Power Laws in Networks

CS 322: (Social and Information) Network Analysis
Jure Leskovec
Stanford University

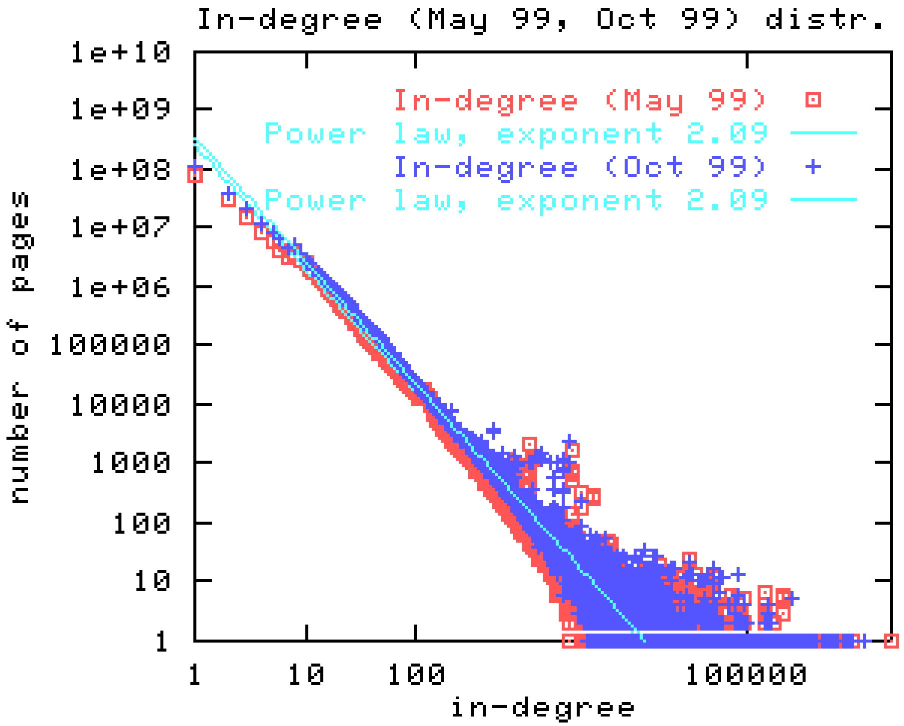
Announcements

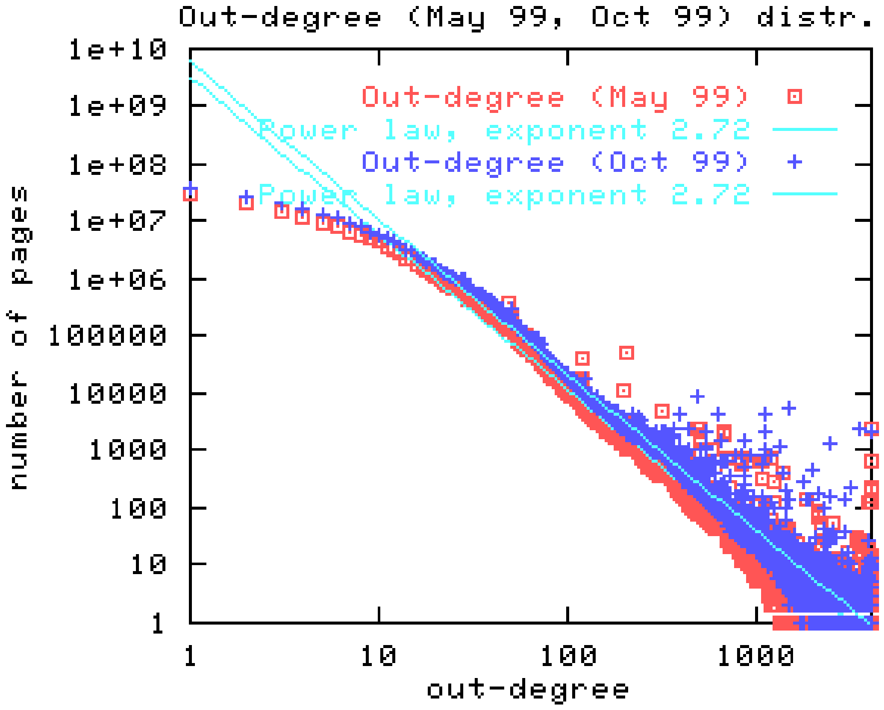
- Return the homework after the class

- Project proposals are due in 1 week
- 2 parts to the proposal:
  - Reaction paper:
    - Read related papers, comment on them
    - Weaknesses, extensions, etc.
  - Proposed work:
    - Put your proposed work in the context of the papers you read

[Broder et al., ‘00]

