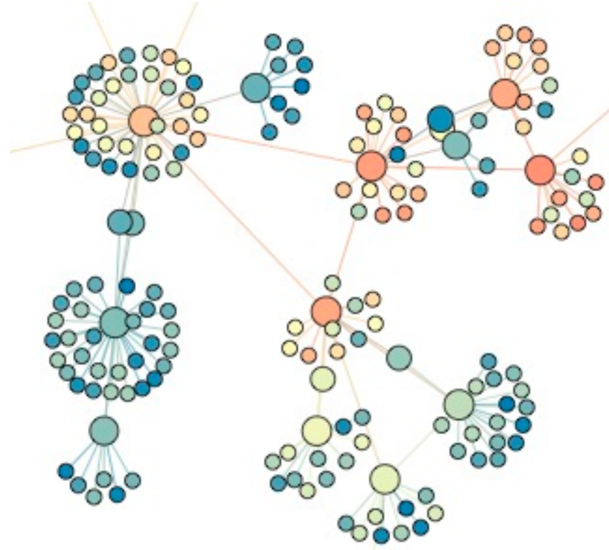


Power laws and preferential attachment

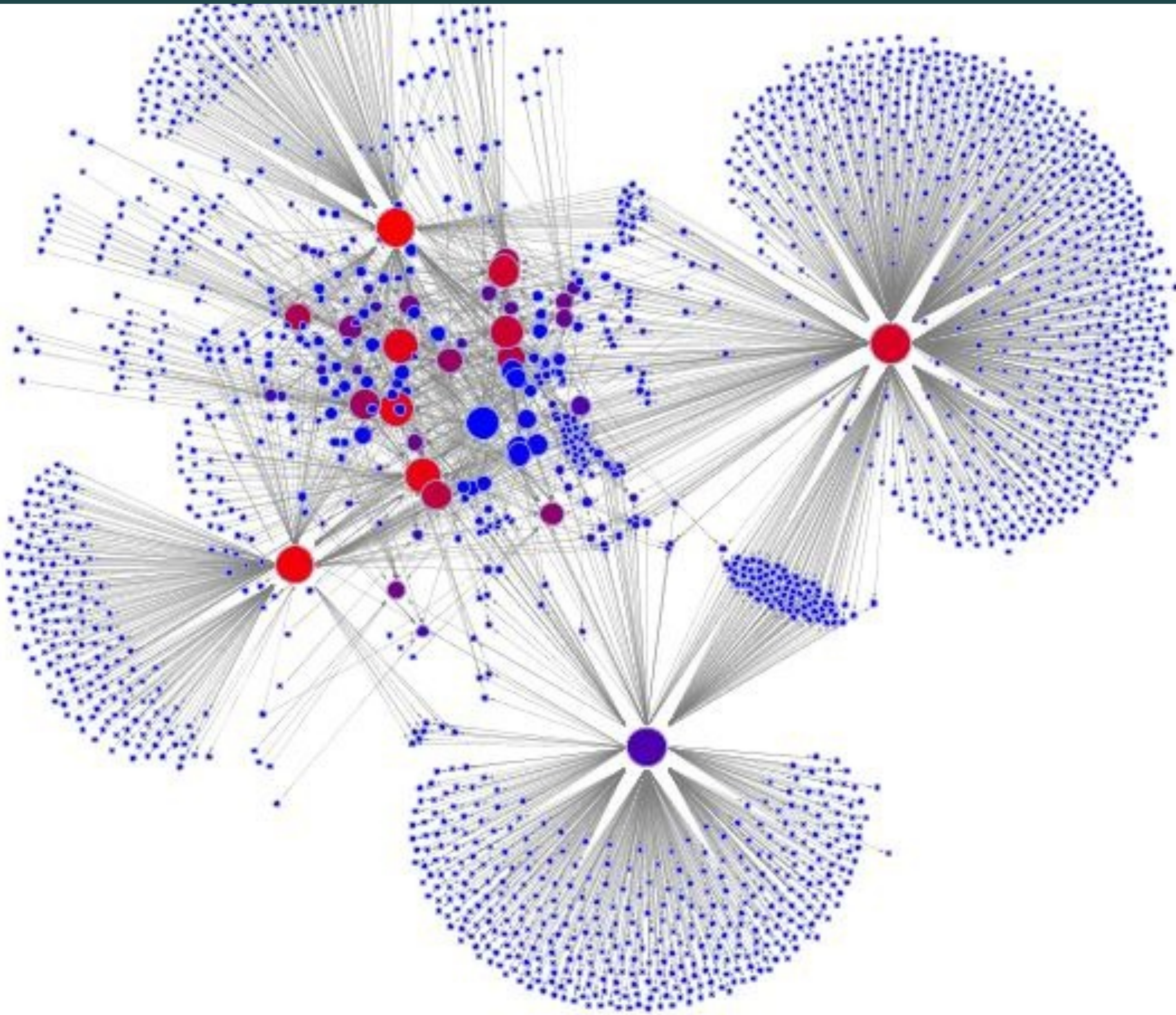
CS224W

From last time: average degree of a neighbor

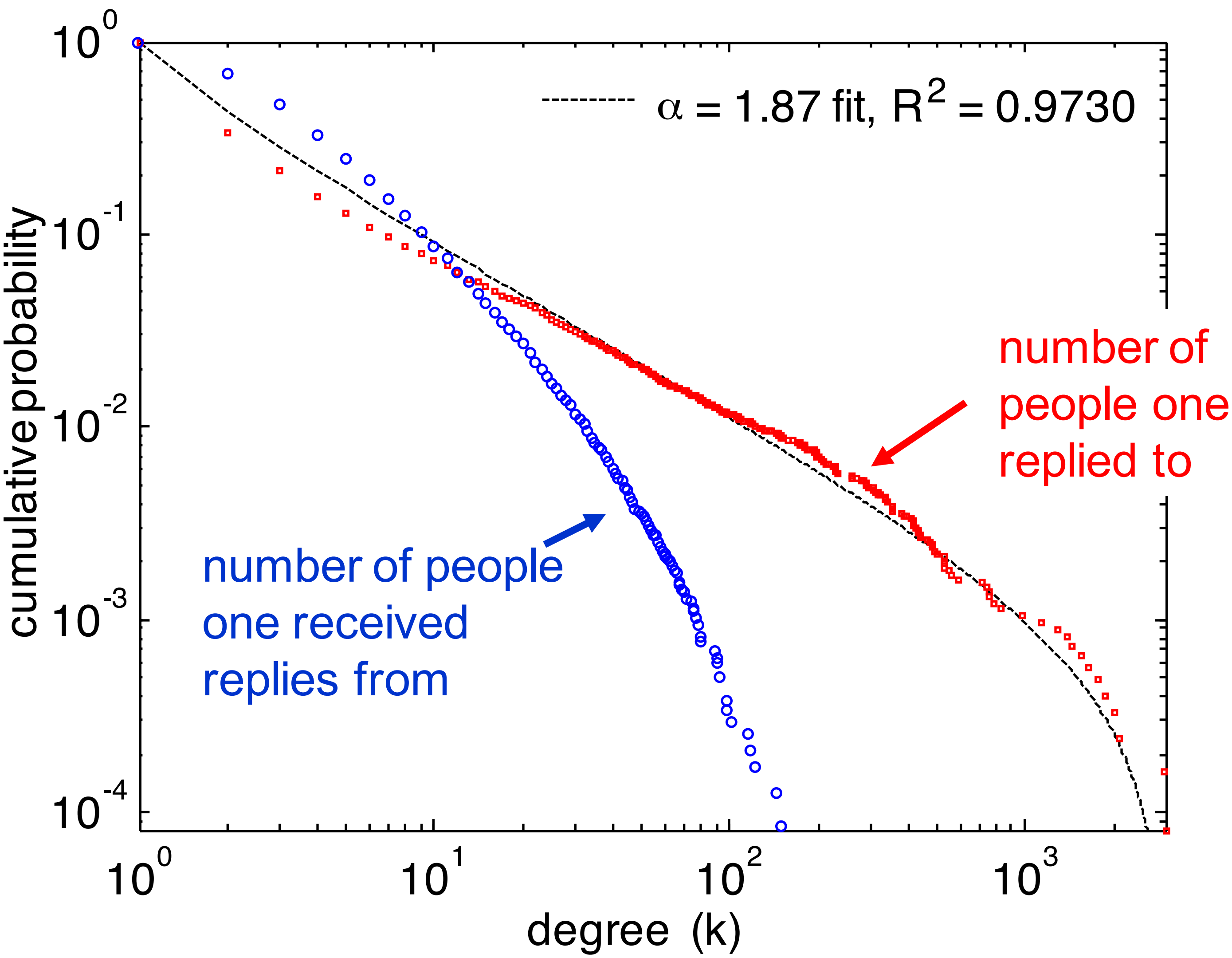
- The probability of our friend having degree k:

(derivation on the board)

Online Question & Answer Forums



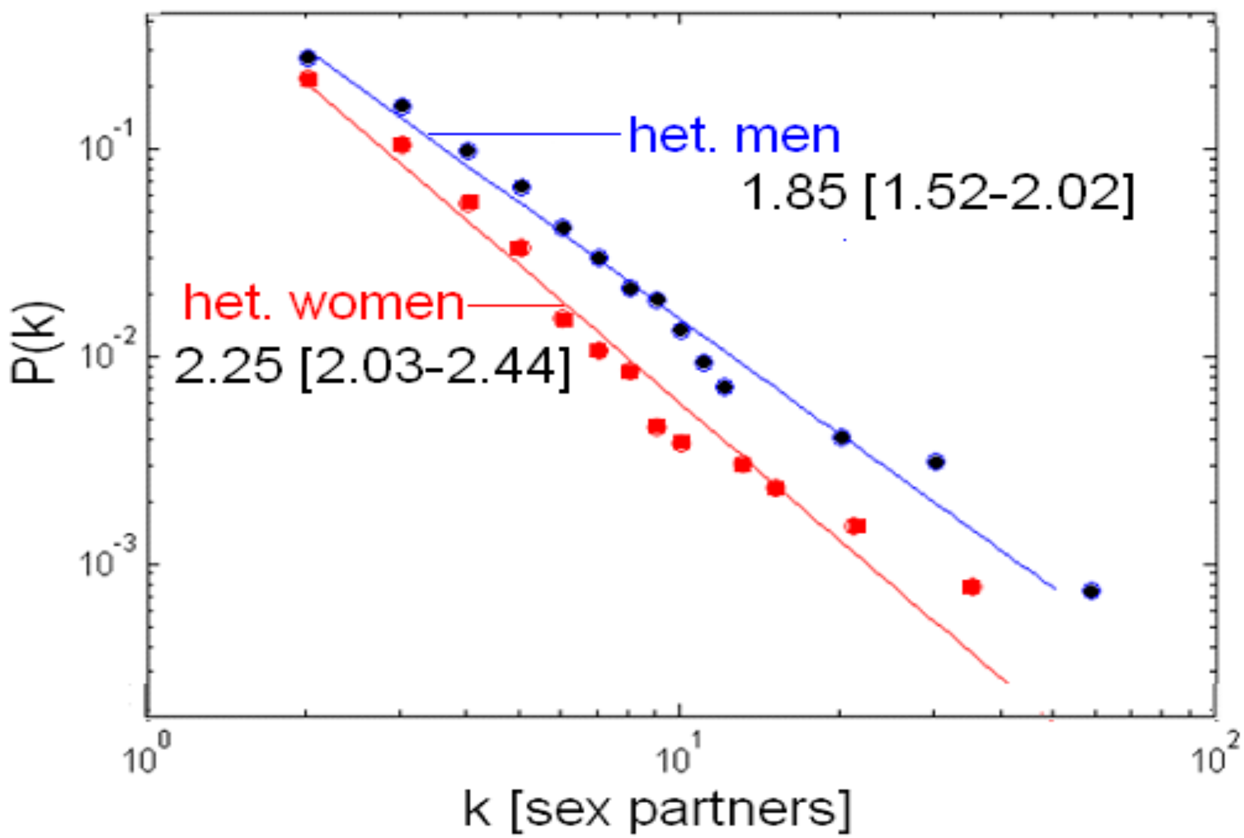
Uneven participation



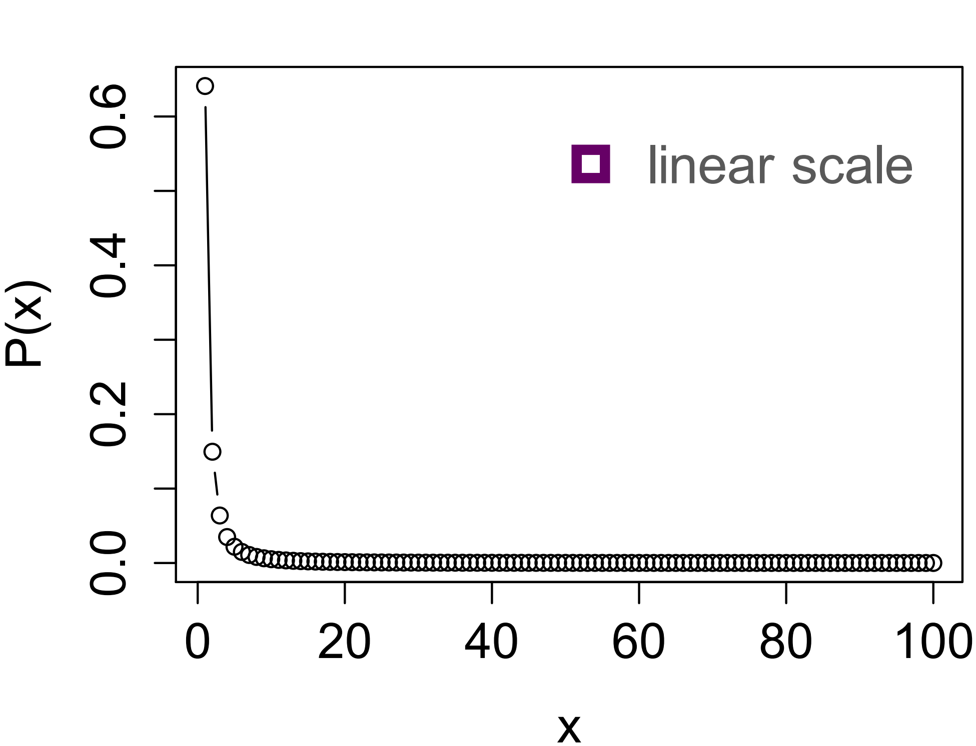
- 'answer people' may reply to thousands of others
- 'question people' are also uneven in the number of repliers to their posts, but to a lesser extent

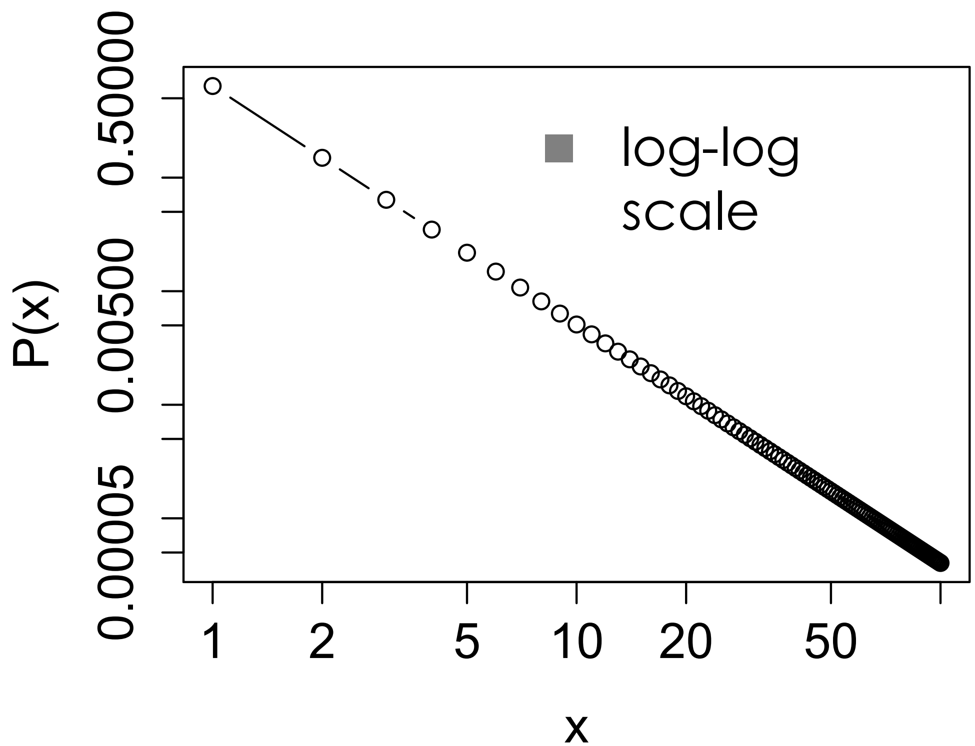
Real-world degree distributions

- Sexual networks
- Great variation in contact numbers



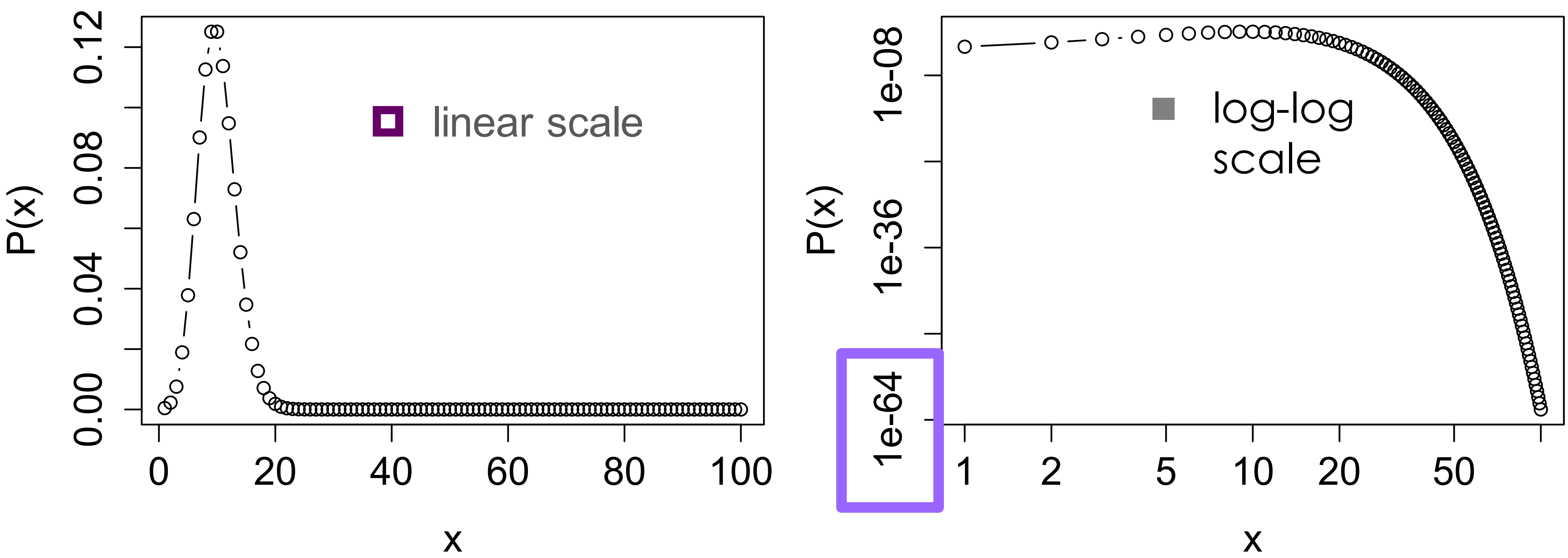
Power-law distribution





- high skew (asymmetry)
- straight line on a log-log plot

Poisson distribution



- little skew (asymmetry)
- curved on a log-log plot

Power law distribution

- Straight line on a log-log plot

$$\ln(p(k)) = c - \alpha \ln(k)$$

- Exponentiate both sides to get that *p(k)*, the probability of observing an node of degree 'k' is given by

