open.michigan

University of Michigan

School of Information
University of Michigan

Power-laws
"Scale free" networks

Reading:

Lada Adamic, Zipf, Power-laws, and Pareto - a ranking tutorial, http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html

M. E. J. Newman, Power laws, Pareto distributions and Zipf's law, Contemporary Physics 46, 323-351 (2005)

Barabasi and Albert, 'Emergence of scaling in random networks', Science 1999

Outline

- Power law distributions
- Fitting
- what kinds of processes generate power laws?
- Barabasi-Albert model for scale-free graphs

random network game

- Around the room once:
  - write your name on an orange square, place it in the cylinder
- Around the room second round:
  - shake the cylinder. Draw a random orange square. Write down the name of the person whose name you drew

- Questions:
  - What does the network look like?
  - What does the degree distribution look like?

2nd random network game

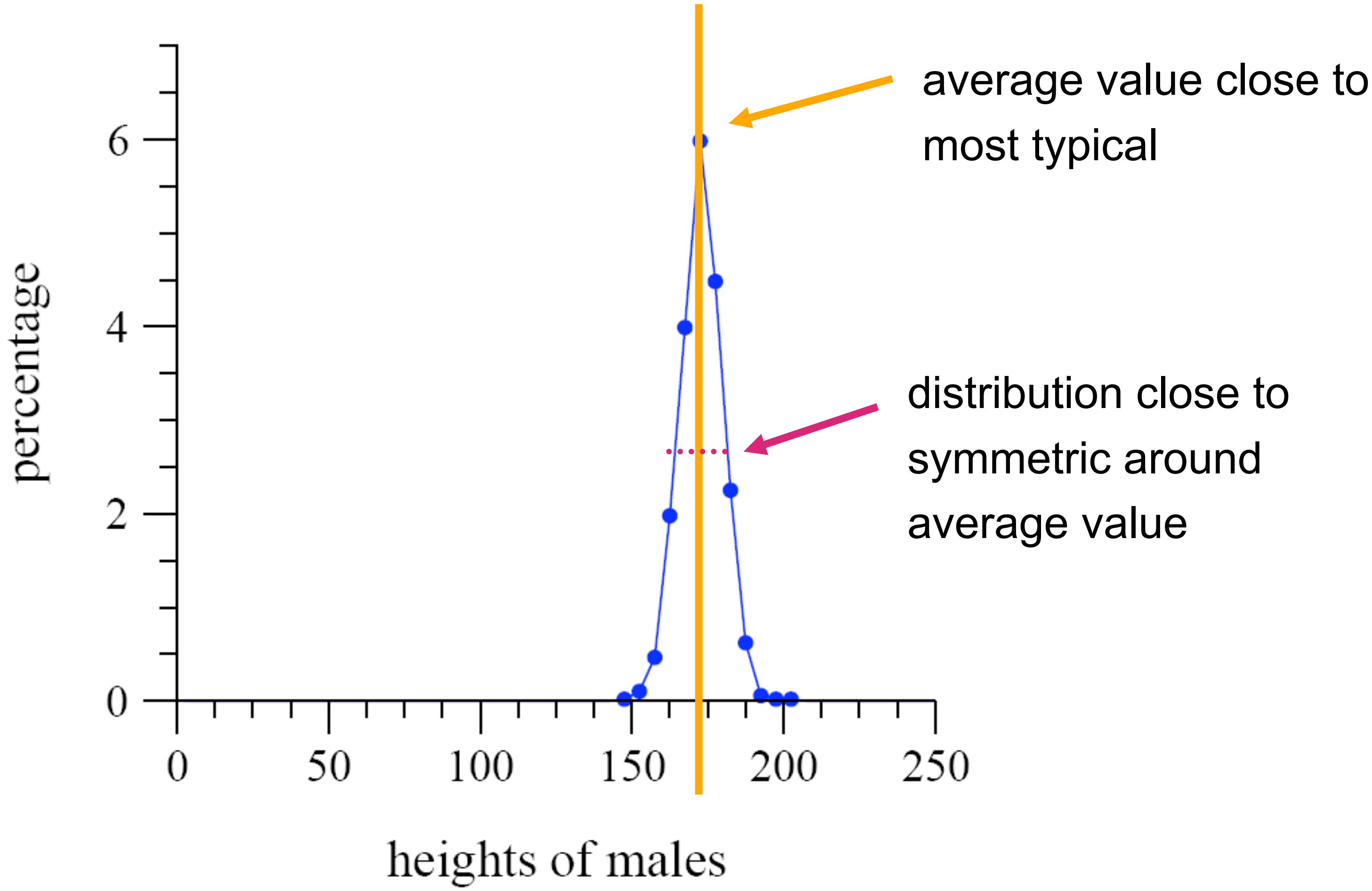
- Around the room once:
  - shake the cylinder. Draw a random square. Write down the name of the person on a new white square, and place both squares back in the cylinder
  - write your name on an orange square, place it in the cylinder

- Questions:
  - What does the network look like?
  - What does the number of squares with a person's name represent?
  - What does the degree distribution look like?
  - How is this process different than the previous one?

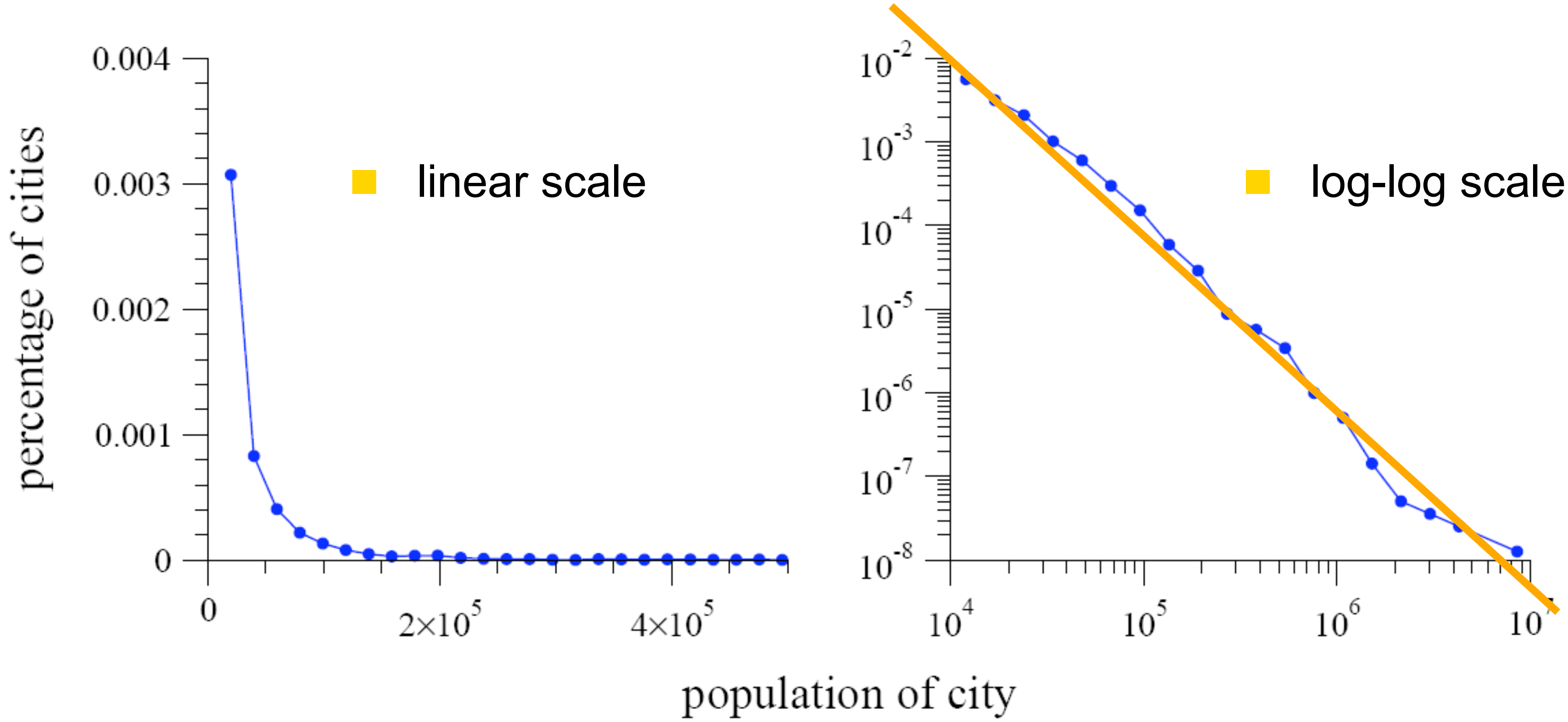
What is a heavy tailed-distribution?