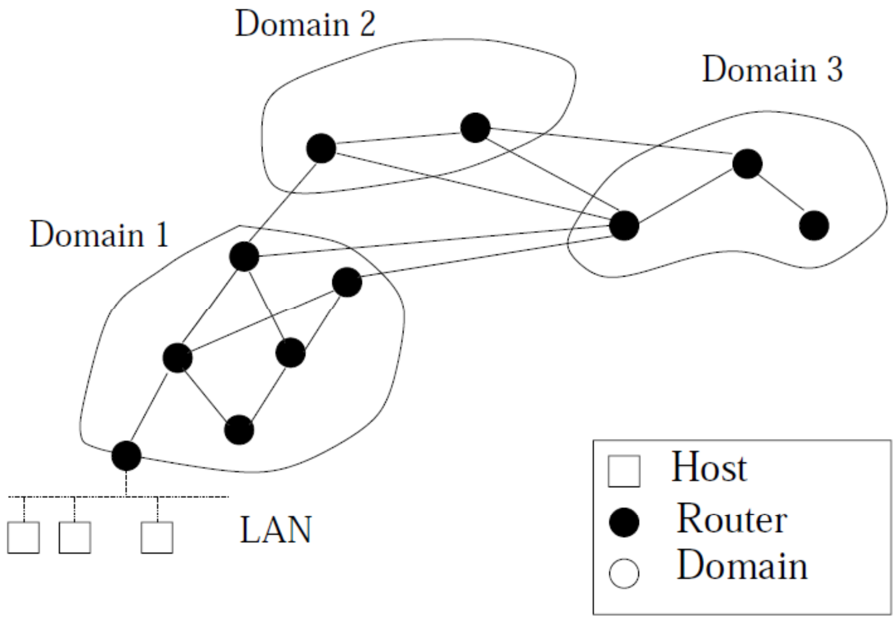
Degree distribution on the Web

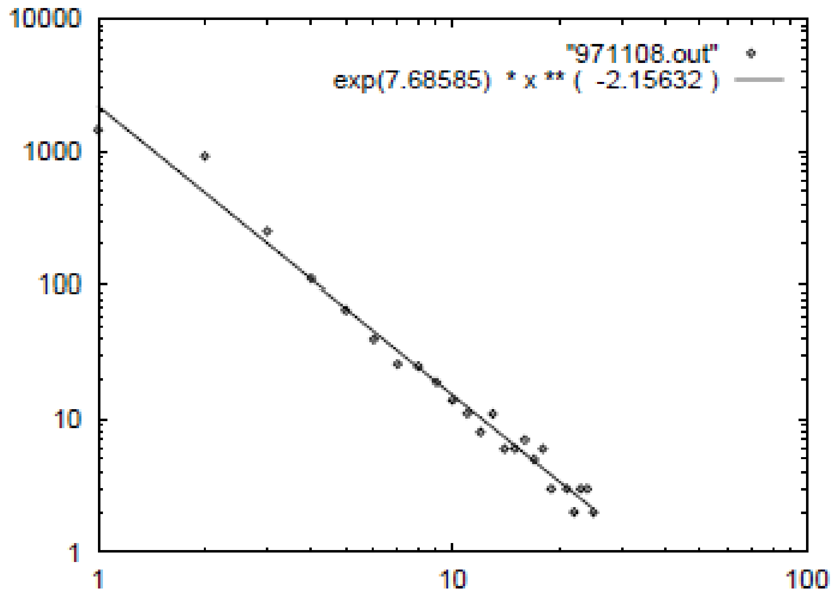




Faloutsos^3

- [Faloutsos, Faloutsos and Faloutsos, 1999]



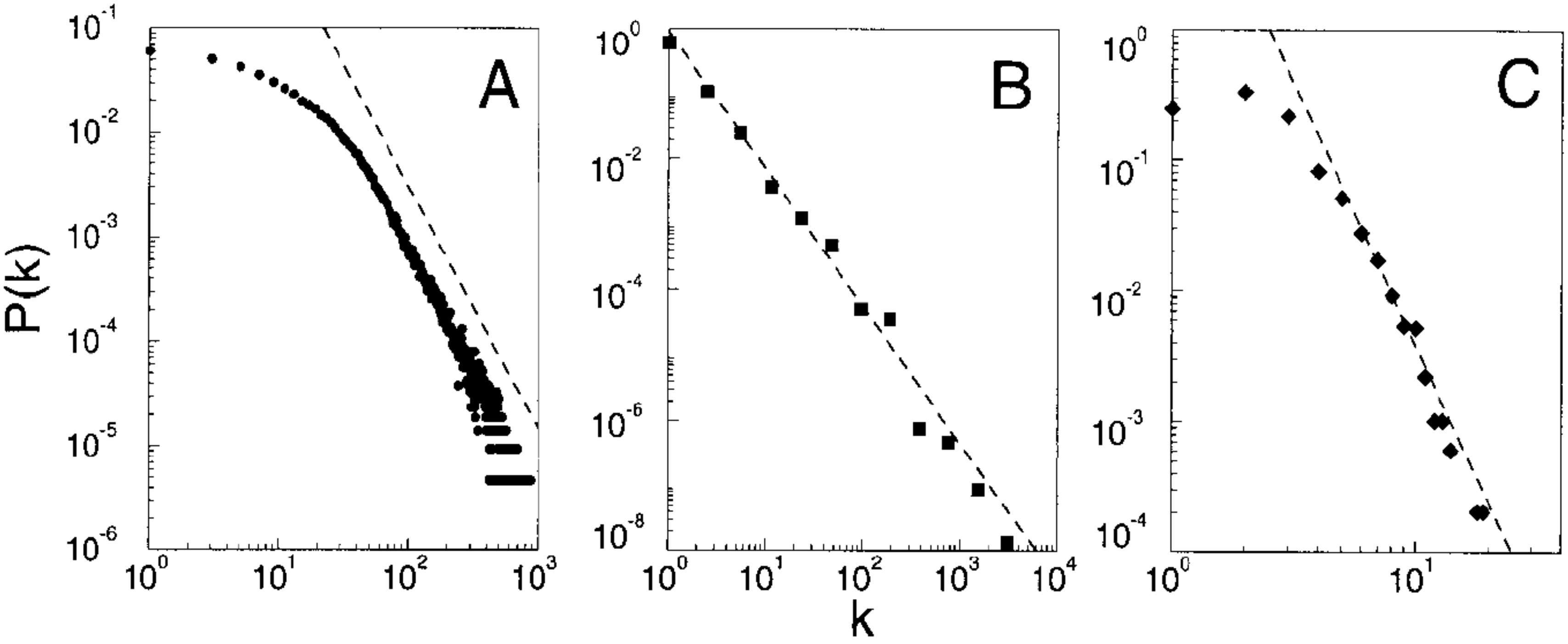


(a) Int-11-97

Internet domain topology

Barabasi&Albert

- [Barabasi-Albert, 1999]



Actor collaborations

Web graph

Power-grid

[Leskovec et al. KDD ‘08]