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

normalization constant (probabilities over all *k* must sum to 1)

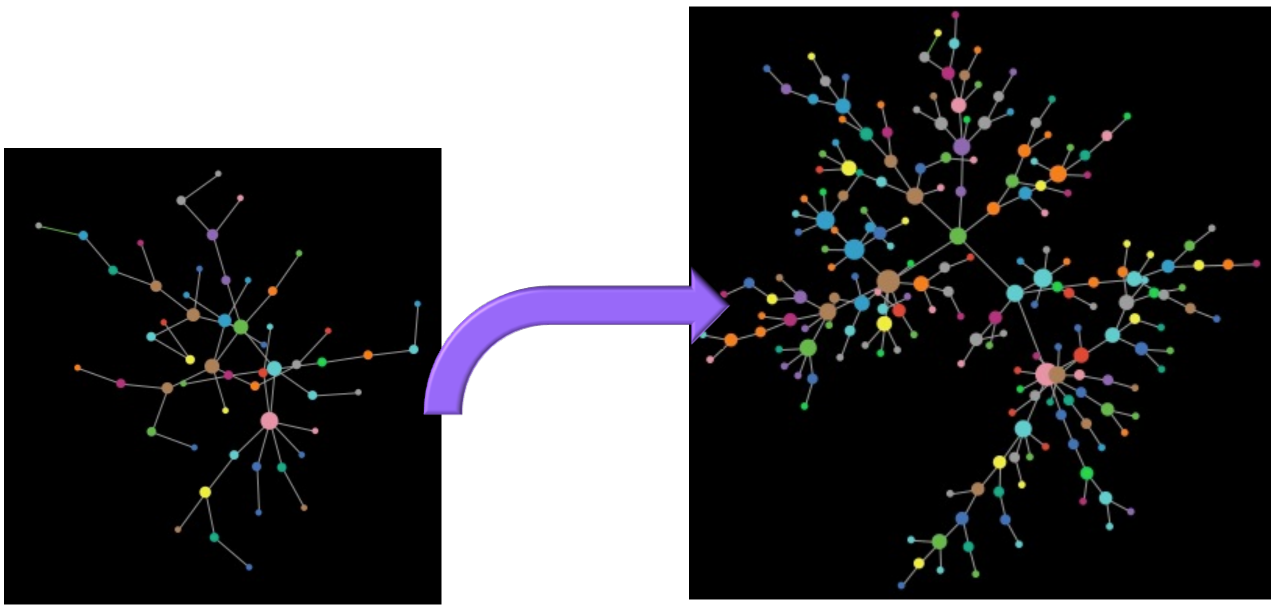
power law exponent α

Quiz Q:

- As the exponent α increases, the downward slope of the line on a log-log plot
  - stays the same
  - becomes milder
  - becomes steeper

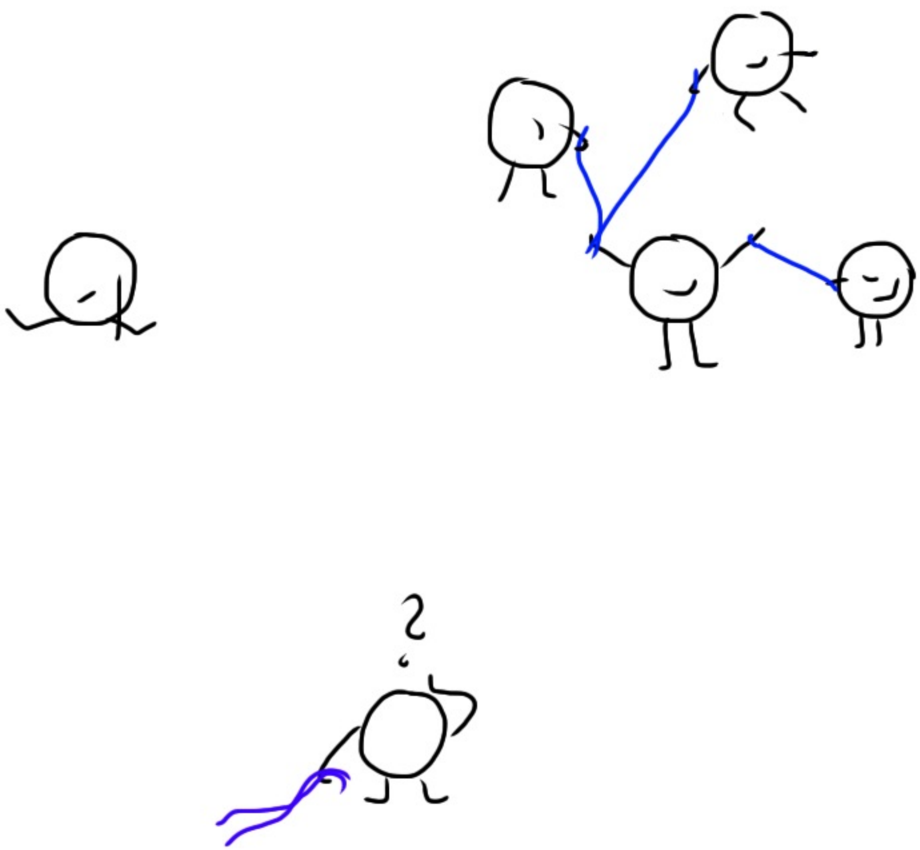
2 ingredients in generating power-law networks

- nodes appear over time (growth)



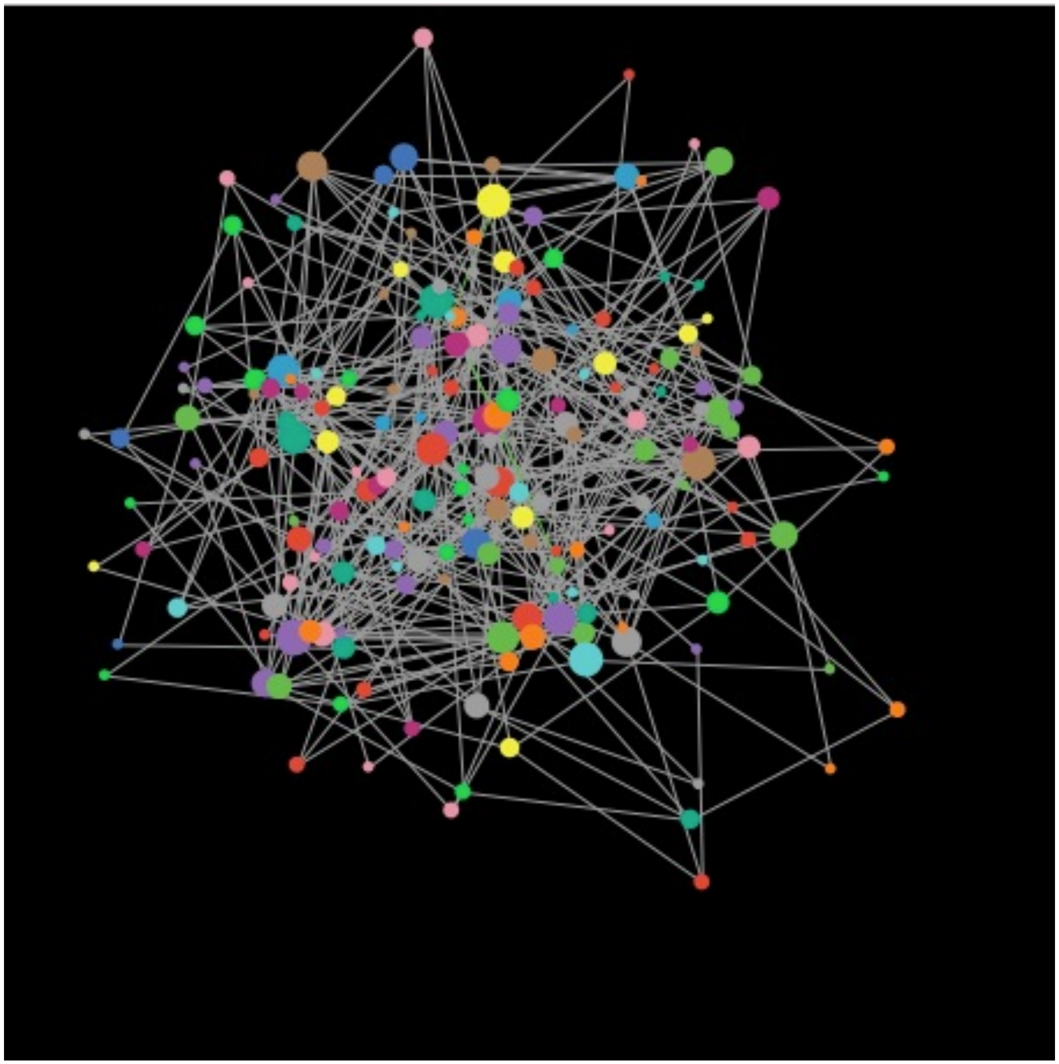
2 ingredients in generating power-law networks

- nodes prefer to attach to nodes with many connections (preferential attachment, cumulative advantage)



Ingredient # 1: growth over time

- nodes appear one by one, each selecting ***m*** other nodes at random to connect to



m = 2

random network growth

- one node is born at each time tick
- at time ***t*** there are ***t*** nodes
- change in degree k_i of node i (born at time i, with $0 < i < t$)

$$\frac{dk_i(t)}{dt} = \frac{m}{t}$$

there are ***m*** new edges being added per unit time (with 1 new node)

the ***m*** edges are being distributed among ***t*** nodes

a node in a randomly grown network

- how many new edges does a node accumulate since it's birth at time *i* until time *t*?
- integrate from *i* to *t*

$$\frac{dk_i(t)}{dt} = \frac{m}{t}$$

to get

$$k_i(t) = m + m\log(\frac{t}{i})$$

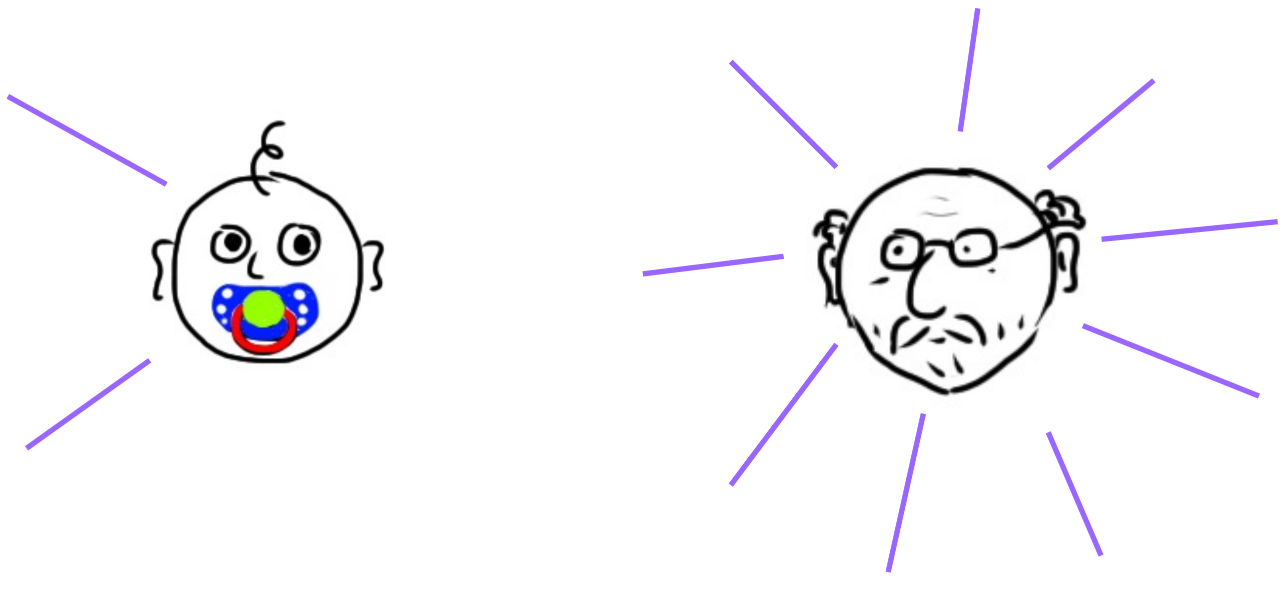
born with ***m*** edges

age and degree

on average $k_i(t) > k_j(t)$

if $i < j$

i.e. older nodes on average have more edges



Quiz Q:

- How could one make the growth model more realistic for social networks?
  - old nodes die
  - some nodes are more sociable
  - friendships vane over time
  - all of the above

growing random networks

Let $\tau(100)$ be the time at which node with degree e.g. 100 is born. The the fraction of nodes that have degree <= 100 is $(t-\tau)/t$

$$k_\tau(t) = m + m\log(\frac{t}{\tau})$$

random growth: degree distribution

- continuing…

$$\log\left(\frac{t}{\tau}\right) = \frac{k-m}{m}$$

$$\frac{\tau}{t} = e^{-\frac{k-m}{m}}$$

exponential distribution in degree

The probability that a node has degree *k* or less is $1-\tau/t$

$$P(k<k') = 1 - e^{-\frac{k'-m}{m}}$$

Quiz Q:

- The degree distribution for a growth model where new nodes attach to old nodes at random will be
  - a curved line on a log-log plot
  - a straight line on a log-log plot

2^{nd} ingredient: preferential attachment

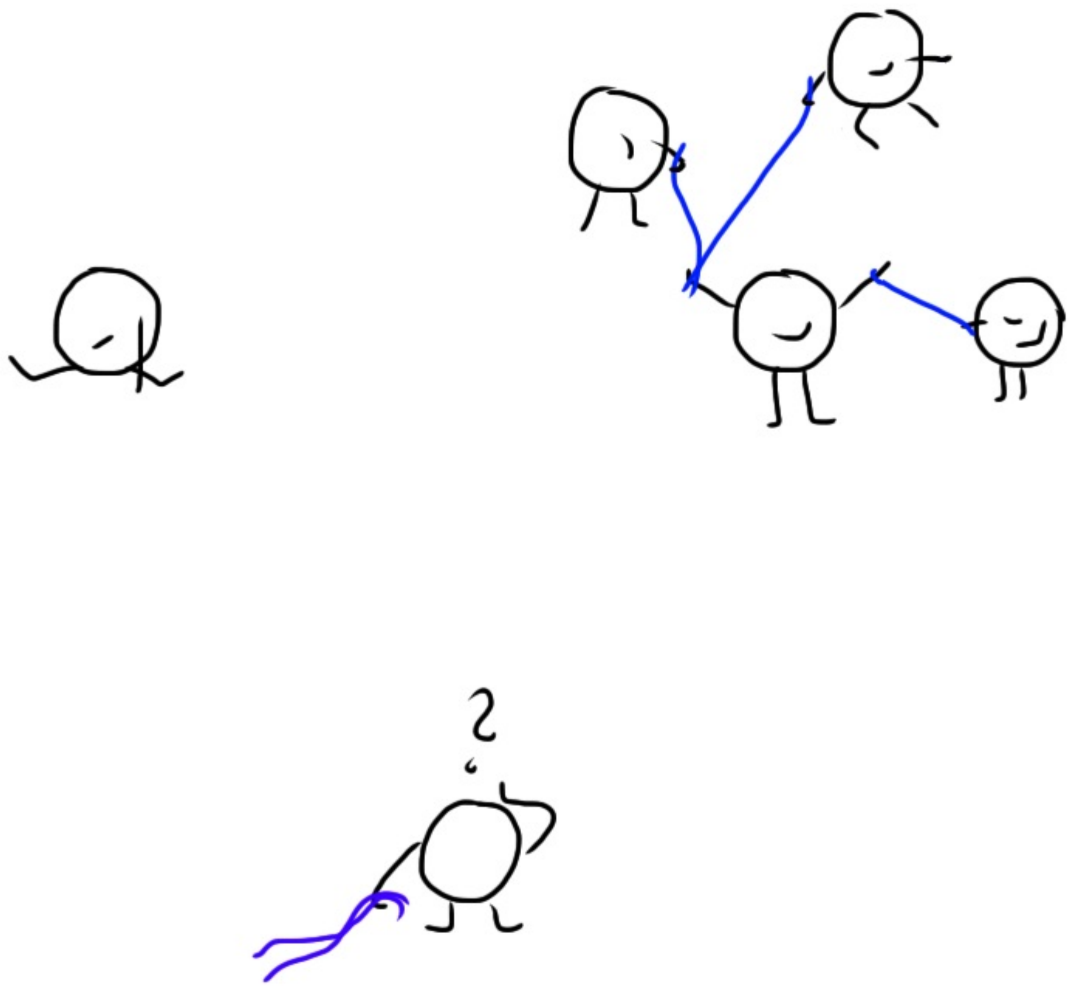
- Preferential attachment:
  - new nodes prefer to attach to well-connected nodes over less-well connected nodes

- Process also known as
  - cumulative advantage
  - rich-get-richer
  - Matthew effect

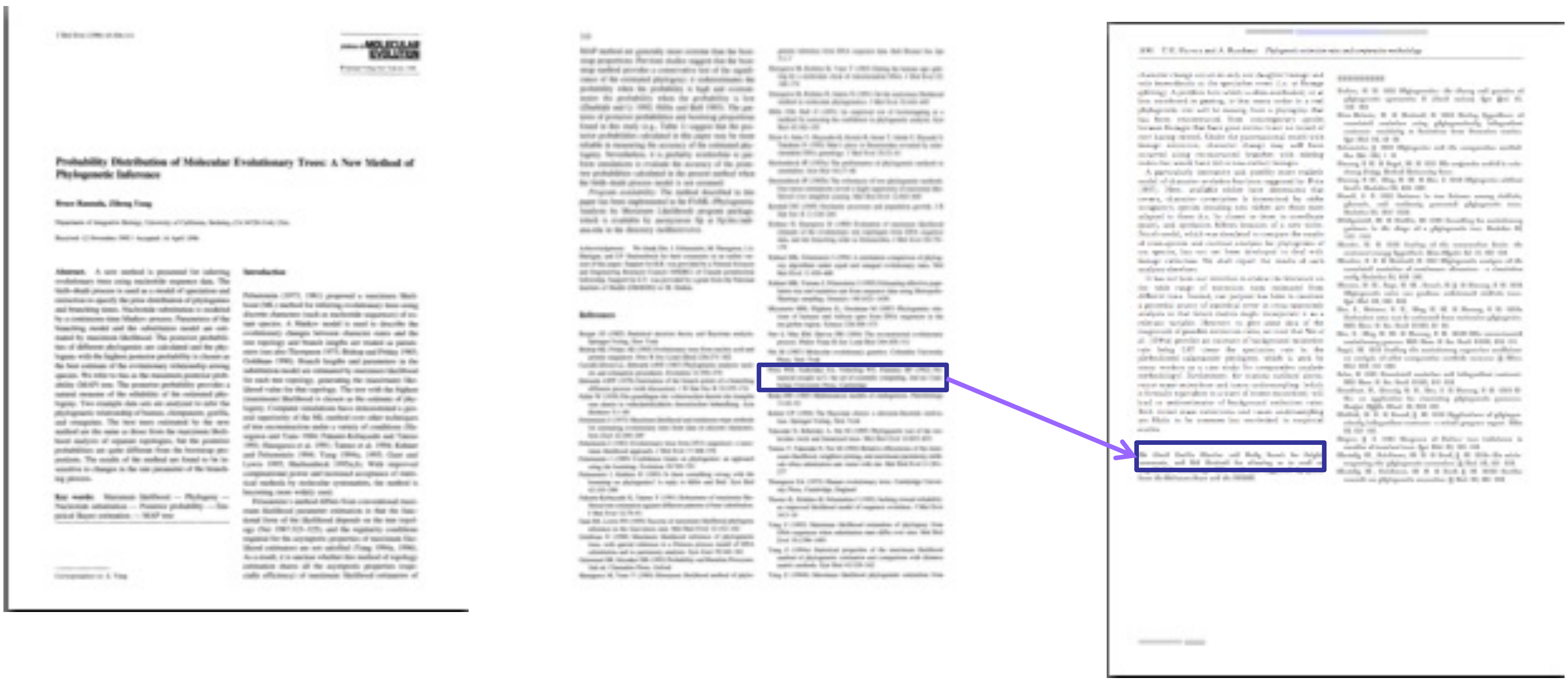
Price's preferential attachment model for citation networks

- [Price 65]
  - each new paper is generated with m citations (mean)
  - new papers cite previous papers with probability proportional to their indegree (citations)
  - what about papers without any citations?
    - each paper is considered to have a "default" citation
    - probability of citing a paper with degree k, proportional to k+1
- Power law with exponent $\alpha = 2+1/m$

Preferential attachment



Cumulative advantage: how?