- 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
- 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*

Normal (also called Gaussian) distribution of human heights



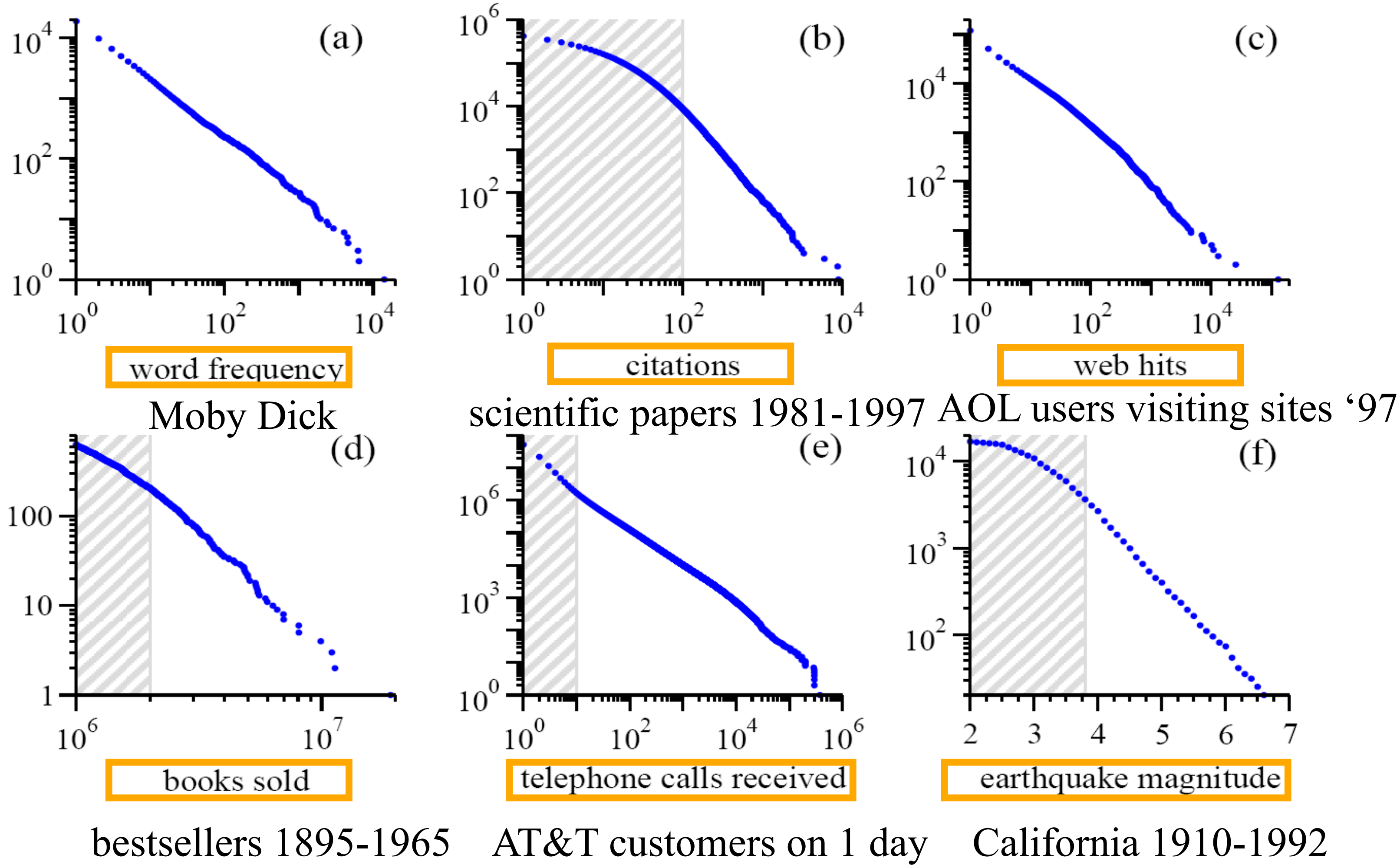
Power-law distribution



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

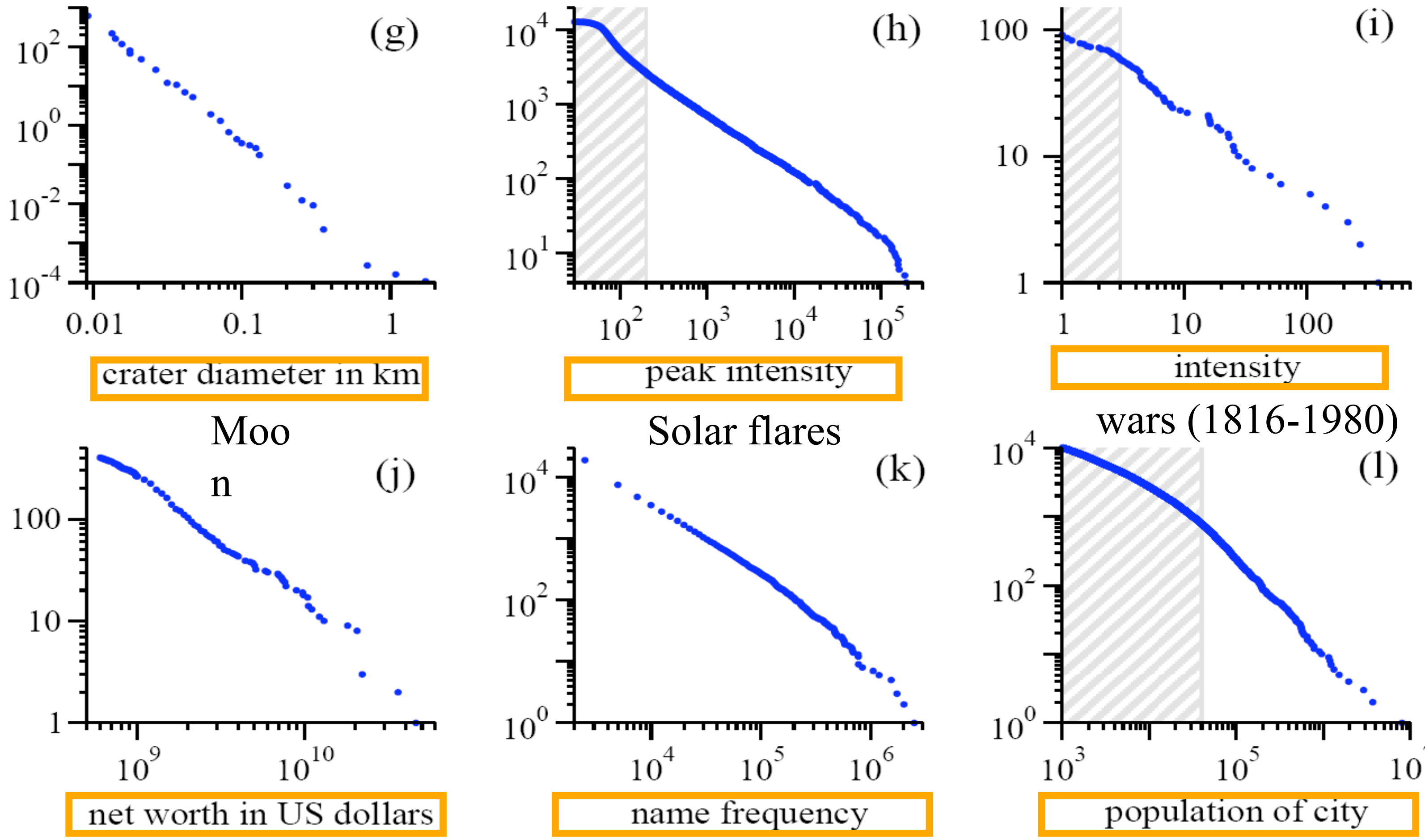
Power laws are seemingly everywhere

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



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)

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

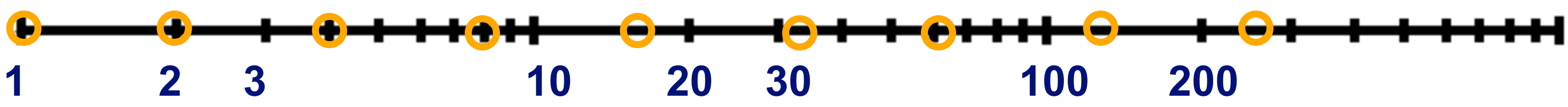
$$p(x) = Cx^{-\alpha}$$

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

power law exponent α

Logarithmic axes

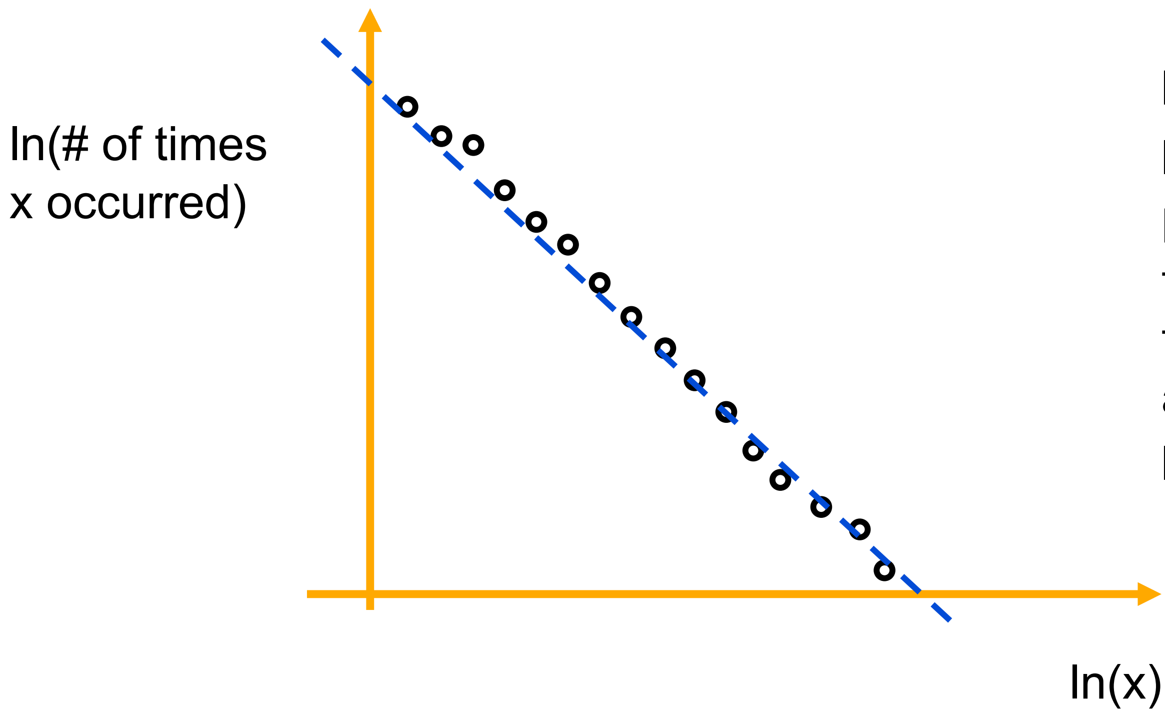
- powers of a number will be uniformly spaced



- $2^0=1$, $2^1=2$, $2^2=4$, $2^3=8$, $2^4=16$, $2^5=32$, $2^6=64$,….

