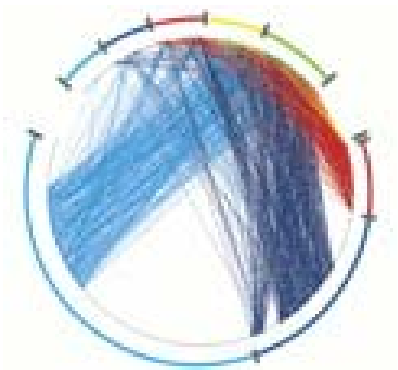
Lecture 11&12: Scale Free Networks and Network Evolution



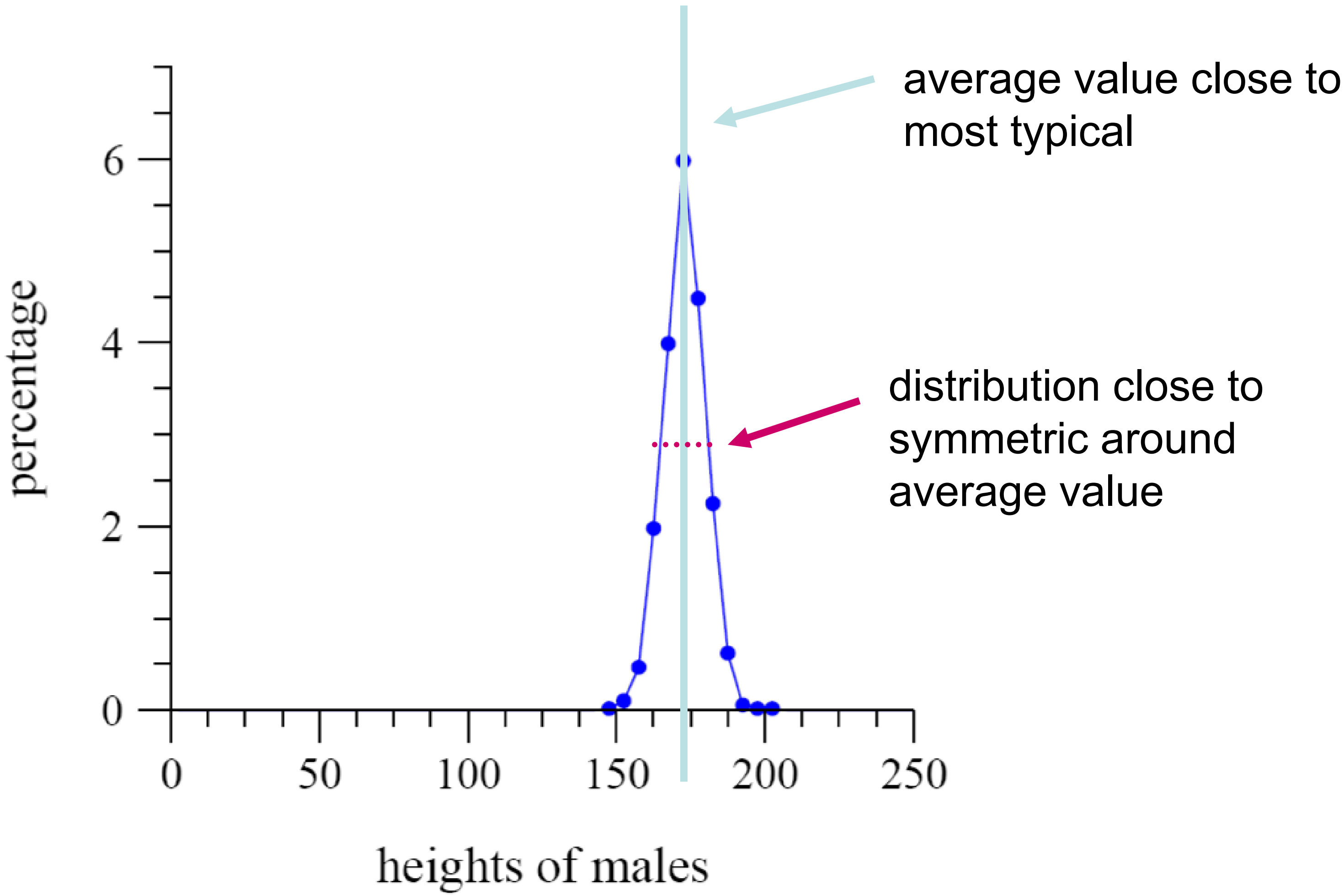
What is a heavy tailed-distribution?

- 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: Tehran 12 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
    - Tehran: pop. 12 million, a village 78, *ratio: 150,000*

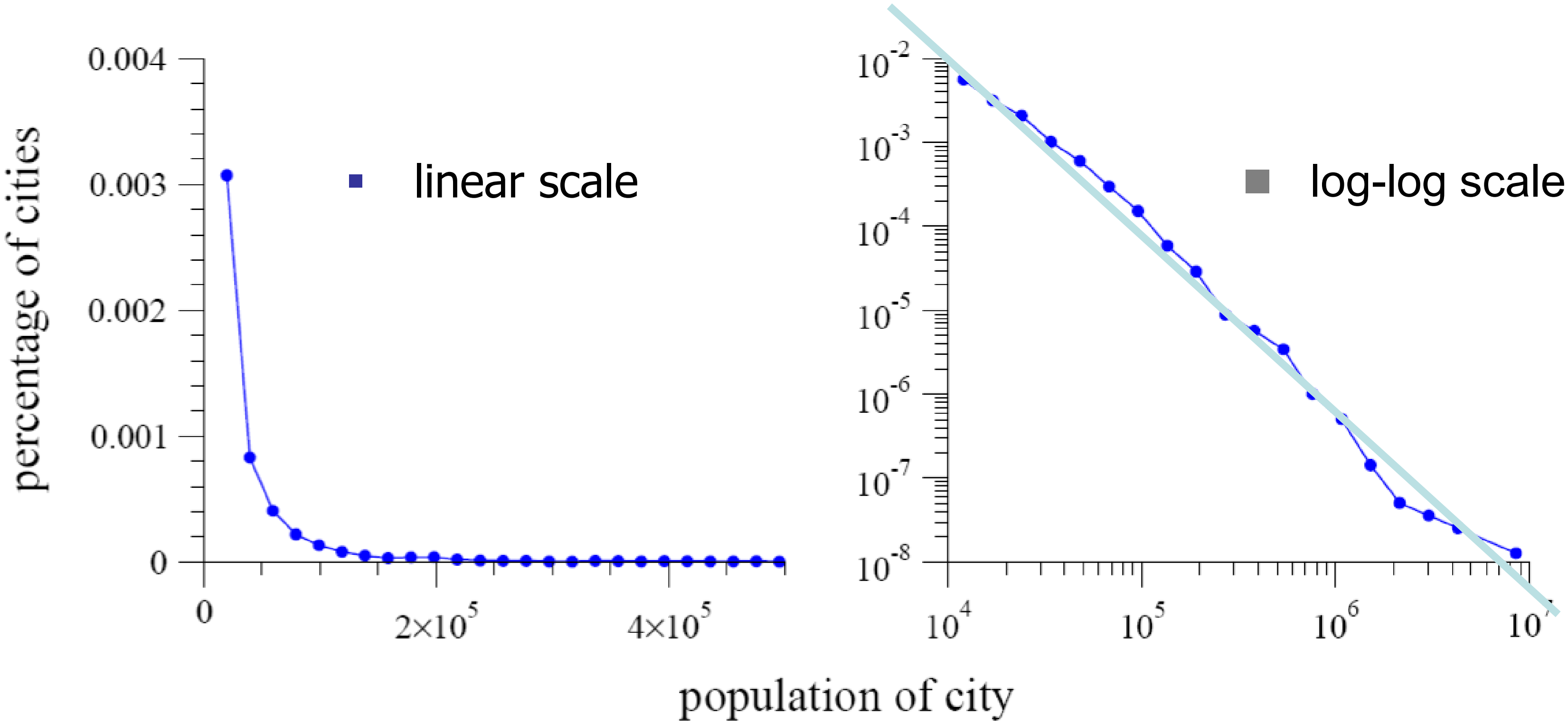
The Heavy Tail

- The power law distribution implies an "infinite variance"
  - The "area" of "big *ks*" in an exponential distribution tend to zero with $k \rightarrow \infty$
  - This is not true for the power law distribution, implying an infinite variance
  - The tail of the distribution counts!!!
- In other words, the power law implies that
  - The probability to have elements very far from the average is not negligible
  - The big numbers counts
- Using an exponential distribution
  - The probability for a Webpage to have more than 100 incoming links, considering the average number of links for page, would be less in the order of 10^{-20}
  - which contradicts the fact that we know a lot of "well linked" sites…