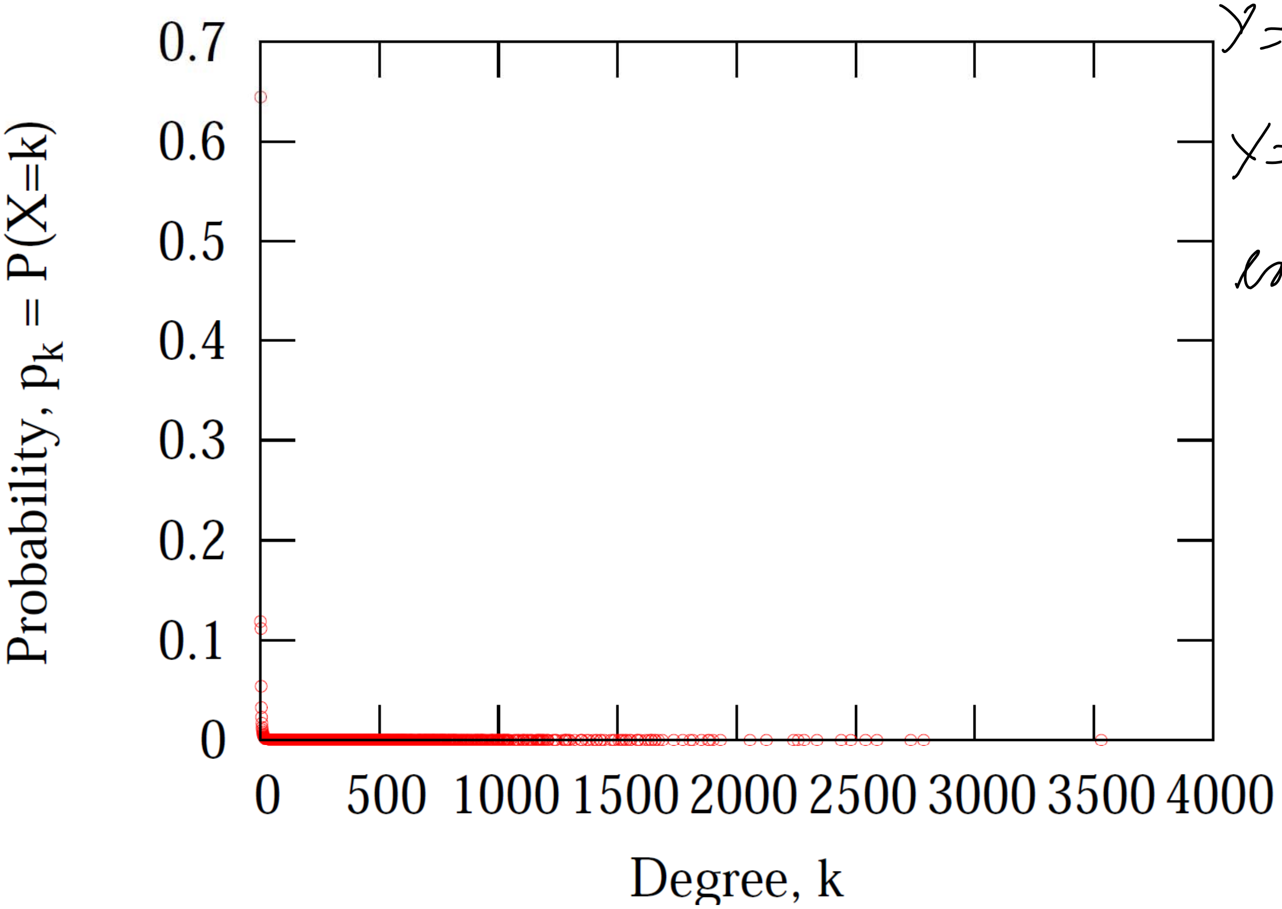
Degrees in real networks

- Take real network plot a histogram of p_k vs. k

$y = x^{-\alpha} \sim \gamma$

$y = e^{-x}$

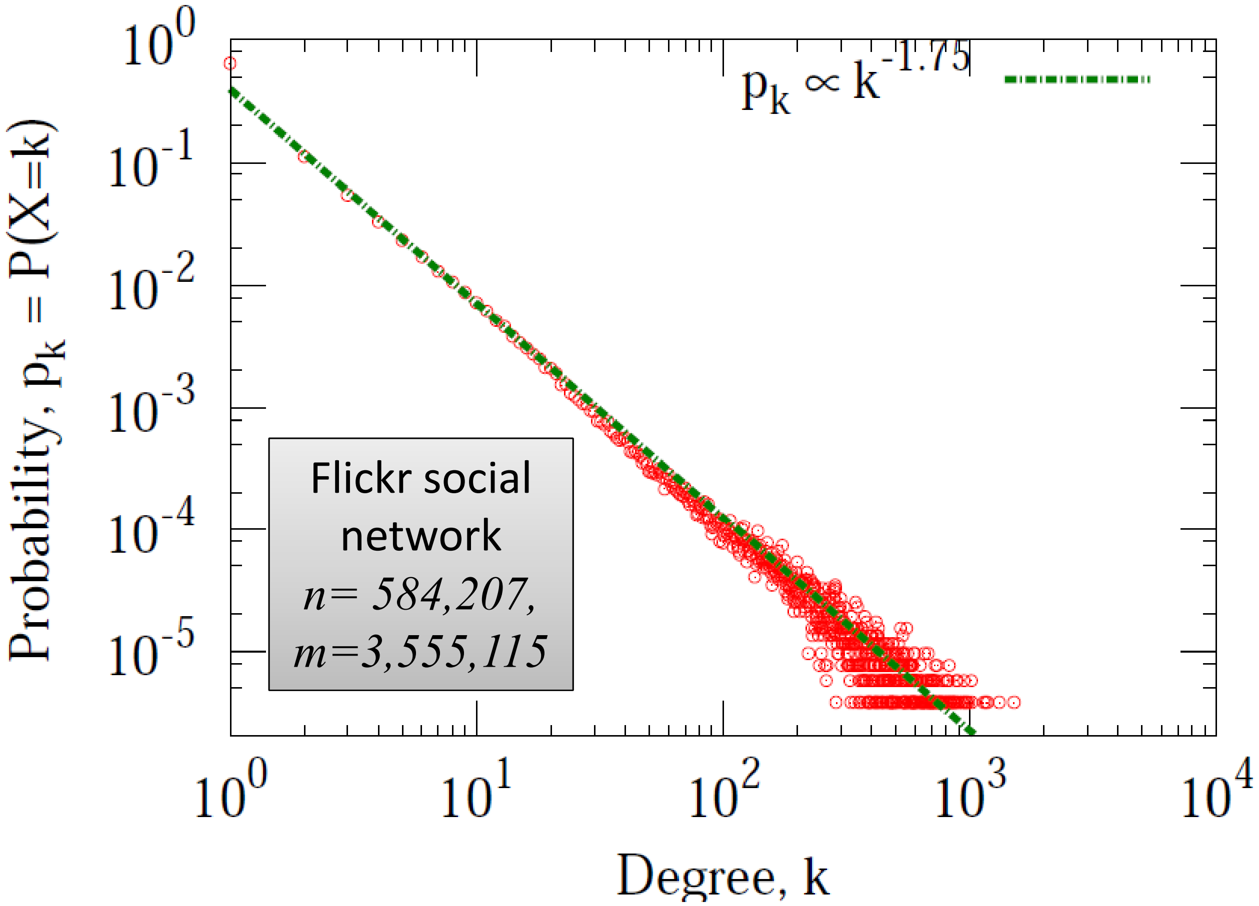
