LECTURE 7: NETWORK FORMATION PROCESSES

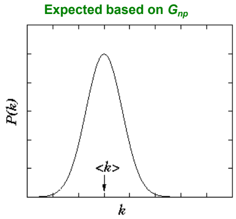
Network Formation Processes

What do we observe that needs explaining

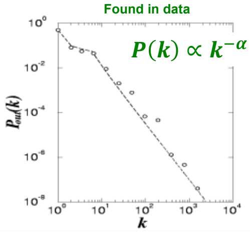
- **Small-world model?**
  - Diameter
  - Clustering coefficient
- **Preferential Attachment:**
  - **Node degree distribution**
    - What fraction of nodes has degree k (as a function of k)?
    - Prediction from simple random graph models: $p(k)$ = exponential function of k
    - **Observation: Power-law:** $p(k) = k^{-\alpha}$

Degree Distributions

Expected based on G_{np}

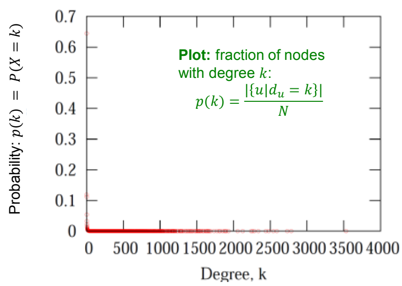


Found in data



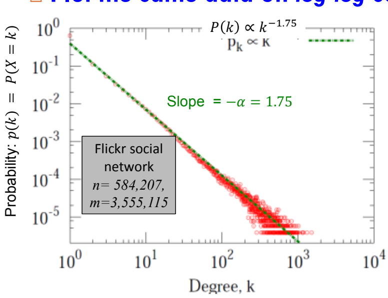
Node Degrees in Networks

- **Take a network, plot a histogram of $P(k)$ vs. k**



Node Degrees in Networks

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

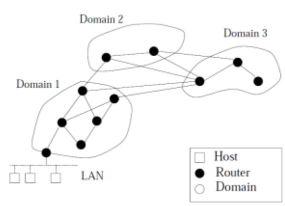


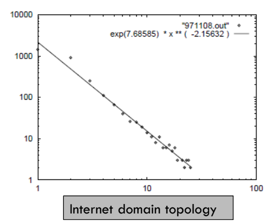
How to distinguish:
$P(k) \propto \exp(-k)$ **vs.** $P(k) \propto k^{-\alpha}$ **?**

Take logarithms:
if $y = f(x) = e^{-x}$ then
$\log(y) = -x$
If $y = x^{-\alpha}$ then
$\log(y) = -\alpha \log(x)$
So, on log-log axis power-law looks like a straight line of slope $-\alpha$!

Node Degrees: Faloutsos[3]

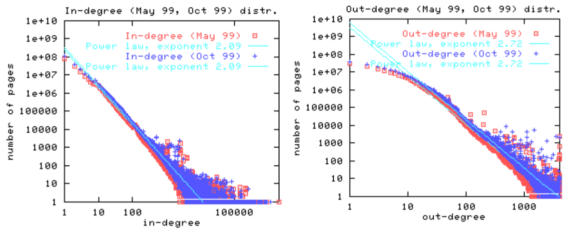
- **Internet Autonomous Systems**
  [Faloutsos, Faloutsos and Faloutsos, 1999]





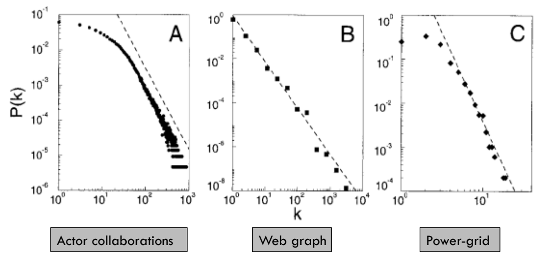
Node Degrees: Web

- **The World Wide Web** [Broder et al., 2000]