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 a because

$\log_{10}(x) = \ln(x)/\ln(10)$

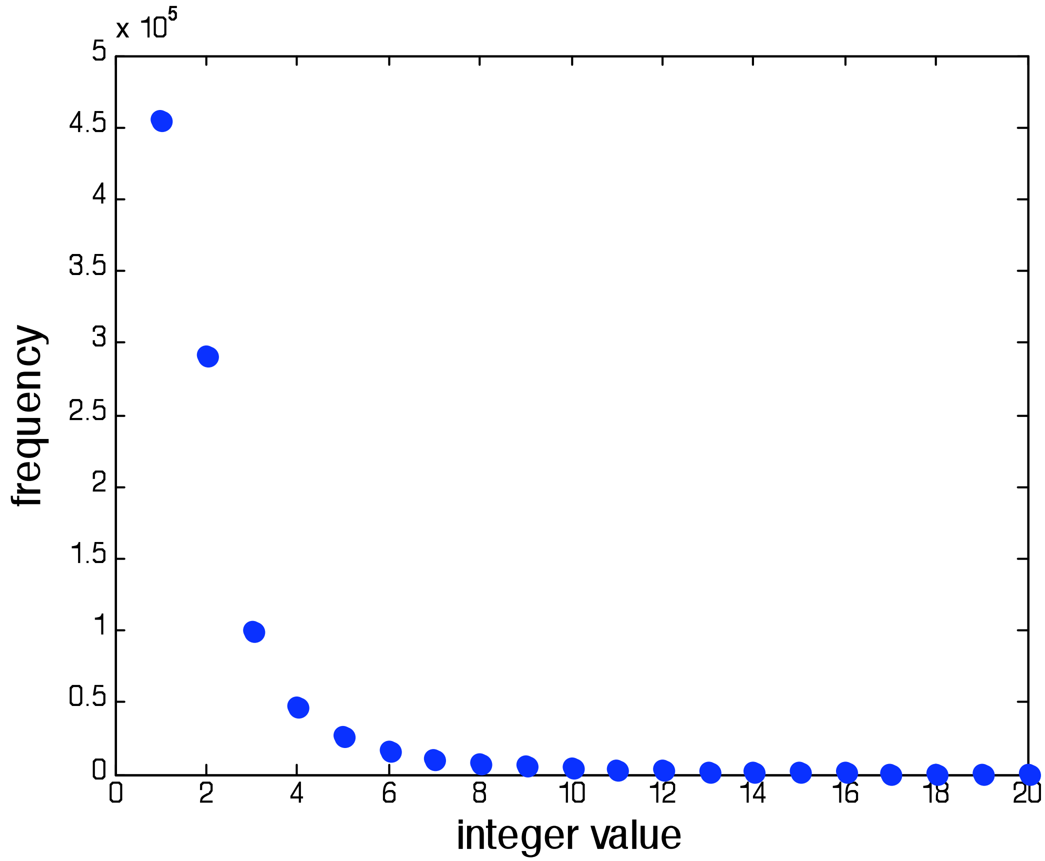
x can represent various quantities, the indegree of a node, the magnitude of an earthquake, the frequency of a word in text

Example on an artificially generated data set

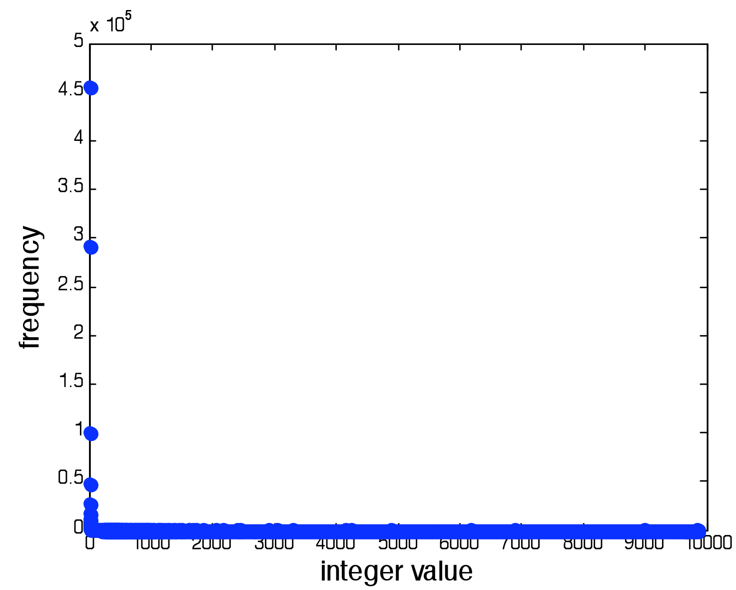
- Take 1 million random numbers from a distribution with $\alpha = 2.5$
- Can be generated using the so-called 'transformation method'
- Generate random numbers r on the unit interval $0 \le r < 1$
- then $x = (1-r)^{-1/(\alpha-1)}$ is a random power law distributed real number in the range $1 \le x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

Linear scale plot of straight bin of the data

- How many times did the number 1 or 3843 or 99723 occur
- Power-law relationship not as apparent
- Only makes sense to look at smallest bins