$\log y = -\alpha \log x$



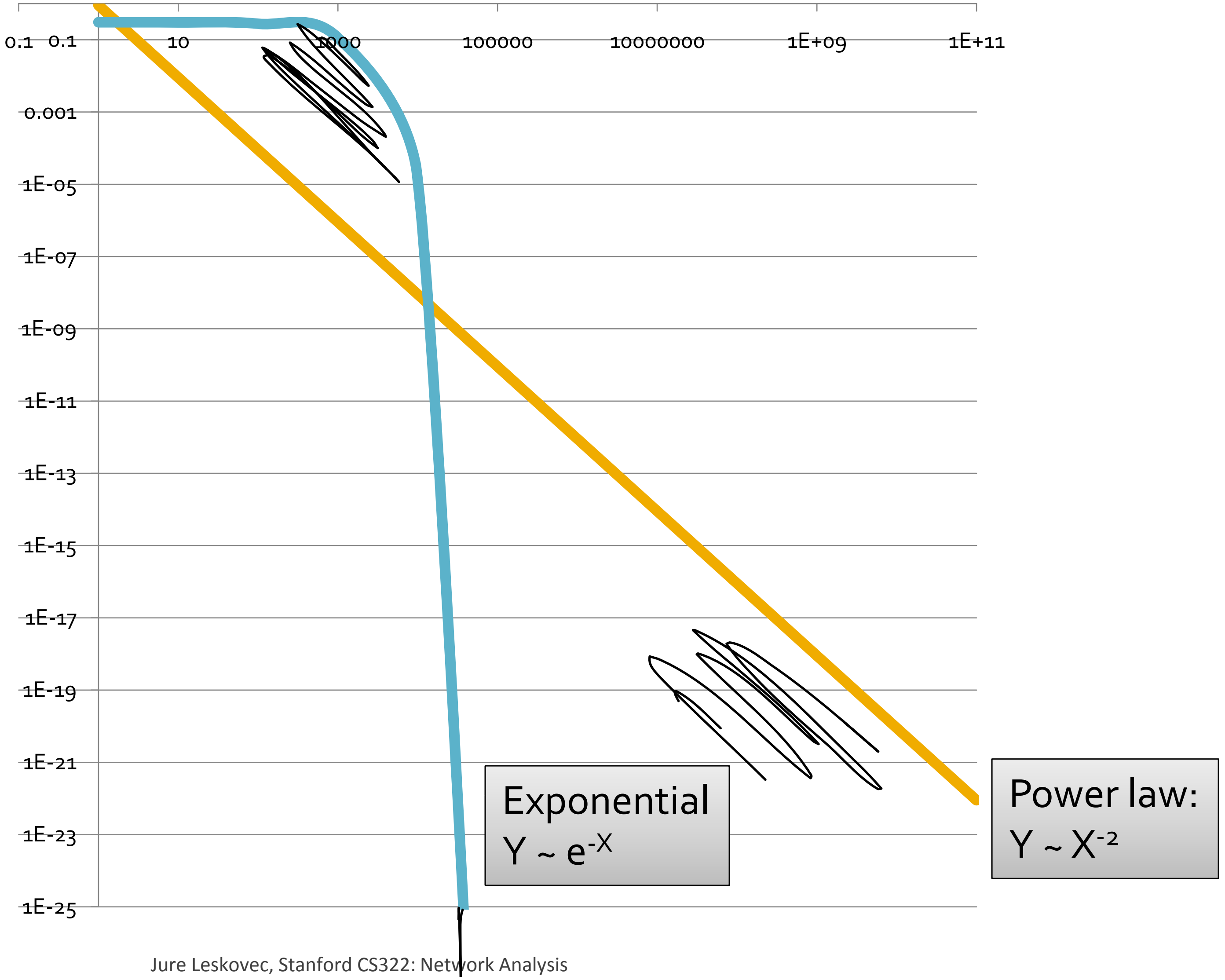
Flickr social network
n= 584,207,
m=3,555,115

Degrees in real networks (2)

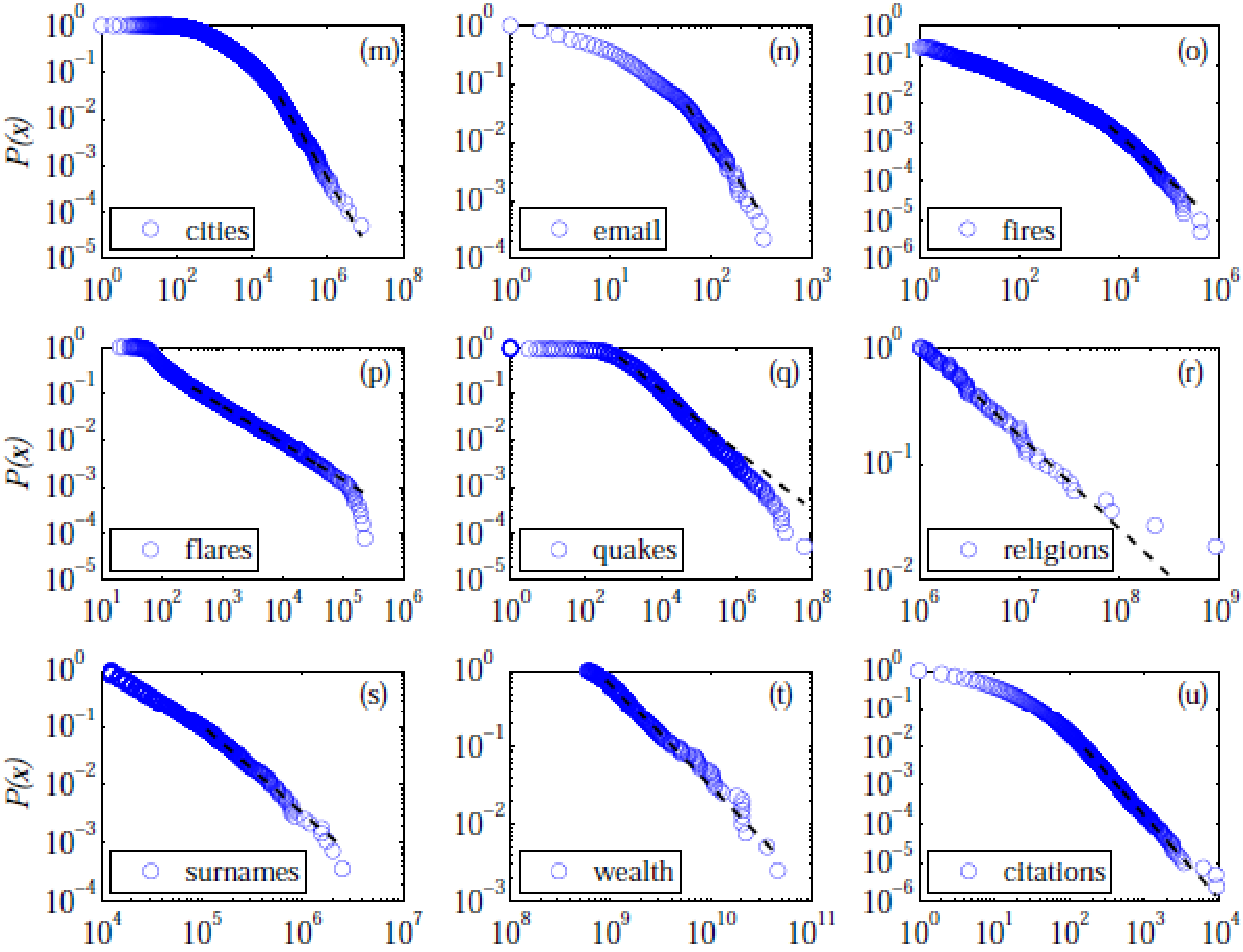
- Plot the same data on *log-log* axis:



Exponential tail vs. Power-law tail



Power-laws are everywhere



Many other quantities follow heavy-tailed distributions

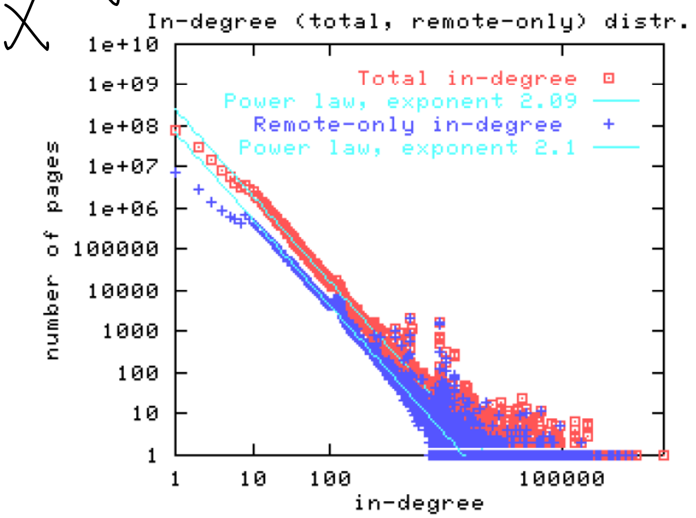
Not everyone likes power-laws ☺

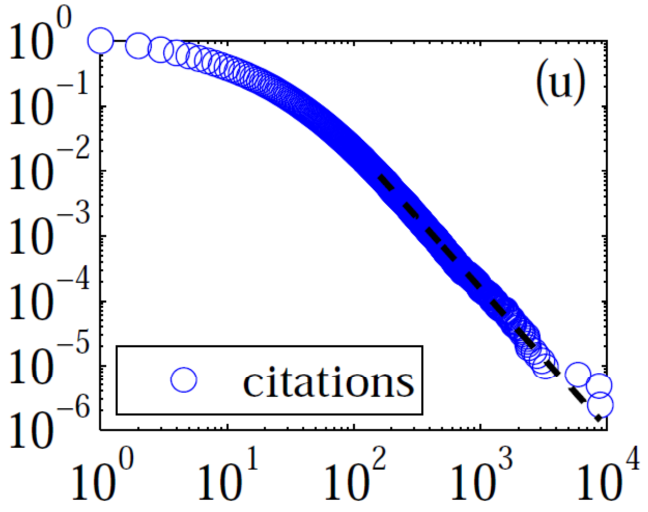
CMU students protesting at the G20 meeting in Pittsburgh in Sept 2009

Power law degree exponents

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

- Power law degree exponent is typically $2 < \alpha < 3$
  - Web graph [Broder et al. 00]:
    - $\alpha_{in} = 2.1$, $\alpha_{out} = 2.4$
  - Autonomous systems [Faloutsos et al. 99]:
    - $\alpha = 2.4$
  - Actor collaborations [Barabasi-Albert 00]:
    - $\alpha = 2.3$
  - Citations to papers [Redner 98]:
    - $\alpha \approx 3$
  - Online social networks [Leskovec et al. 07]:
    - $\alpha \approx 2$





[Clauset-Shalizi-Newman 2007]

Mathematics of Power-laws

- What is the normalizing constant?

$p(x) = C \cdot x^{-\alpha}$ $C = ?$

$1 = \int_{x_{min}}^{\infty} p(x)\,dx = C \int_{x_{min}}^{\infty} x^{-\alpha}\,dx$

$$= \frac{C}{\alpha - 1}\left[x^{-\alpha+1}\right]_{x_{min}}^{\infty}$$

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

[Clauset-Shalizi-Newman 2007]

Mathematics of Power-laws

- $E[x] = \int_{x_{min}}^{\infty} x\, p(x)\, dx = C \int_{x_{min}}^{\infty} x^{-\alpha+1}\, dx$

$$= \frac{C}{2-\alpha} \left[x^{-\alpha+2} \right]_{x_{min}}^{\infty}$$

$$E(X) = \frac{\alpha-1}{\alpha-2} X_{min}$$

- Tails are heavy:
  - If $\alpha \leq 2$: $\mathrm{E}[x] = \infty$
  - If $\alpha \leq 3$: $\mathrm{Var}[x] = \infty$