Normal distribution of human heights

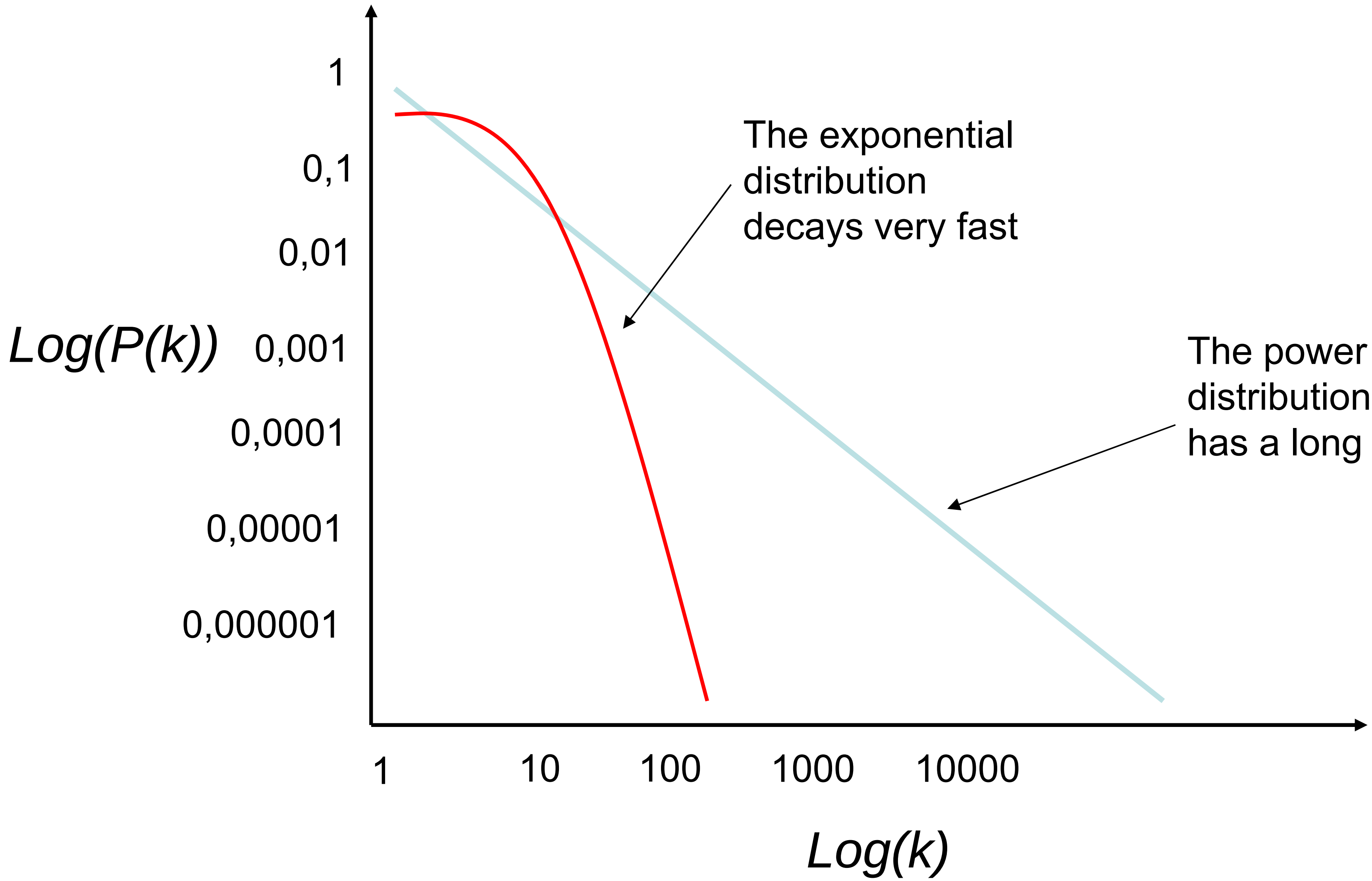


Power-law distribution

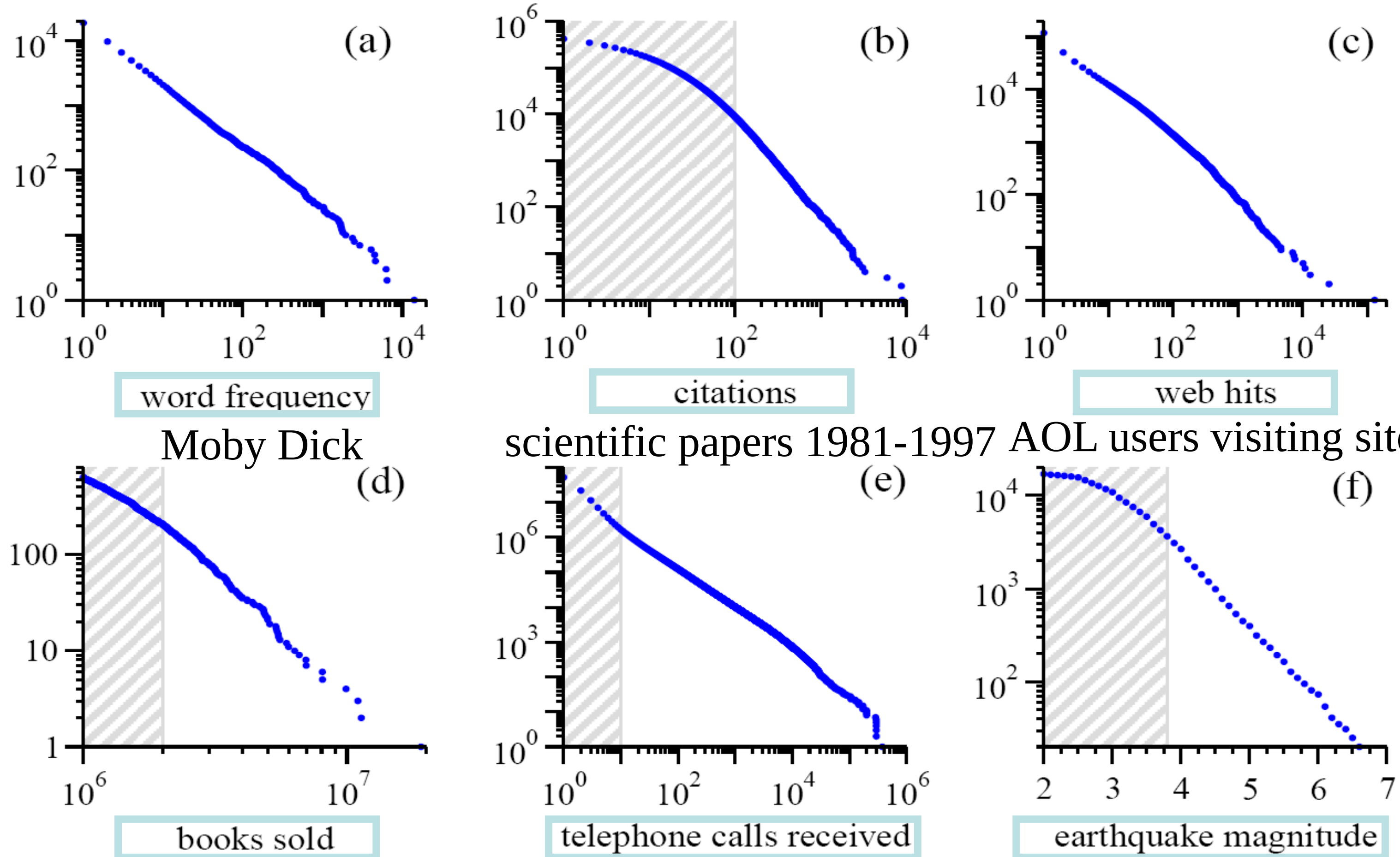


- high skew (asymmetry)
- straight line on a log-log plot

Power-law vs. Exponential distribution



Power laws everywhere

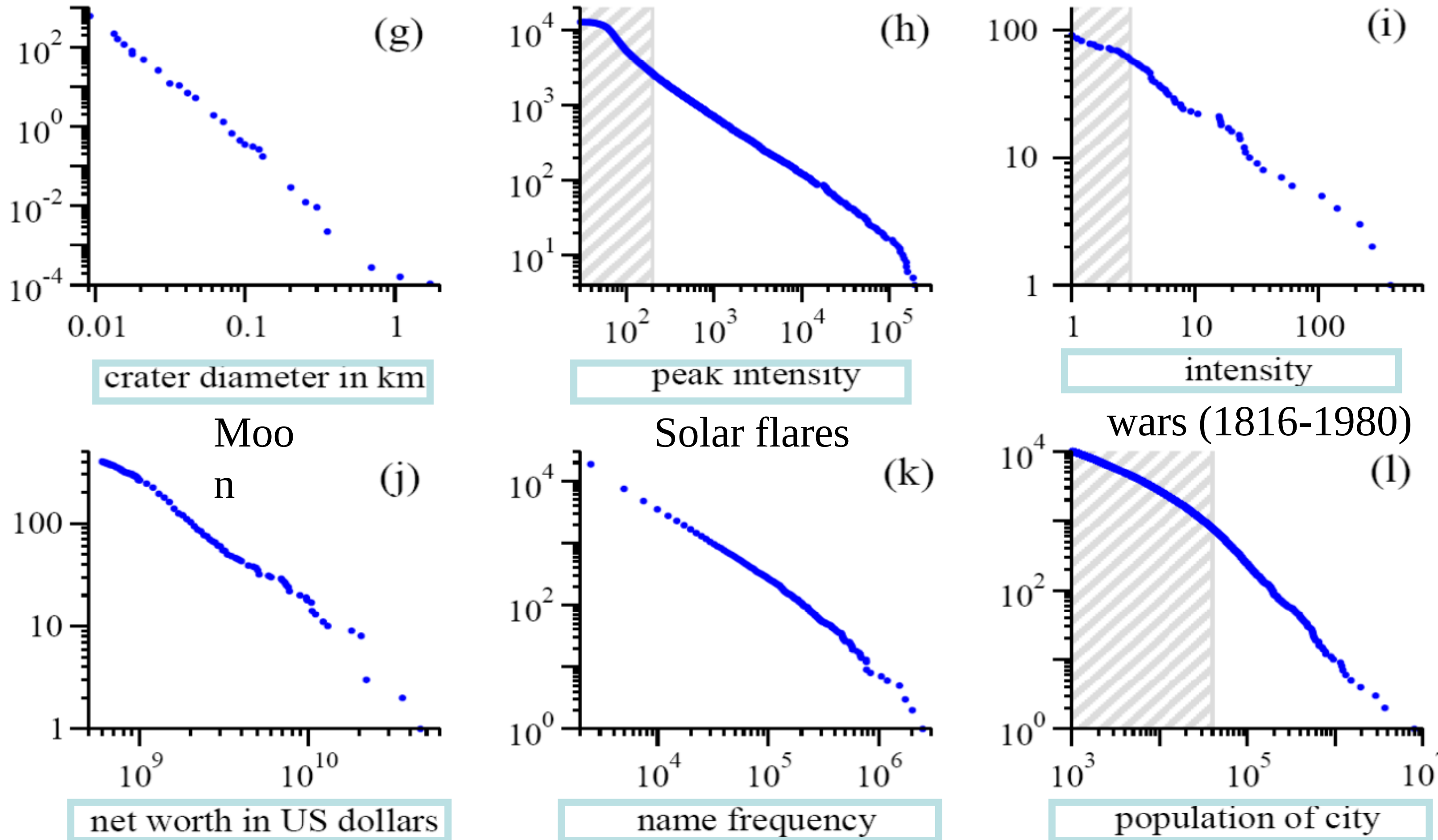


Moby Dick — scientific papers 1981-1997 — AOL users visiting sites '97

bestsellers 1895-1965 — AT&T customers on 1 day — California 1910-1992

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)

The Power-law in real networks

Average k

Power law exponents

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}	ℓ_{real}	ℓ_{rand}	ℓ_{pow}	Reference
WWW	325 729	4.51	900	2.45	2.1	11.2	8.32	4.77	Albert, Jeong, and Barabási 1999
WWW	4×10^7	7		2.38	2.1				Kumar *et al.*, 1999
WWW	2×10^8	7.5	4000	2.72	2.1	16	8.85	7.61	Broder *et al.*, 2000
WWW, site	260 000				1.94				Huberman and Adamic, 2000
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2	4	6.3	5.2	Faloutsos, 1999
Internet, router*	3888	2.57	30	2.48	2.48	12.15	8.75	7.67	Faloutsos, 1999
Internet, router*	150 000	2.66	60	2.4	2.4	11	12.8	7.47	Govindan, 2000
Movie actors*	212 250	28.78	900	2.3	2.3	4.54	3.65	4.01	Barabási and Albert, 1999
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2	4	2.12	1.95	Newman, 2001b
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1	6	5.01	3.86	Barabási *et al.*, 2001
Co-authors, math.*	70 975	3.9	120	2.5	2.5	9.5	8.2	6.53	Barabási *et al.*, 2001
Sexual contacts*	2810			3.4	3.4				Liljeros *et al.*, 2001
Metabolic, *E. coli*	778	7.4	110	2.2	2.2	3.2	3.32	2.89	Jeong *et al.*, 2000
Protein, *S. cerev.**	1870	2.39		2.4	2.4				Jeong, Mason, *et al.*, 2001
Ythan estuary*	134	8.7	35	1.05	1.05	2.43	2.26	1.71	Montoya and Solé, 2000
Silwood Park*	154	4.75	27	1.13	1.13	3.4	3.23	2	Montoya and Solé, 2000
Citation	783 339	8.57			3				Redner, 1998
Phone call	53×10^6	3.16		2.1	2.1				Aiello *et al.*, 2000
Words, co-occurrence*	460 902	70.13		2.7	2.7				Ferrer i Cancho and Solé, 2001
Words, synonyms*	22 311	13.48		2.8	2.8				Yook *et al.*, 2001b

The Ubiquity of the Power Law

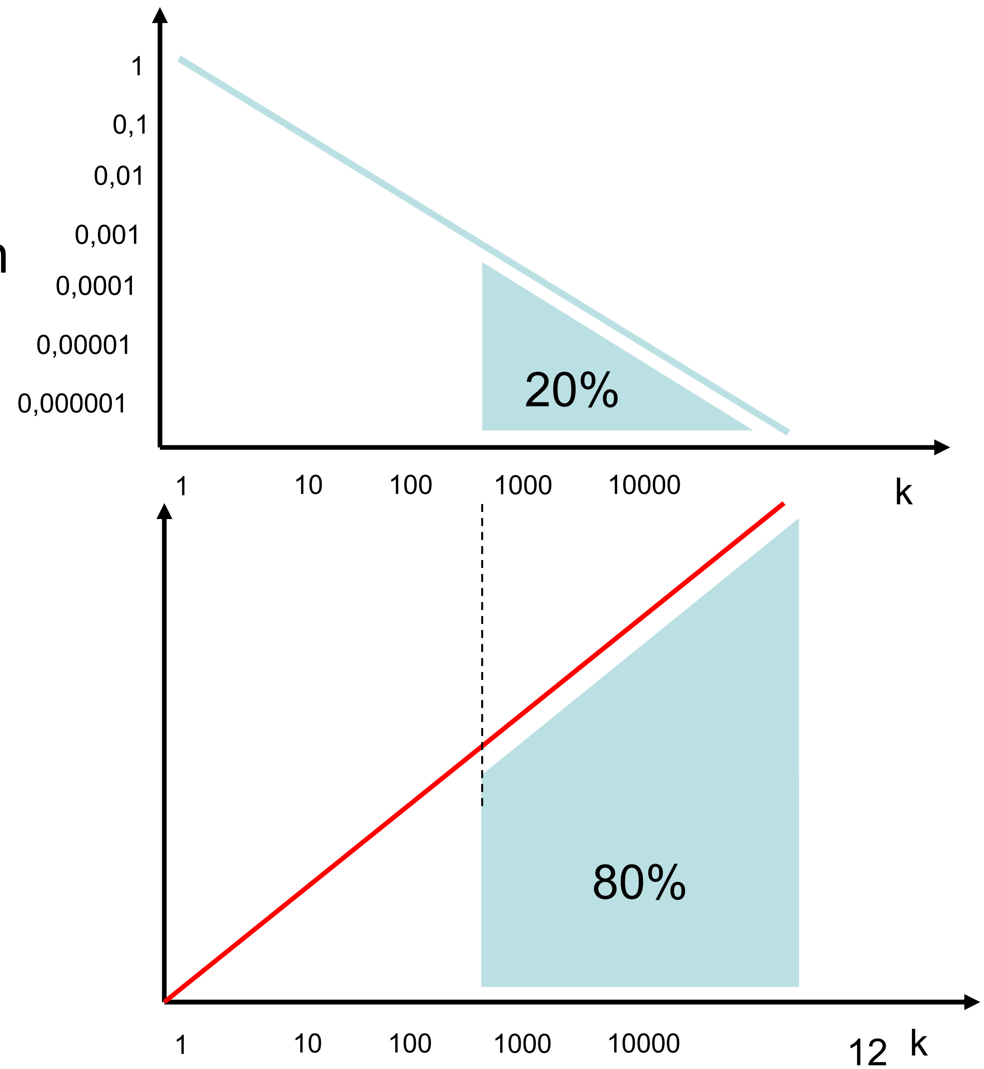
- The previous table includes not only technological networks
  - Most real systems and events have a probability distribution that does not follow the "normal" distribution
  - And obeys to a power law distribution
- Examples, in addition to technological and social networks
  - The distribution of size of files in file systems
  - The distribution of network latency in the Internet
  - The networks of protein interactions (a few protein exists that interact with a large number of other proteins)
  - The power of earthquakes: statistical data tell us that the power of earthquakes follow a power-law distribution
  - The size of rivers: the size of rivers in the world is power law
- The power law distribution is the "normal" distribution for complex systems (i.e., systems of interacting autonomous components)
  - We see later how it can be derived…

The 20-80 Rule

- It's a common "way of saying"
  - But it has scientific foundations
  - For all those systems that follow a power law distribution
- Examples
  - The 20% of the Web sites gests the 80% of the visits (actual data: 15%-85%)
  - The 20% of the Internet routers handles the 80% of the total Internet traffic
  - The 20% of world industries hold the 80% of the world's income
  - The 20% of the world population consumes the 80% of the world's resources
  - The 20% of the earthquakes caused the 80% of the victims
  - The 20% of the rivers in the world carry the 80% of the total sweet water
  - The 20% of the proteins handles the 80% of the most critical metabolic processes
- Does this derive from the power law distribution? YES!