- copying mechanism
- visibility

Barabasi-Albert model

- First used to describe skewed degree distribution of the World Wide Web

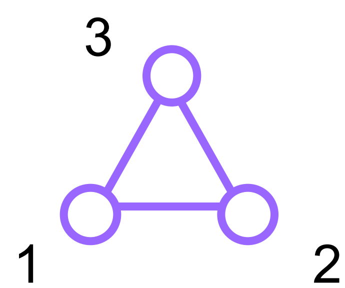
- Each node connects to other nodes with probability proportional to their degree
  - the process starts with some initial subgraph
  - each new node comes in with *m* edges
  - probability of connecting to node *i*

$$\Pi(i) = m \frac{k_i}{\sum_j k_j}$$

- Results in power-law with exponent $\alpha = 3$

Basic BA-model

- Very simple algorithm to implement
  - start with an initial set of m_0 fully connected nodes
    - e.g. $m_0 = 3$



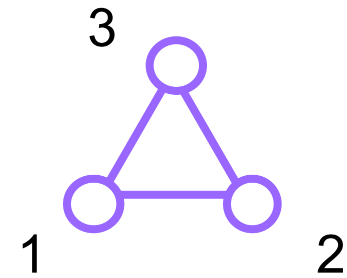
  - now add new vertices one by one, each one with exactly m edges
  - each new edge connects to an existing vertex in proportion to the number of edges that vertex already has → ***preferential attachment***
  - easiest if you keep track of edge endpoints in one large array and select an element from this array at random
    - the probability of selecting any one vertex will be proportional to the number of times it appears in the array – which corresponds to its degree

1 1 2 2 2 3 3 4 5 6 6 7 8 ….

generating BA graphs – cont’d

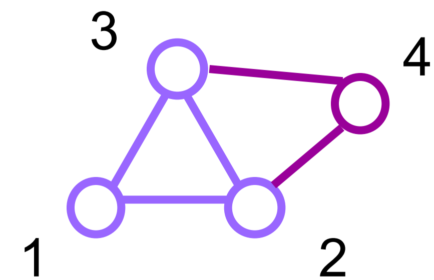
- To start, each vertex has an equal number of edges (2)
  - the probability of choosing any vertex is 1/3

1 1 2 2 3 3

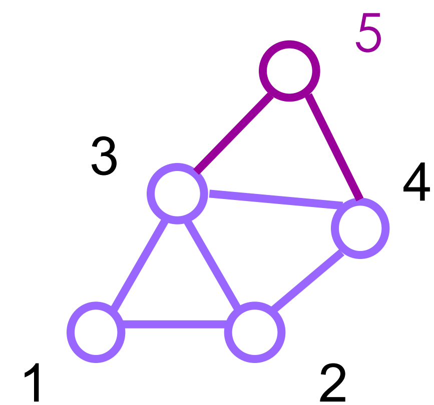


- We add a new vertex, and it will have m edges, here take m=2
  - draw 2 random elements from the array – suppose they are 2 and 3

1 1 2 2 2 3 3 3 4 4



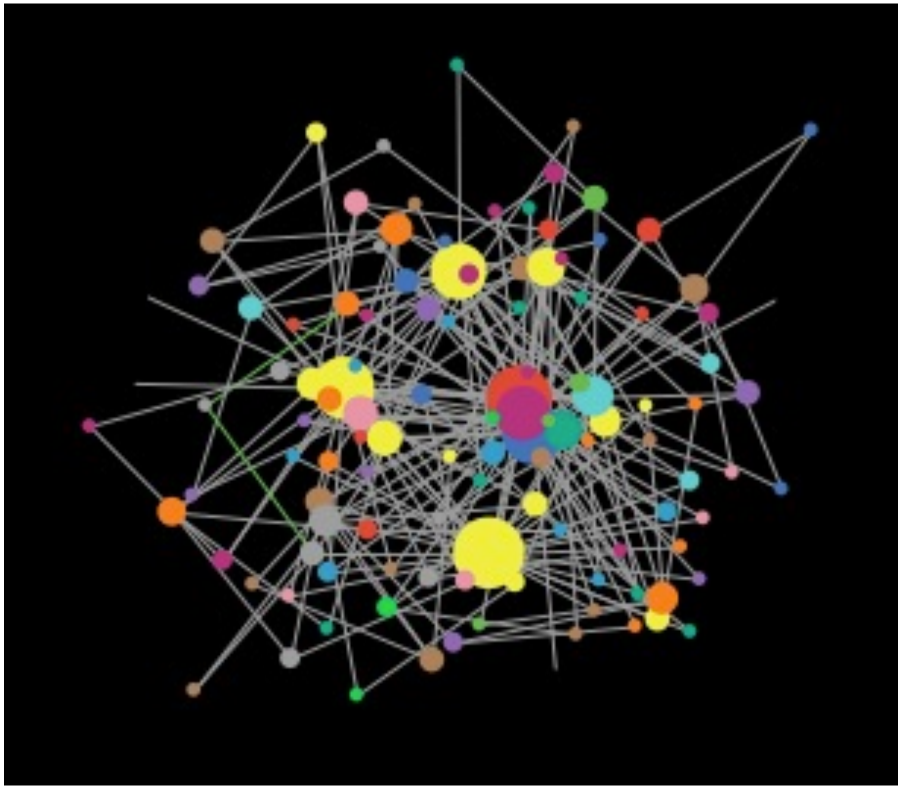
- Now the probabilities of selecting 1,2,3,or 4 are 1/5, 3/10, 3/10, 1/5



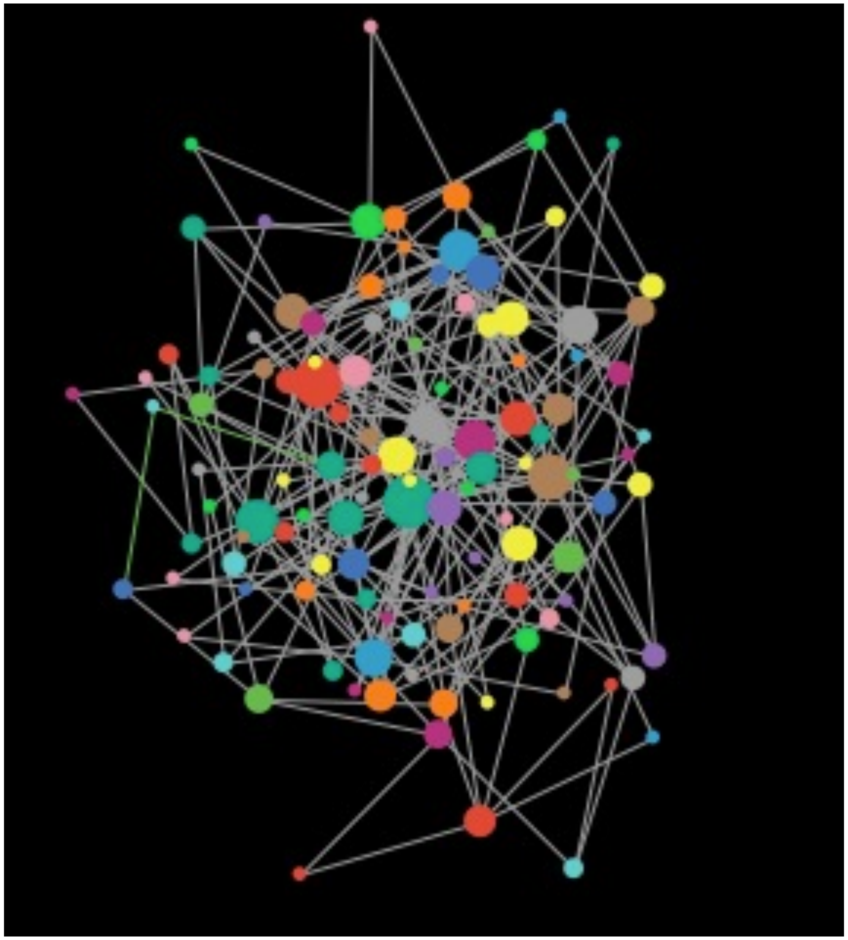
- Add a new vertex, draw a vertex for it to connect from the array
  - etc.

1 1 2 2 2 3 3 3 3 4 4 4 5 5

after a while...

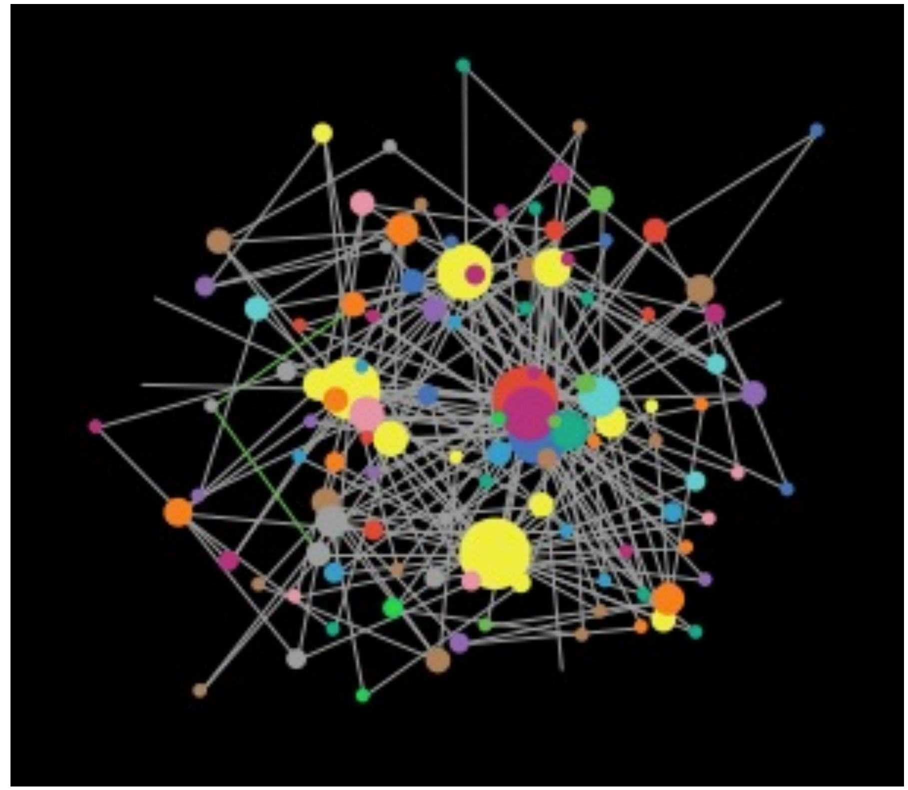


contrasting with random (non-preferential) growth



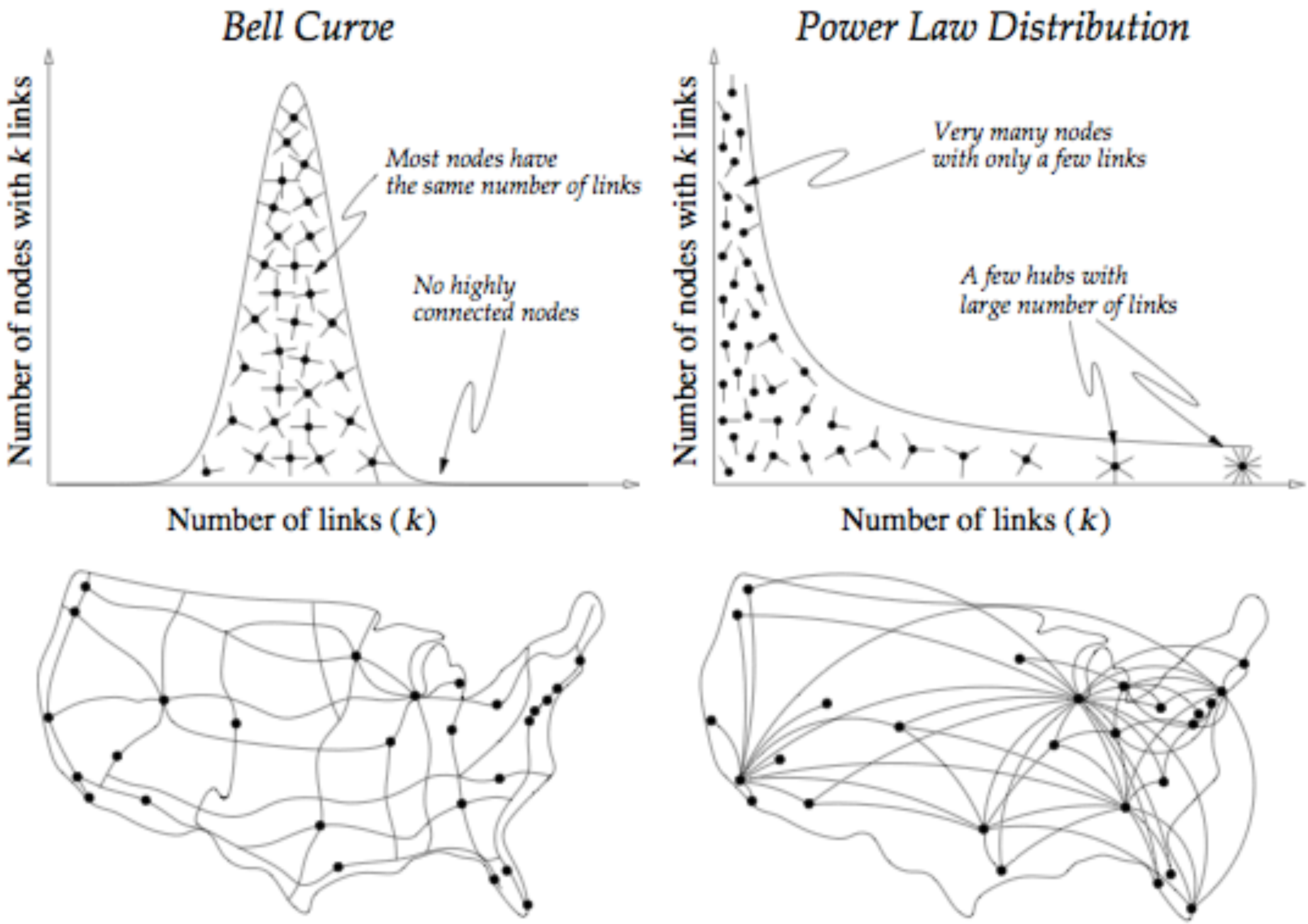
random

m = 2



preferential

Exponential vs. Power-Law



mean field approximation

- probability that node ***i*** acquires a new link at time ***t***

$$\frac{dk_i(t)}{dt} = m\frac{k_i}{2tm} = \frac{k_i}{2t} \quad \text{with} \quad k_i(i) = m$$

$$k_i(t) = m\left(\frac{t}{i}\right)^{1/2}$$

BA model degree distribution

- time of birth of node of degree k′: τ

$$\frac{\tau}{t} = \left(\frac{m}{k'}\right)^2$$

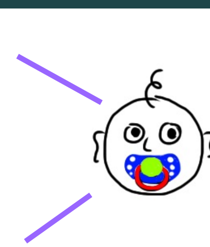
$$P(k < k') = 1 - \frac{m^2}{k'^2}$$

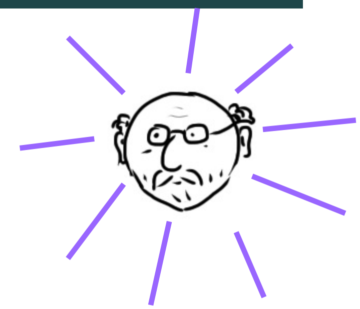
$$p(k) = \frac{2m^2}{k^3}$$

Properties of the BA graph

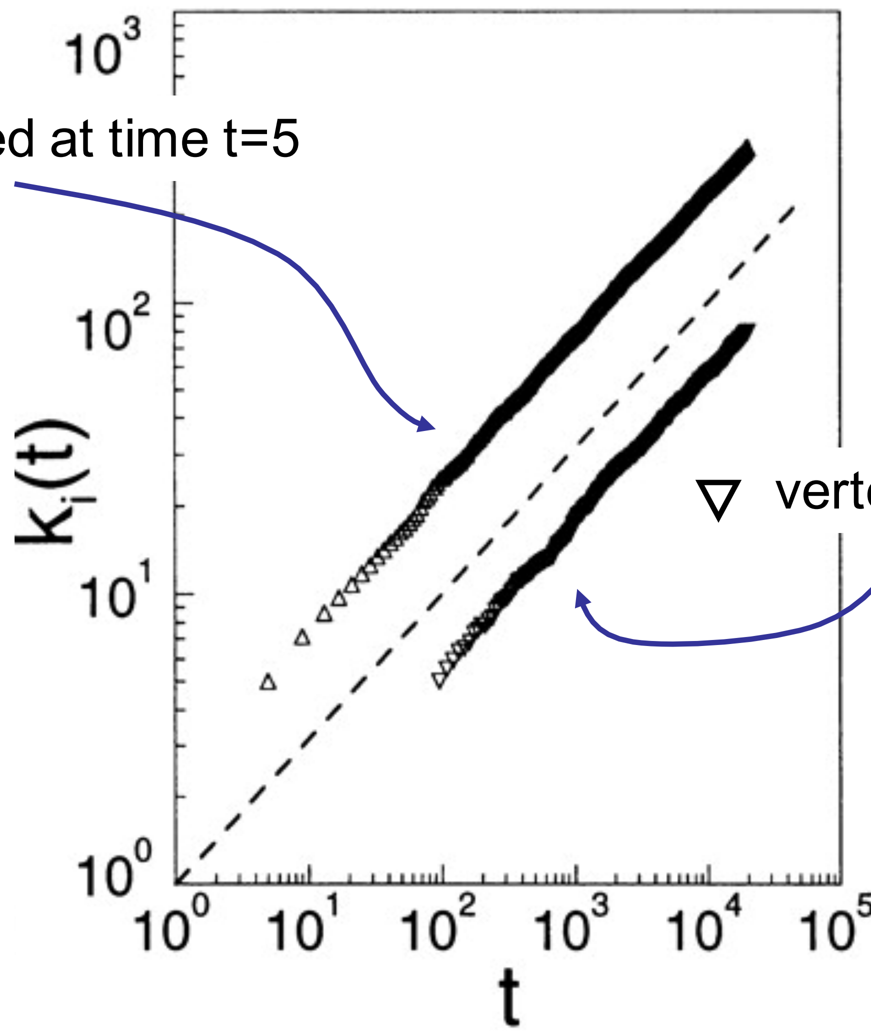
- The distribution is power-law with exponent $\alpha = 3$
  $P(k) = 2\ m^2/k^3$
- The graph is connected
  - Every new vertex is born with a link or several links (depending on whether m = 1 or m > 1)
  - It then connects to an 'older' vertex, which itself connected to another vertex when it was introduced
  - And we started from a connected core
- The older are richer
  - Nodes accumulate links as time goes on, which gives older nodes an advantage since newer nodes are going to attach preferentially – and older nodes have a higher degree to tempt them with than some new kid on the block

