizing networks
- More complex than single numbers
- Many standard networks are known to have “standard” degree distributions
- Gives ways to incorporate notions of “popularity” and understand them

Degree distributions in networks

- As a function of *k*, what fraction of pages in the network have *k* links?

- A histogram

Degree Distribution in Random networks

- Suppose we take a random network
- What does the degree distribution look like?

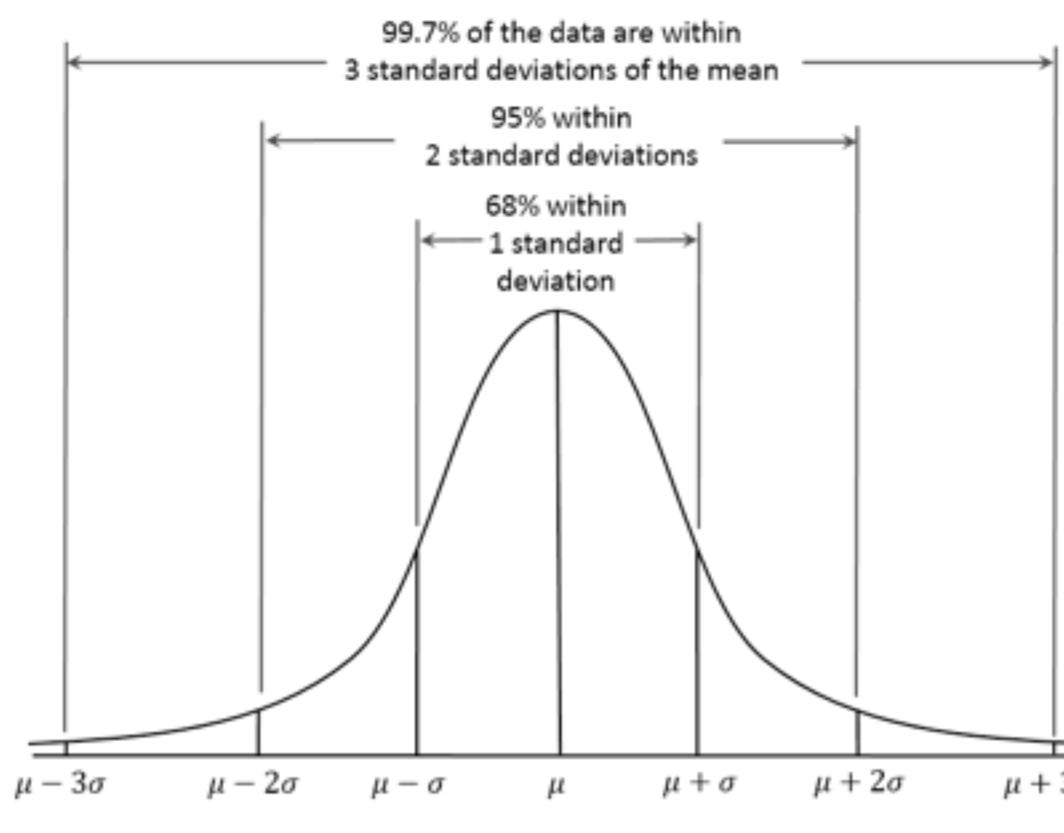
Normal distribution and central limit theorem

- Central limit theorem: The distribution of sum (or average) of n independent random quantities approaches a normal distribution with increasing n.

- Applies to edges on a particular vertex

- Normal distribution:

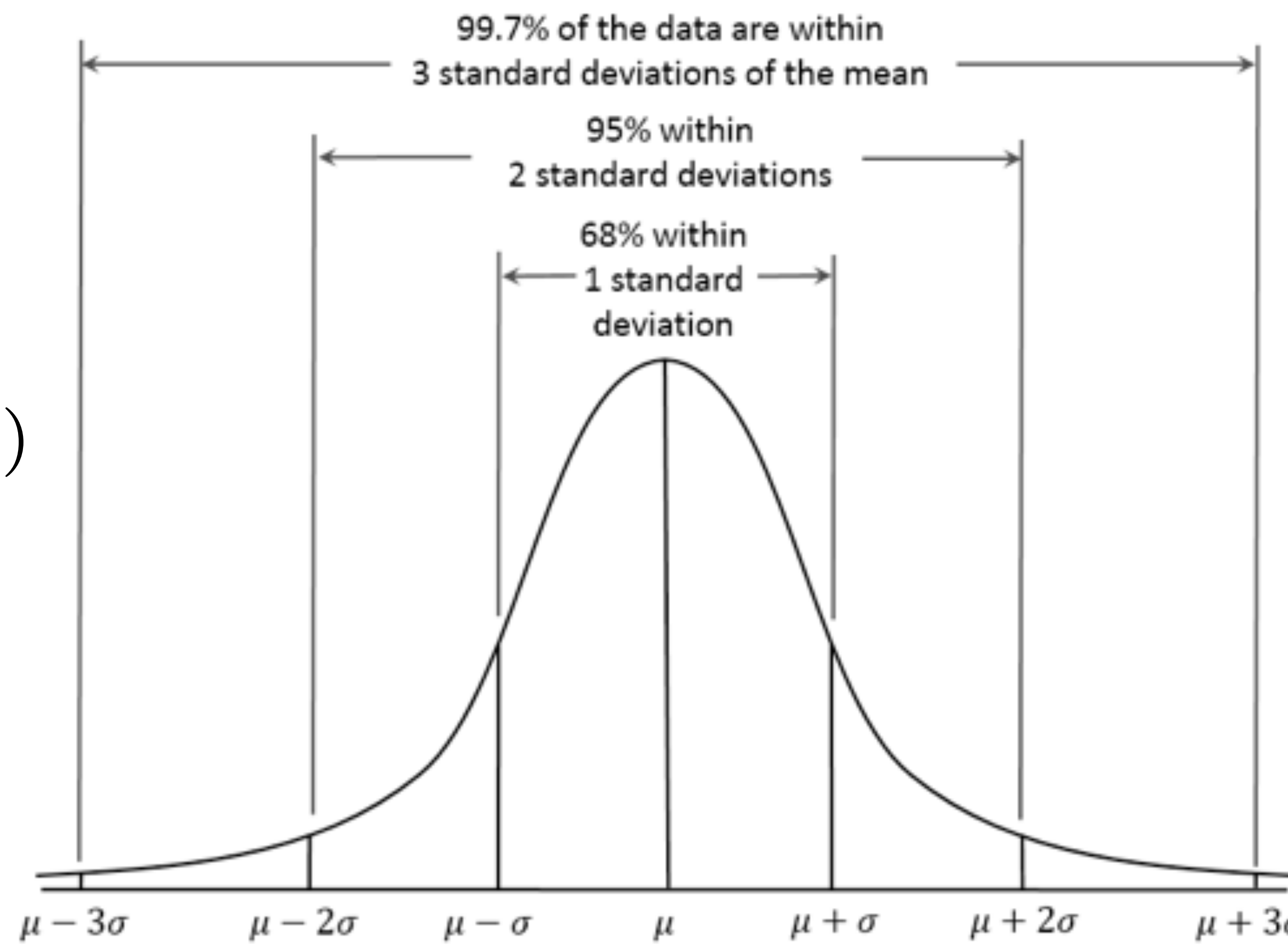
$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma)}$$



Normal distribution

- The probability density drops exponentially with distance from the mean.