The 20-80 Rule Unfolded

- The 20% of the population
  - Remember the area represents the amount of population in the distribution
- Get the 80% of the resources
  - In fact, it can be found that the “amount of resources” (i.e., the amount of links in the network) is the integral of *P(k)*k*, which is nearly linear

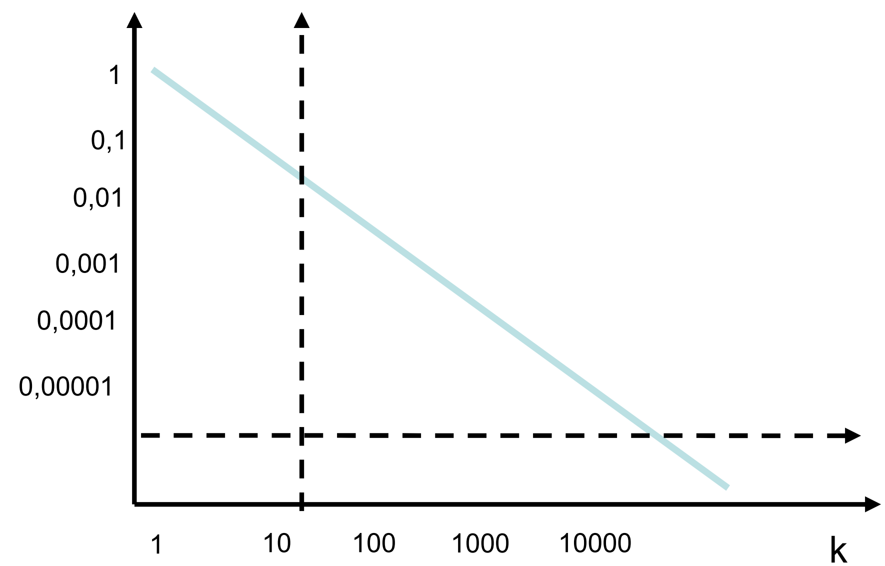


Hubs and Connectors

- Scale free networks exhibit the presence of nodes that
  - Act as hubs, i.e., as point to which most of the other nodes connects to
  - Act as connectors, i.e., nodes that make a great contributions in getting great portion of the network together
  - "smaller nodes" exists that act as hubs or connectors for local portion of the network
- This may have notable implications

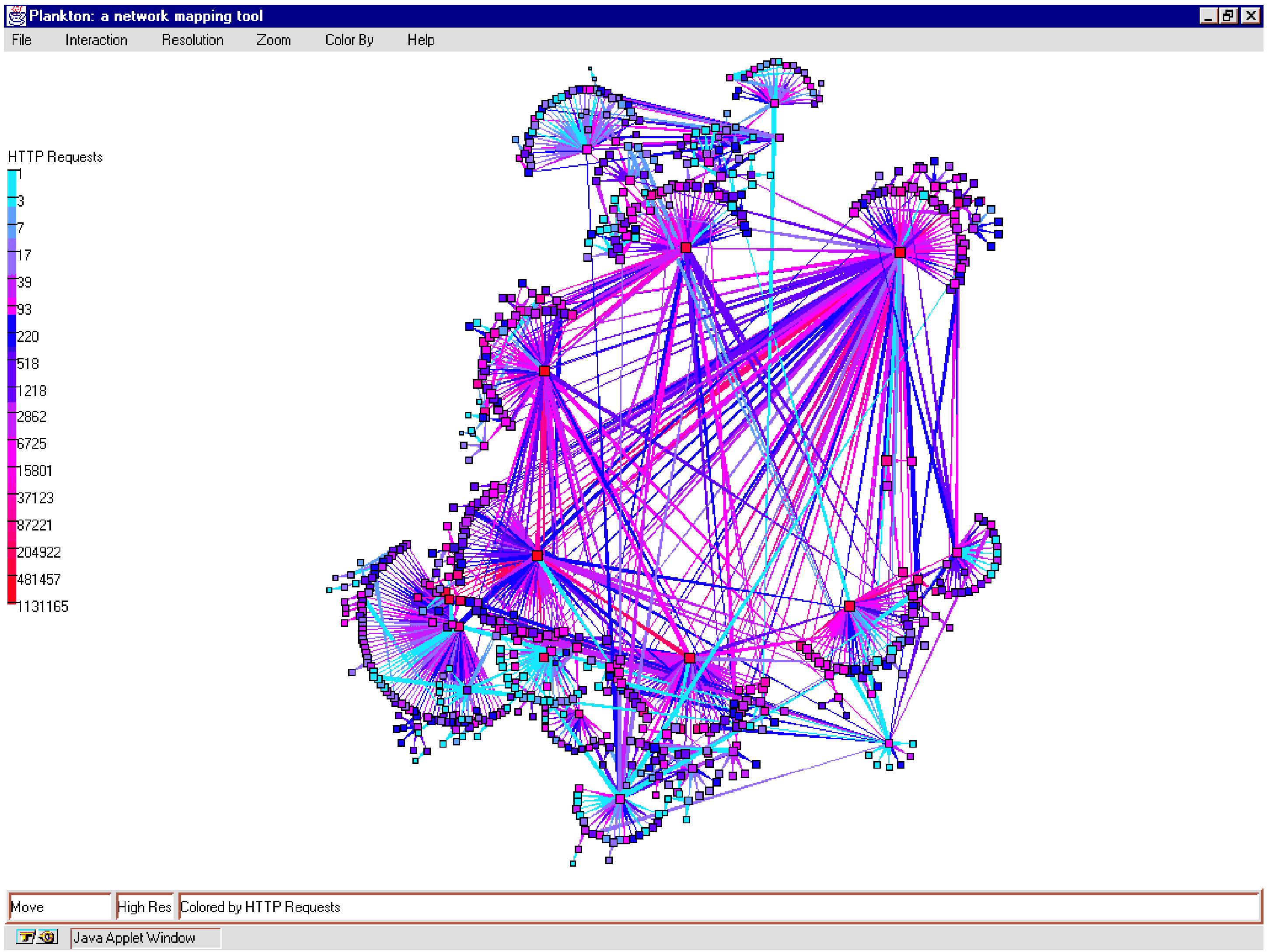
Why "Scale-Free" Networks

- Why networks following a power law distribution for links are called "**scale free**"?
  - Whatever the scale at which we observe the network
  - The network looks the same, i.e., it looks similar to itself
- The overall properties of the network are preserved independently of the scale
- In particular:
  - If we cut off the details of a network – skipping all nodes with a limited number of links – the network will preserve its power-law structure
  - If we consider a sub-portion of any network, it will have the same overall structure of the whole network



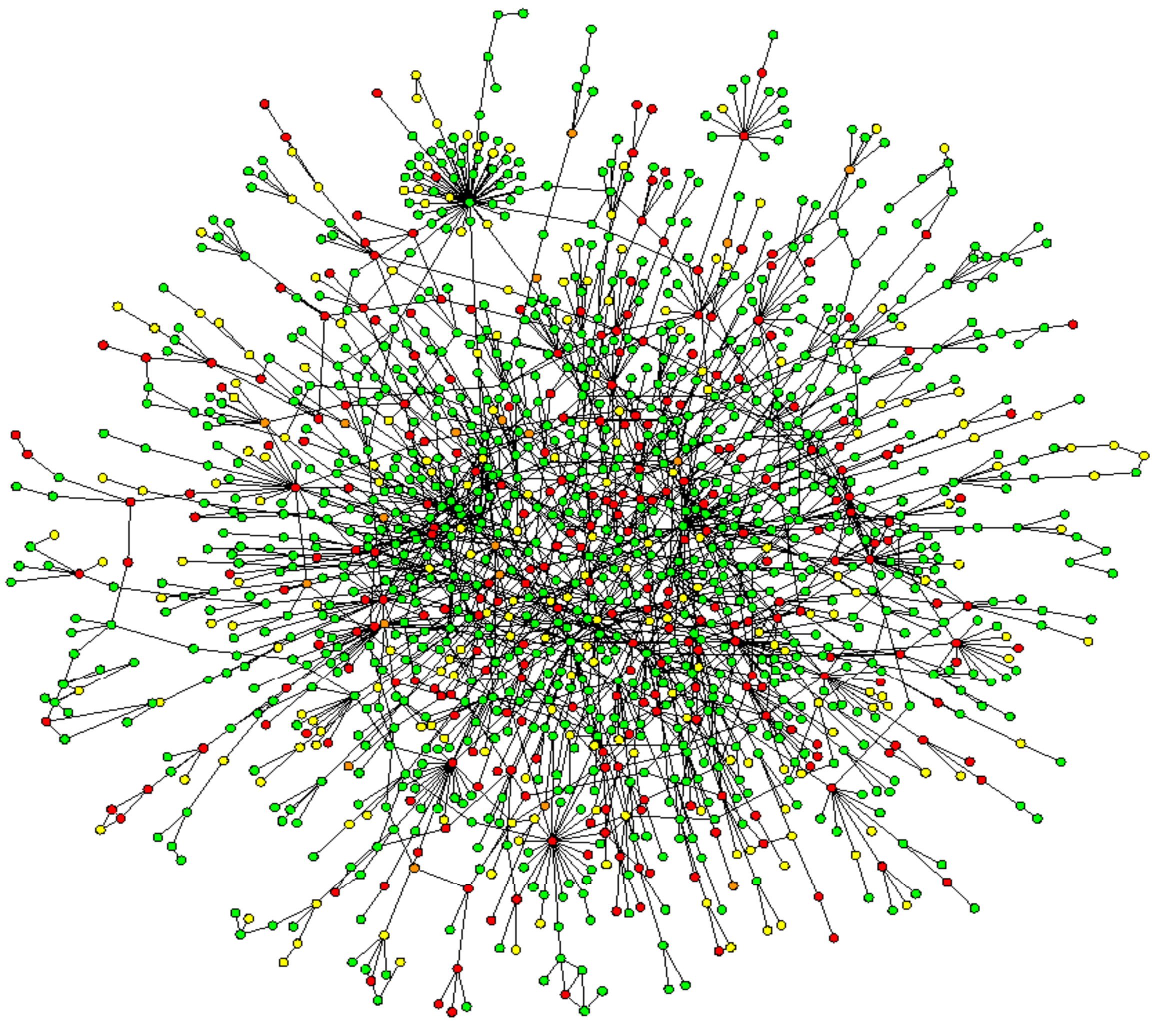
How do they look like?

Web Cache Network



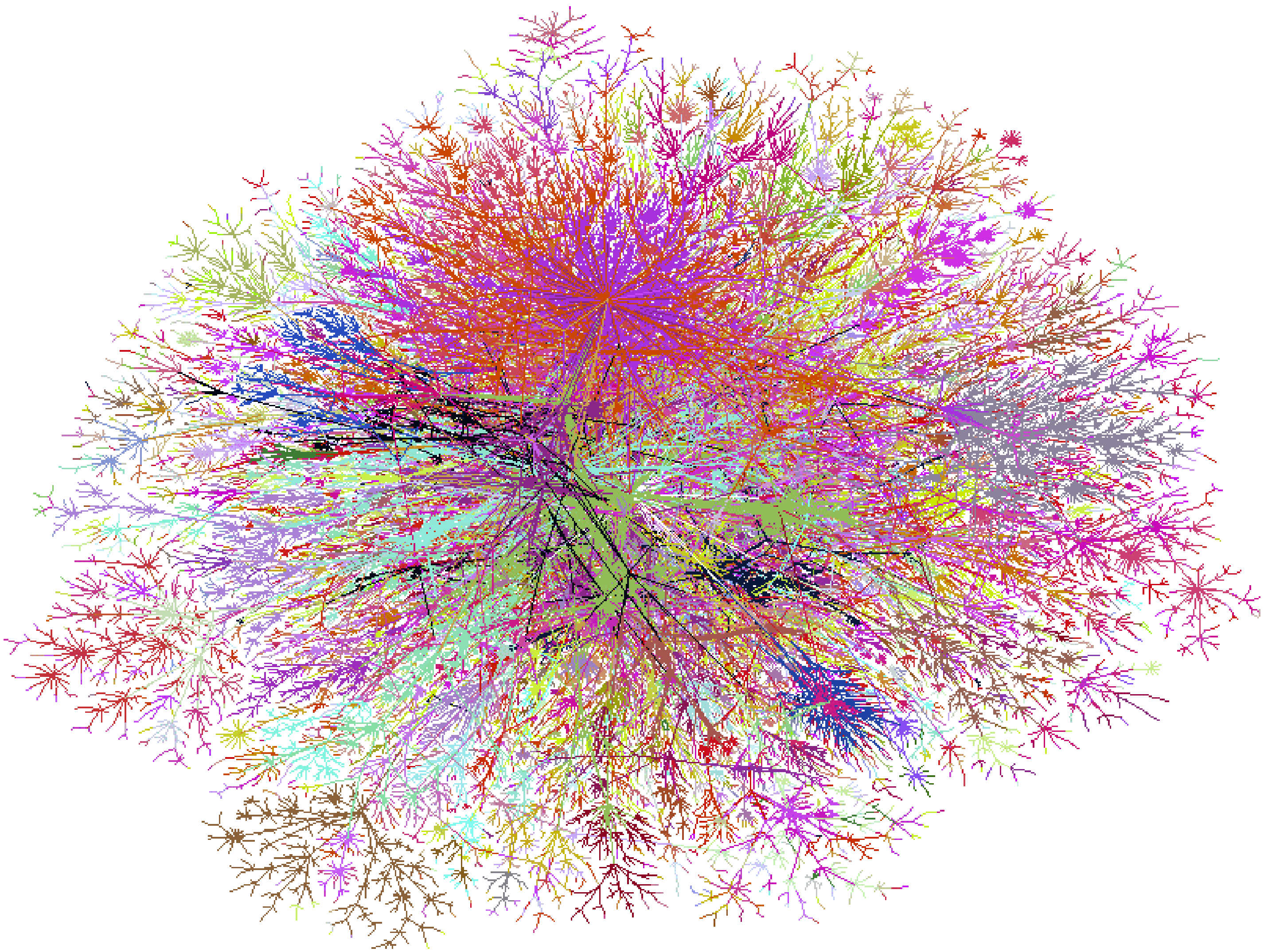
How do they look like?

Protein
Network



How do they look like?