Young vs. old in BA model



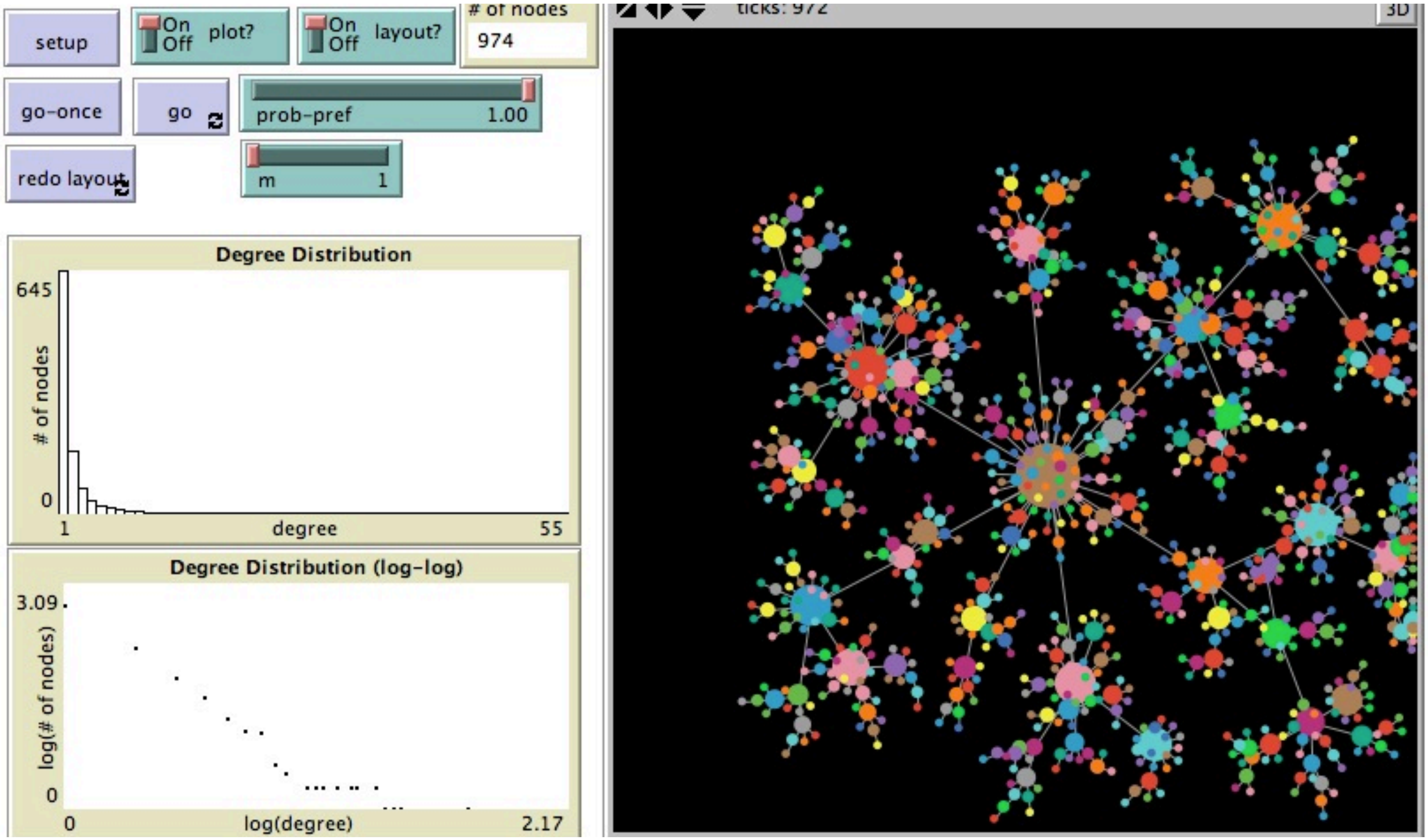


△ vertex introduced at time t=5



▽ vertex introduced at time t=95

try it yourself



http://web.stanford.edu/class/cs224w/NetLogo/RAndPrefAttachment.nlogo

Quiz Q:

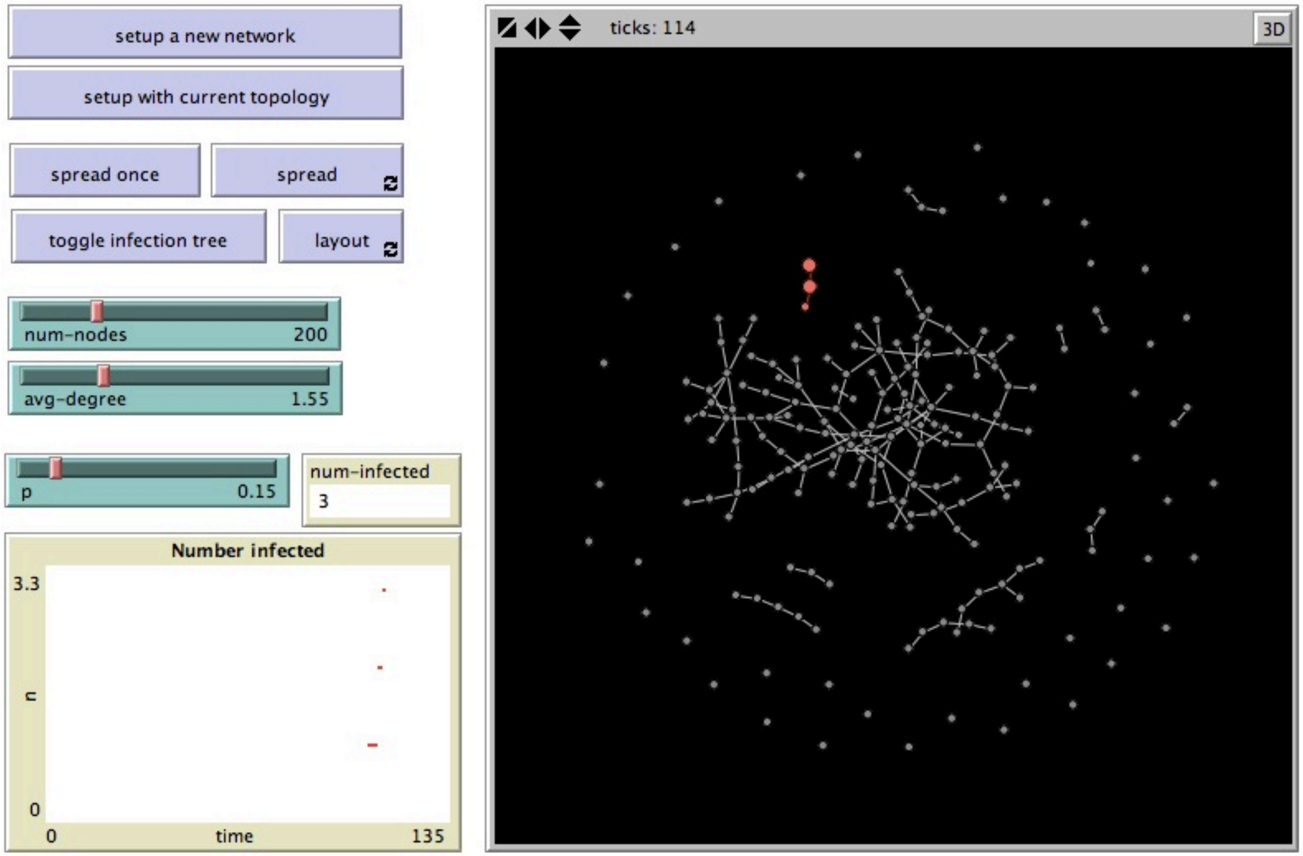
- Relative to the random growth model, the degree distribution in the preferential attachment model
  - resembles a power-law distribution less
  - resembles a power-law distribution more

Summary: growth models

- Most networks aren't 'born', they are made.
- Nodes being added over time means that older nodes can have more time to accumulate edges
- Preference for attaching to 'popular' nodes further skews the degree distribution toward a power-law

Implications for diffusion

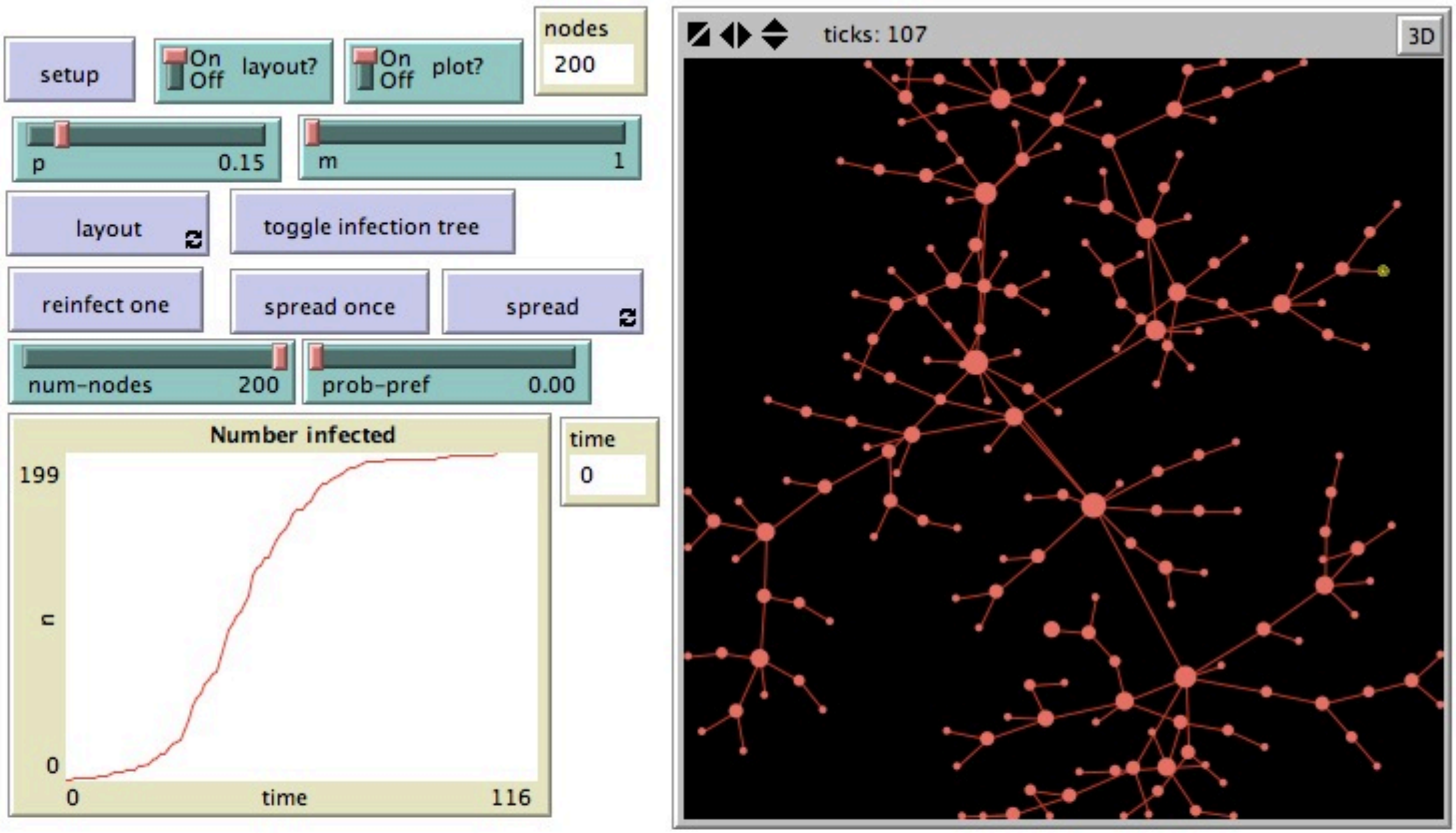
- How does the size of the giant component influence diffusion?



http://web.stanford.edu/class/cs224w/NetLogo/BADiffusion.nlogo

Implications for diffusion

- How do growth and preferential attachment influence diffusion?

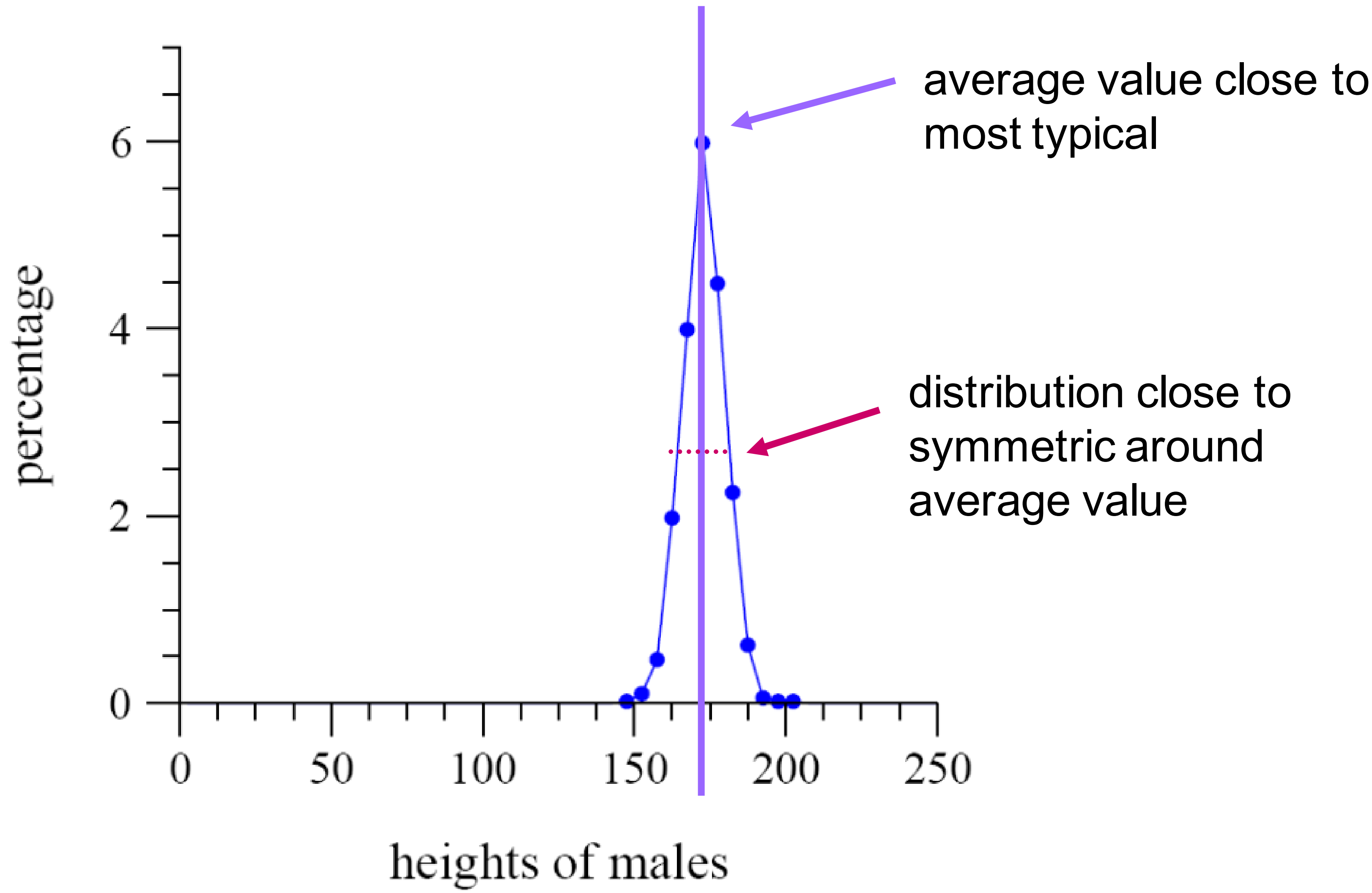


http://web.stanford.edu/class/cs224w/NetLogo/BADiffusion.nlogo

Heavy tails: right skew

- Right skew
  - normal distribution (not heavy tailed)
    - e.g. heights of human males: centered around 180cm (5’11’’)
  - Zipf’s or power-law distribution (heavy tailed)
    - e.g. city population sizes: NYC 8 million, but many, many small towns

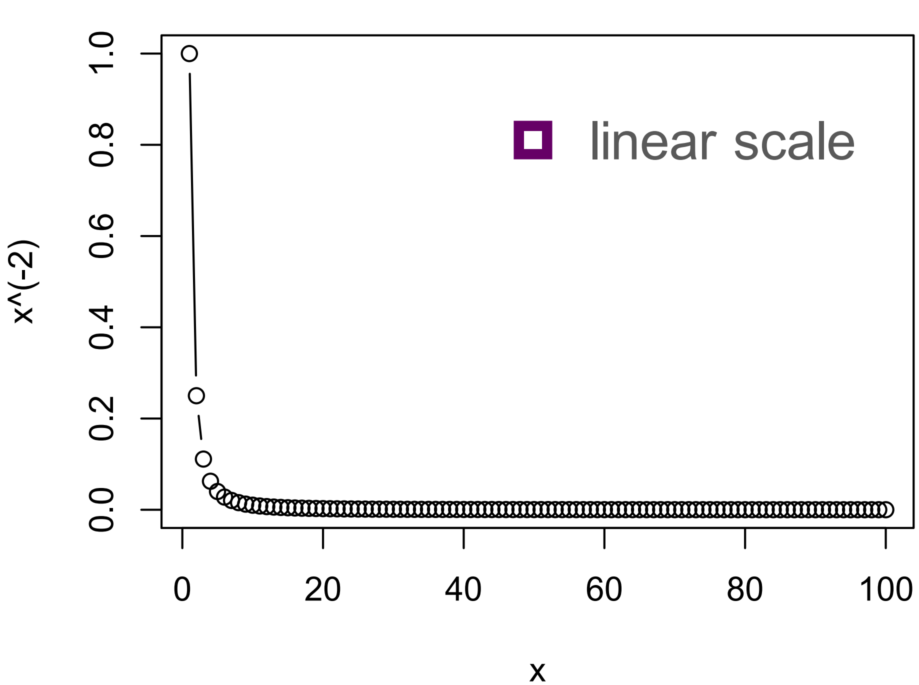
Normal distribution (human heights)

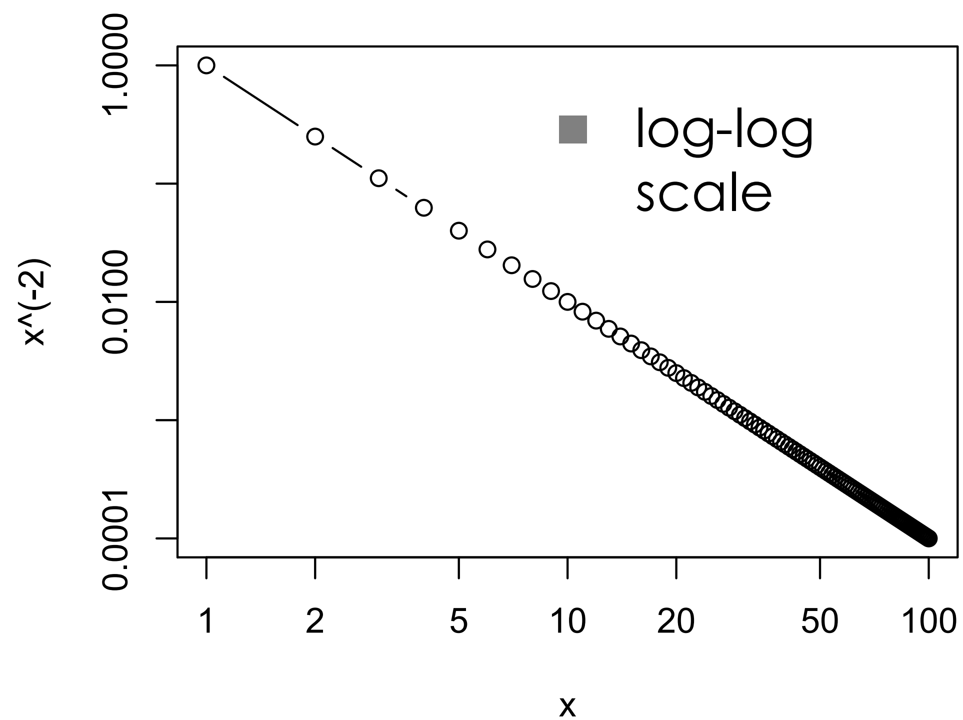


Heavy tails: max to min ratio

- High ratio of max to min
  - human heights
    - tallest man: 272cm (8’ 11”), shortest man: (1’ 10”) *ratio: 4.8*
      from the Guinness Book of world records
  - city sizes
    - NYC: pop. 8 million, Duffield, Virginia pop. 52, *ratio: 150,000*