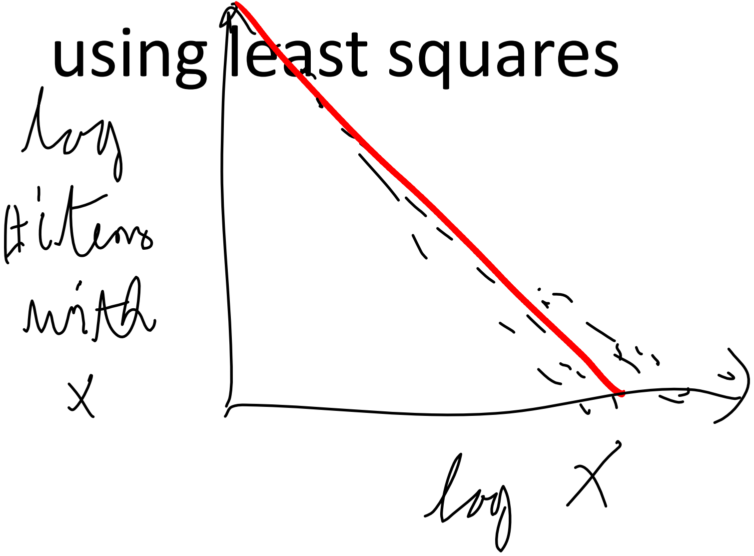
x_i is degree of node i

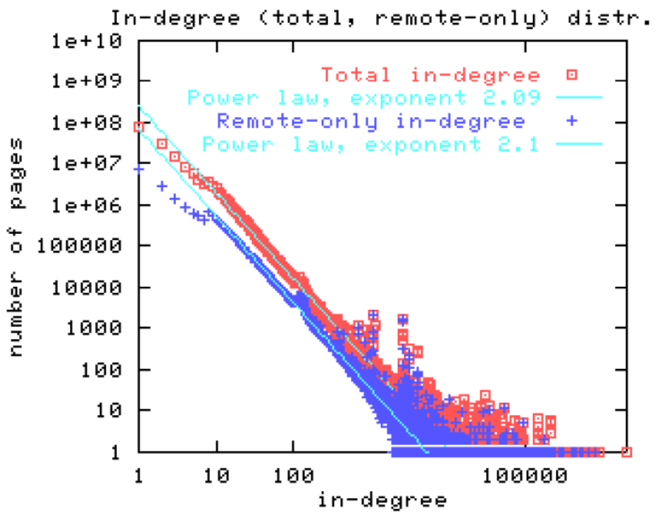
[Clauset-Shalizi-Newman 2007]

Estimating α

- Estimating α from data:

BAD! 1. Fit a line on log-log axis using least squares





[Clauset-Shalizi-Newman 2007]

Estimating α

- Estimating α from data:

Ok

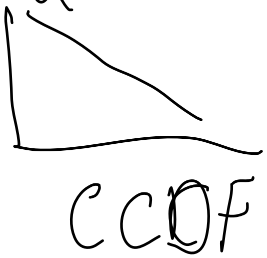
2. Plot Complementary CDF $P(X>x)$

   Then $\alpha=1+\alpha^*$ where α^* is the slope of $P(X>x)$.

   e.i., if $P(X=x)\ x^{-\alpha}$ then $P(X>x)=x^{-(\alpha-1)}$

$$P(X>x) = \sum_{j=x}^{\infty} p(x) = \int_{x}^{\infty} c\, j^{-\alpha}\, dj$$

$$= \frac{c}{\alpha} x^{-(\alpha-1)}$$



[Clauset-Shalizi-Newman 2007]

Estimating α

- Estimating power-law exponent α from data:

Ok 3. Use MLE: $1+n\left[\sum_{i=1}^{n}\ln\frac{x_i}{x_{min}}\right]^{-1}$ x_i is degree of node i

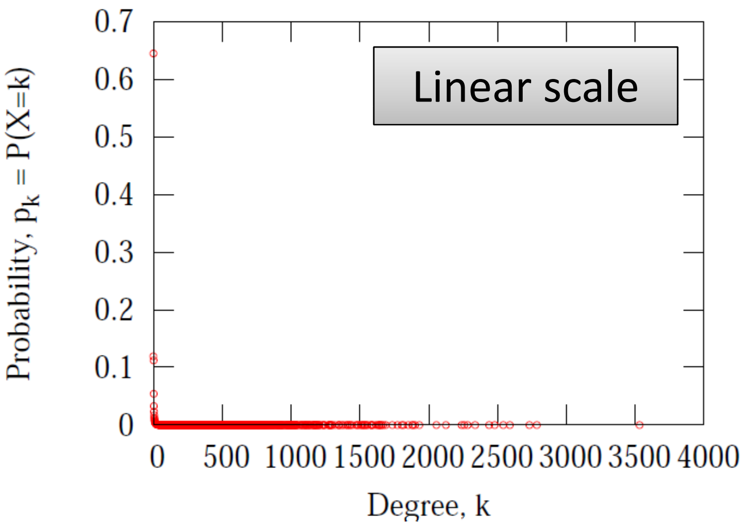
$$p(x) = \frac{\alpha-1}{x_{min}}\left(\frac{x}{x_{min}}\right)^{-\alpha}$$

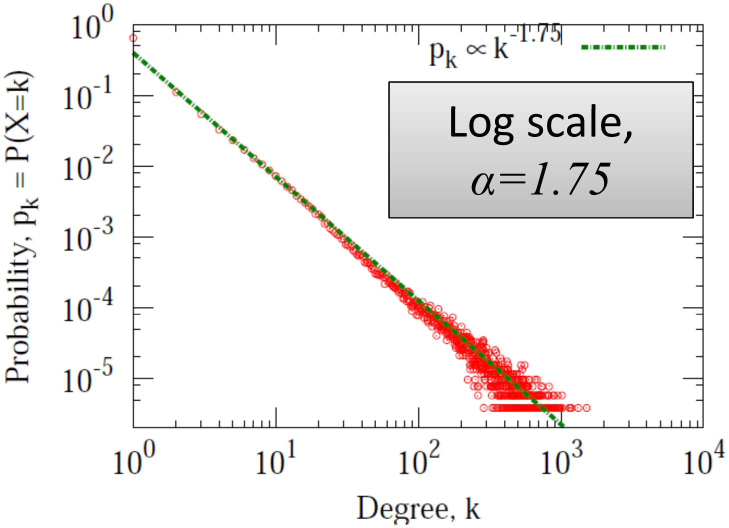
$$\mathcal{L}(\alpha) = \log\prod_i p(x_i) = \sum_i^n \log p(x_i)$$

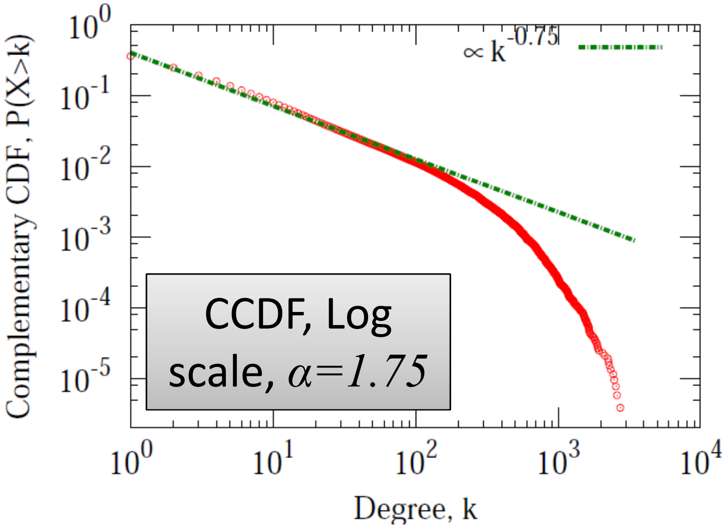
$$= \sum_i^n \left[\log(\alpha-1) - \log x_{min} - \alpha\log\left(\frac{x_i}{x_{min}}\right)\right]$$