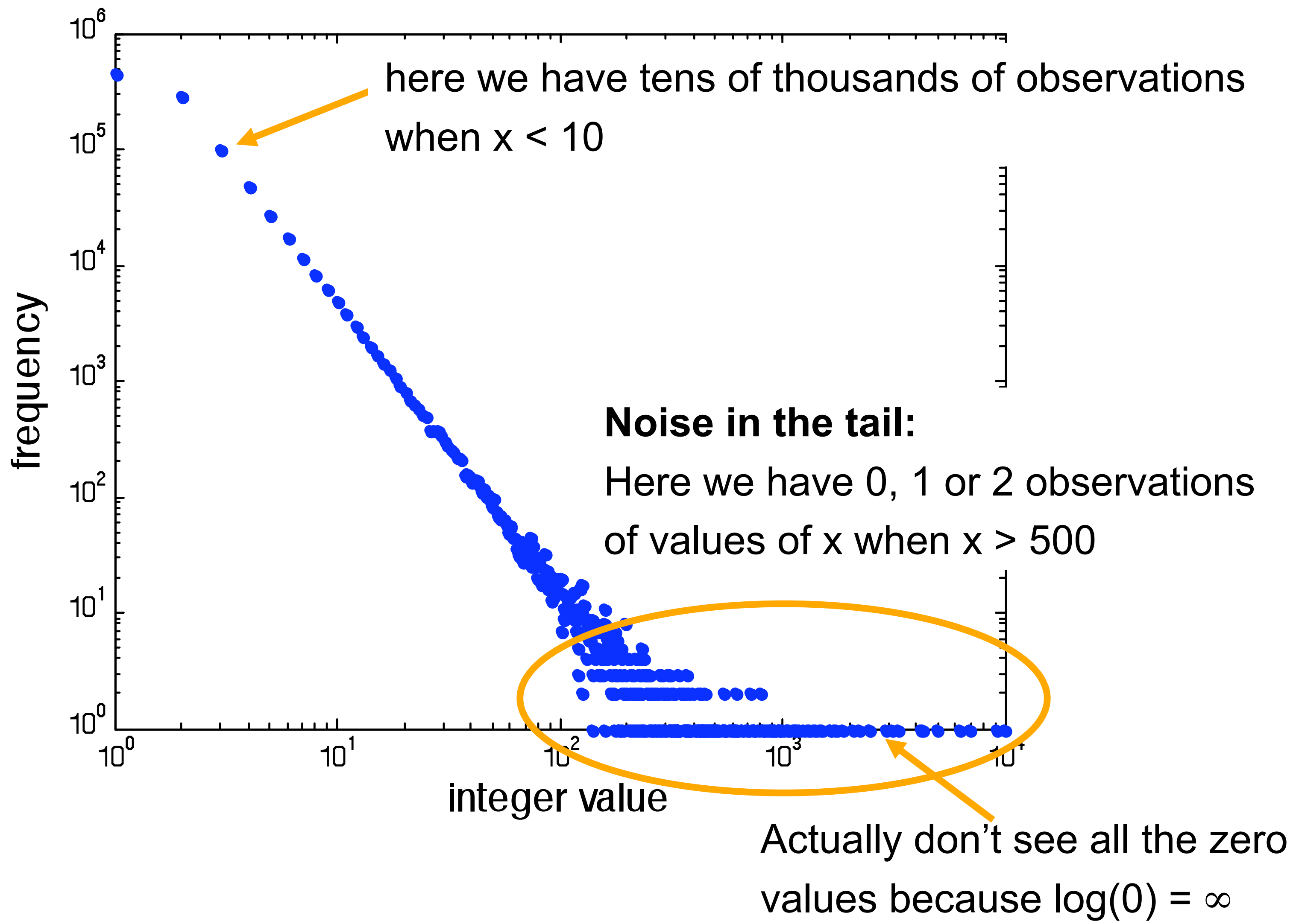
first few bins



whole range

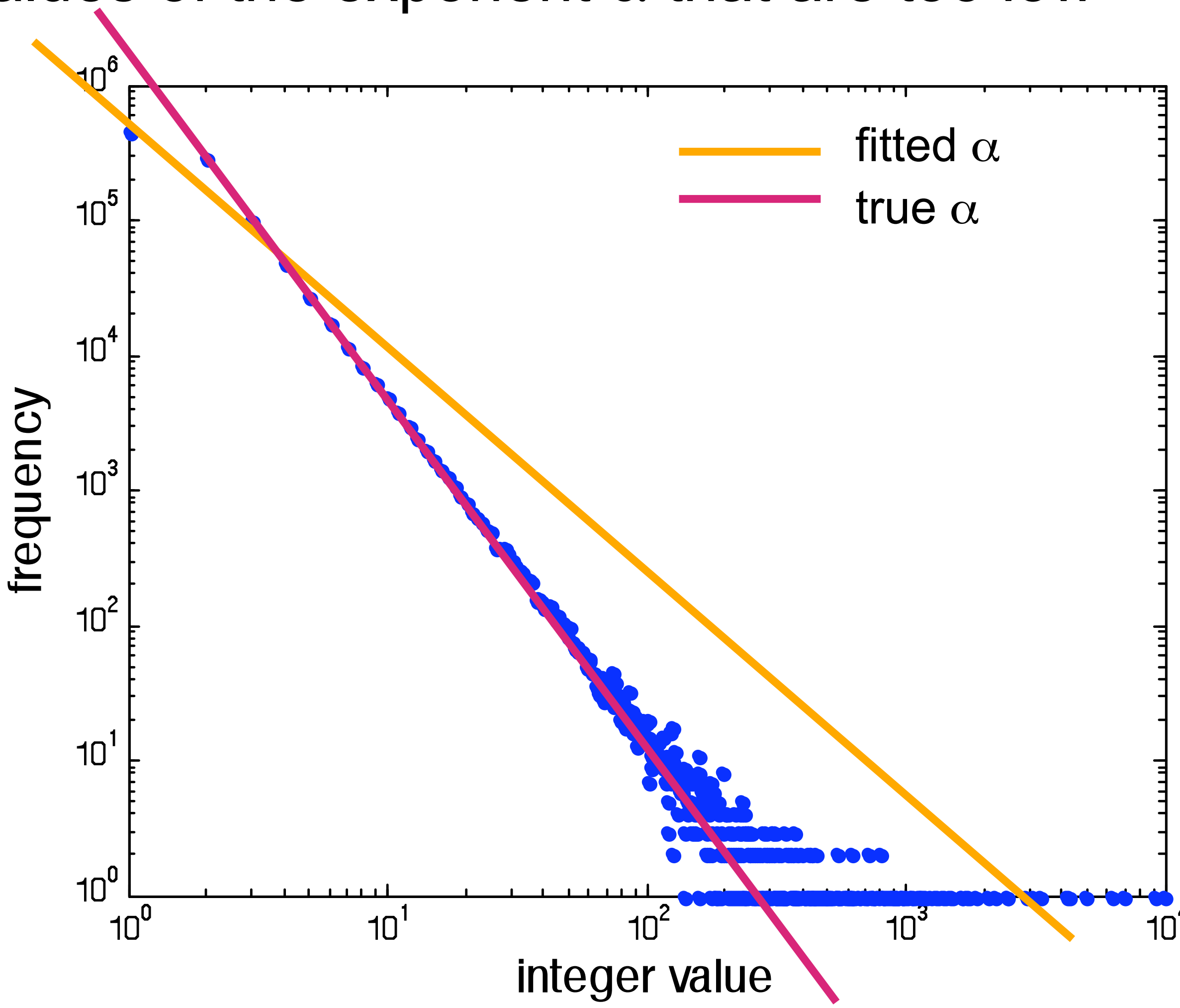
Log-log scale plot of straight binning of the data

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



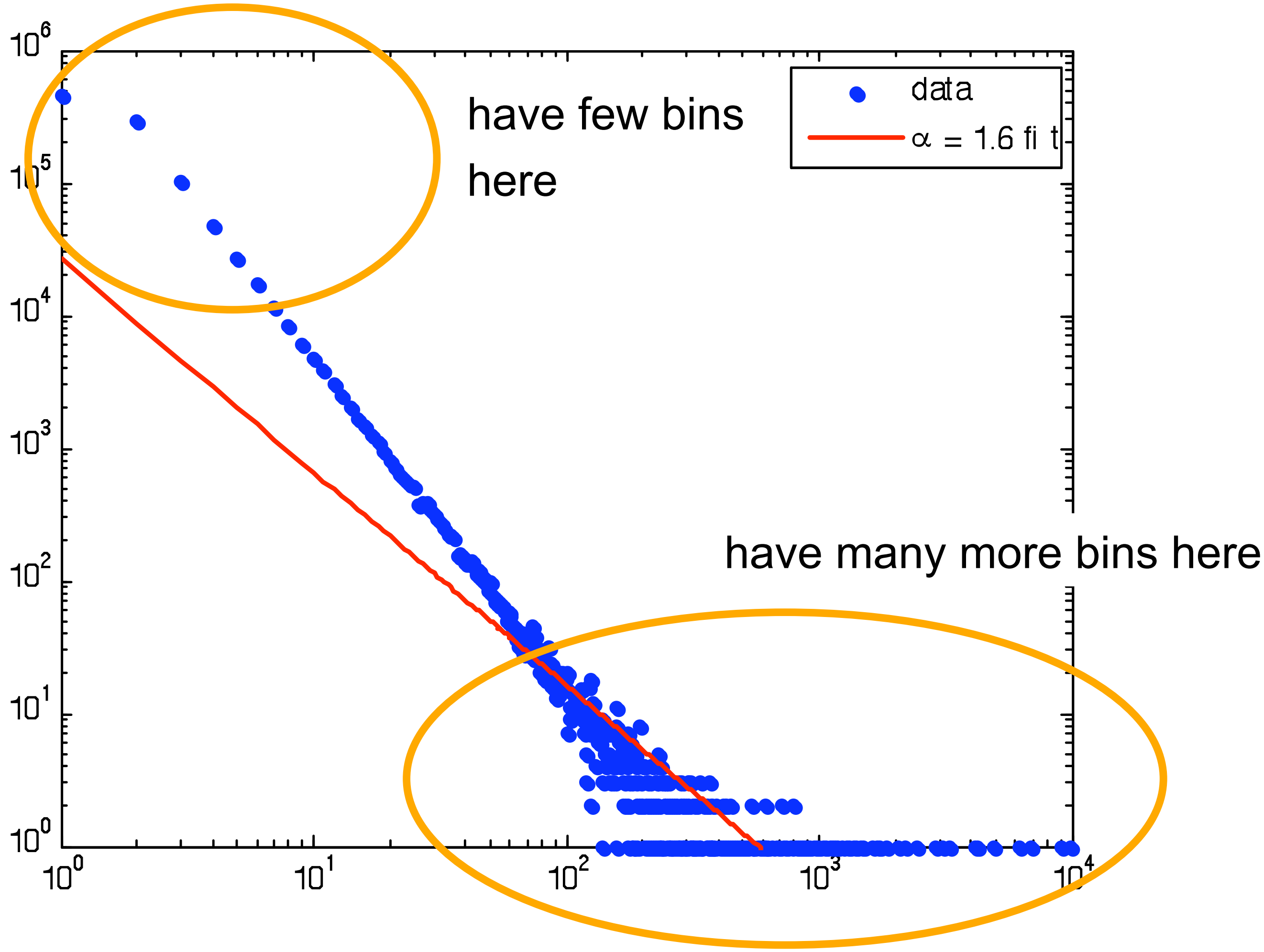
Log-log scale plot of straight binning of the data