Power-law distribution

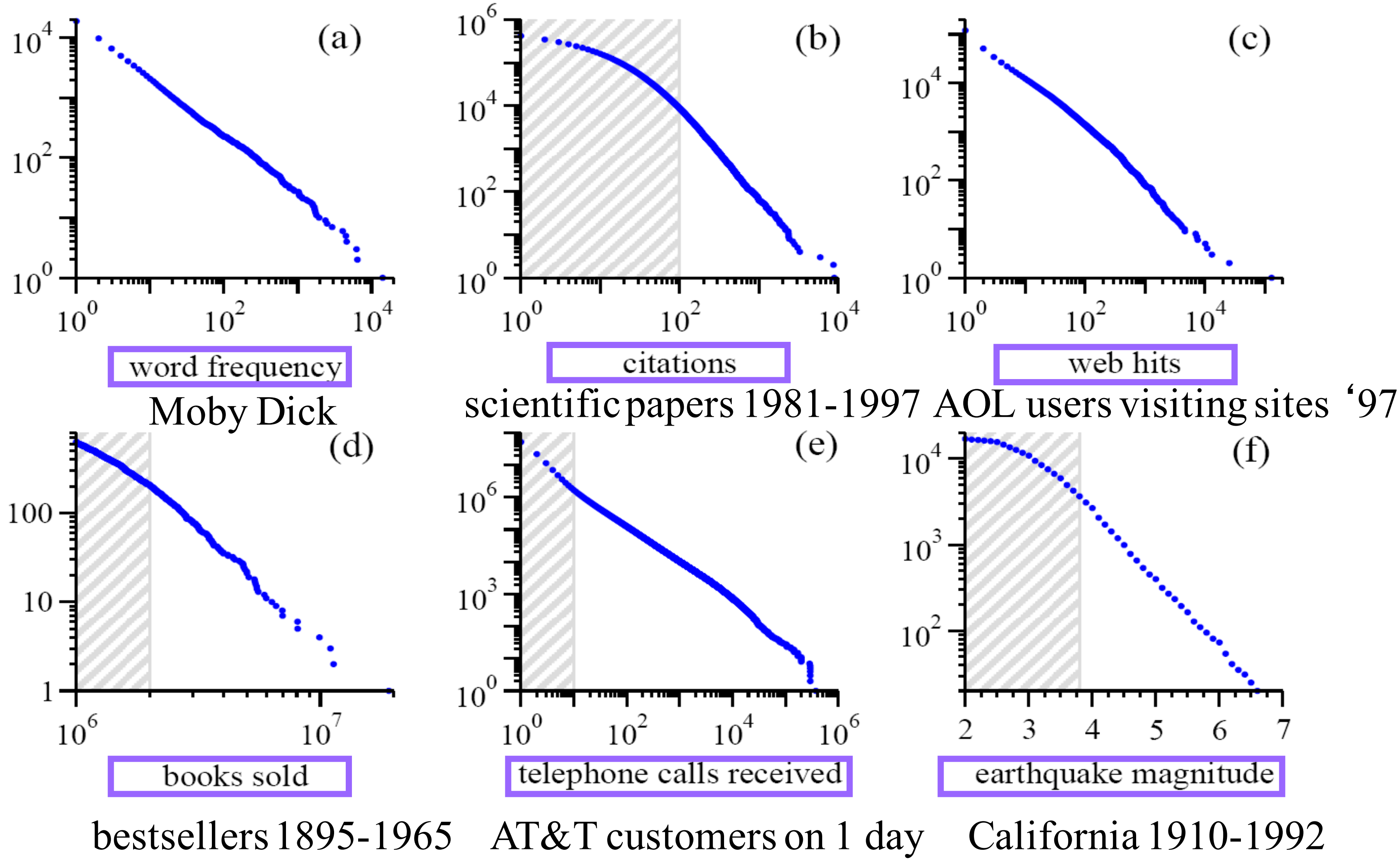




- high skew (asymmetry)
- straight line on a log-log plot

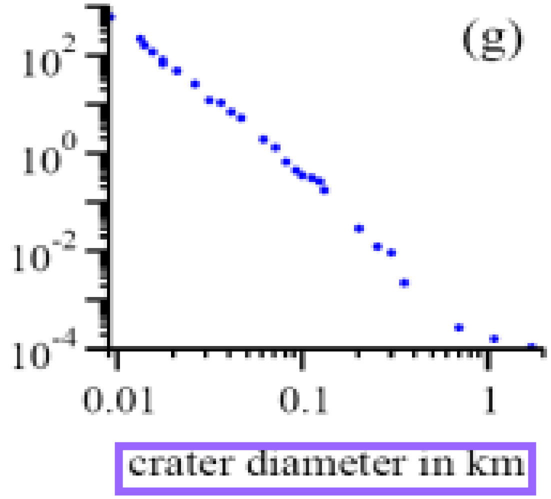
Power laws are seemingly everywhere

note: these are cumulative distributions, more about this in a bit…

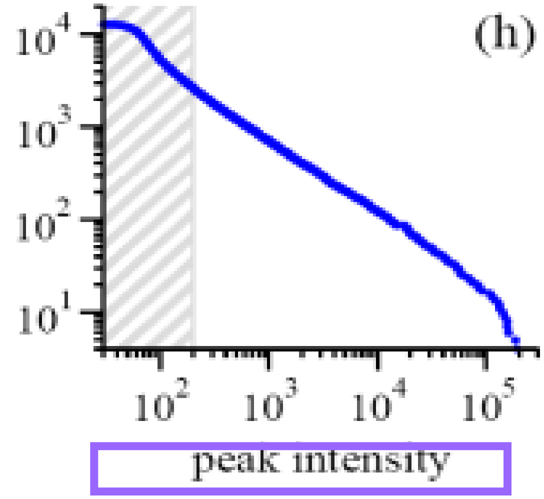


Source:MEJ Newman, ’Power laws, Pareto distributions and Zipf’s law’, *Contemporary Physics* **46**, 323–351 (2005)

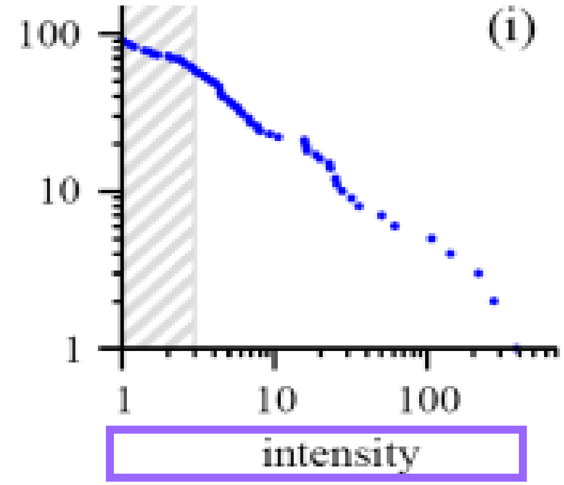
Yet more power laws



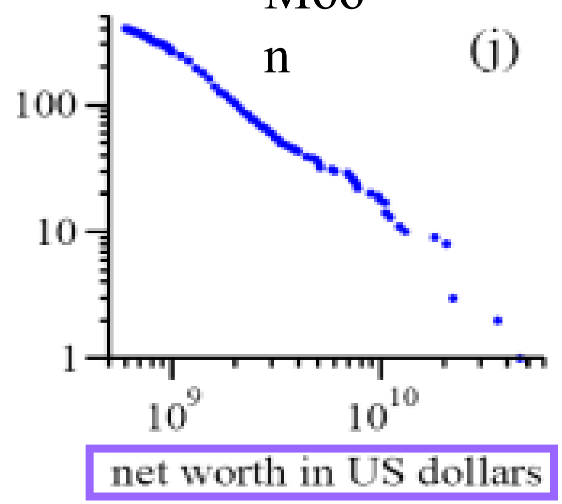
Moon



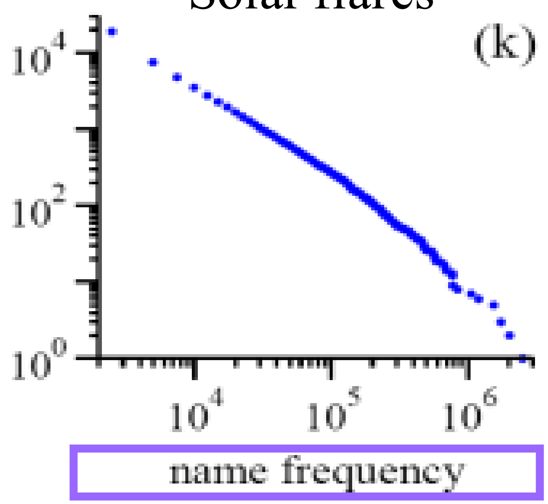
Solar flares



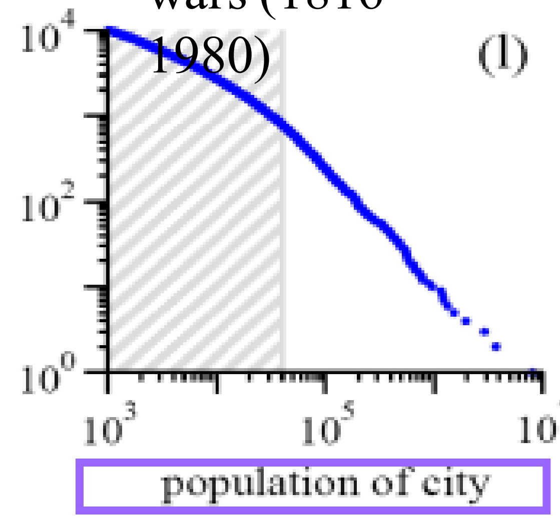
wars (1816-1980)



richest individuals 2003



US family names 1990

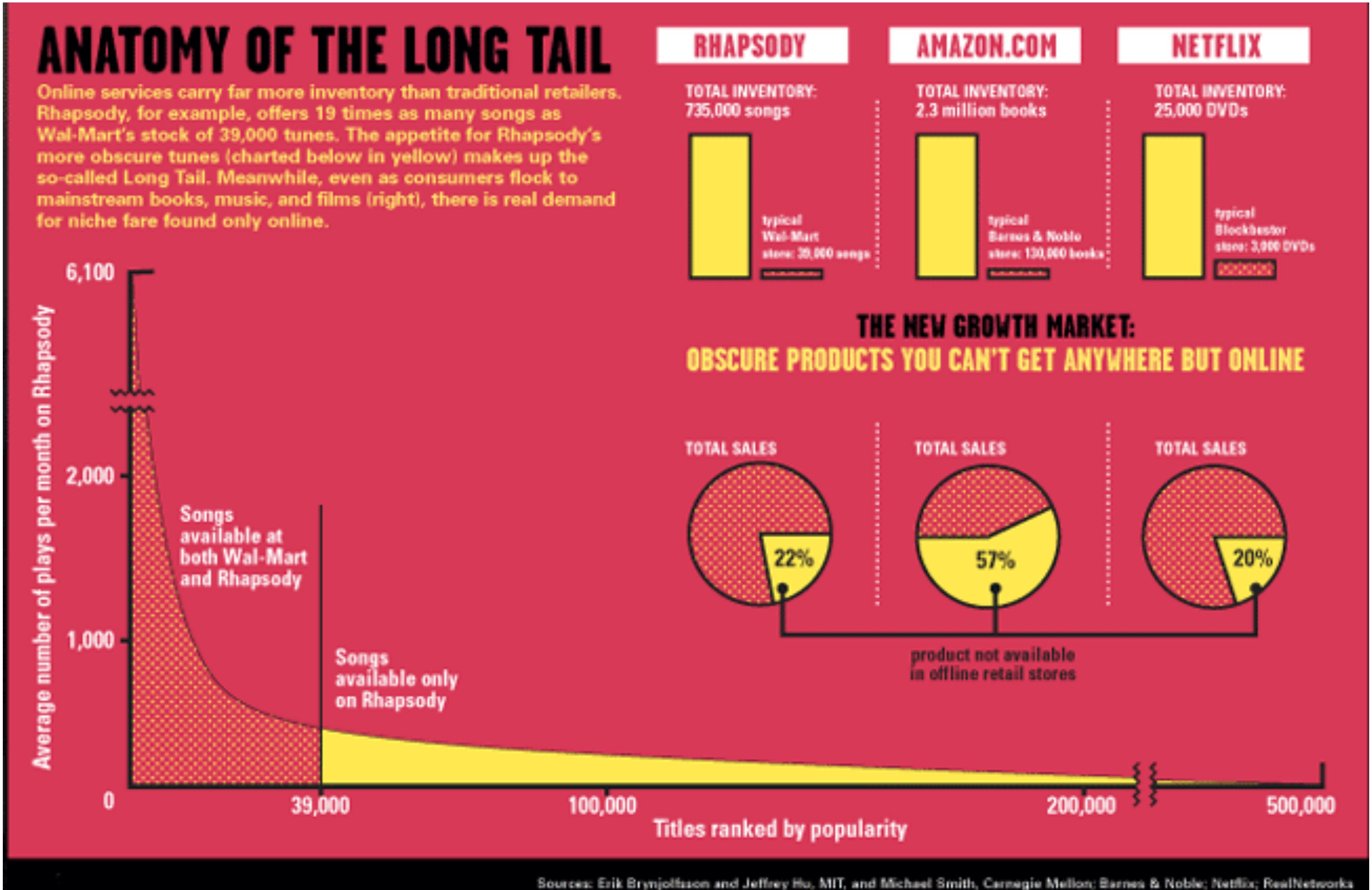


US cities 2003

Source:MEJ Newman, 'Power laws, Pareto distributions and Zipf's law', *Contemporary Physics* **46**, 323–351 (2005)

[Chris Anderson, Wired, 2004]

Anatomy of the Long Tail



Power law distribution

- Straight line on a log-log plot

$$\ln(p(x)) = c - \alpha \ln(x)$$

- Exponentiate both sides to get that *p(x)*, the probability of observing an item of size 'x' is given by

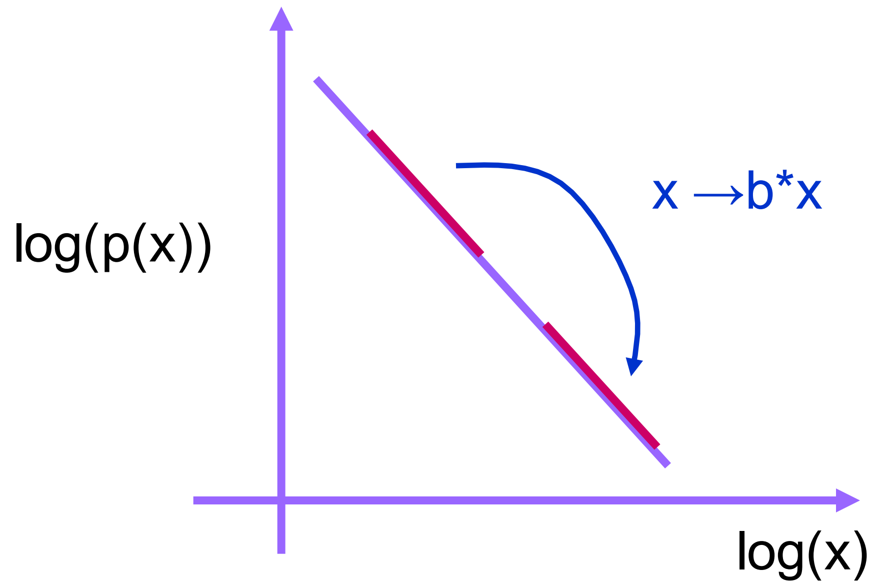
$$p(x) = Cx^{-\alpha}$$

normalization constant (probabilities over all *x* must sum to 1)

power law exponent α

What does it mean to be scale-free?

- A power law looks the same no mater what scale we look at it on (2 to 50 or 200 to 5000)
- Only true of a power-law distribution!
- $p(bx) = g(b)\, p(x)$ – Scale-free definition: shape of the distribution is unchanged except for a multiplicative constant
- $p(bx) = (bx)^{-\alpha} = b^{-\alpha}\, x^{-\alpha}$



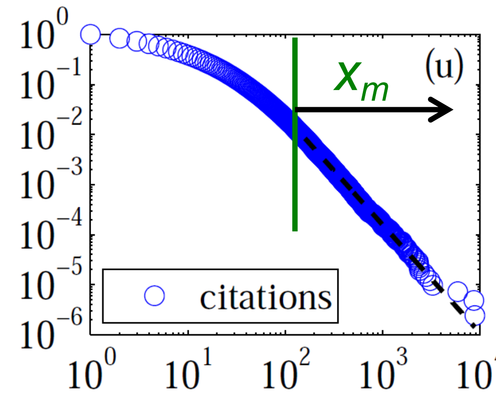
Popular distributions to try and fit

Name	Distribution $p(x) = Cf(x)$ $f(x)$	C
Power law	$x^{-\alpha}$	$(\alpha - 1)x_{\min}^{\alpha-1}$
Power law with cutoff	$x^{-\alpha}\mathrm{e}^{-\lambda x}$	$\frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha,\lambda x_{\min})}$
Exponential	$\mathrm{e}^{-\lambda x}$	$\lambda \mathrm{e}^{\lambda x_{\min}}$
Stretched exponential	$x^{\beta-1}\mathrm{e}^{-\lambda x^{\beta}}$	$\beta\lambda \mathrm{e}^{\lambda x_{\min}^{\beta}}$
Log-normal	$\frac{1}{x}\exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\right]$	$\sqrt{\frac{2}{\pi\sigma^2}}\left[\operatorname{erfc}\left(\frac{\ln x_{\min}-\mu}{\sqrt{2}\sigma}\right)\right]^{-1}$

Mathematics of Power-laws

- **What is the normalizing constant?**

$$p(x) = Z\, x^{-\alpha} \quad Z = ?$$



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

- $p(x)$ is a distribution: $\int p(x)dx = 1$

Continuous approximation

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

Need: $\alpha > 1$!

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

Integral:

$$\int (ax)^n = \frac{(ax)^{n+1}}{a(n+1)}$$

Mathematics of Power-laws

- **What's the expected value of a power-law random variable *X*?**

- $E[X] = \int_{x_m}^{\infty} x\, p(x) dx = Z \int_{x_m}^{\infty} x^{-\alpha+1} dx$

- $= \frac{Z}{2-\alpha}\left[x^{2-\alpha}\right]_{x_m}^{\infty} = \frac{(\alpha-1)x_m^{\alpha-1}}{-(\alpha-2)}\left[\infty^{2-\alpha} - x_m^{2-\alpha}\right]$

Need: α > 2 !