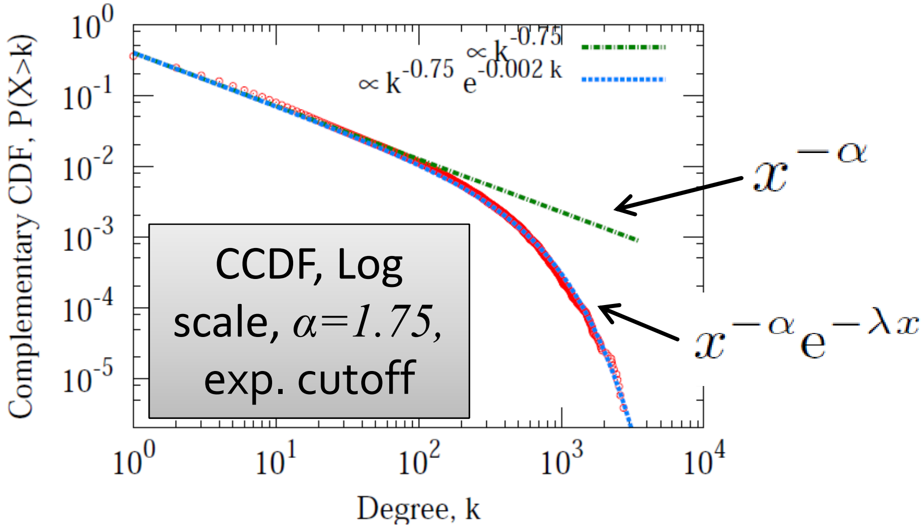
$$\frac{d\mathcal{L}}{d\alpha} = 0 \Rightarrow \frac{n}{\alpha-1} - \sum_i^n \log\left(\frac{x_i}{x_{min}}\right) \Rightarrow \hat{\alpha} = 1+n\left[\sum \log\frac{x_i}{x_{min}}\right]^{-1}$$

Flickr: Degree exponent







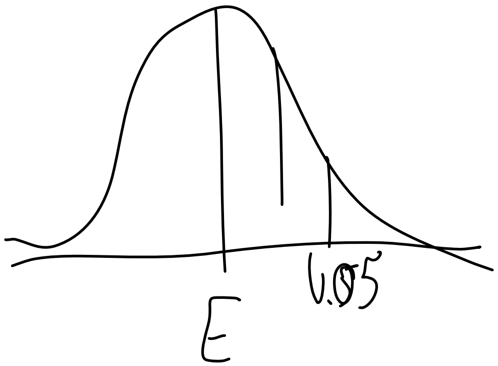


Why are power-laws surprising

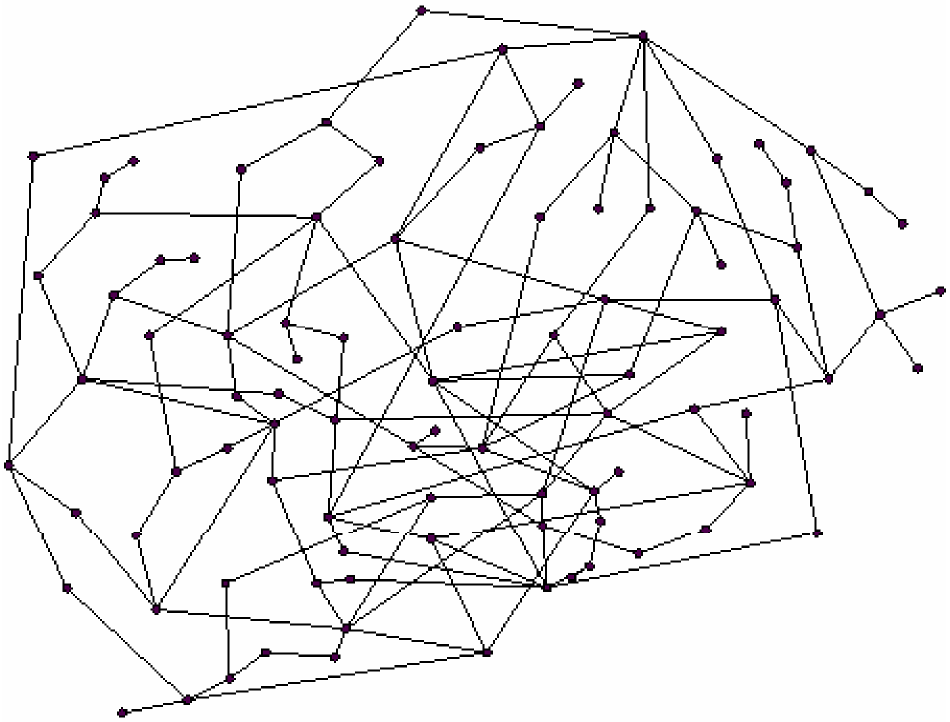
- Can not arise from sums of independent events
  - Recall: in G_{np} each pair of nodes in connected independently with prob. p

X_i ... deg of node i

$$E(X_i) = E\left[\left(\sum_u X_u\right)\right] = \sum_{i}^{n-1} E[X_u] = \underline{\underline{p(n-1)}}$$