The Internet
Routers



Power law distribution

- Straight line on a log-log plot

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

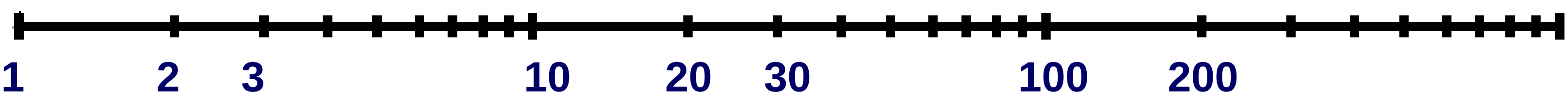
- Exponentiate both sides to get that *p(x)*, the probability of observing an item of size 'x' is given by

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

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

power law exponent α

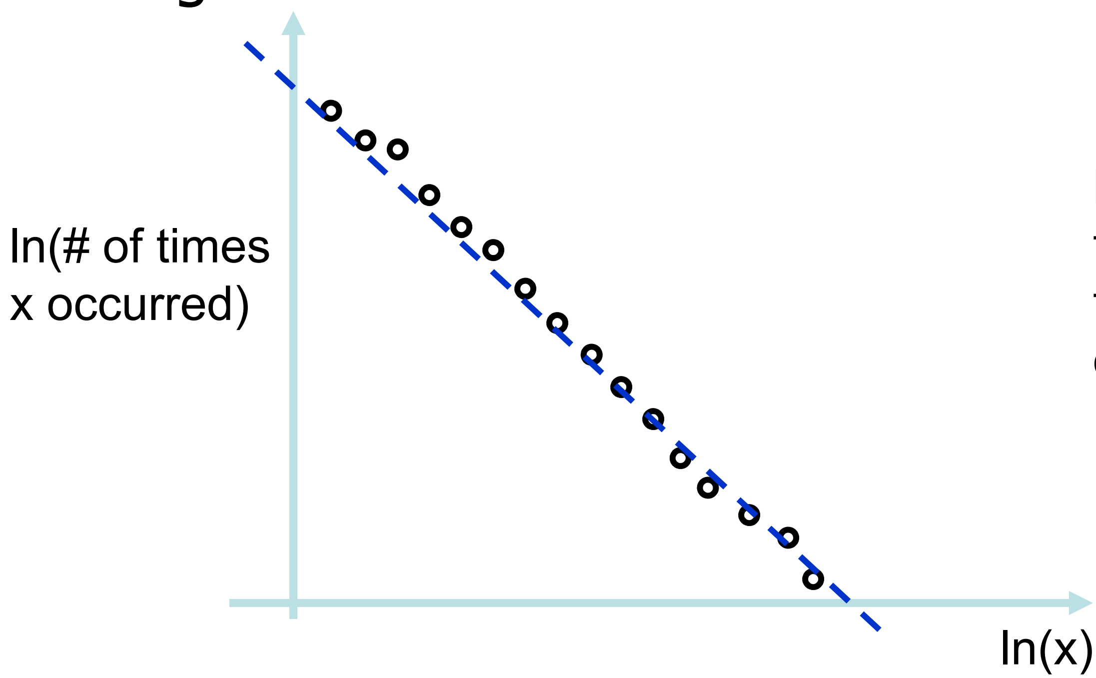
- powers of a number will be uniformly spaced



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

Fitting power-law distributions

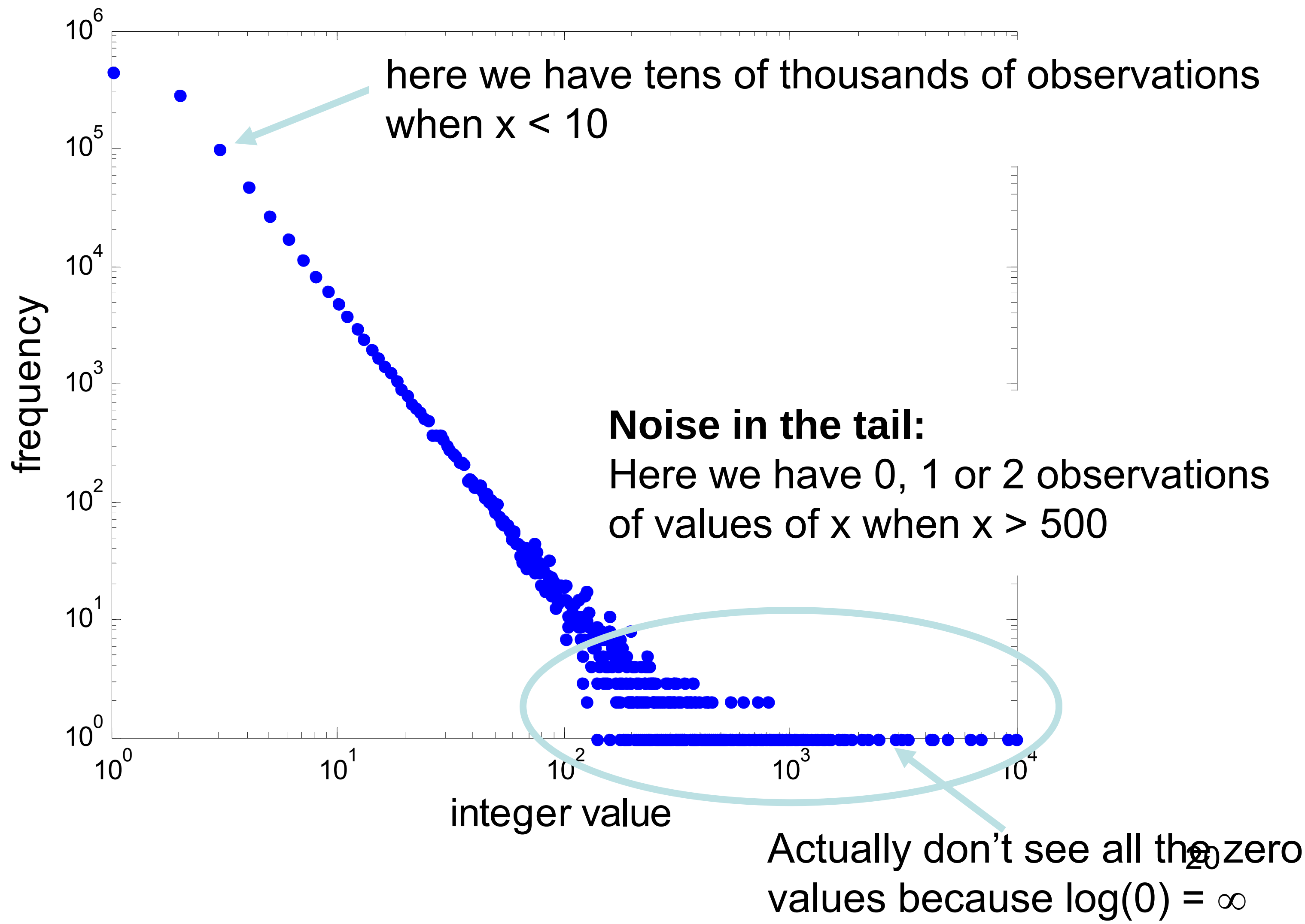
- Most common and not very accurate method:
  - Bin the different values of x and create a frequency histogram



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

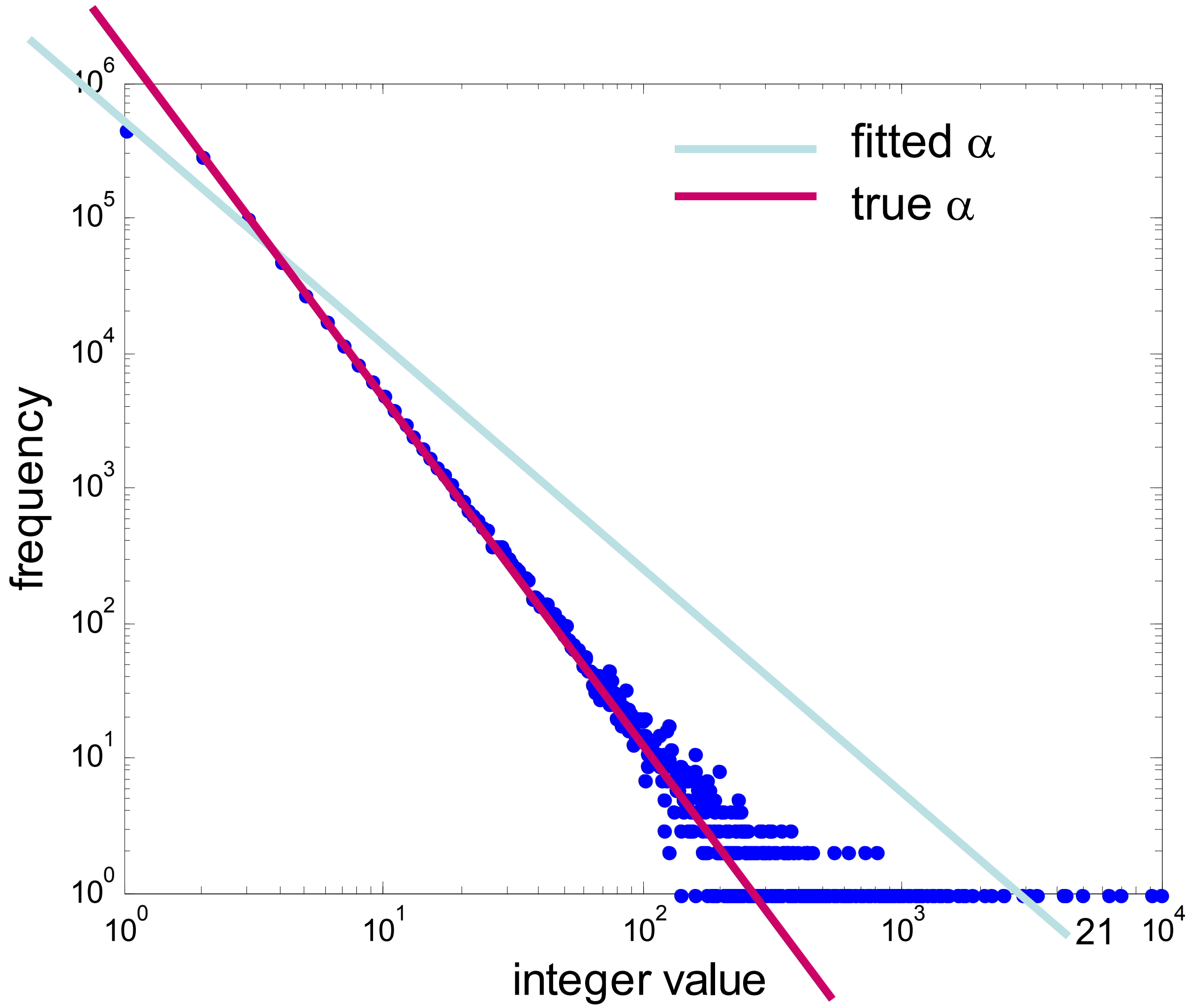
x can represent various quantities, the indegree of a node, the magnitude of an earthquake, the frequency of a word in text

Log-log scale plot of straight binning of the data



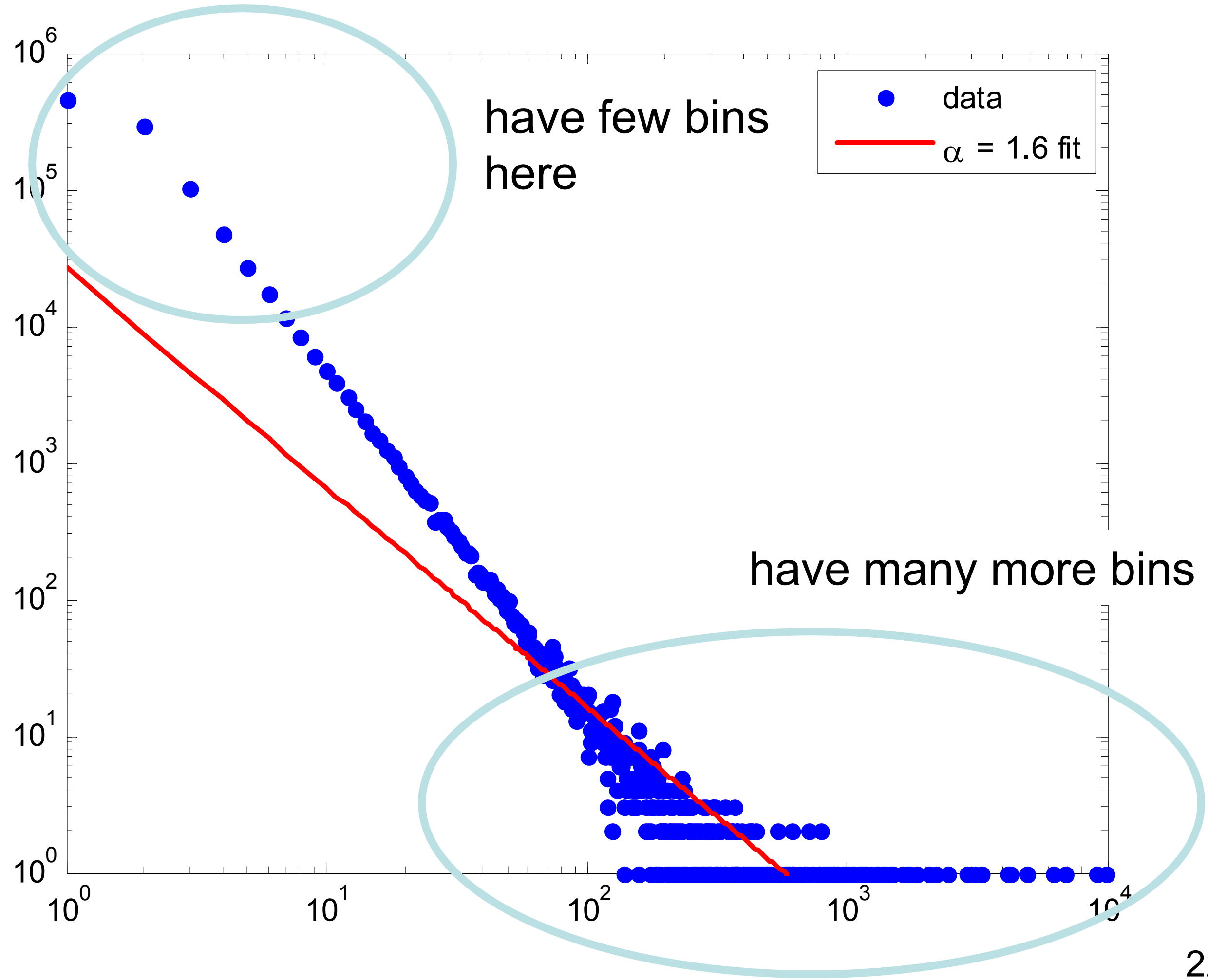
Log-log scale plot of straight binning of the data