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

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

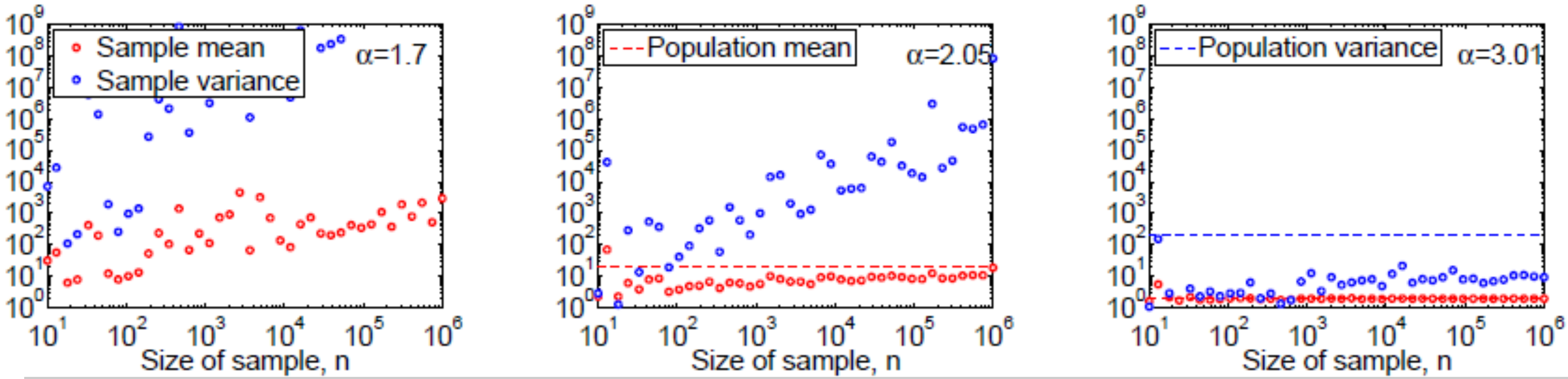
- **Power-laws have infinite moments!**

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

  - If $\alpha \leq 2 : E[X] = \infty$
  - If $\alpha \leq 3 : Var[X] = \infty$
    - **Average is meaningless, as the variance is too high!**

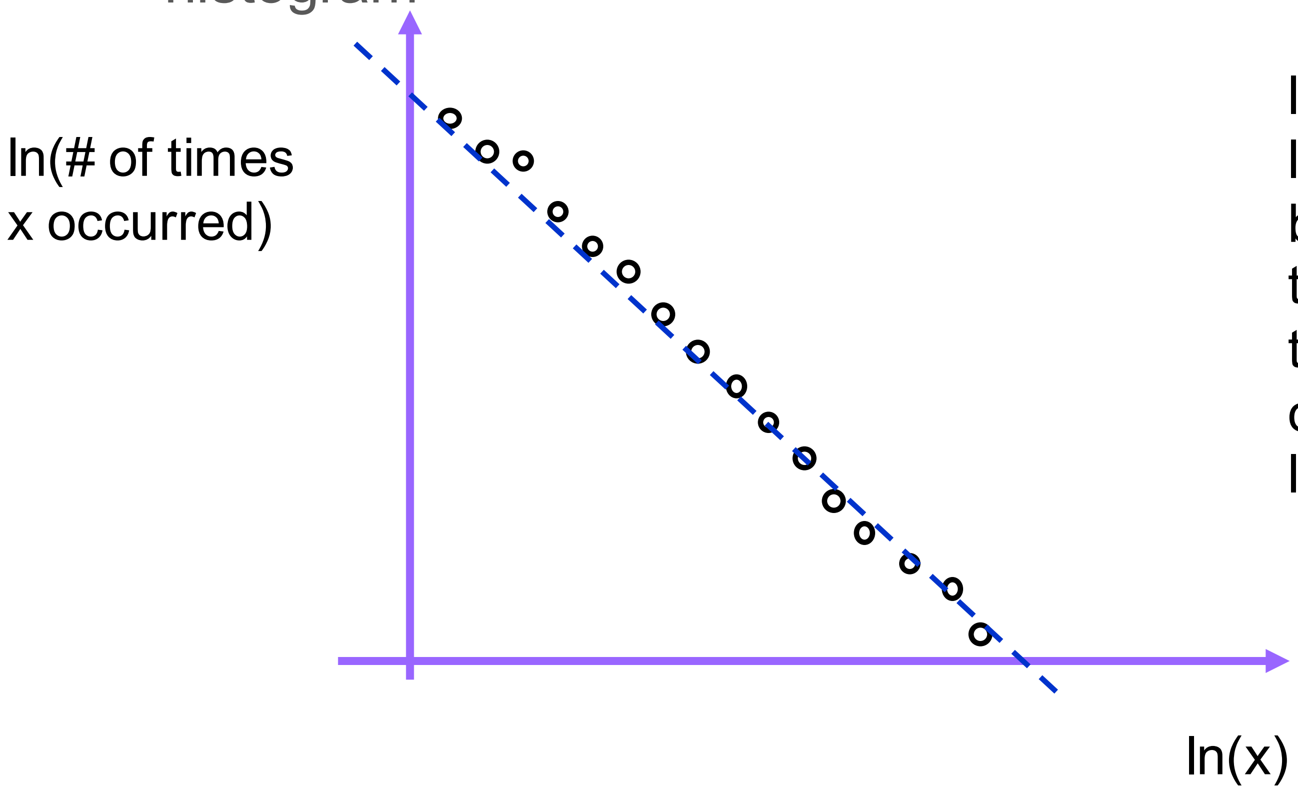
In real networks **2 < α < 3** so:
E[X] = const
Var[X] = ∞

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



Fitting power-law distributions

- Most common and not very accurate method:
  - Bin the different values of x and create a frequency histogram



$\ln(x)$ is the natural logarithm of x, but any other base of the logarithm will give the same exponent of α because $\log_{10}(x) = \ln(x)/\ln(10)$

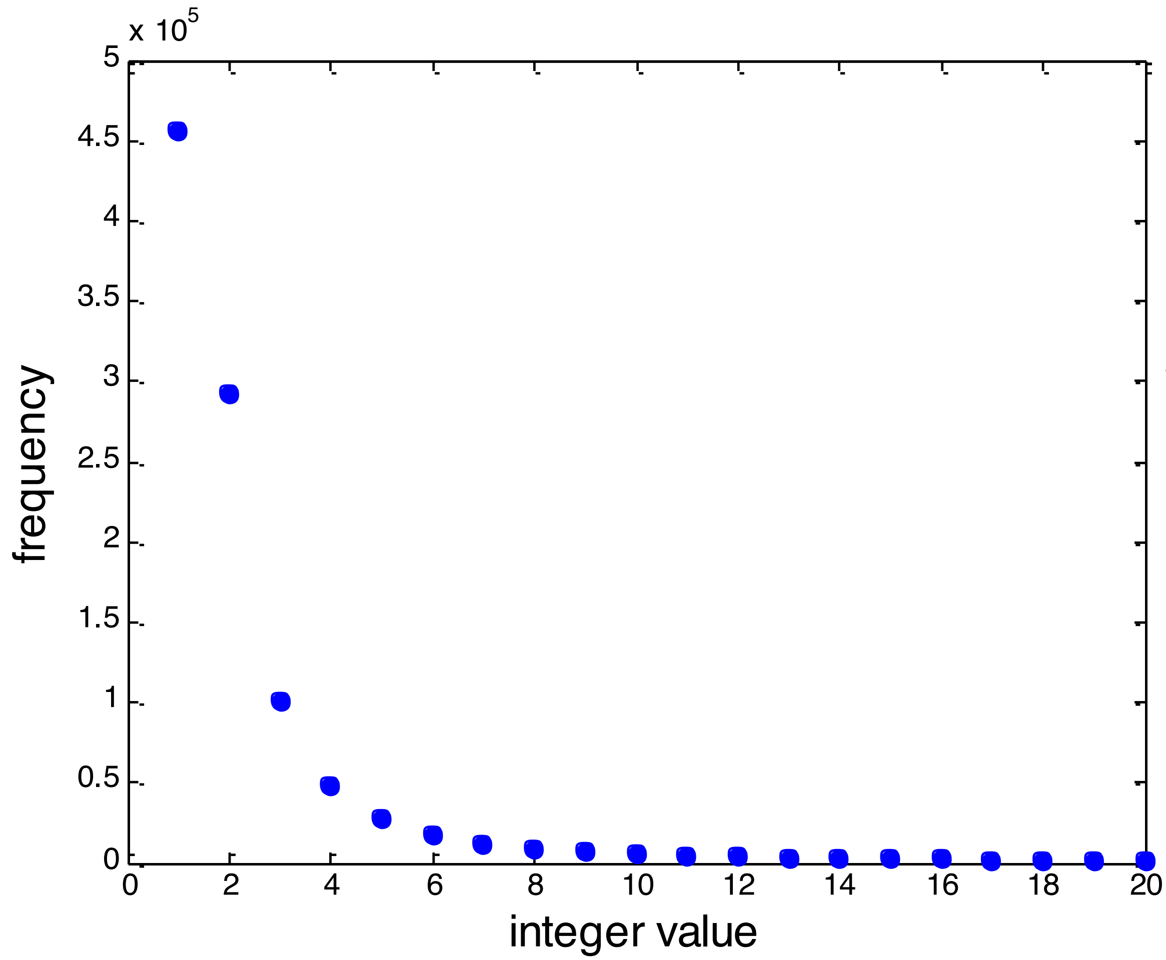
x can represent various quantities, the indegree of a node, the magnitude of an earthquake, the frequency of a word in text

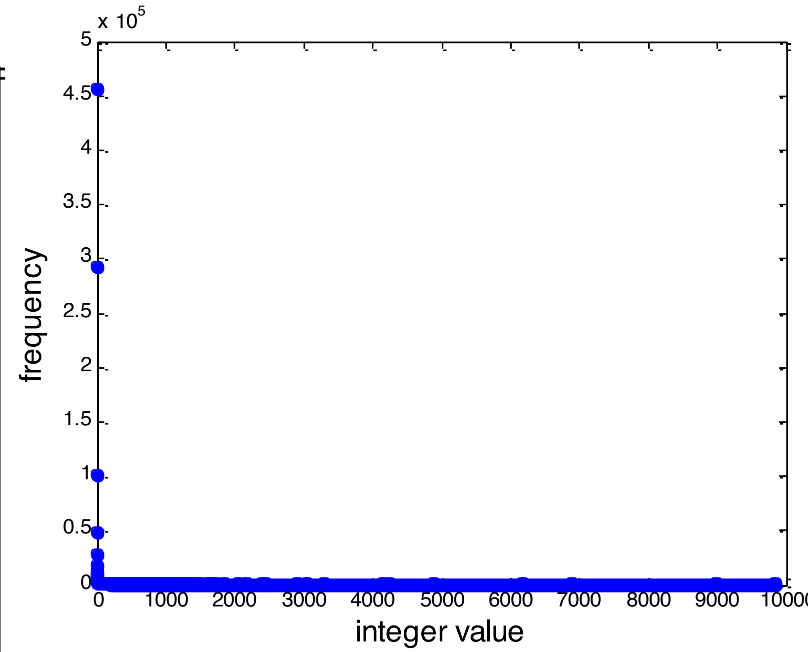
Example on an artificially generated data set

- Take 1 million random numbers from a distribution with $\alpha = 2.5$
- Can be generated using the so-called 'transformation method'
- Generate random numbers r on the unit interval $0 \leq r < 1$
- then $x = (1-r)^{-1/(\alpha-1)}$ is a random power law distributed real number in the range $1 \leq x < \infty$

Linear scale plot of straight bin of the data

- Number of times 1 or 3843 or 99723 occurred
- Power-law relationship not as apparent
- Only makes sense to look at smallest bins



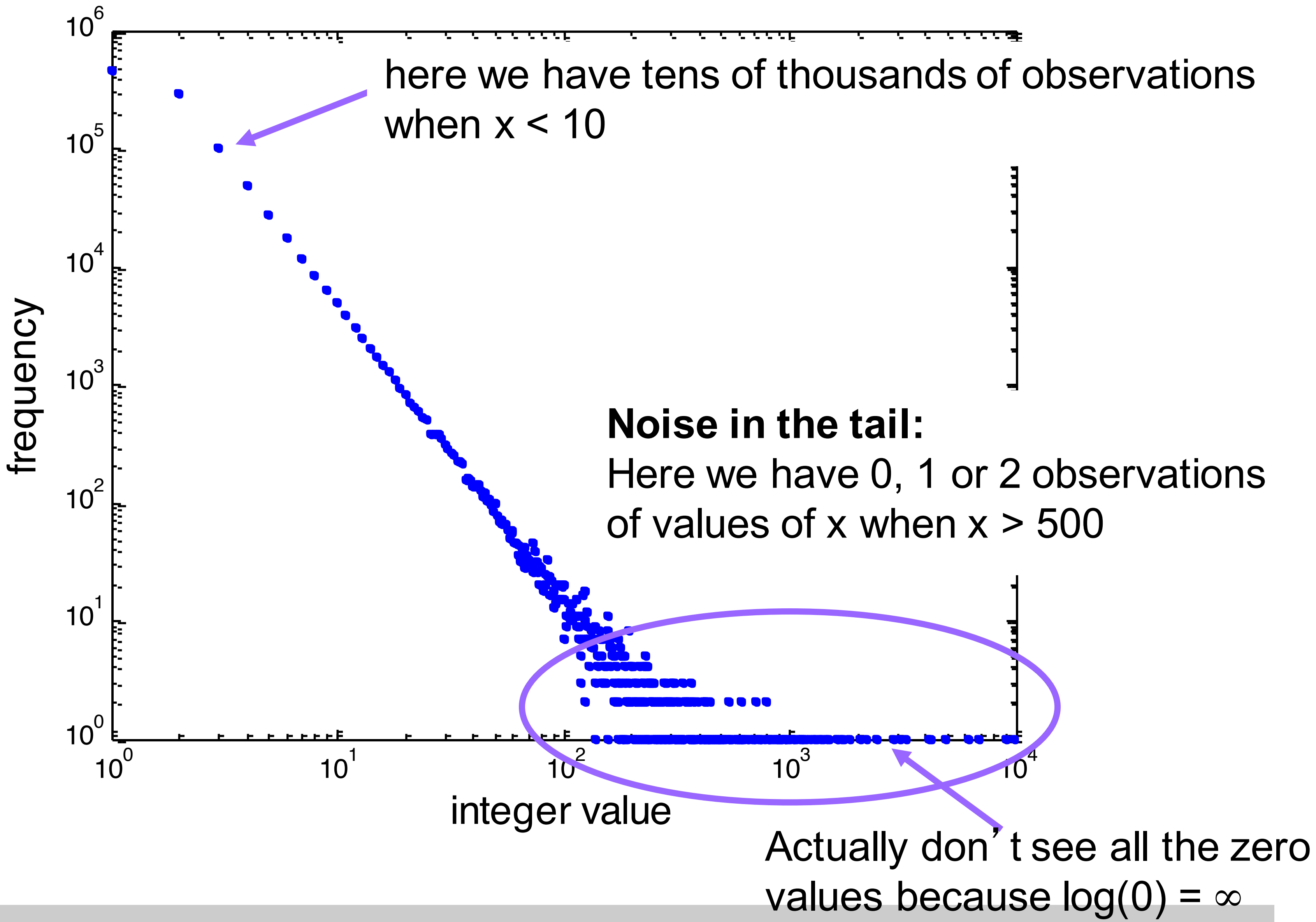


whole range

first few bins

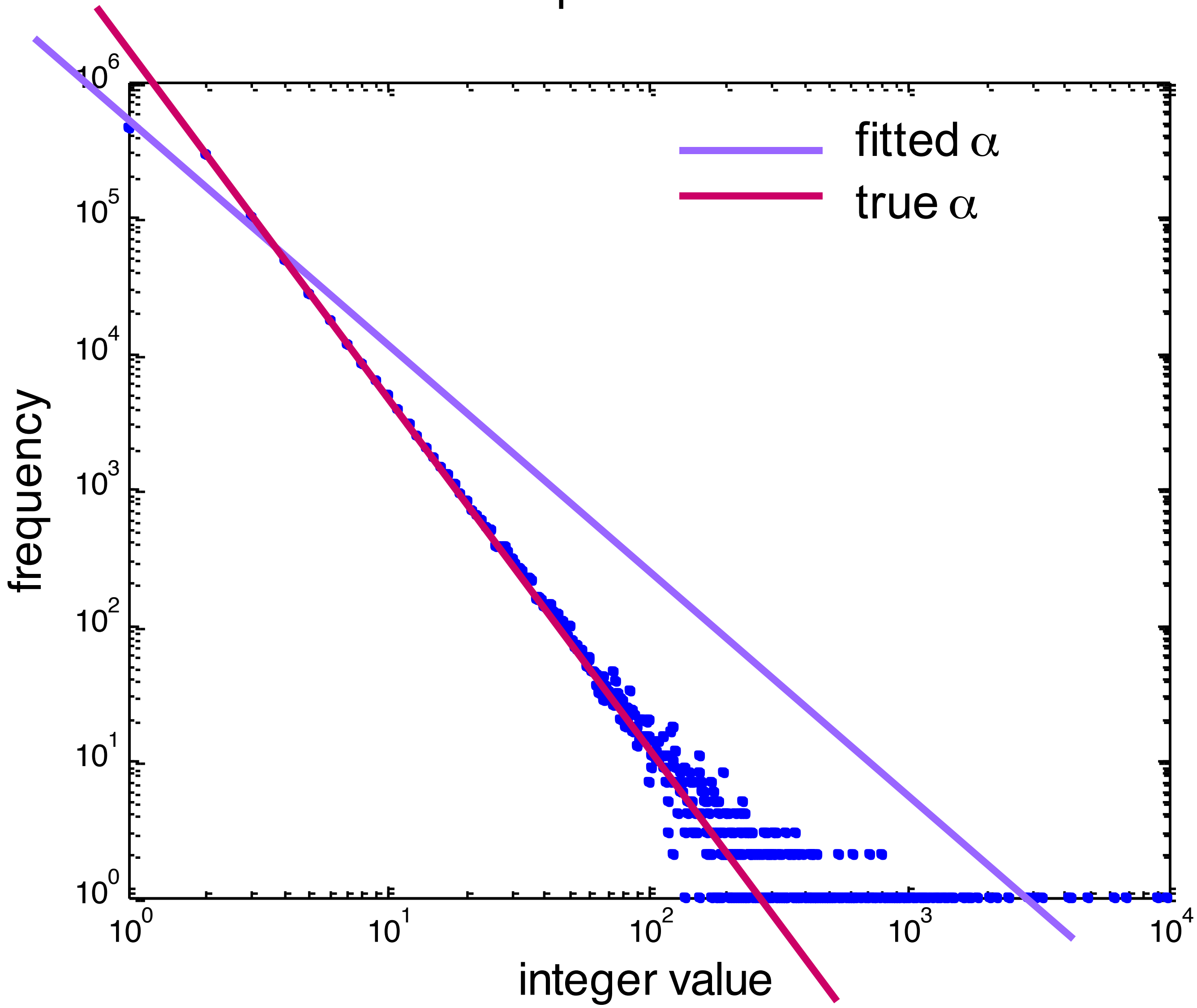
Log-log scale plot of simple binning of the data