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma)}$$



Degree distribution in www

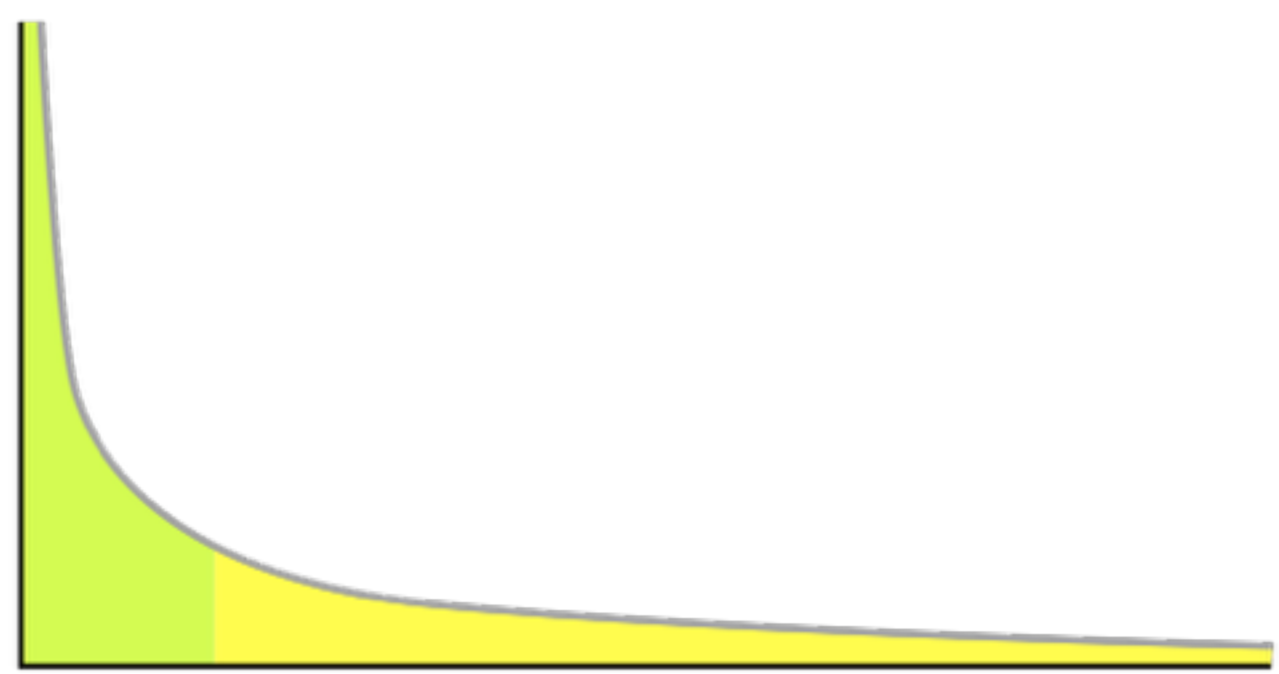
- Suppose we take a real network like the world wide web, and compute degree distribution. What does that look like?
- Let's try.

Degree distribution in www

- Usually for www snapshots, number of nodes with in-degree k is approximately proportional to:

$$\frac{1}{k^2}$$

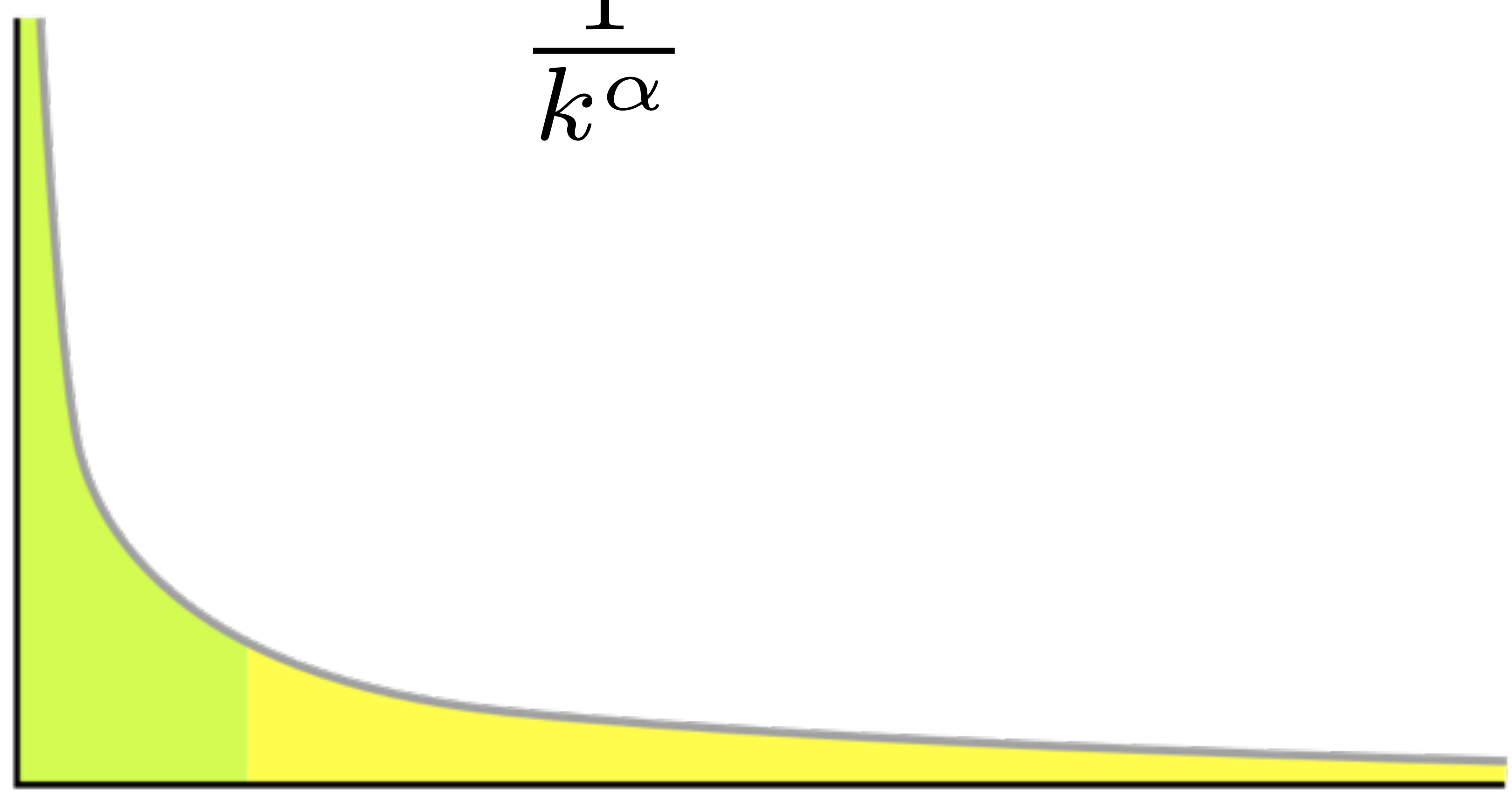
- Usually, for www the exponent is slightly larger than 2



Power law

- The variable concerned — degree or popularity etc changes as (for some constant α):