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

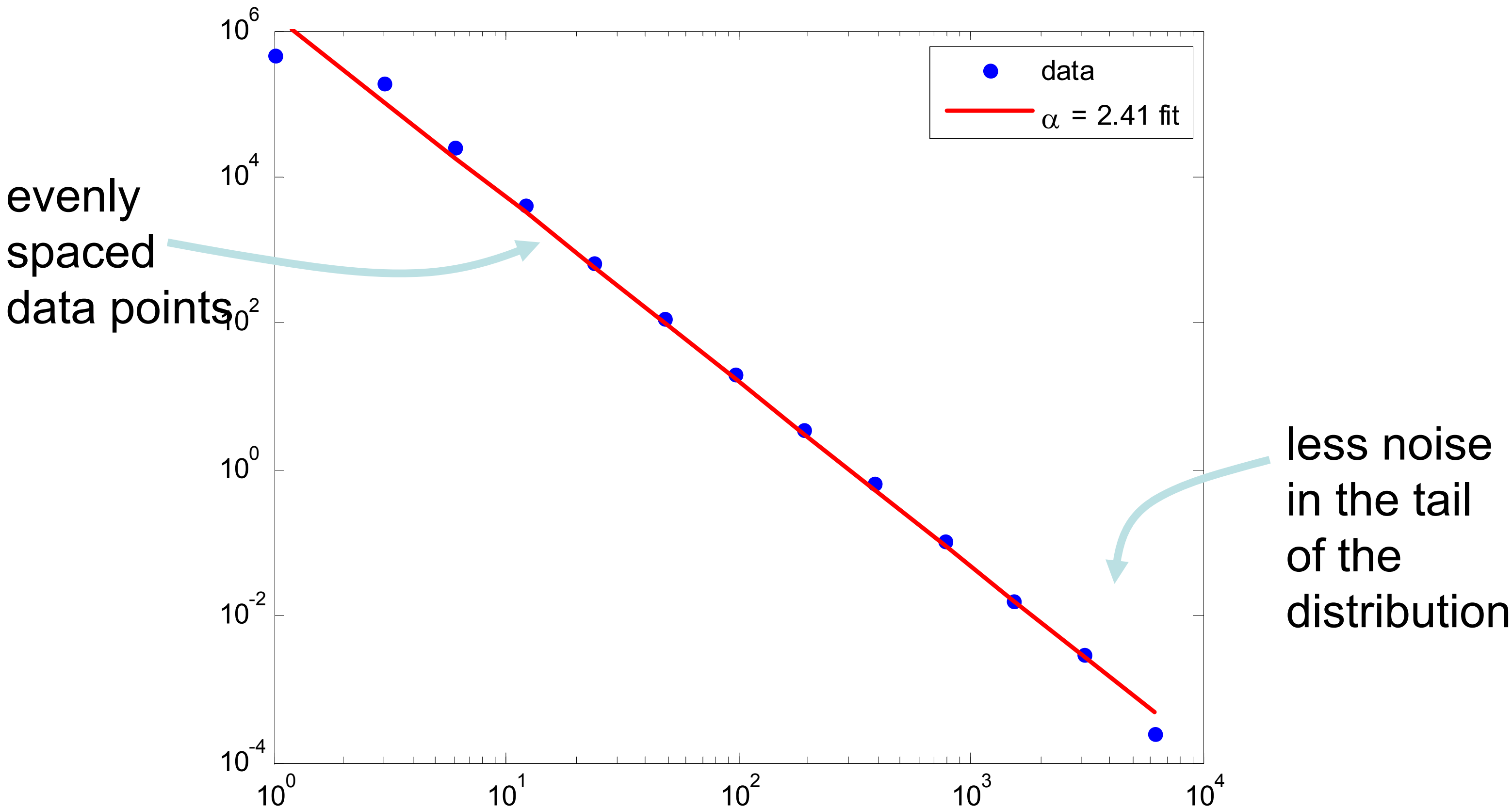


What goes wrong with straightforward binning

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
- 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**

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

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 where 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

Some exponents for real world data

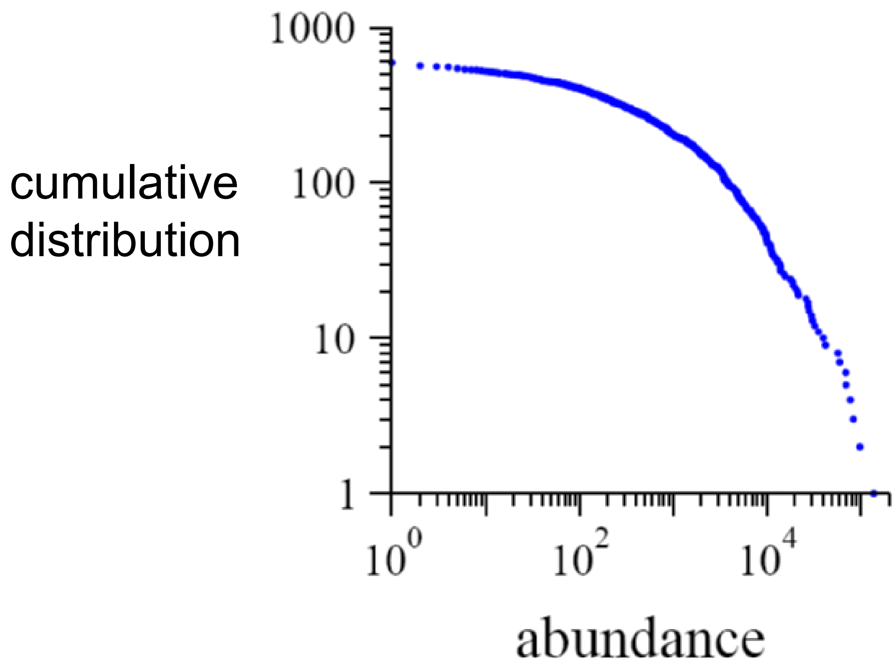
	x_{min}	exponent α
frequency of use of words	1	2.20
number of citations to papers	100	3.04
number of hits on web sites	1	2.40
copies of books sold in the US	2 000 000	3.51
telephone calls received	10	2.22
magnitude of earthquakes	3.8	3.04
diameter of moon craters	0.01	3.14
intensity of solar flares	200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

But, 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)

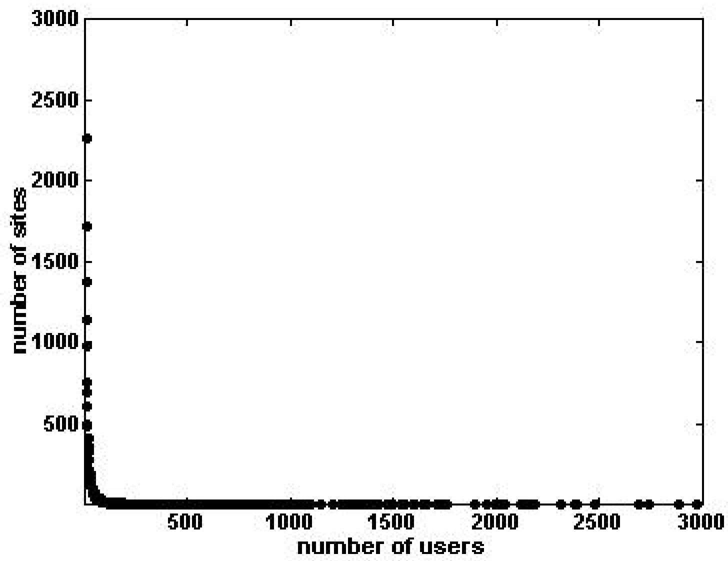
Source:MEJ Newman, 'Power laws, Pareto distributions and Zipf's law', *Contemporary Physics* **46**, 323–351 (2005)

Not every network is power law distributed

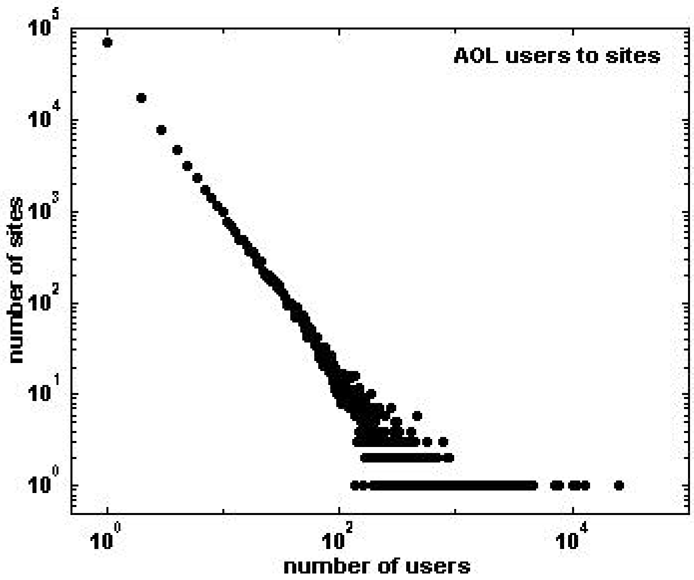
- reciprocal, frequent email communication
- power grid
- water distribution network
- company directors
- …