- Same bins, but plotted on a log-log scale



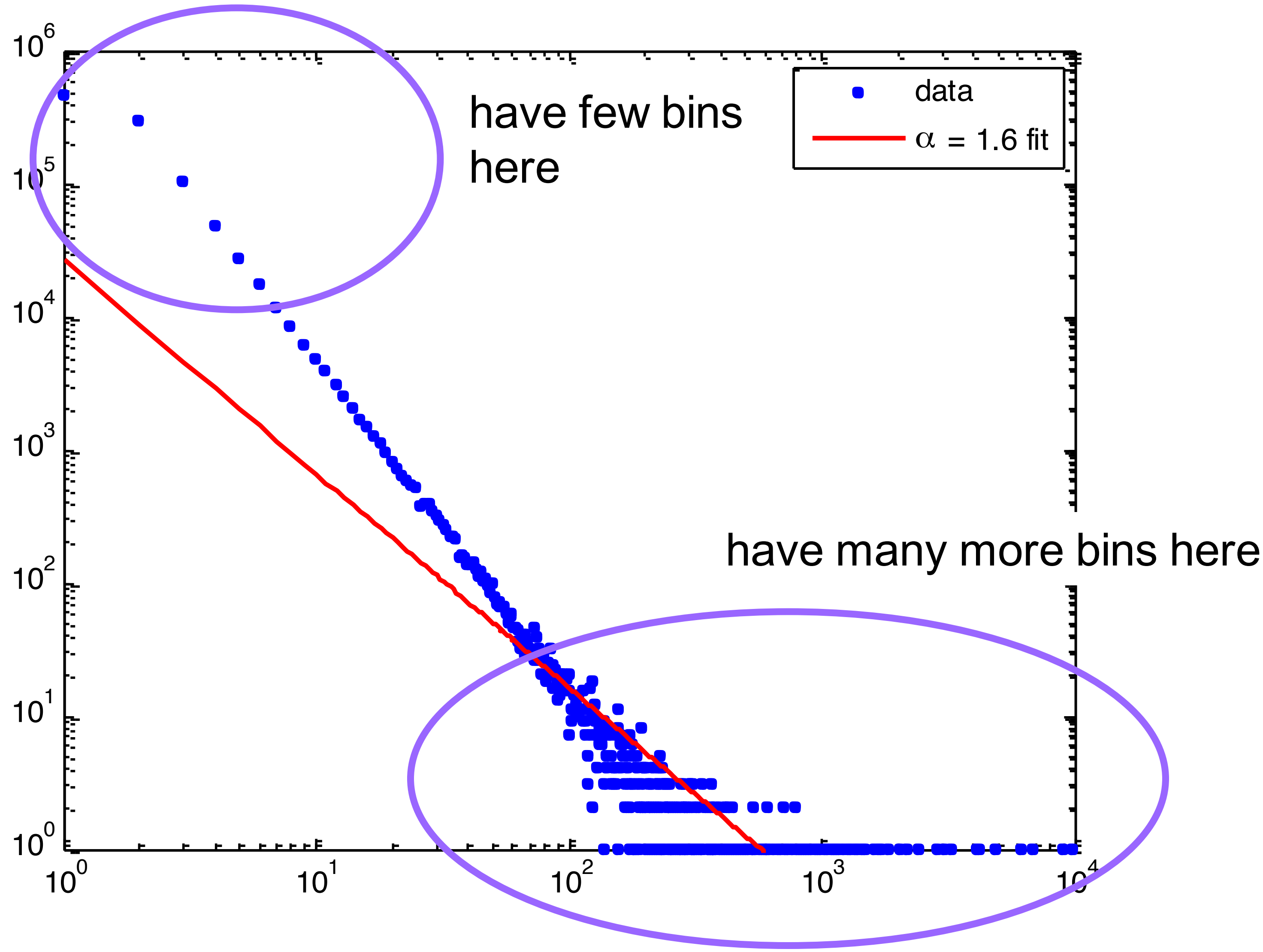
Log-log scale plot of straight binning of the data

- Fitting a straight line to it via least squares regression will give values of the exponent α that are too low



What goes wrong with straightforward binning

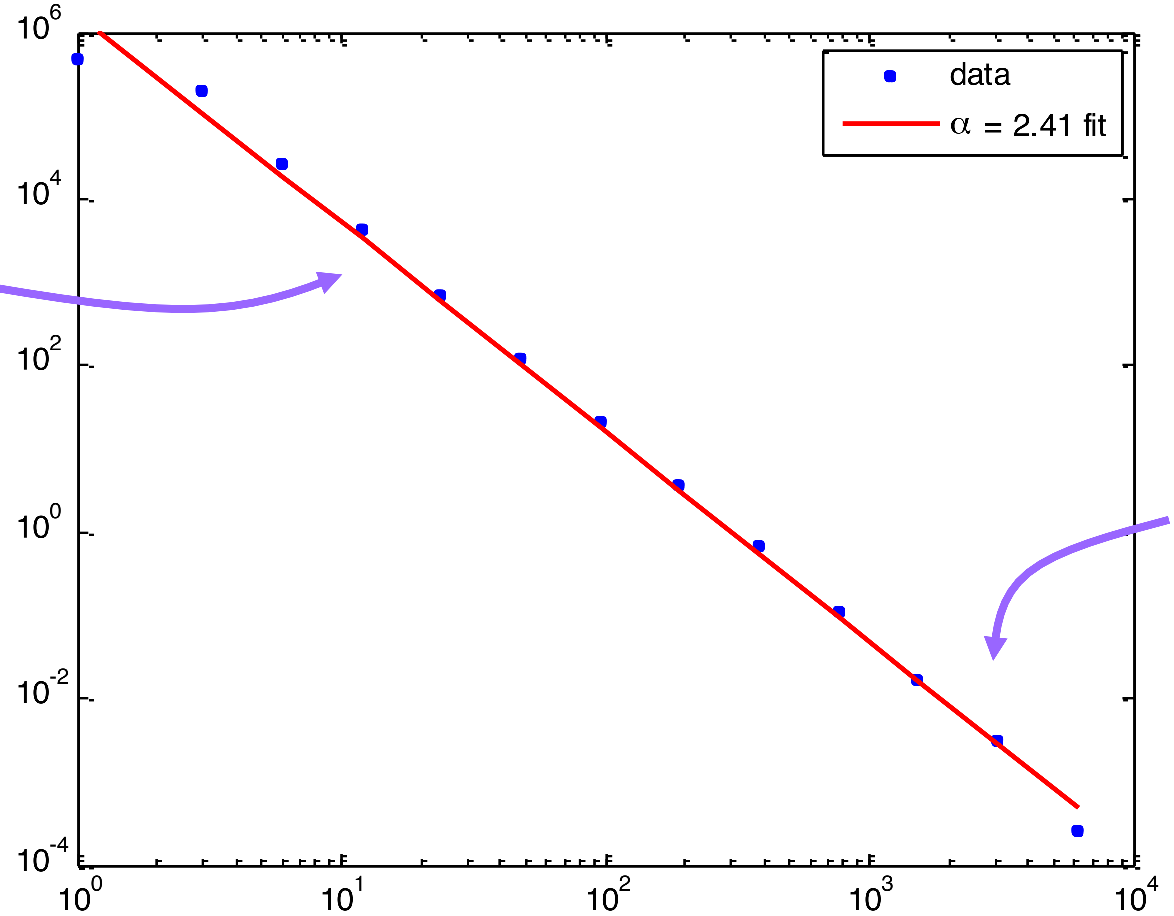
- Noise in the tail skews the regression result



First solution: logarithmic binning

- bin data into exponentially wider bins:
  - 1, 2, 4, 8, 16, 32, …
- normalize by the width of the bin

evenly spaced datapoints



less noise in the tail of the distribution

- disadvantage: binning smoothes out data but also loses information

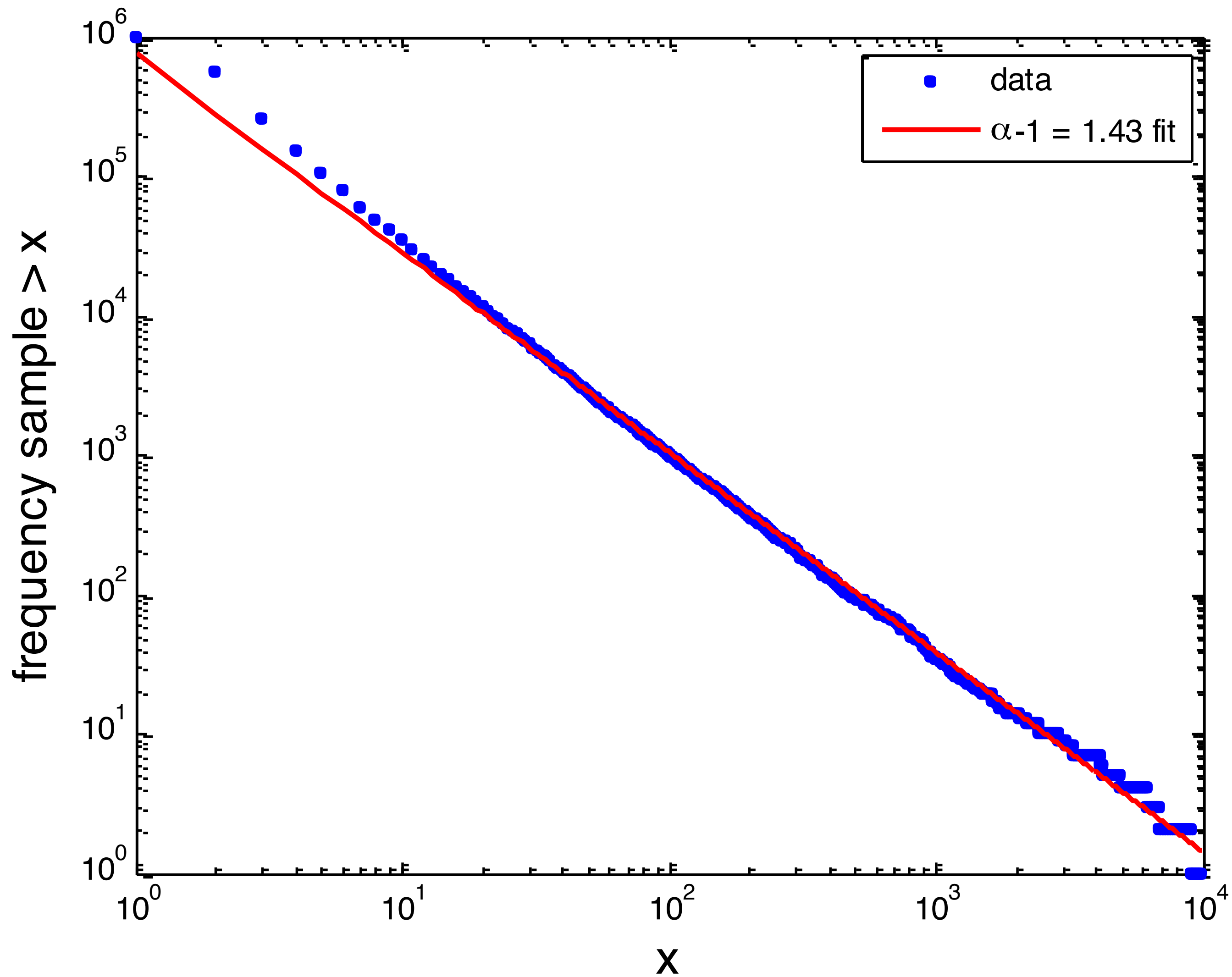
Second solution: cumulative binning

- No loss of information
  - No need to bin, has value at each observed value of x
- But now have cumulative distribution
  - i.e. how many of the values of x are at least X
  - The cumulative probability of a power law probability distribution is also power law but with an exponent **α - 1**

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

Fitting via regression to the cumulative distribution

- fitted exponent (2.43) much closer to actual (2.5)

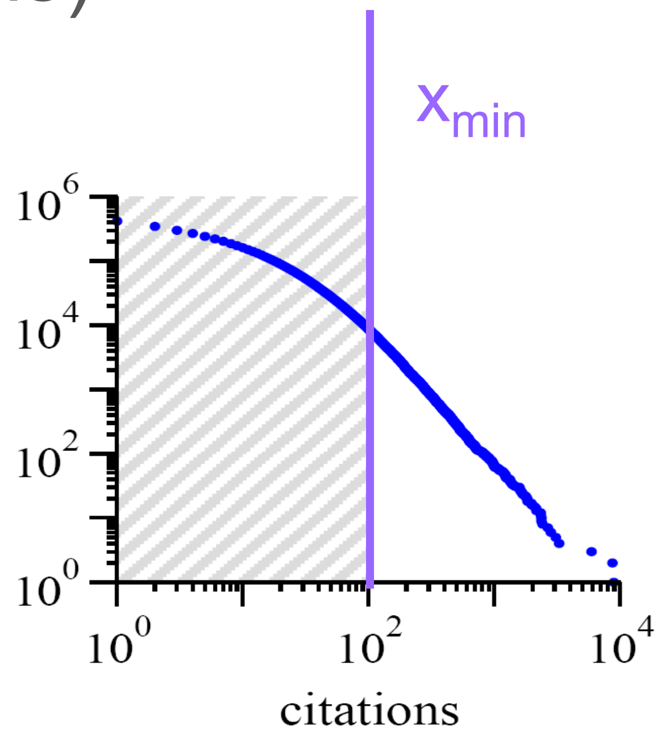


Where to start fitting?

- some data exhibit a power law only in the tail
- after binning or taking the cumulative distribution you can fit to the tail
- so need to select an x_{min} the value of x where you think the power-law starts
- certainly x_{min} needs to be greater than 0, because $x^{-\alpha}$ is infinite at $x = 0$

Example:

- Distribution of citations to papers
- power law is evident only in the tail (x_{min} > 100 citations)



Source:MEJ Newman, 'Power laws, Pareto distributions and Zipf's law', *Contemporary Physics* **46**, 323–351 (2005)

Maximum likelihood fitting – best

- You have to be sure you have a power-law distribution (this will just give you an exponent but not a goodness of fit)

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

- x_i are all your data points, and you have *n* of them
- for our data set we get $\alpha = 2.503$ – pretty close!

Some exponents for real world data

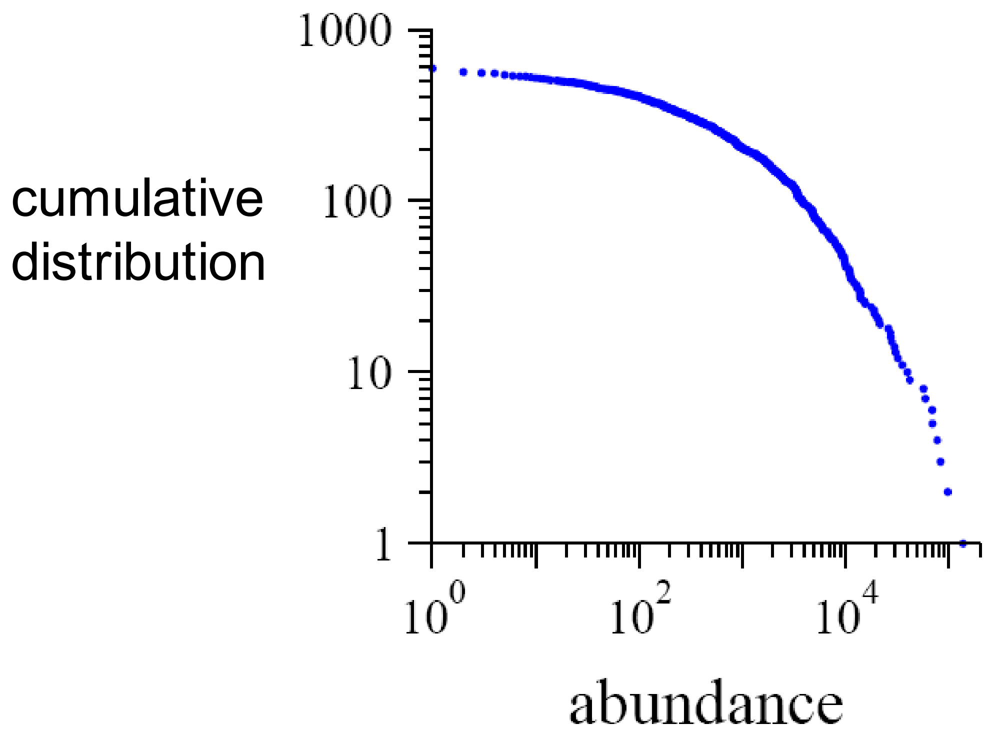
	x_{min}	exponent α
frequency of use of words	1	2.20
number of citations to papers	100	3.04
number of hits on web sites	1	2.40
copies of books sold in the US	2 000 000	3.51
telephone calls received	10	2.22
magnitude of earthquakes	3.8	3.04
diameter of moon craters	0.01	3.14
intensity of solar flares	200	1.83
intensity of wars	3	1.80
net worth of Americans	$600m	2.09
frequency of family names	10 000	1.94
population of US cities	40 000	2.30

Many real world networks are power law

	exponent α (in/out degree)
film actors	2.3
telephone call graph	2.1
email networks	1.5/2.0
sexual contacts	3.2
WWW	2.3/2.7
internet	2.5
peer-to-peer	2.1
metabolic network	2.2
protein interactions	2.4

Hey, not everything is a power law

- number of sightings of 591 bird species in the North American Bird survey in 2003.



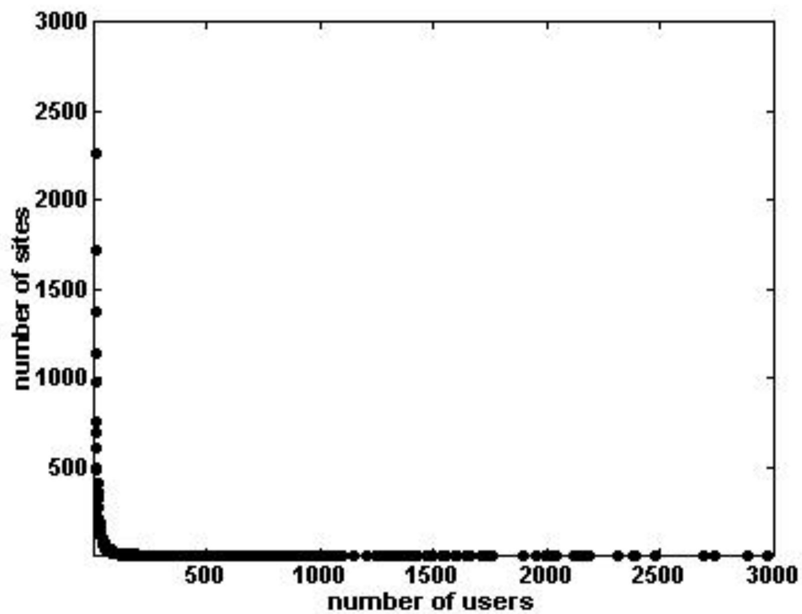
- another example:
  - size of wildfires (in acres)

Source:MEJ Newman, 'Power laws, Pareto distributions and Zipf's law', *Contemporary Physics* **46**, 323–351 (2005)

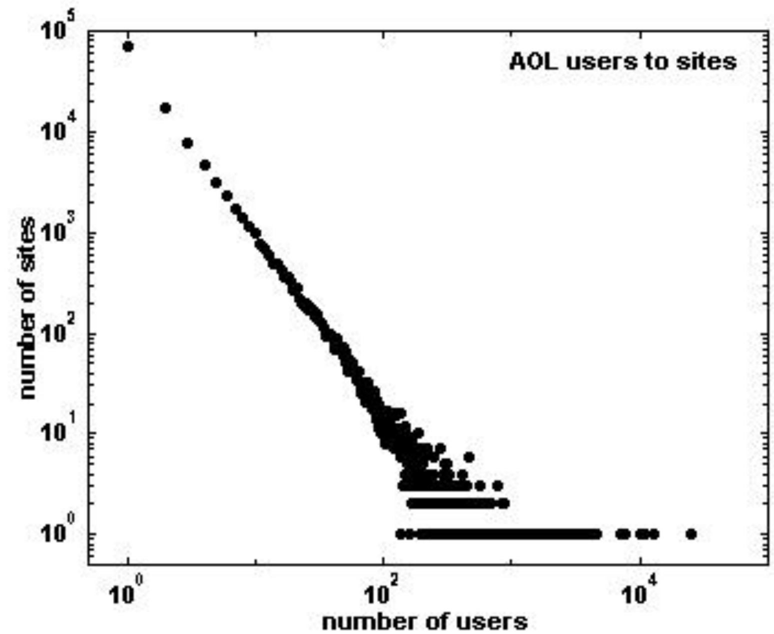
Not every network is power law distributed

- reciprocal, frequent email communication
- power grid
- Roget's thesaurus
- company directors…

Example on a real data set: number of AOL visitors to different websites back in 1997