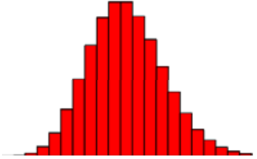


Random vs. Scale-free network

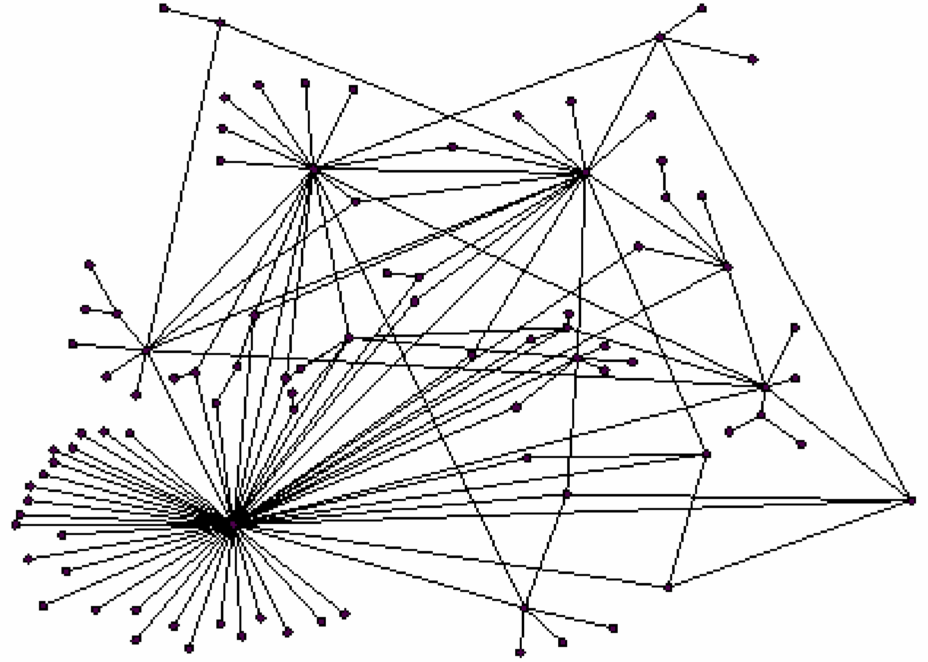


Random network

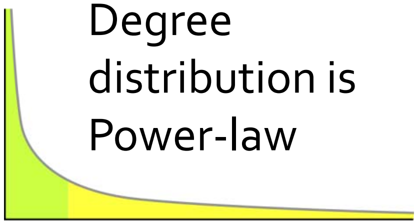
(Erdos-Renyi random graph)



Degree distribution is Binomial



Scale-free (power-law) network



Degree distribution is Power-law

Function is scale free if:

$f(ax) = c\,f(x)$

What is a good model?

- What is a good model that gives rise to power-law degree distributions?

- What is the analog of central limit theorem for power-laws?

Model: Preferential attachment

- Preferential attachment [Price 1965, Albert-Barabasi 1999]:
  - Nodes arrive in order
  - A new node j creates m out-links
  - Prob. of linking to a previous node i is proportional to its degree d_i

$$P(j \rightarrow i) = \frac{d_i}{\sum_k d_k}$$

Rich-get-richer

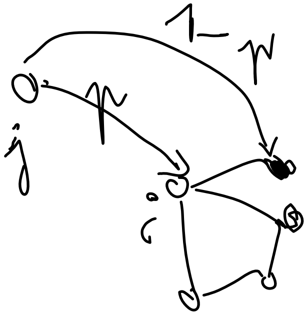
- New nodes are more likely to link to nodes that already have high degree

- Herbert Simon's result
  - Power-laws arise from "Rich get richer" (cumulative advantage)

- Examples [Price 65]:
  - Citations: new citations of a paper are proportional to the number it already has

[Mitzenmacher, '03]

Rich-get-richer



- Pages are created in order 1,2,3,…,N
- When page j is created it produces a link to earlier webpages:

  1) With prob. p, page j creates a link to a page i chosen uniformly at random (from among all earlier pages)

  2) With prob. 1-p, page j chooses a page i uniformly at random (from among all earlier pages) and creates a link to the page i points to.

Note this is same as saying:

2)With prob. 1-p, page j creates a link to the page i with prob. proportional to d_i (the degree of i)

The model givens power-laws

- Claim: the described model generates networks where the fraction of nodes with degree d scales as:

$$d^{-\left(1+\frac{1}{q}\right)}$$

where

$$q = 1-p$$

Continuous approximation (1)

- Degree $d_i(t)$ of node i ($i=1,2,\ldots,n$) is a continuous quantity and it grows deterministically as a function of time t.

- Analyze $d_i(t)$ – continuous degree of node i at time $t \geq i$.

Continuous approximation (2)

What do we know? $d_i(i) = 0$

- Initial condition: $d_i(t)=0$, when $t=i$ (i just arrived)
- Expected change of $d_i(t)$ over time:
  - Node i gains an in-link at step $t+1$ only if a link from a newly created node $t+1$ points to it.
  - What's the prob. of this event?
    - With prob. p node $t+1$ links to a random node: links to i with prob. $1/t$
    - With prob $1\text{-}p$ node $t+1$ links preferentially: links to i with prob. $d_i(t)/t$
  - So: prob. node $t+1$ links to i is

$$p \frac{1}{t} + (1-p) \frac{d_i(t)}{2t}$$

Continuous approximation (3)

- At time t we have t nodes
- What is the rate of growth of d_i?

$q = 1 - p$

$$\frac{d\,d_i}{dt} = p\,\frac{1}{t} + (1-p)\frac{d_i}{t}$$

$$\frac{d\,d_i}{dt} = \frac{p + q\,d_i}{t}$$

What is the rate of growth of d_i?

$$\frac{d\,d_i}{dt} = \frac{p+q\,d_i}{t} \Rightarrow \frac{1}{p+q\,d_i}\,\frac{d\,d_i}{dt} = \frac{1}{t}$$

$$\int \frac{1}{p+q\,d_i}\,\frac{d\,d_i}{dt}\,dt = \int \frac{1}{t}\,dt \Rightarrow \log(p+q\,d_i) = q\log t \underset{+C}{}$$

$$p+q\,d_i = A\,t^{q} \qquad A = e^{c}$$

$$\underline{d_i(t) = \frac{1}{q}\left(A\,t^{q} - p\right)}$$

What is the constant A?

- We know: $d_i(i) = 0$

$$\frac{1}{q}\left(A i^q - r\right) = 0 \Rightarrow A = \frac{r}{i^q}$$

$$d_i(t) = \frac{1}{q}\left[\frac{r}{i^q} t^q - r\right]$$

$$d_i(t) = \frac{r}{q}\left[\left(\frac{t}{i}\right)^q - 1\right]$$

Degree distribution

- Fraction of nodes with degree $>d$ at time t

$$\frac{m}{q}\left[\left(\frac{t}{i}\right)^{q}-1\right]>d$$

$$\Rightarrow\ i \le \underbrace{t}_{\uparrow\ \#\text{nodes}}\ \underbrace{\left[\frac{q}{m}\cdot d+1\right]^{-\frac{1}{q}}}_{\text{fraction}}$$

Degree distribution

- Fraction of nodes with degree exactly d at time t?

$$\left[\frac{q}{p}d+1\right]^{-\frac{1}{q}}$$

$$= \frac{1}{q}\,\frac{q}{p}\left[\frac{q}{p}d+1\right]^{\overbrace{-\frac{1}{q}-1}^{\alpha}} \propto d^{-\left(1+\frac{1}{q}\right)}$$

$$= d^{-\alpha}$$

$$\Rightarrow\ \alpha = \underline{1+\frac{1}{q}} = 1+\frac{1}{(1-q)}$$