$$\frac{1}{k^{\alpha}}$$

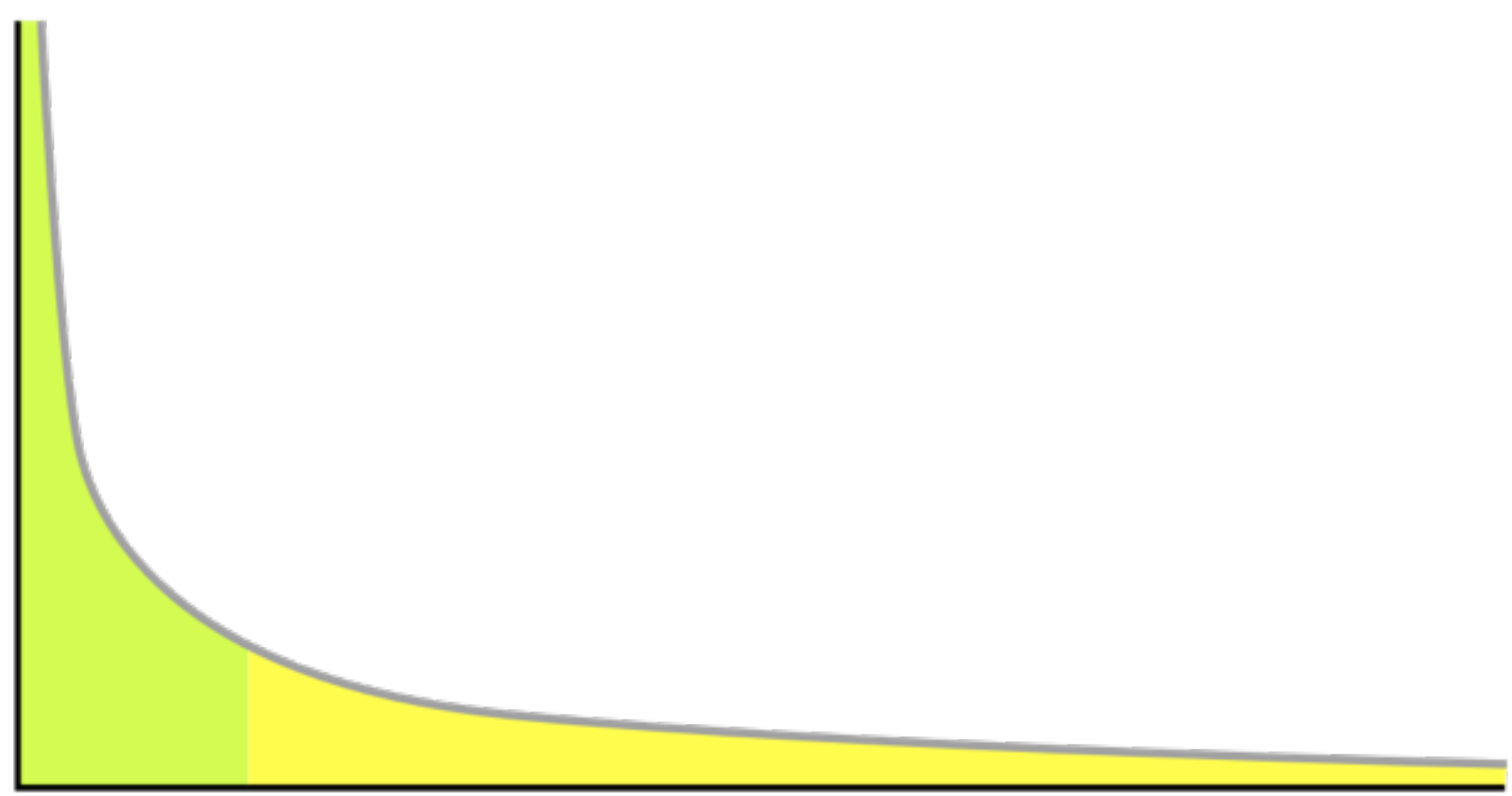


Normal distribution vs power law in networks:

- Normal distribution drops exponentially. That is very fast.
  - Ignoring constants: $P(k) \propto e^{-(k)}$
  - The probability of a node having a high degree (like 100) is small
- Power law drops slower
  - Ignoring constants: $P(k) \propto k^{-\alpha}$
  - Therefore probability of a node having high degree (like 100) is not so small