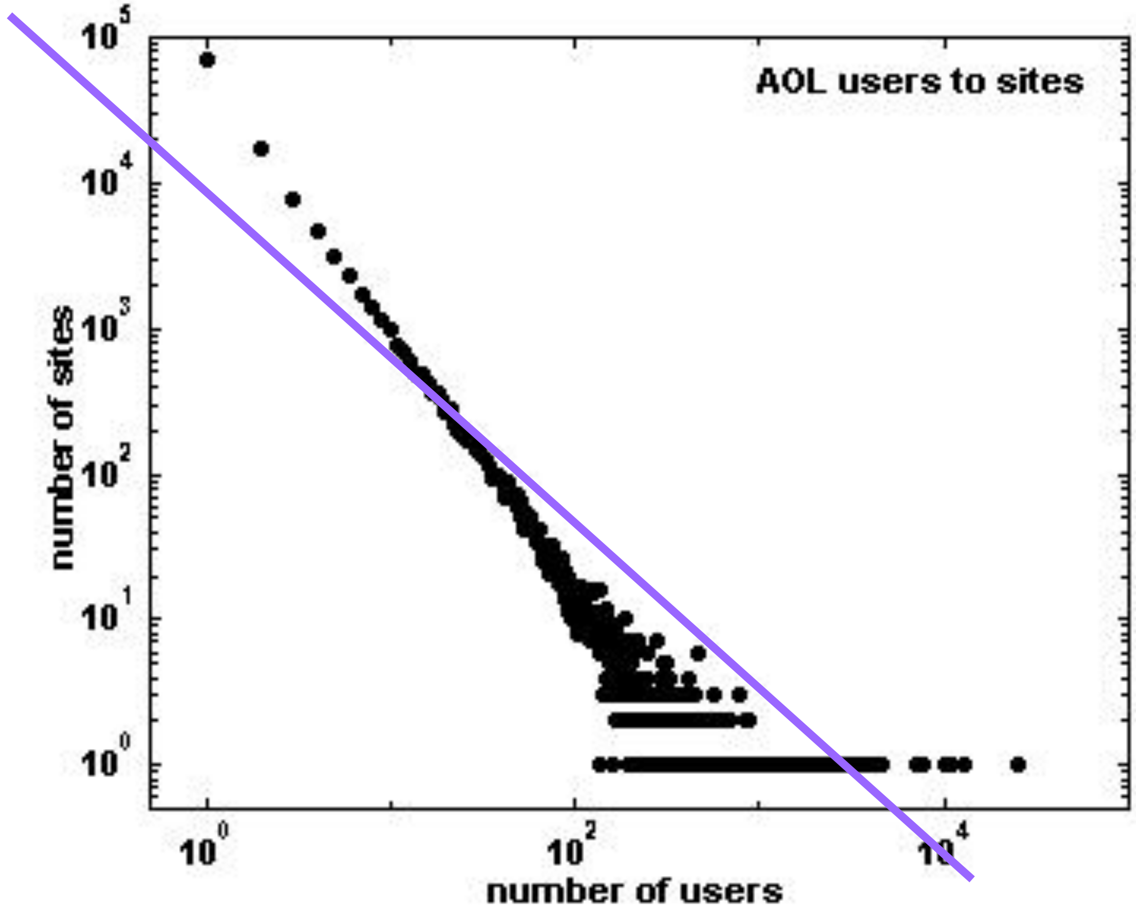
simple binning on a linear scale



simple binning on a log-log scale

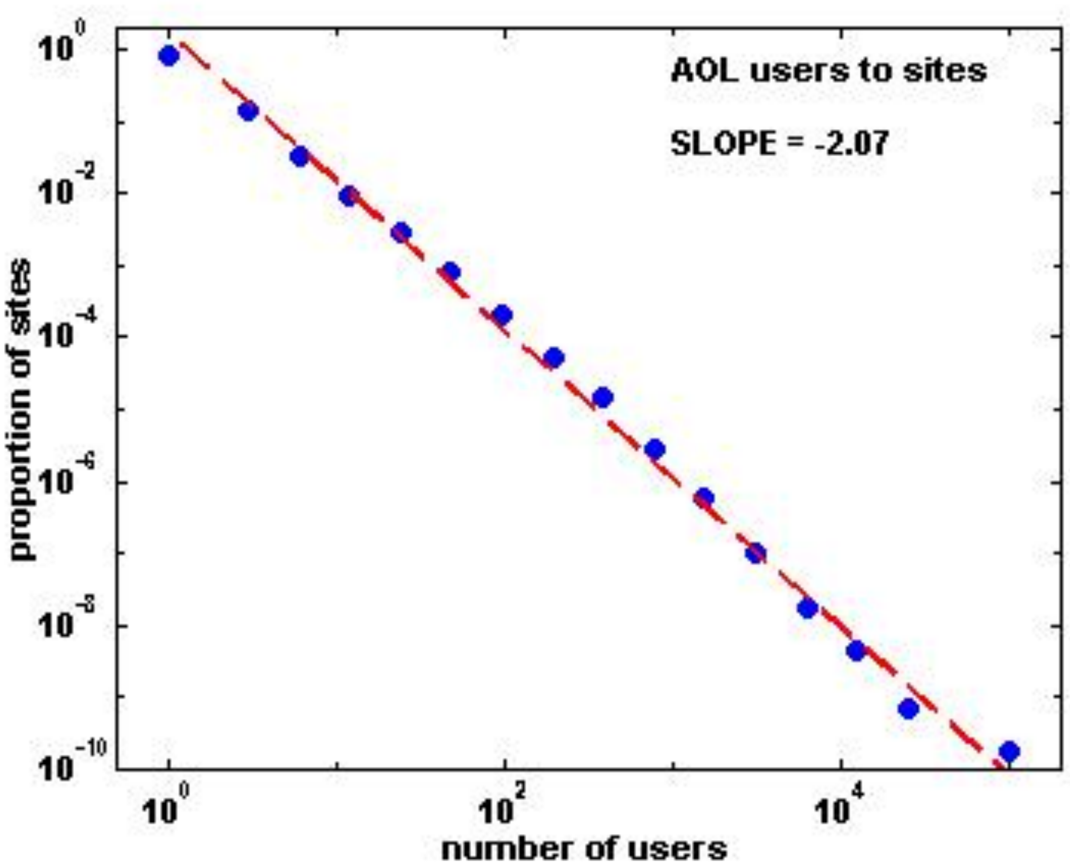
trying to fit directly…

- direct fit is too shallow: $\alpha = 1.17\ldots$



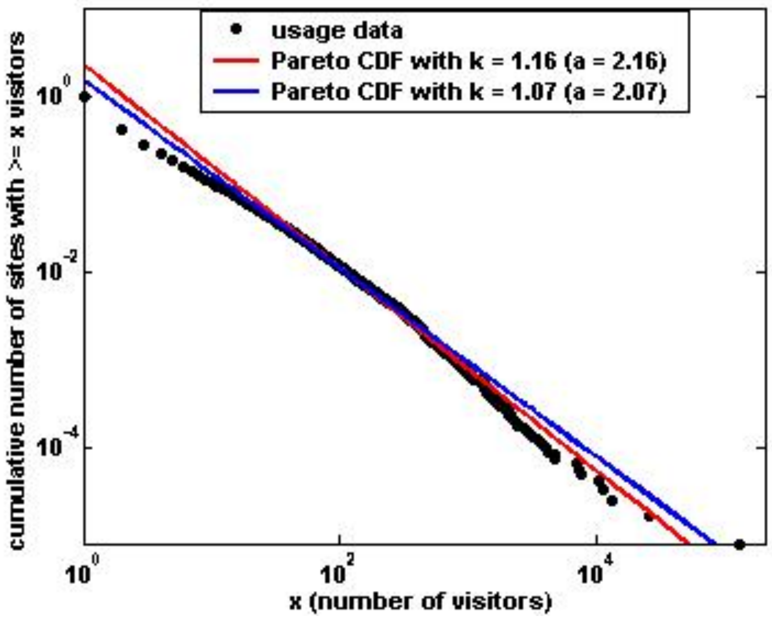
Binning the data logarithmically helps

- select exponentially wider bins
  - 1, 2, 4, 8, 16, 32, ….



Or we can try fitting the cumulative distribution

- Shows perhaps 2 separate power-law regimes that were obscured by the exponential binning
- Power-law tail may be closer to 2.4

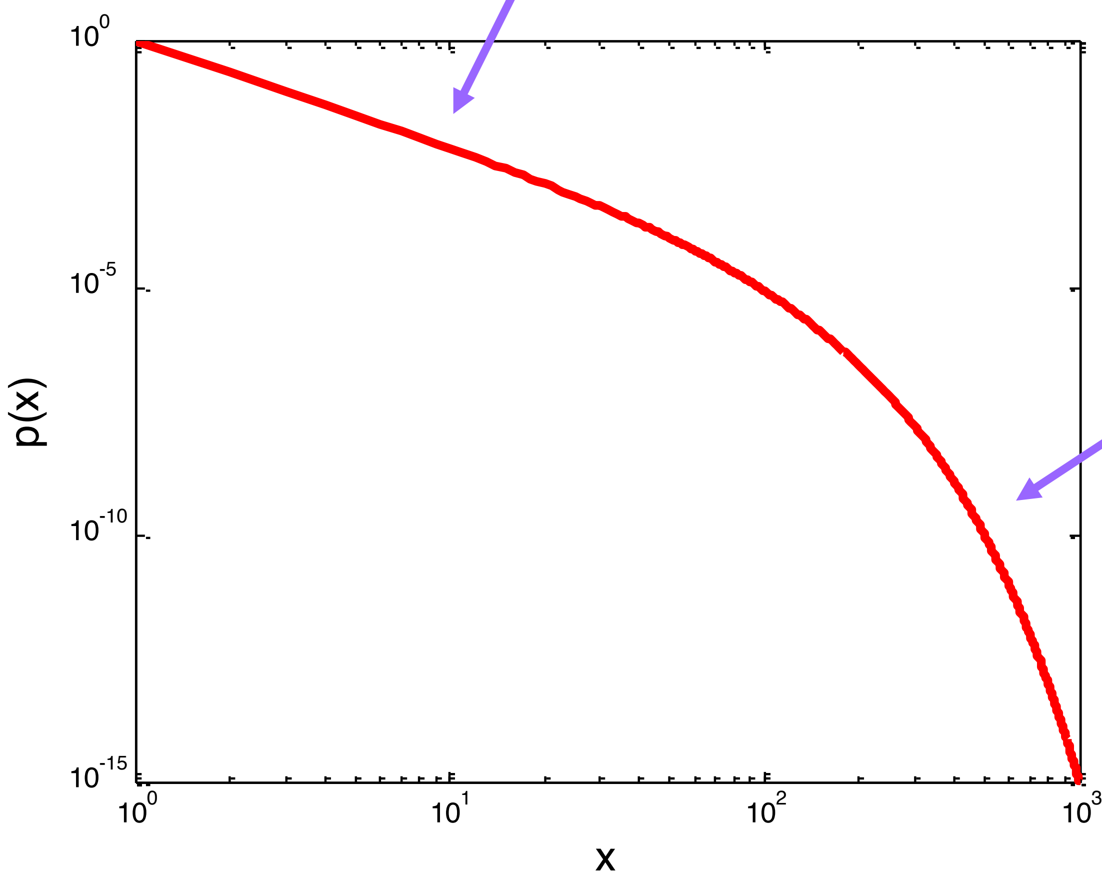


Another common distribution: power-law with an exponential cutoff

- $p(x) \sim x^{-a} e^{-x/\kappa}$

starts out as a power law

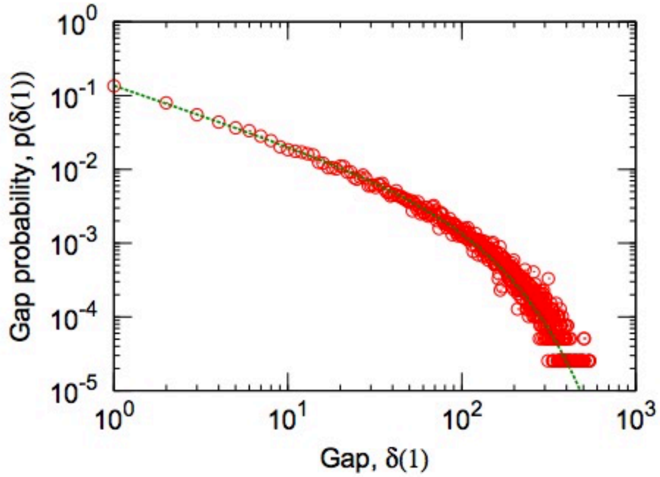
ends up as an exponential



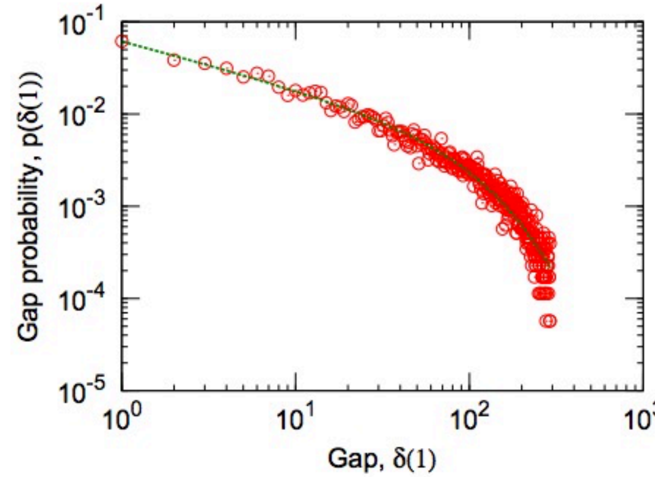
but could also be a lognormal or double exponential…

example: time between edge initiations

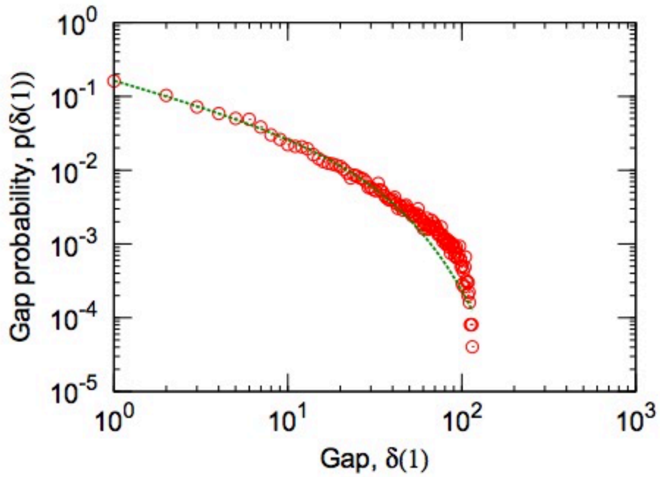
Q: Why is the cutoff present?



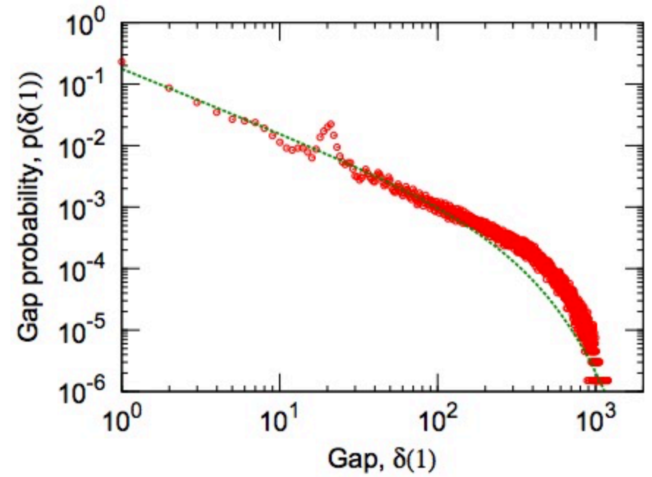
(a) FLICKR



(b) DELICIOUS



(c) ANSWERS



(d) LINKEDIN

Leskovec et al., KDD'08

Zipf &Pareto: what they have to do with power-laws

- Zipf
  - George Kingsley Zipf, a Harvard linguistics professor, sought to determine the 'size' of the 3rd or 8th or 100th most common word.
  - Size here denotes the frequency of use of the word in English text, and not the length of the word itself.
  - Zipf's law states that the size of the r'th largest occurrence of the event is inversely proportional to its rank:

$\boldsymbol{y \sim r^{-\beta}}$, with β close to unity.

So how do we go from Zipf to Pareto?

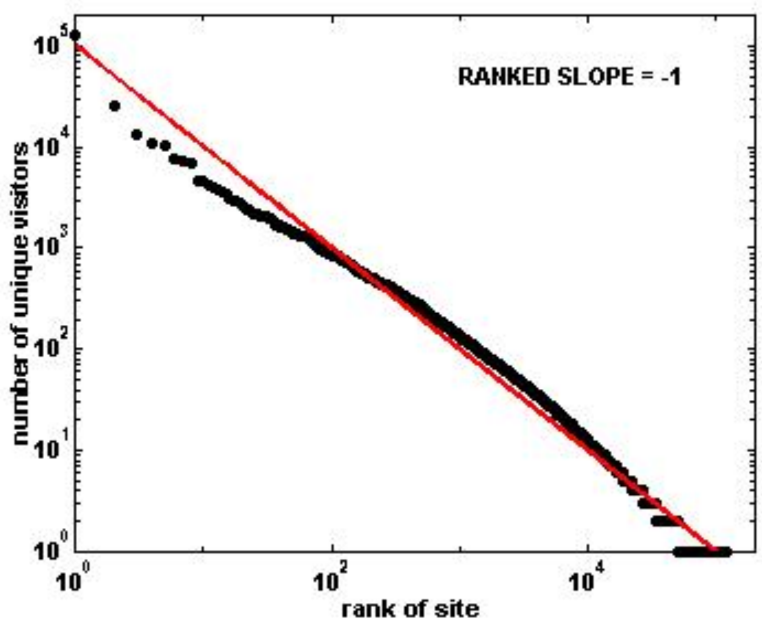
- The phrase "The ***r*** th largest city has ***n*** inhabitants" is equivalent to saying "***r*** cities have ***n*** or more inhabitants".

- This is exactly the definition of the Pareto distribution, except the x and y axes are flipped. Whereas for Zipf, ***r*** is on the x-axis and ***n*** is on the y-axis, for Pareto, ***r*** is on the y-axis and ***n*** is on the x-axis.

- Simply inverting the axes, we get that if the rank exponent is β, i.e.
  $n \sim r^{-\beta}$ for Zipf, (n = income, r = rank of person with income n)
  then the Pareto exponent is $1/\beta$ so that
  $r \sim n^{-1/\beta}$ (n = income, r = number of people whose income is n or higher)

Zipf's law & AOL site visits

- Deviation from Zipf's law
  - slightly too few websites with large numbers of visitors:



Zipf’s Law and city sizes (~1930) [2]

Rank(k)	City	Population (1990)	Zips’s Law $10{,}000{,}000/k$	Modified Zipf’s law: (Mandelbrot) $5{,}000{,}000/\left(k-\frac{2}{5}\right)^{3/4}$
1	Now York	7,322,564	10,000,000	7,334,265
7	Detroit	1,027,974	1,428,571	1,214,261
13	Baltimore	736,014	769,231	747,693
19	Washington DC	606,900	526,316	558,258
25	New Orleans	496,938	400,000	452,656
31	Kansas City	434,829	322,581	384,308
37	Virgina Beach	393,089	270,270	336,015
49	Toledo	332,943	204,082	271,639
61	Arlington	261,721	163,932	230,205
73	Baton Rouge	219,531	136,986	201,033
85	Hialeah	188,008	117,647	179,243
97	Bakersfield	174,820	103,270	162,270

not any more

slide: Luciano Pietronero

80/20 rule

- The fraction W of the wealth in the hands of the richest P of the the population is given by

$$W = P^{(\alpha-2)/(\alpha-1)}$$

- Example: US wealth: $\alpha = 2.1$
  - richest 20% of the population holds 86% of the wealth

Wrap up on power-laws

- Power-laws are cool and intriguing
- But make sure your data is actually power-law before boasting