- 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

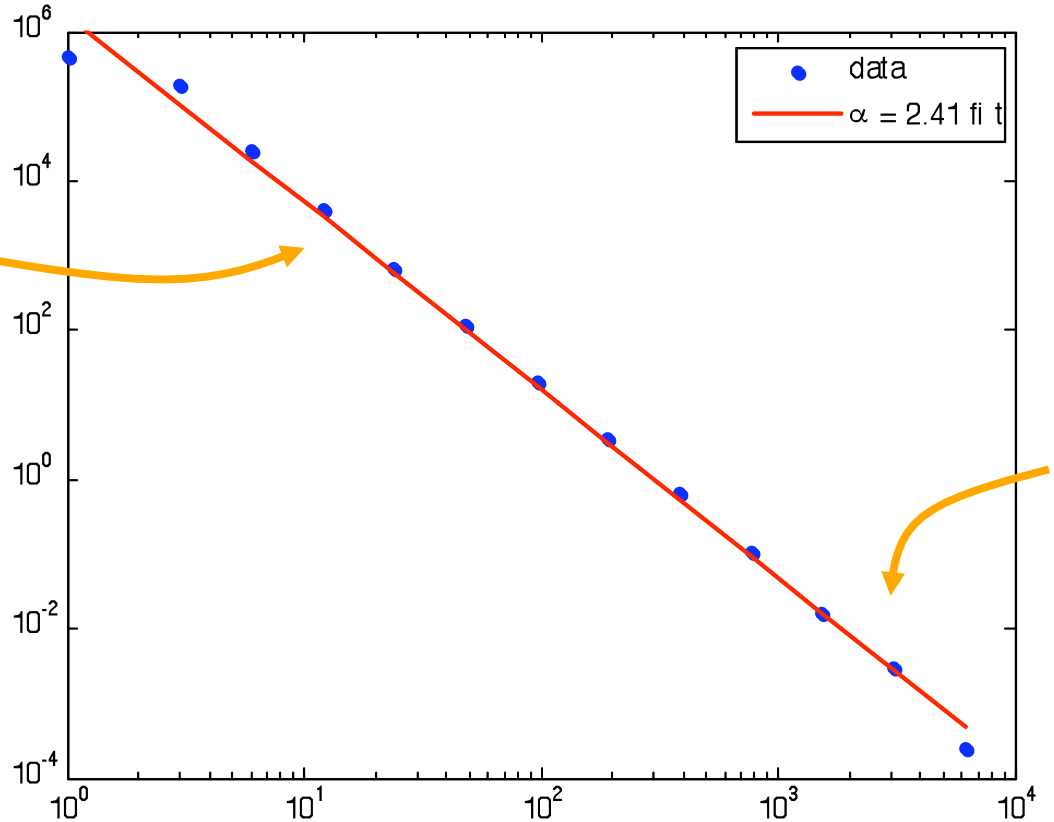
- Noise in the tail skews the regression result



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

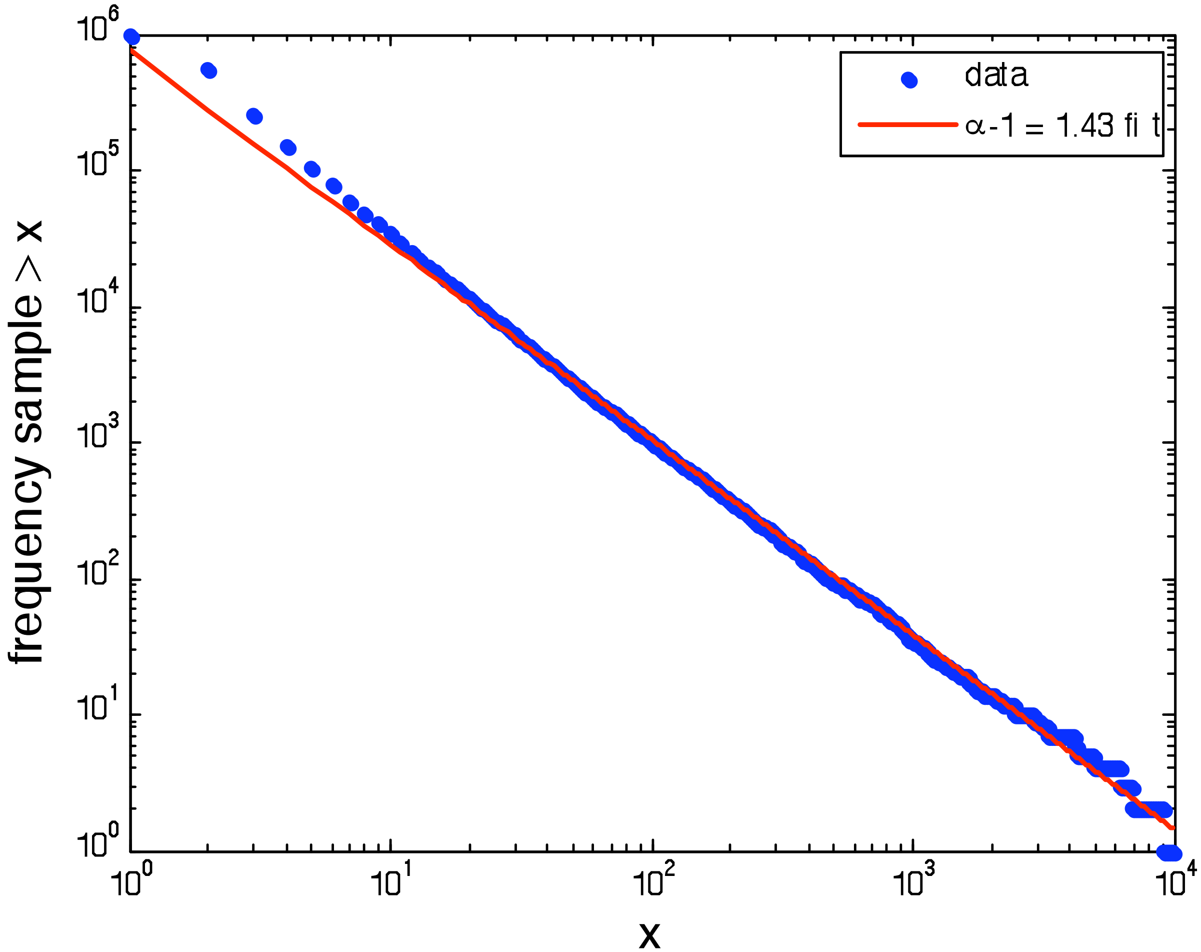
- No loss of information
  - No need to bin, has value at each observed value of x
- But now have cumulative distribution
  - i.e. how many of the values of x are at least X

  - The cumulative probability of a power law probability distribution is also power law but with an exponent **α - 1**

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

Fitting via regression to the cumulative distribution

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

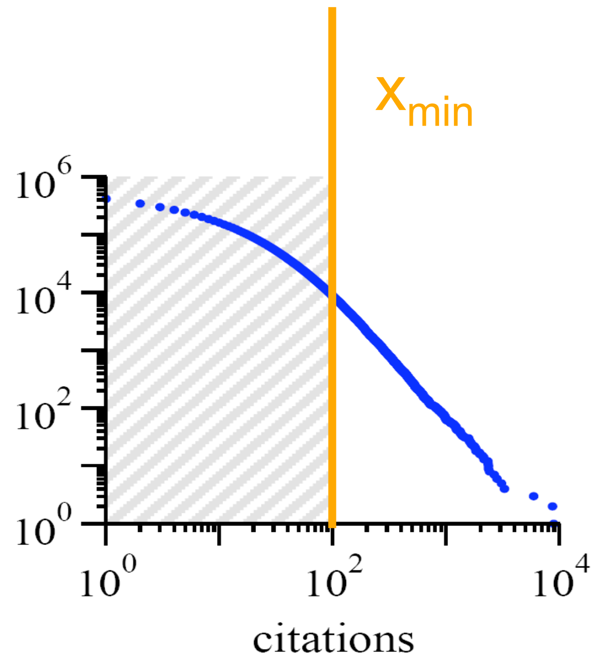


Where to start fitting?

- 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 datapoints, and you have *n* of them
- for our data set we get α = 2.503 – pretty close!

Some exponents for real world data

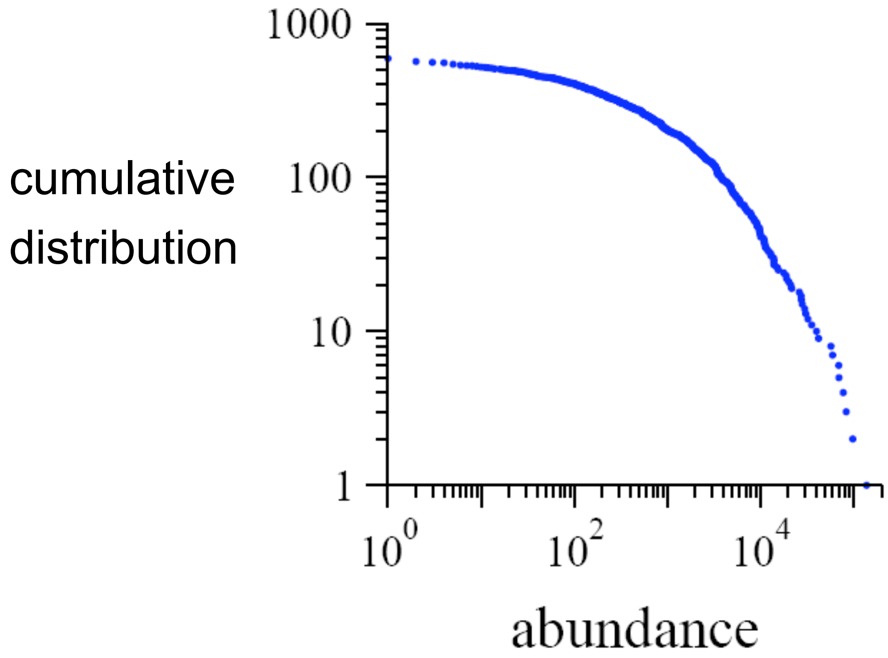
	x_{min}	exponent α
frequency of use of words	1	2.20
number of citations to papers	100	3.04
number of hits on web sites	1	2.40
copies of books sold in the US	2 000 000	3.51
telephone calls received	10	2.22
magnitude of earthquakes	3.8	3.04
diameter of moon craters	0.01	3.14
intensity of solar flares	200	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.