Example

number of AOL visitors to different websites back in 1997



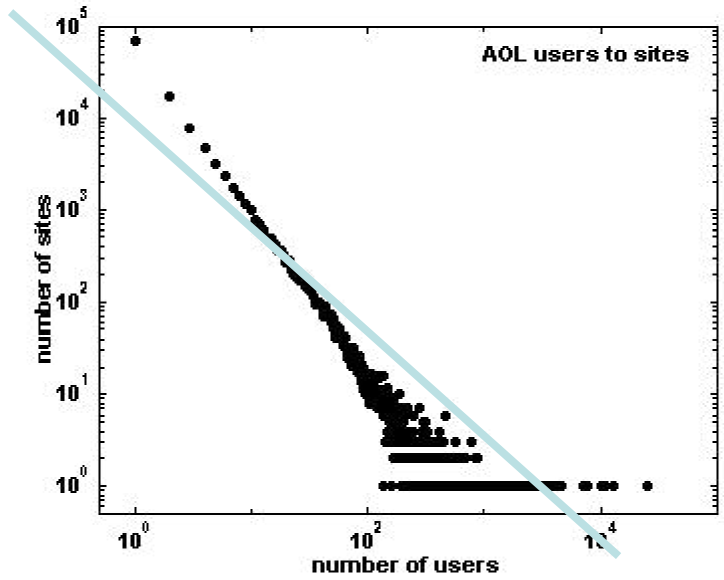
simple binning on a linear scale



simple binning on a log-log scale

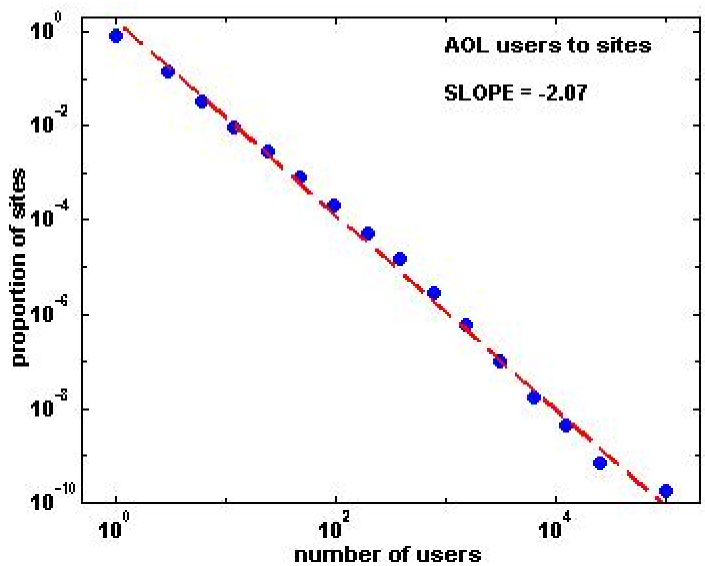
Trying to fit directly…

direct fit is too shallow: α = 1.17…



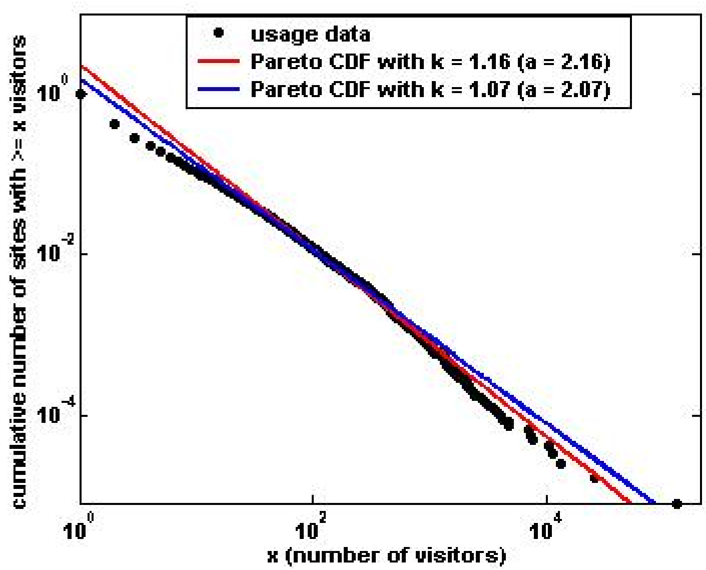
Binning the data logarithmically helps

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



Or we can try fitting the cumulative distribution

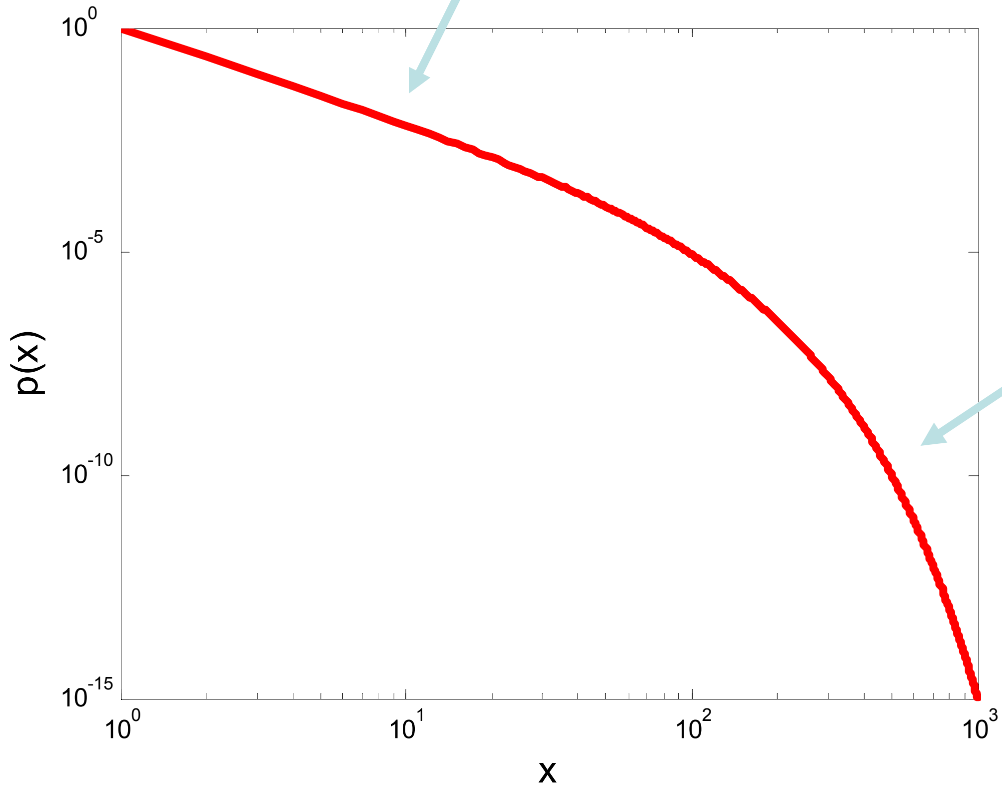
- Shows perhaps 2 separate power-law regimes that were obscured by the exponential binning
- Power-law tail may be closer to 2.4



Another common distribution: power-law with an exponential cutoff

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

starts out as a power law



ends up as an exponential

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:

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

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

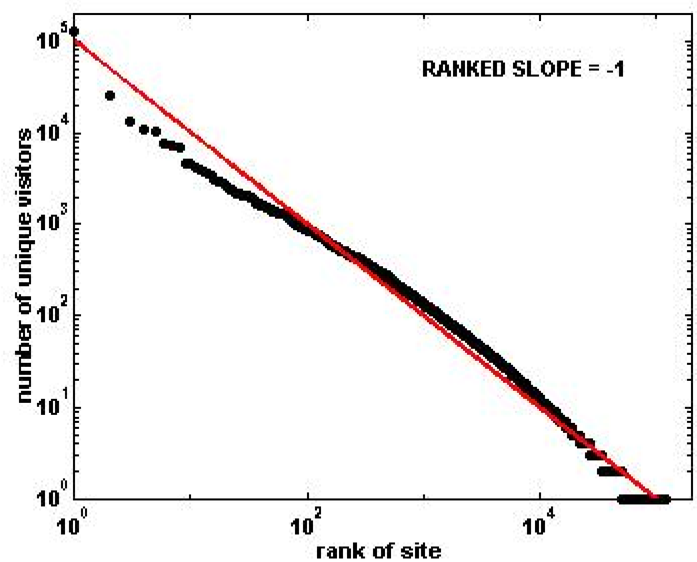
- Pareto
  - The Italian economist Vilfredo Pareto was interested in the distribution of income.
  - Pareto's law is expressed in terms of the cumulative distribution (the probability that a person earns X or more).

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

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

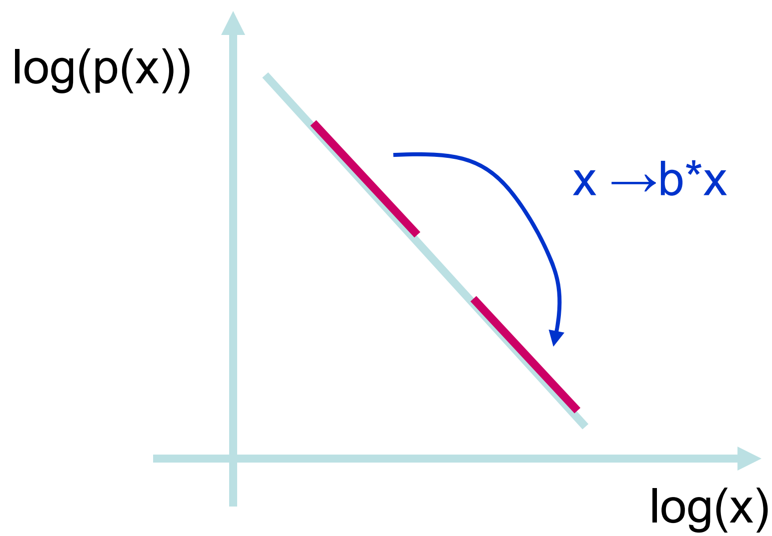
Zipf's law & AOL site visits

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



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
- $p(bx) = (bx)^{-\alpha} = b^{-\alpha}\, x^{-\alpha}$



Growing Networks

- In general, networks are not static entities
- They grow, with the continuous addition of new nodes
  - The Web, the Internet, acquaintances, the scientific literature, etc.
  - Thus, edges are added in a network with time
- The probability that a new node connects to another existing node may depend on the characteristics of the existing node
  - This is not simply a random process of independent node additions
  - But there could be “preferences” in adding an edge to a node
  - E.g.,. Google, a well known and reliable Internet router, a famous scientist,
  - Both of these could attract more link…

Evolving Networks

- More in general…
  - Networks grow AND
  - Network evolve
- The evolution may be driven by various forces
  - Connection age
  - Connection satisfaction
- What matters is that connections can change during the life of the network
  - Not necessarily in a random way
  - But following characteristics of the network
- Let’s start with the growing process …

Preferential Attachment in Networks

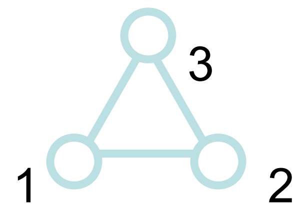
- Real-world networks are often power-law though
- First considered by [Price 65] as a model for citation networks
  - each new paper is generated with m citations (mean)
  - new papers cite previous papers with probability proportional to their indegree (citations)
  - what about papers without any 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

Barabasi-Albert model

- Undirected model: each node connects to other nodes with probability proportional to their degree
  - the process starts with some initial subgraph (m_0 all-all connected nodes)
  - each node comes with m edges
  - the probability of tipping the new node to the old ones is proportional to the degrees of old nodes
  - is a kind of preferential attachment algorithm
  - After t time steps, the network will have $n=t+m_0$ nodes and $M=m_0+mt$ edges
- It can be shown that this leads to a power law network!

Basic BA-model

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



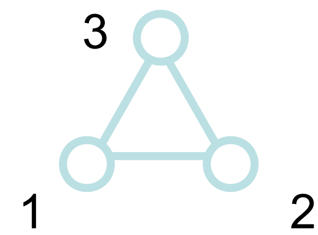
1 1 2 2 2 3 3 4 5 6 6 7 8 ….

  - 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

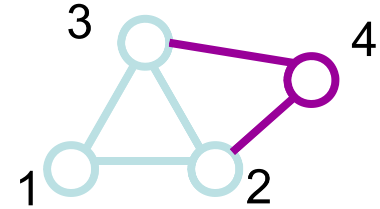
- To start, each vertex has an equal number of edges (2)
  - the probability of choosing any vertex is 1/3

1 1 2 2 3 3



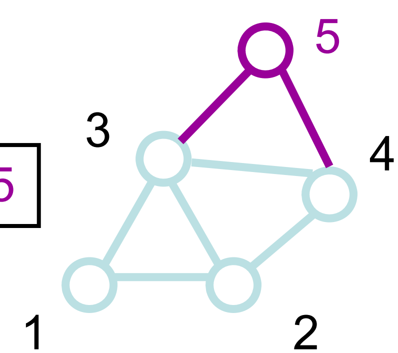
- We add a new vertex, and it will have m edges, here take m=2
  - draw 2 random elements from the array – suppose they are 2 and 3

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.

Proof of the scale-freeness

- Assume for simplicity that the degree k_i for any node i is a continuous variable
- The probability of the tipping a node to node i is

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

- Because of the assumptions, k_i is expected to grow proportionally to $\Pi(k_i)$, that is to its probability of having a new edge
- Consequently, and because m edges are attached at each time, k_i should obey the differential equation aside

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{n-1} k_j}$$

Proof of the scale-freeness

- The sum: $$\sum_{j=1}^{n-1} k_j$$
- Goes over all nodes except the new ones
- Thus, it results in: $$\sum_{j=1}^{n-1} k_j = 2mt - m$$
- Remember that the total number of edges is almost *mt* and that here an edge is counted twice
- Substituting in the differential equation

$$\frac{\partial k_i}{\partial t} = m \frac{k_i}{\sum_{j=1}^{n-1} k_j} = m \frac{k_i}{2mt - m} \approx \frac{k_i}{2t}$$

Proof of the scale-freeness

- We have now to solve this equation: $$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}$$
  - That is, we have to find a $k_i(t)$ function such as its derivative is equal to: itself, divided by 2t
- We now show this is: $$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}; \quad \text{with } \beta = \frac{1}{2}$$
- In fact:

$$\frac{\partial}{\partial t}\left(m\left(\frac{t}{t_i}\right)^{\beta}\right) = \frac{1}{2}\frac{m}{t_i^{\beta}}\frac{1}{t^{\beta}} = \frac{1}{2}\frac{m}{t_i^{\beta}}\frac{1}{t^{\beta}}\frac{t^{\beta}}{t^{\beta}} = \frac{m}{2}\frac{t^{\beta}}{t_i^{\beta}}\frac{1}{t^{2\beta}} = \frac{k_i(t)}{2t}$$

- where we also consider the initial condition $k_i(t_i)=m$, where t_i is the time at which node i has arrived

Proof of the scale-freeness

- The $k_i(t)$ function that we have calculated shows that the degree of each node grown with a power law with time
- Now, let's calculate the probability that a node has a degree $k_i(t)$ smaller than k
- We have:

$$P[k_i(t) < k] = P\left[m \frac{t^{\beta}}{t_i^{\beta}} < k \right] = P\left[m^{\frac{1}{\beta}} \frac{t^{\beta\frac{1}{\beta}}}{t_i^{\beta\frac{1}{\beta}}} < k^{\frac{1}{\beta}} \right] =$$

$$= P\left[m^{\frac{1}{\beta}} \frac{t}{t_i} < k^{\frac{1}{\beta}} \right] = P\left[t_i > \frac{m^{\frac{1}{\beta}} t}{k^{\frac{1}{\beta}}} \right]$$

Proof of the scale-freeness

- Now let's remember that we add nodes at each time interval
- Therefore, the probability t_j for a node, that is the probability for a node to have arrived at time t_i is a constant and is:

$$P(t_i) = \frac{1}{t + m_0}$$

- Substituting this into the previous probability distribution:

$$P[k_i(t) < k] = P\left[t_i > \frac{m^{\frac{1}{\beta}} t}{k^{\frac{1}{\beta}}} \right] = 1 - P\left[t_i \leq \frac{m^{\frac{1}{\beta}} t}{k^{\frac{1}{\beta}}} \right] = 1 - \frac{m^{\frac{1}{\beta}} t}{k^{\frac{1}{\beta}} (t + m_0)}$$

Proof of the scale-freeness

- Now given the probability distribution:

$$P\left[k_i(t) < k\right]$$

- Which represents the probability that a node i has less than k link

$$P(k) = \frac{\partial P\left[k_i(t) < k\right]}{\partial k}$$

- The probability that a node has exactly k link can be derived by the derivative of the probability distribution

$$P(k) = \frac{\partial P\left[k_i(t) < k\right]}{\partial k} = \frac{\partial}{\partial k}\left(1 - \frac{m^{\frac{1}{\beta}} t}{k^{\frac{1}{\beta}}(t + m_0)}\right) = \frac{2m^{\frac{1}{\beta}} t}{m_0 + t} \frac{1}{k^{\frac{1}{\beta}+1}}$$

Conclusion of the Proof

- Given P(k):

$$P(k) = \frac{2m^{\frac{1}{\beta}}t}{m_0 + t}\frac{1}{k^{\frac{1}{\beta}+1}}$$

- After a while, that is for t→∞

$$P(k) \approx 2m^{\frac{1}{\beta}}k^{-\frac{1}{\beta}-1} = 2m^{\frac{1}{\beta}}k^{-\gamma} \quad \text{where } \gamma = \frac{1}{\beta}+1 = 3$$

- That is, **we have obtained a power law probability density**, with an exponent which is independent of any parameter (being the only initial parameter m)

Probability Density for a Random Network

- In a random network model, each new node that attach to the network attachs its edges independently of the current situation
  - Thus, all the events are independent
- The probability for a node to have a certain number of edges attached is thus a "normal", exponential, distribution
- It can be easily found, using standard statistical methods that:

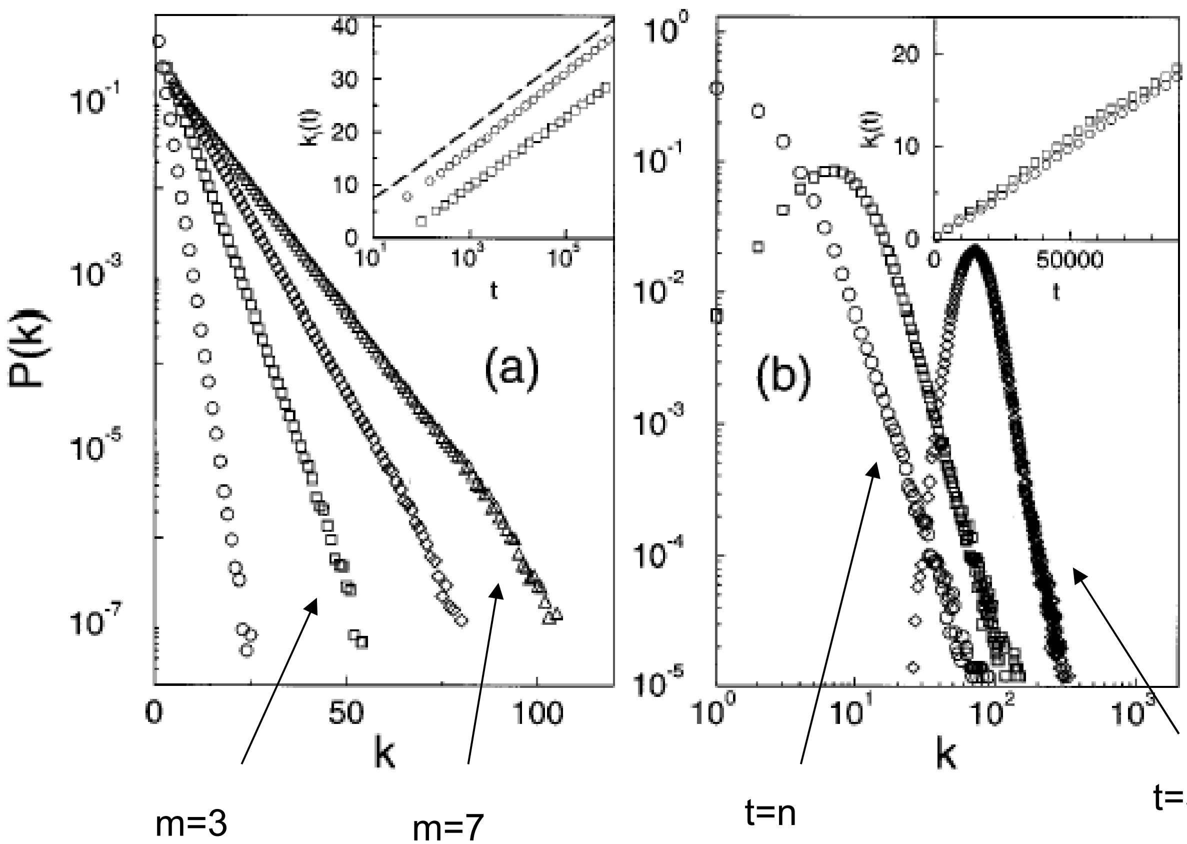
$$P(k) = \frac{1}{m} e^{-\frac{k}{m}}$$

BA Model vs. Random Networks

- See the difference for the evolution of the Barabasi-Albert model vs. the Random Network model (from Barabasi and Albert, Reviews of Modern Physics 2002)

Barabasi-Albert Model
n=800000

Simulations performed with various values of m



Random network model for n=10000

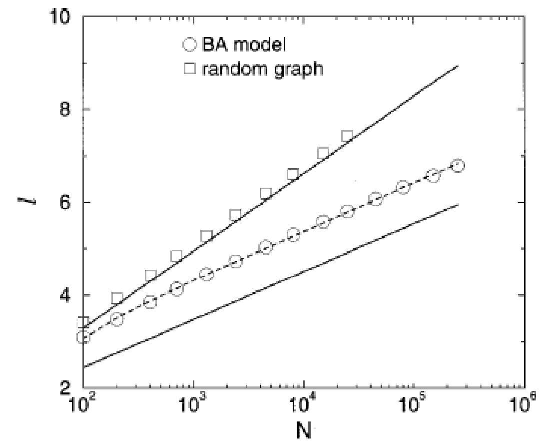
The degree distribution gradually becomes a normal one with passing time

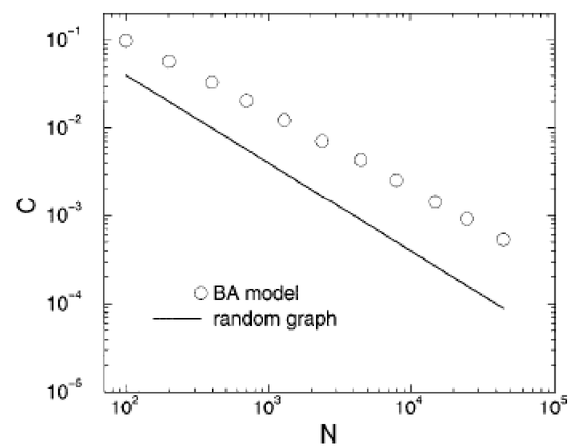
Generality of the BA Model

- In its simplicity, the BA model captures the essential characteristics of a number of phenomena
  - In which events determining "size" of the individuals in a network are not independent from each other
  - Leading to a power law distribution
- So, it can somewhat explain why the power law distribution is as ubiquitous as the normal Gaussian distribution
- Examples
  - **Gnutella (the first decentralized P2P network)**: a peer which has been there for a long time, has already collected a strong list of acquaintances, so that any new node has higher probability of getting aware of it
  - **Rivers**: the eldest and biggest a river, the more it has probability to break the path of a new river and get its water, thus becoming even bigger
  - **Industries**: the biggest an industry, the more its capability to attract clients and thus become even bigger
  - **Earthquakes**: big stresses in the earth plaques can absorb the effects of small earthquakes, this increasing the stress further. A stress that will eventually end up in a dramatic earthquakes
  - **Richness**: the rich I am, the more I can exploit my money to make new money → "RICH GET RICHER"

Additional Properties of the BA Model

- Characteristic Path Length
  - It can be shown (but it is difficult) that the BA model has a length proportional to log(n)/log(log(n))
  - Which is even shorter than in random networks
  - And which is often in accord with – but sometimes underestimates –experimental data
- Clustering
  - There are no analytical results available
  - Simulations shows that in scale-free networks the clustering decreases with the increases of the network order
  - As in random graph, although a bit less
  - This is not in accord with experimental data!



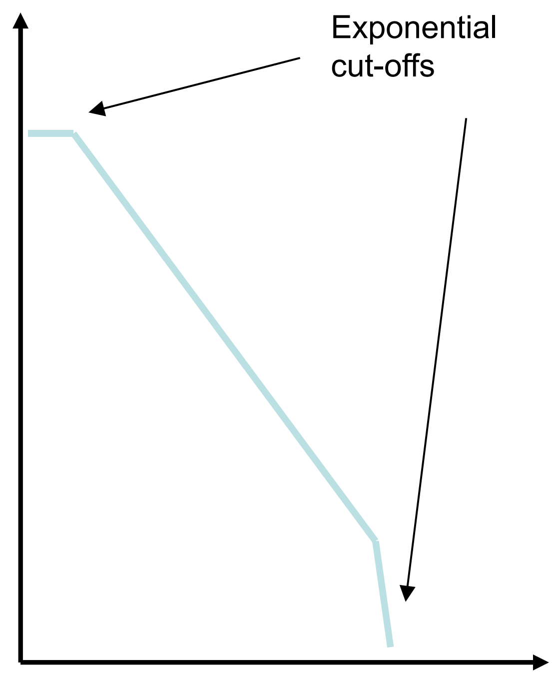


Problems of the BA Model

- The BA model is a nice one, but is not fully satisfactory!
- The BA model does not give satisfactory answers with regard to clustering
  - While the small world model of Watts and Strogatz does!
  - So, there must be something wrong with the model..
- The BA model predicts a fixed exponent of 3 for the power law
  - However, real networks shows exponents between 1 and 3
  - So, there most be something wrong with the model

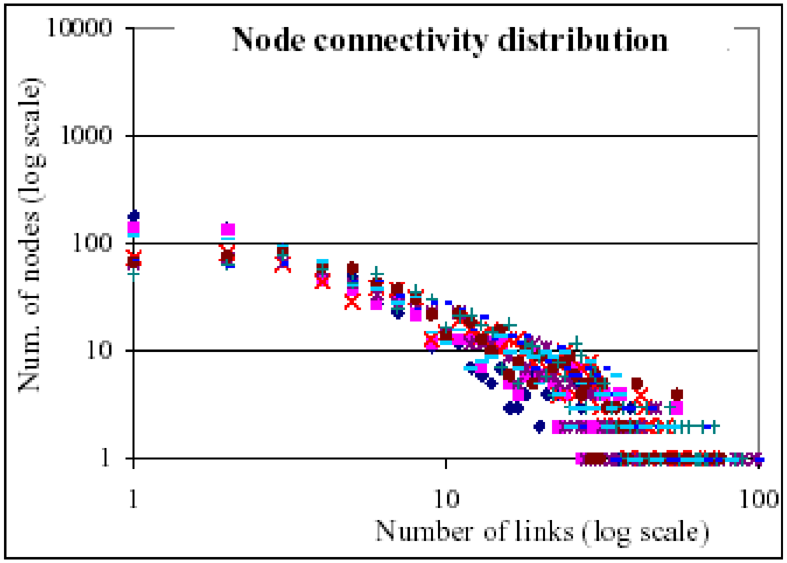
Problems of the BA Model

- An additional problem is that real networks are not "completely" power law
  - They exhibit a so-called ***exponential cut-off***
  - After having obeyed the power-law for a large amount of k
  - For very large k, the distribution suddenly becomes exponential
- In general
  - The distribution has still a "heavy tailed" compared to standard exponential distribution
  - However, such tail is not infinite
- This can be explained because
  - The number of resources (i.e., of links) that an individual (i.e., a node) can sustain (i.e., can properly handled) is often limited
  - So, there can be no individual that can sustain any large number of resources
  - Vice versa, there could be a minimal amount of resources a node can have
- The Barabasi-Albert model does not predict this



Exponential Cut-offs in Gnutella

- Gnutella is a network with exponential cut-offs
- That can be easily explained
  - A node cannot connect to the network without having a minimal number of connections
  - A node cannot sustain an excessive number of TCP connections



Variations on the BA Model: Non-linear Preferential Attachments

- One can consider non-linear models for preferential attachment
  - E.g. $\Pi(k) \propto k^{\alpha}$
- However, it can be shown that these models destroy the power-law nature of the network

Variations on the BA Model: Evolving Networks

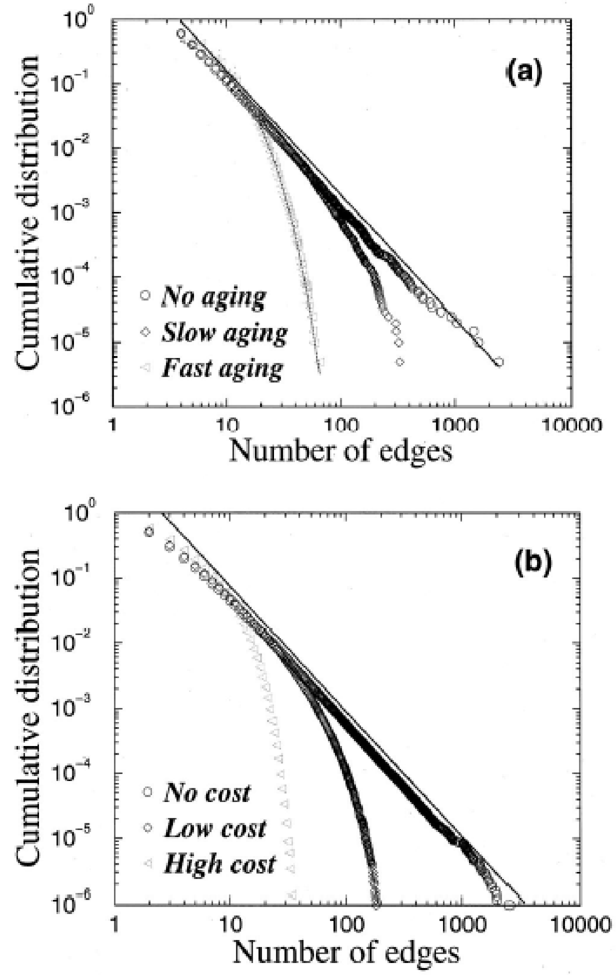
- The problems of the BA Model may depend on the fact that networks not only grow but also evolve
  - The BA model does not account for evolutions following the growth
- Which may be indeed frequent in real networks, otherwise
  - Google would have never replaced Altavista
  - All new Routers in the Internet would be unimportant ones
  - A Scientist would have never the chance of becoming a highly-cited one
- A sound theory of evolving networks is still missing
  - Still, we can start from the BA model and adapt it to somehow account for network evolution
  - And obtain a bit more realistic model

Variations on the BA Model: Edges Rewiring

- By coupling the model for node additions
  - Adding new nodes at new time interval
- One can consider also mechanisms for edge rewiring
  - E.g., adding some edges at each time interval
  - Some of these can be added randomly
  - Some of these can be added based on preferential attachment
- Then, it is possible to show (Albert and Barabasi, 2000)
  - That the network evolves as a power law with an exponent that can vary between 2 and infinity
  - This enables explaining the various exponents that are measured in real networks

Variations on the BA Model: Aging and Cost

- One can consider in real networks (Amaral et al., 2000):
- Node Aging
  - The possibility of hosting new links decreased with the "age" of the node
  - E.g. nodes get tired or out-of-date
- Link cost
  - The cost of hosting new link increases with the number of links
  - E.g., for a Web site this implies adding more computational power, for a router this means buying a new powerful router
- These two models explain the "exponential cut-off" in power law networks



Variations on the BA Model: Fitness

- One can also consider in real networks:
- Not all nodes are equal, but some nodes "fit" better specific network characteristics
  - E.g. Google has a more effective algorithm for page-indexing and ranking
  - A new scientific paper may be indeed a breakthrough
- In terms of preferential attachment, this implies that
  - The probability for a node of attracting links is proportional to some fitness parameter μ_i
  - See the formula below
- It can be shown that the fitness model for preferential attachment enables even very young nodes to attract a lot of links

$$\Pi(k_i) = \frac{\mu_i k_i}{\sum_j \mu_j k_j}$$

Evolving networks

- dynamic appearance/disappearance of individual nodes and links
  - new links (university email network over time)
  - team assembly (coauthor & collaborator networks)
  - evolution of affiliation network related to social networks (online groups, CS conferences)
- evolution of aggregate metrics:
  - densification & shrinking diameters (internet, citation, authorship, patents)
  - models:
    - community structure
    - forest file model
- What events can occur to change a network over time?
- What properties do you expect to remain roughly constant?
- What properties do you expect to change?