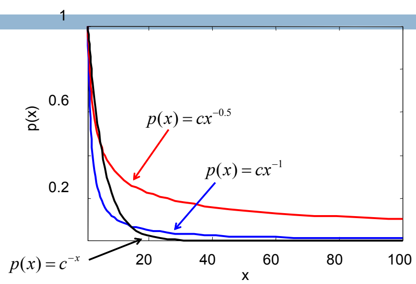


Node Degrees: Barabasi&Albert

- **Other Networks** [Barabasi-Albert, 1999]



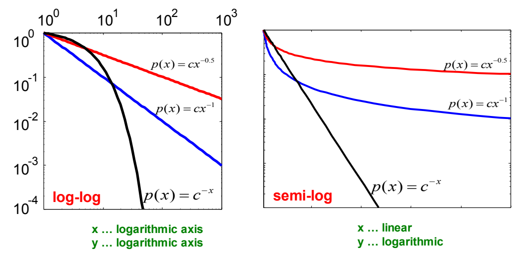
Exponential vs. Power-Law



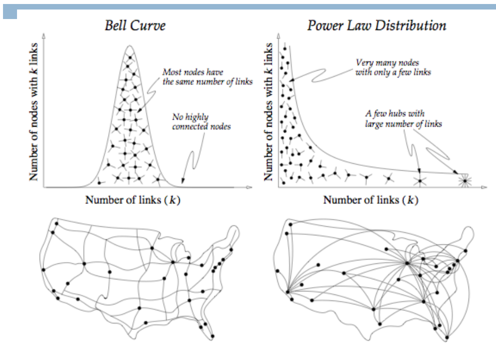
- Above a certain x value, the power law is always higher than the exponential!

Exponential vs. Power-Law

- **Power-law vs. Exponential on log-log and log-lin scales**

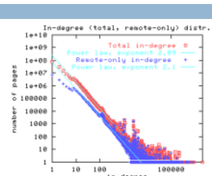


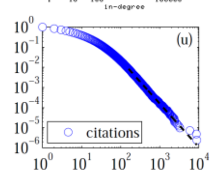
Exponential vs. Power-Law



Power-Law Degree Exponents

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





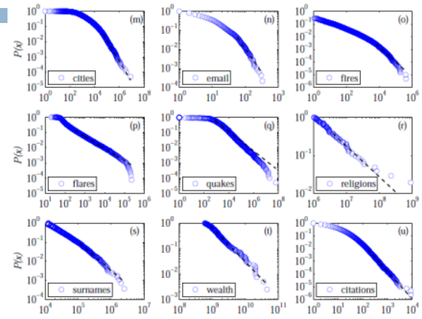
Scale-Free Networks

13

- **Definition:**
  Networks with a power law tail in their degree distribution are called "scale-free networks"
- **Where does the name come from?**
  - **Scale invariance:** There is no characteristic scale
  - **Scale-free function:** $\boldsymbol{f(ax) = a^{\lambda} f(x)}$
    - Power-law function: $f(ax) = a^{\lambda} x^{\lambda} = a^{\lambda} f(x)$

Power-Laws are Everywhere

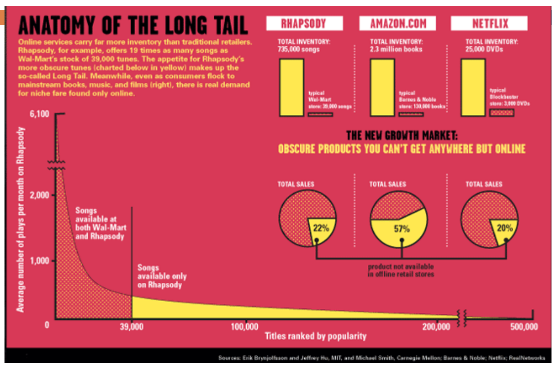
14



Many other quantities follow heavy-tailed distributions

Anatomy of the Long Tail

15



Not Everyone Likes Power-Laws ☺

CMU grad-students at the G20 meeting in Pittsburgh in Sept 2009

MATHEMATICS OF POWER-LAWS

Heavy Tailed Distributions

18

- **Degrees are heavily skewed:**

  Distribution $P(X > x)$ is **heavy tailed if:**

$$\lim_{x \to \infty} \frac{P(X > x)}{e^{-\lambda x}} = \infty$$

- **Note:**
  - **Normal PDF:** $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}}$
  - **Exponential PDF:** $p(x) = \lambda e^{-\lambda x}$
    - **then** $P(X > x) = 1 - P(X \le x) = e^{-\lambda x}$

  are not heavy tailed!