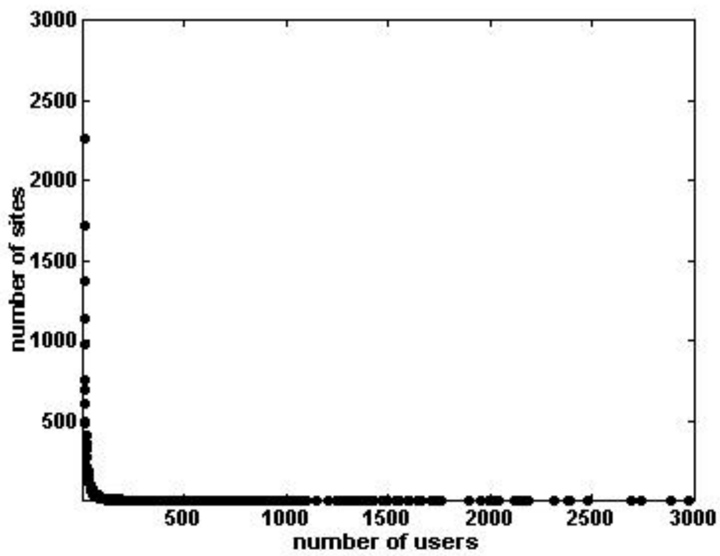
- another example:
  - size of wildfires (in acres)

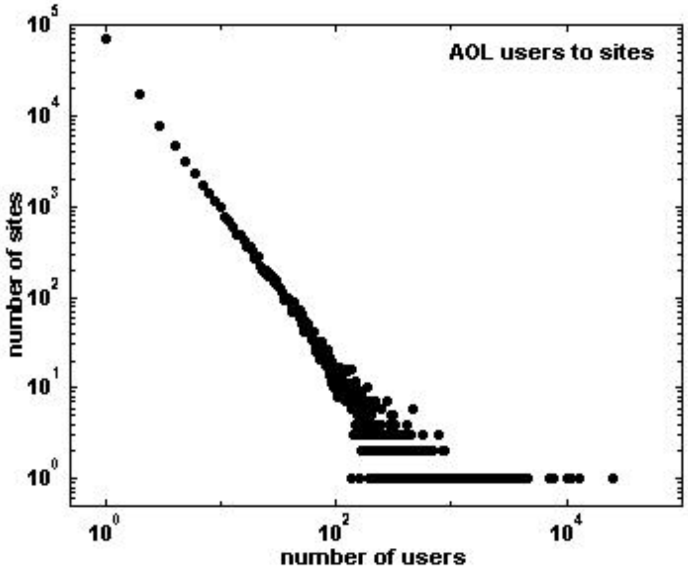
Source:MEJ Newman, 'Power laws, Pareto distributions and Zipf's law', *Contemporary Physics* **46**, 323–351 (2005)

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



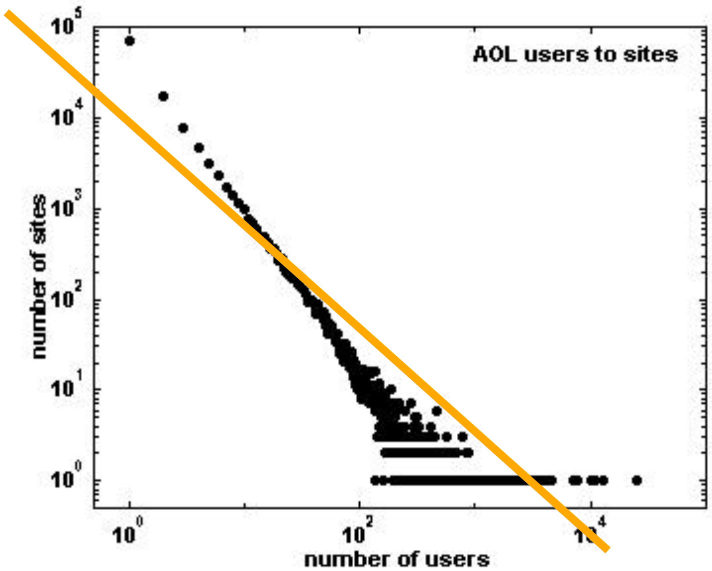


simple binning on a linear scale

simple binning on a log-log scale

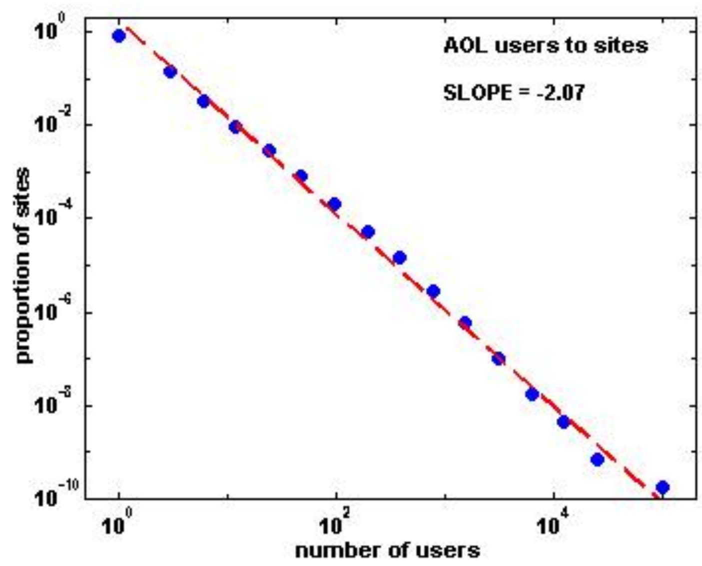
trying to fit directly...

- direct fit is too shallow: α = 1.17…



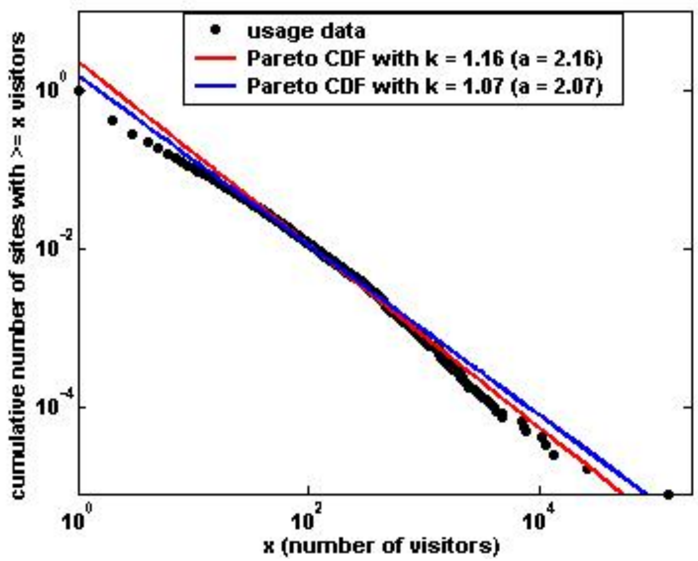
Binning the data logarithmically helps

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



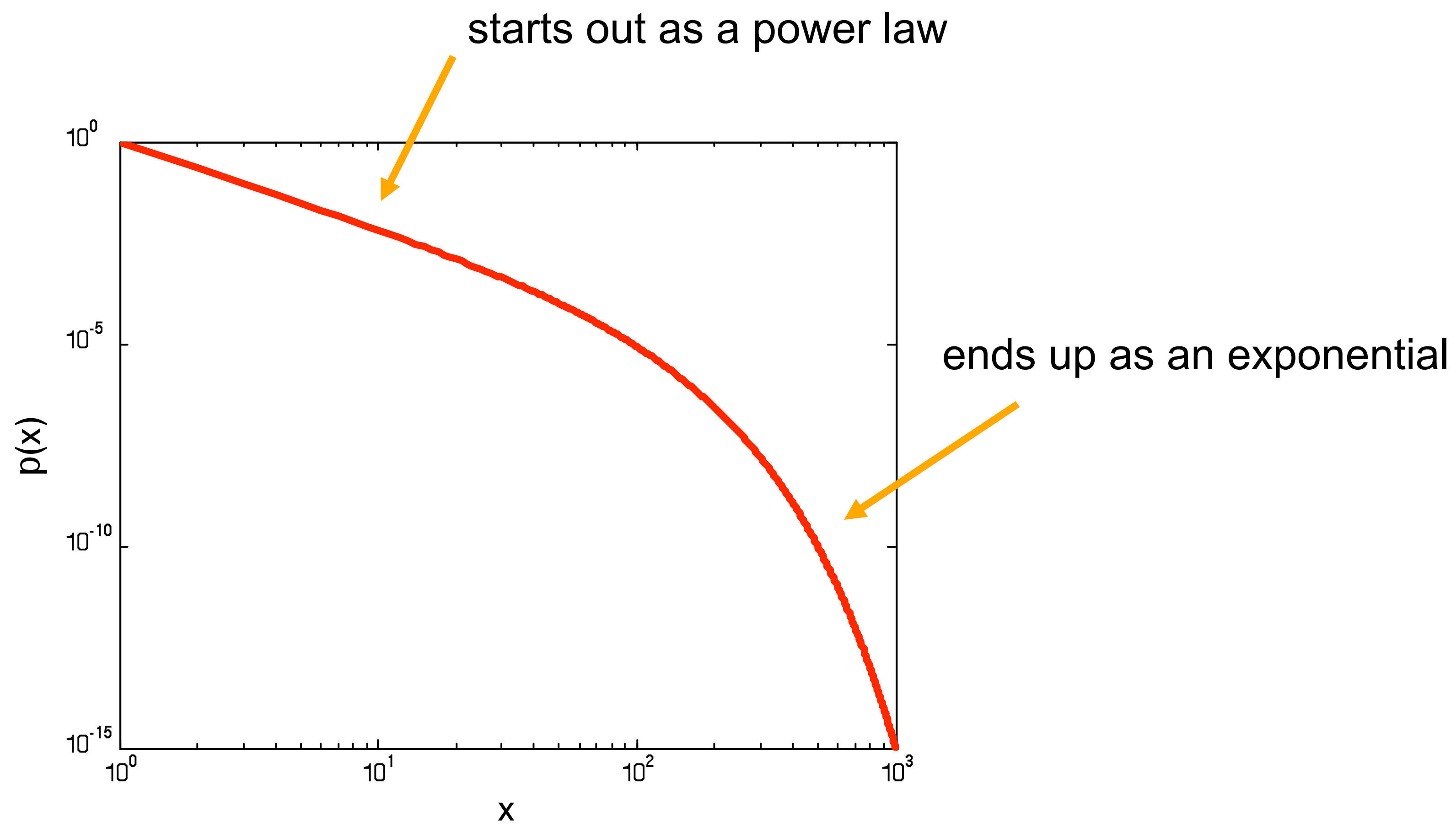
Or we can try fitting the cumulative distribution