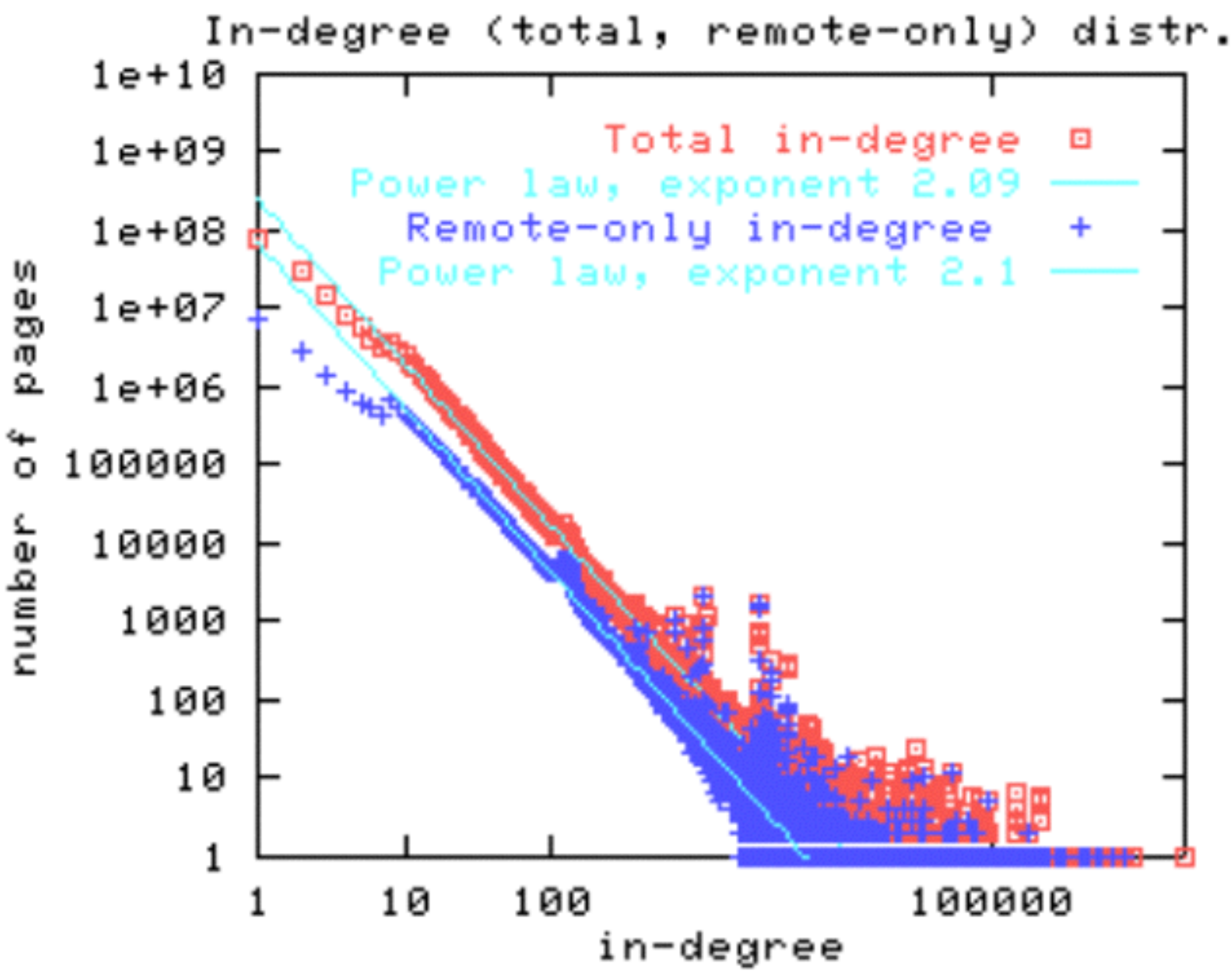
Power law networks

- There are a fair number of “hubs” : heavy tailed distribution
  - Nodes that are very well connected
  - Important in Social networks: there are many popular people
    - Influence spread of epidemics
    - Influences strategies for product placement/advertising…



log-log plots

log-log plots are nice & straight



log log plots are not so nice!

- The “straight” part needs to extend quite a few orders of magnitude
- Fitting the straight line to determine the right coefficient alpha is not trivial due to non-linear nature of data
- Beware: log-normal distributions can look similar to power law.

Mean of a power law distribution

- The mean is finite iff $\alpha > 2$

- Thus, average degrees on www should remain finite as www grows

- May not be the case in other types of networks

Preferential attachment mechanism

- We need a “model” i.e. a way to think about the creation of www that fits with the power law distribution

- Idea: older and established (popular) sites are likely to have more links to them (yahoo, google…)

  - So how about: When a new page arrives, it links to older pages in proportion to their popularity

  - When a new link is created on a new page, randomly to older pages with probability of hitting a page x proportional to current popularity of x (number of links to x)

Preferential attachment model

- Takes a parameter p: $0 \leq p \leq 1$
- On a new page, create k links as follows:
- When creating a new link:
  - With probability p
    - Assign it with preferential attachment mechanism
  - With probability 1-p
    - Assign it with uniform random probability

Preferential attachment model

- Takes into consideration that popularity is not the only force behind link creation.
- The randomly assigned links model other reasons for link creation.
- Can be proven to produce power law. see [Kempe lecture notes, 2011]
- Produces same exponent as www for p~0.9

Other reasons for power law

- Optimization:
  - Power law found in linguistics (word lengths): most frequent words are short
    - Mandelbrot, Zipf : emerges from need for efficient communication
- Random processes:
  - Press space with probability p, else press a random letter key
  - This will produce a power law distribution of word lengths

How realistic are preferential attachment graphs?

Diameter

- Preferential attachment networks have small diameter

Expander properties

- What do you think happens for real power law networks?
- What about preferential attachment networks?