Heavy Tailed Distributions

19

- **Various names, kinds and forms:**
  - Long tail, Heavy tail, Zipf's law, Pareto's law
- **Heavy tailed distributions:**
  - **P(x) is proportional to:**

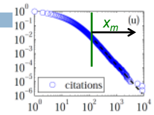
power law	$P(x) \propto \quad x^{-\alpha}$
power law with cutoff	$x^{-\alpha} e^{-\lambda x}$
stretched exponential	$x^{\beta-1} e^{-\lambda x^{\beta}}$
log-normal	$\frac{1}{x} \exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\right]$

Mathematics of Power-laws

20

- **What is the normalizing constant?**

  $p(x) = Z\, x^{-\alpha}$ $\quad Z = ?$
  - $p(x)$ is a distribution: $\int p(x)dx = 1$



$p(x)$ diverges as $x \to 0$ so x_m is the minimum value of the power-law distribution $x \in [x_m, \infty]$

Continuous approximation

- $1 = \int_{x_{min}}^{\infty} p(x)dx = Z \int_{x_m}^{\infty} x^{-\alpha} dx$
- $= -\frac{Z}{\alpha-1}\left[x^{-\alpha+1}\right]_{x_m}^{\infty} = -\frac{Z}{\alpha-1}\left[\infty^{1-\alpha} - {x_m}^{1-\alpha}\right]$
- $\Rightarrow Z = (\alpha-1)x_m^{\alpha-1}$

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

Mathematics of Power-laws

21

- **What's the expectation of a power-law random variable x?**
- $E[x] = \int_{x_m}^{\infty} x\, p(x)dx = z\int_{x_m}^{\infty} x^{-\alpha+1}dx$
- $= -\frac{Z}{2-\alpha}\left[x^{2-\alpha}\right]_{x_m}^{\infty} = -\frac{(\alpha-1)x_m^{\alpha-1}}{2-\alpha}\left[\infty^{2-\alpha} - x_m^{2-\alpha}\right]$

Need: α > 2 !

$$\Rightarrow E[x] = \frac{\alpha-1}{\alpha-2}x_m$$

Power-law density:

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

$Z = \frac{\alpha-1}{x_m^{1-\alpha}}$

Mathematics of Power-Laws

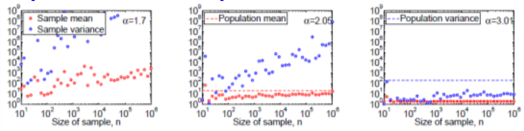
22

- **Power-laws have infinite moments!**

  $$E[x] = \frac{\alpha-1}{\alpha-2}x_m$$
  - If $\alpha \le 2 : E[x] = \infty$
  - If $\alpha \le 3 : Var[x] = \infty$
    - **Average is meaningless, as the variance is too high!**

In real networks **2 < α < 3** so:
E[x] = const
Var[x] = ∞

- **Sample average of *n* samples from a power-law with exponent α:**



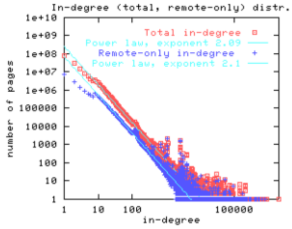
Estimating Power-Law Exponent α

23

Estimating α from data:

- **(1)** Fit a line on log-log axis using least squares:
  - **Solve $arg\min_{\alpha}(\log(y) - \alpha\log(x))^2$**

BAD!