- 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

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



but could also be a lognormal or double exponential…

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.

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

- Pareto
  - The Italian economist Vilfredo Pareto was interested in the distribution of income.
  - Pareto's law is expressed in terms of the cumulative distribution (the probability that a person earns X or more).

$$P[X > x] \sim x^{-k}$$

  - Here we recognize *k* as just α -1, where α is the power-law exponent

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.

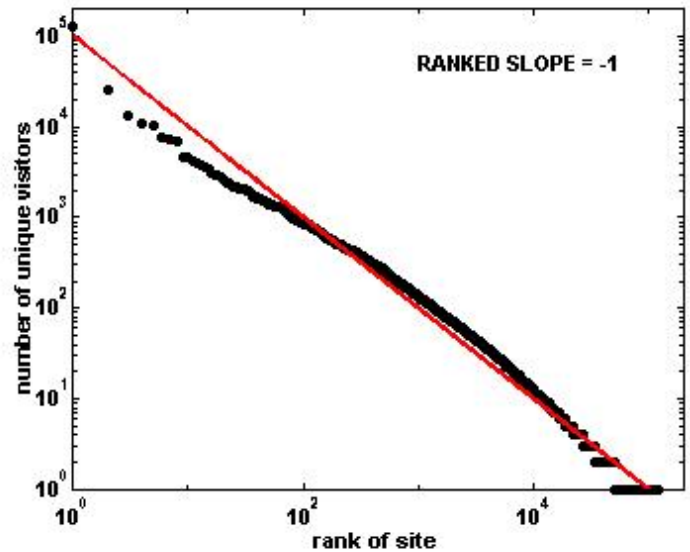
  $n \sim r^{-\beta}$ for Zipf, (n = income, r = rank of person with income n)

  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

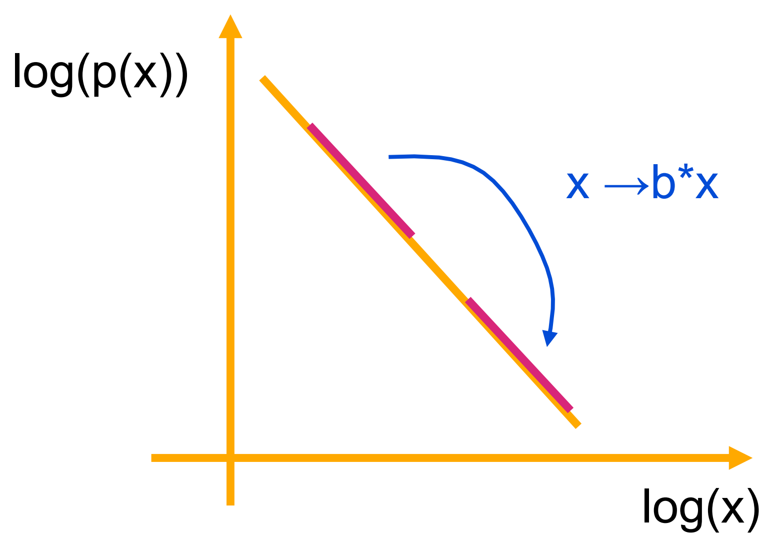
- The fraction W of the wealth in the hands of the richest P of the the population is given by

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

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

What does it mean to be scale free?

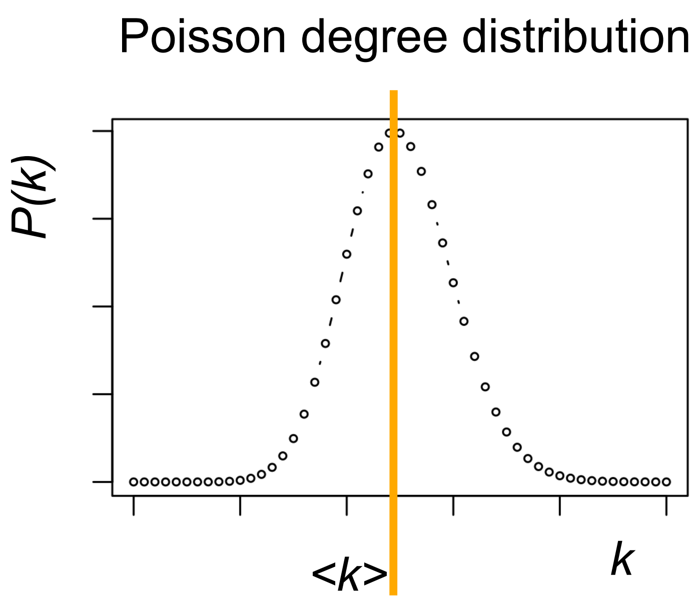
- A power law looks the same no mater what scale we look at it on (2 to 50 or 200 to 5000)
- Only true of a power-law distribution!
- $p(bx) = g(b)\, p(x)$ – shape of the distribution is unchanged except for a multiplicative constant
- $p(bx) = (bx)^{-\alpha} = b^{-\alpha}\, x^{-\alpha}$



Back to networks:
skewed degree distributions

Simplest random network

- Erdos-Renyi random graph: each pair of nodes is equally likely to be connected, with probability p.
- p = 2*E/N/(N-1)
- Poisson degree distribution is narrowly distributed around <k> = p*(N-1)



Random graph model

- The degree distribution is given by
  - coinflips to see how many people you'll be connected to, one coin flip per each of the (n-1) other nodes
  - probability *p*, of connecting