Estimating Power-Law Exponent α

24

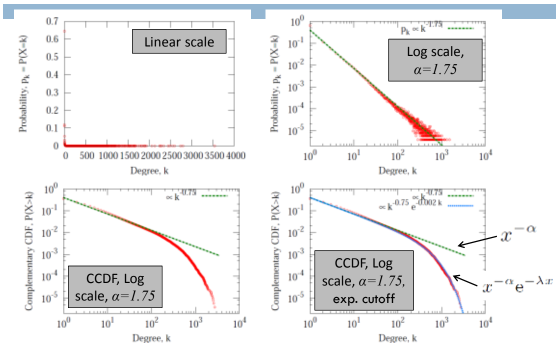
Estimating α from data:

OK!

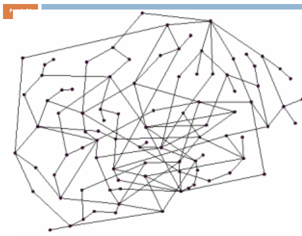
- Plot **Complementary CDF (CCDF)** $P(X \ge x)$. Then the estimated $\alpha = 1 + \alpha'$ where α' is the slope of $P(X > x)$.
- **If $p(x) = P(X = x) \propto x^{-\alpha}$ then $P(X \ge x) \propto x^{-(\alpha-1)}$**
  - $P(X \ge x) = \sum_{j=x}^{\infty} p(j) \approx = \int_x^{\infty} Z\, j^{-\alpha} dj =$
  - $= \frac{Z}{1-\alpha}\left[j^{1-\alpha}\right]_x^{\infty} = \frac{Z}{1-\alpha}x^{-(\alpha-1)}$

Flickr: Fitting Degree Exponent

25

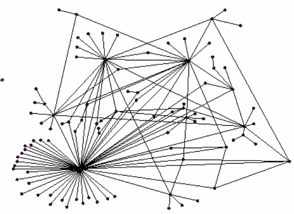


Random vs. Scale-free network

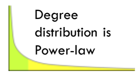


Random network
(Erdos-Renyi random graph)

Degree distribution is Binomial



Scale-free (power-law) network



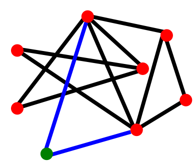
MODEL: PREFERENTIAL ATTACHMENT

Model: Preferential attachment

28

- **Preferential attachment**
  [Price '65, Albert-Barabasi '99, Mitzenmacher '03]
  - Nodes arrive in order 1,2,…,n
  - At step j, let d_i be the degree of node $i < j$
  - A new node j arrives and creates m out-links
  - Prob. of j linking to a previous node i is **proportional to degree d_i of node i**

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



Rich Get Richer

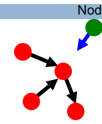
29

- **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 to a paper are proportional to the number it already has

The Exact Model

30

We will analyze the following model:



- Nodes arrive in order $1,2,3,\ldots,n$
- When node j is created it makes a single out-link to an earlier node i chosen:
  - **1)** With prob. p, j links to i chosen uniformly at random (from among all earlier nodes)
  - **2)** With prob. $1-p$, node j chooses node i uniformly at random and links to a node i points to.
    - **This is same as saying:** With prob. $1-p$, node j links to node u with prob. proportional to d_u (the in-degree of u)
    - **Our graph is directed:** Every node has out-degree 1.

The Model Givens Power-Laws

31

- **Claim:** The described model generates networks where the fraction of nodes with in-degree k scales as:

$$P(d_i = k) \propto k^{-(1+\frac{1}{q})}$$

where q=1-p

So we get power-law degree distribution with exponent:

$$\alpha = 1 + \frac{1}{1-p}$$

Preferential attachment: Good news

32

- **Preferential attachment gives power-law degrees**
- Intuitively reasonable process
- Can tune p to get the observed exponent
  - On the web, $P[\text{node has degree } d] \sim d^{-2.1}$
  - $2.1 = 1+1/(1-p)$ → $\underline{p \sim 0.1}$

There are also other network formation mechanisms that generate scale-free networks:

- Random surfer model [Blum-Mugizi]
- Copying model [Kleinberg et al.]
- Forest Fire model [Leskovec et al.]