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

$$p_k = \frac{z^k e^{-z}}{k!}$$

Binomial

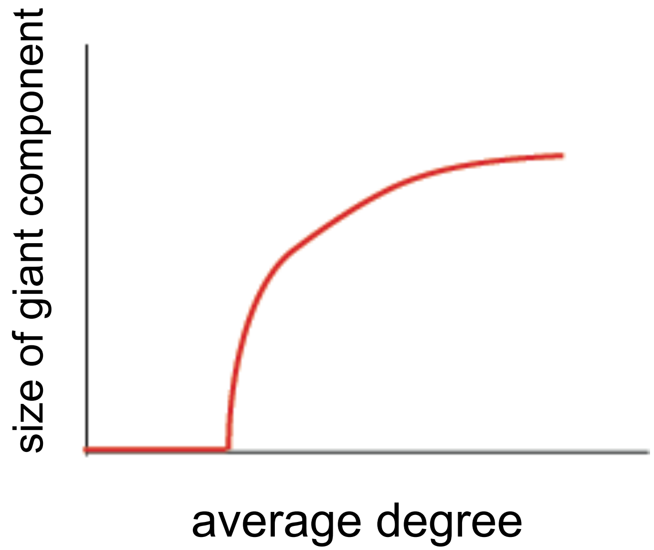
↓ limit *p* small

Poisson

↓ limit large *n*

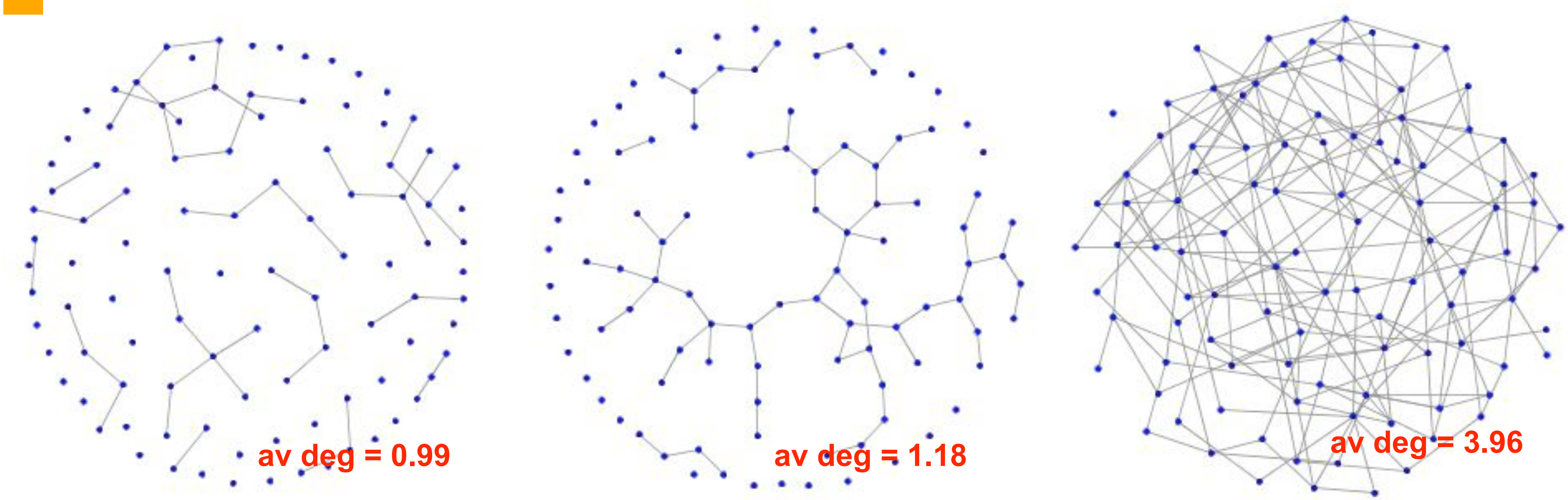
Normal

Percolation threshold in Erdos-Renyi Graphs



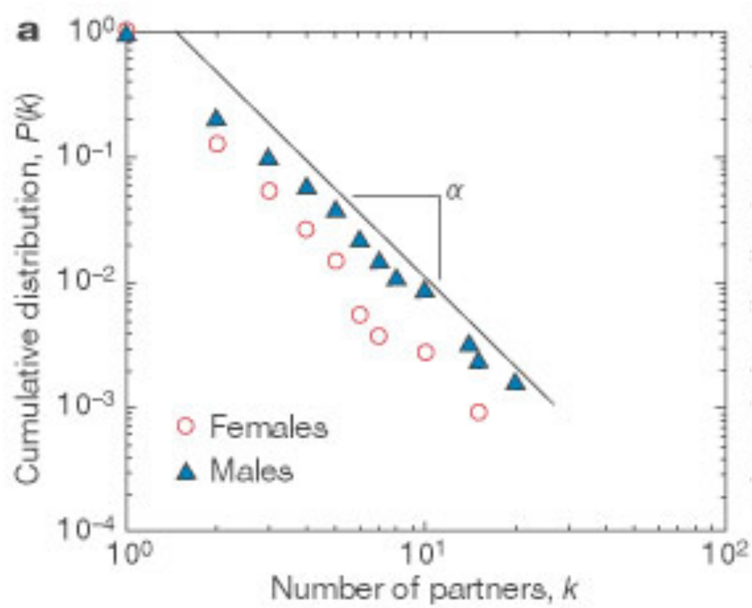
Percolation theshold: how many edges need to be added before the giant component appears?

As the average degree increases to z = 1, a giant component suddenly appears



Real world networks are often power law though…

- Sexual networks
- Most individuals report 1-2 partners in the past 12 months, but some…



Source: The web of human sexual contacts, Liljeroset al.,Nature 411, 907-908(21 June 2001)

Preferential Attachment in Networks

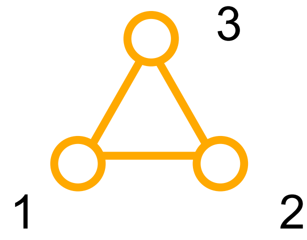
- First considered by [Price 65] as a model for citation networks
  - each new paper is generated with m citations (mean)
  - new papers cite previous papers with probability proportional to their indegree (citations)
  - what about papers without any citations?
    - each paper is considered to have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

Barabasi-Albert model

- Undirected(?) model: each node connects to other nodes with probability proportional to their degree
  - the process starts with some initial subgraph
  - each node comes with m edges
- 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. $m_0 = 3$



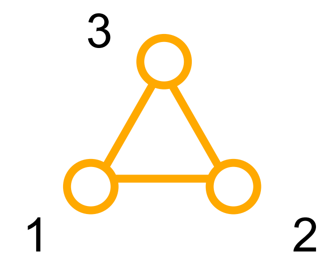
1 1 2 2 2 3 3 4 5 6 6 7 8 ….

  - 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

generating BA graphs – cont'd

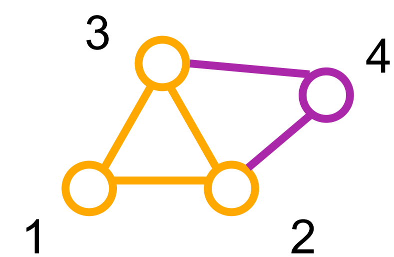
- To start, each vertex has an equal number of edges (2)
  - the probability of choosing any vertex is 1/3

1 1 2 2 3 3

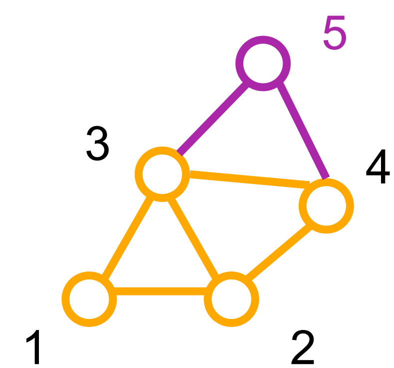


- We add a new vertex, and it will have m edges, here take m=2
  - draw 2 random elements from the array – suppose they are 2 and 3

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



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

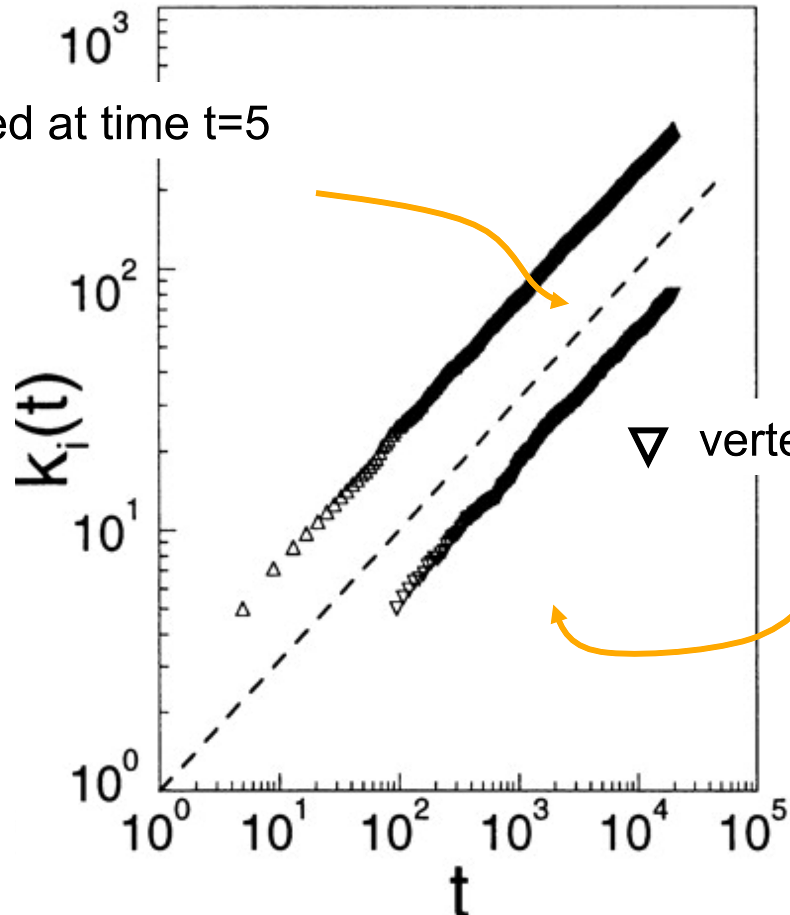
- Add a new vertex, draw a vertex for it to connect from the array
  - etc.

Properties of the BA graph

- The distribution is scale free with exponent $\alpha = 3$
  $P(k) = 2\ m^2/k^3$
- The graph is connected
  - Every vertex is born with a link (m= 1) or several links (m > 1)
  - It connects to older vertices, which are part of the giant component
- The older are richer
  - Nodes accumulate links as time goes on
  - preferential attachment will prefer wealthier nodes, who tend to be older and had a head start

Time evolution of the connectivity of a vertex in the BA model

△ vertex introduced at time t=5



▽ vertex introduced at time t=95

- Younger vertex does not stand a chance:
- at t=95 older vertex has ~ 20 edges, and younger vertex is starting out with 5
- at t ~ 10,000 older vertex has 200 edges and younger vertex has 50

Source: Barabasi and Albert, 'Emergence of scaling in random networks', Science 1999.

thoughts

- BA networks are not clustered.
  Can you think of a growth model of having preferential attachment and clustering at the same time?
- What would the network look like if nodes are added over time, but not attached preferentially?
- What other processes might give rise to power law networks?

wrap up

- power law distributions are everywhere
- there are good and bad ways of fitting them
- some distributions are not power-law
- preferential attachment leads to power law networks…
- … but it's not the whole story, and not the only way of generating them

Lab:

generating scale free network with Pajek

generating scale free networks with NetLogo