Empirical analysis of an evolving social network

- Gueorgi Kossinets & Duncan J. Watts
  - Science, Jan. 6th, 2006
- The data
  - university email logs
  - sender, recipient, timestamp
    - no content
  - 43,553 undergraduate and graduate students, faculty, staff
  - filtered out messages with more than 4 recipients (5% of messages)
  - 14,584,423 messages remaining sent over a period of 355 days (2003-2004 school year)

How does one choose new acquaintances?

- triadic closure: choose a friend of friend
- homophily: choose someone with similar interests
- proximity: choose someone who is close spatially and with whom you spend a lot of time
- seek novel information and resources
  - connect outside of circle of acquaintances
  - span structural holes between people who don't know each other
- sometimes social ties also dissolve
  - avoid conflicting relationships
  - reason for tie is removed: common interest, activity

Weighted ties

$$w_{ij}(t,\tau) = \sqrt{m_{ij} m_{ji}} / \tau$$

- w_{ij} = weight of the tie between individuals i and j
- m = # of messages from i to j in the time period between (t-τ) and t
- "geometric rate" – because rates are multiplied together
  - high if email is reciprocated
  - low if mostly one-way
- τ serves as a relevancy horizon (30 days, 60 days…)
- 60 days chosen as window in study because rate of tie formation stabilizes after 60 days
- sliding window: compare networks day by day (but each day represents an overlapping 60 day window)

Cyclic closure & focal closure

shortest path distance between i and j

$$P_{new}(d_{ij}, s_{ij}) = \sum_{t=61}^{270} M_{new}(d_{ij}, s_{ij}, t) \,/\, \sum_{t=61}^{270} M(d_{ij}, s_{ij}, t)$$

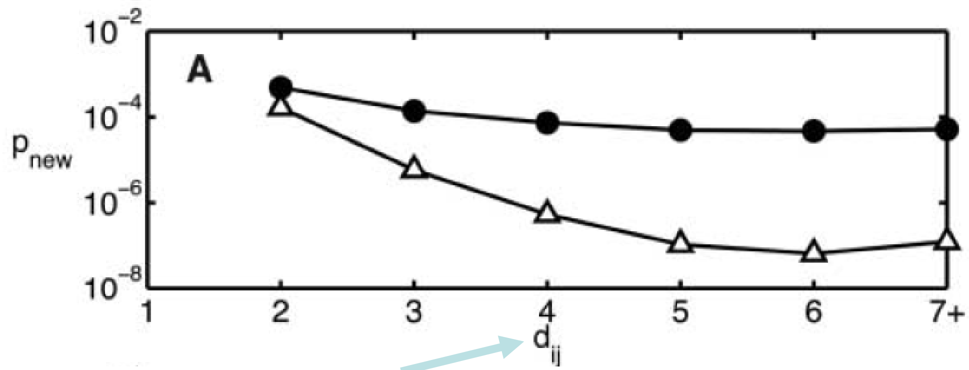
new ties that appeared on day t

ties that were there in the past 60 days

number of common foci, i.e. classes

● pairs that attend one or more classes together

△ do not attend classes together



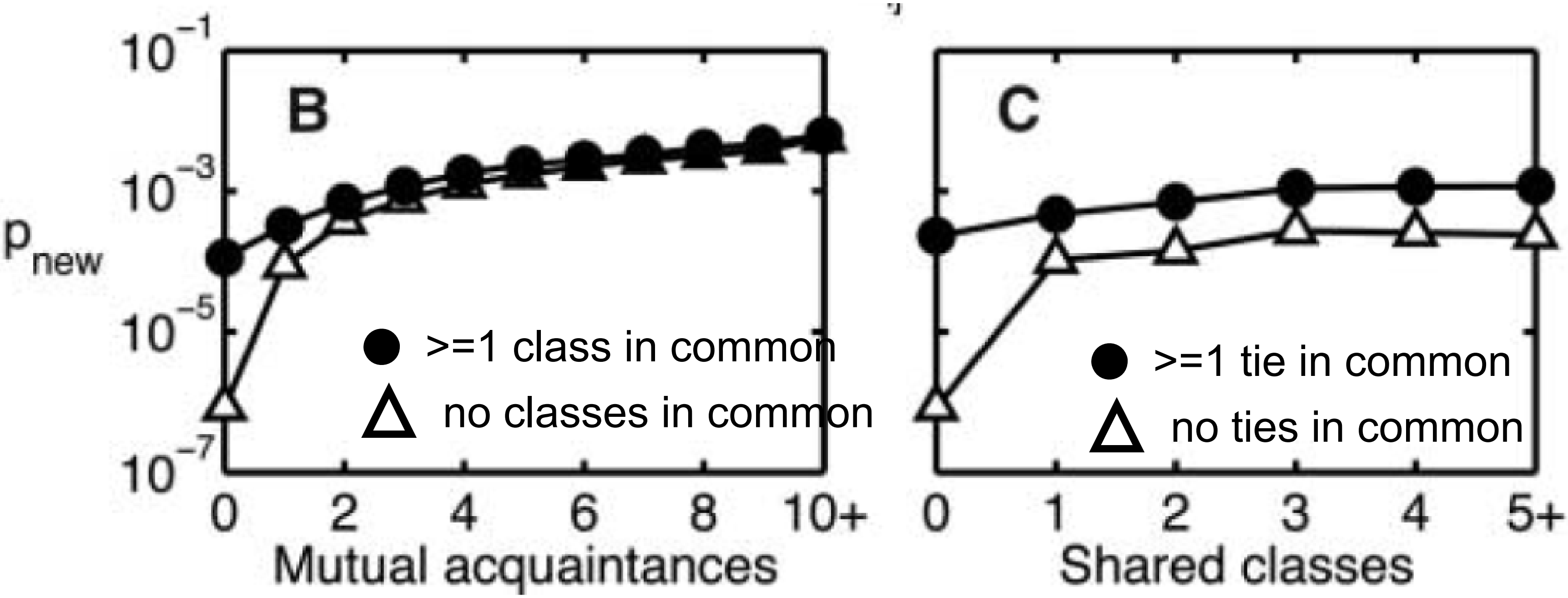
distance between two people in the email graph

- Individuals who share at least one class are *three times* more likely to start emailing each other if they have an email contact in common
- If there is no common contact, then the probability of a new tie forming is lower, but ~ 140 times more likely if the individuals share a class than if they don't

Source: Empirical Analysis of an Evolving Social Network; Gueorgi Kossinets and Duncan J. Watts , 2006, Science 311 (5757), 88.

triads vs. # foci

- Having 1 tie or 1 class in common yield equal probability of a tie forming
- probability increases significantly for additional acquaintances, but rises modestly for additional foci



Source: Empirical Analysis of an Evolving Social Network; Gueorgi Kossinets and Duncan J. Watts, 2006, Science 311 (5757), 88.

the strength of ties

- the stronger the ties, the greater the likelihood of triadic closure
- bridges are on average weaker than other ties
- *but*, bridges are more unstable:
  - may get stronger, become part of triads, or disappear

Group Formation in Large Social Networks

- Backstrom, Huttenlocher, Kleinberg, Lan @ KDD 2006
- data:
  - LiveJournal
  - DBLP

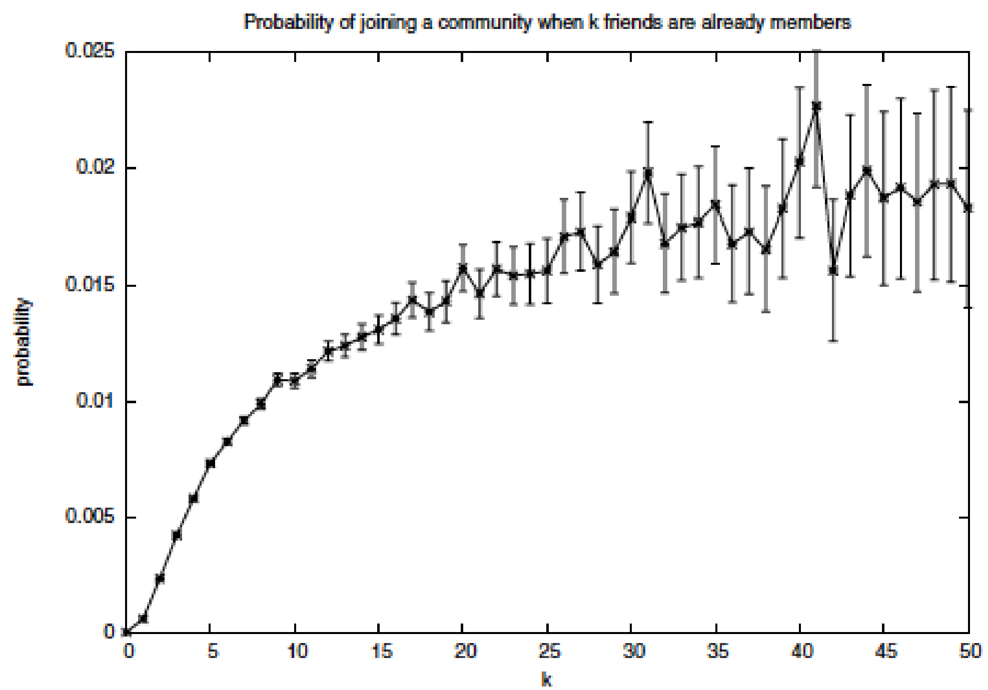


Figure 1: The probability p of joining a LiveJournal community as a function of the number of friends k already in the community. Error bars represent two standard errors.

Source: Backstrom, L., D. Huttenlocher, J. Kleinberg, and Et. Group formation in large social networks: Membership, growth, and evolution.

if it's a "group" of friends that have joined…

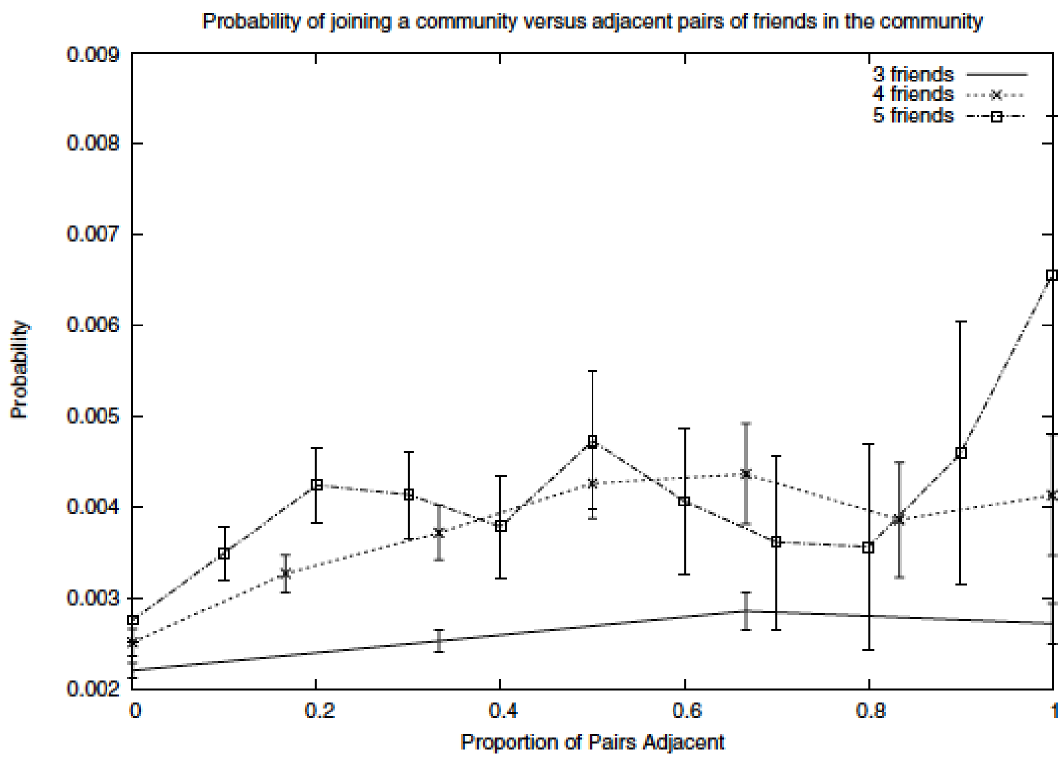


Figure 4: The probability of joining a LiveJournal community as a function of the internal connectedness of friends already in the community. Error bars represent two standard errors.

Source: Backstrom, L., D. Huttenlocher, J. Kleinberg, and Et. Group formation in large social networks: Membership, growth, and evolution.

but community growth is slower if entirely cliquish…

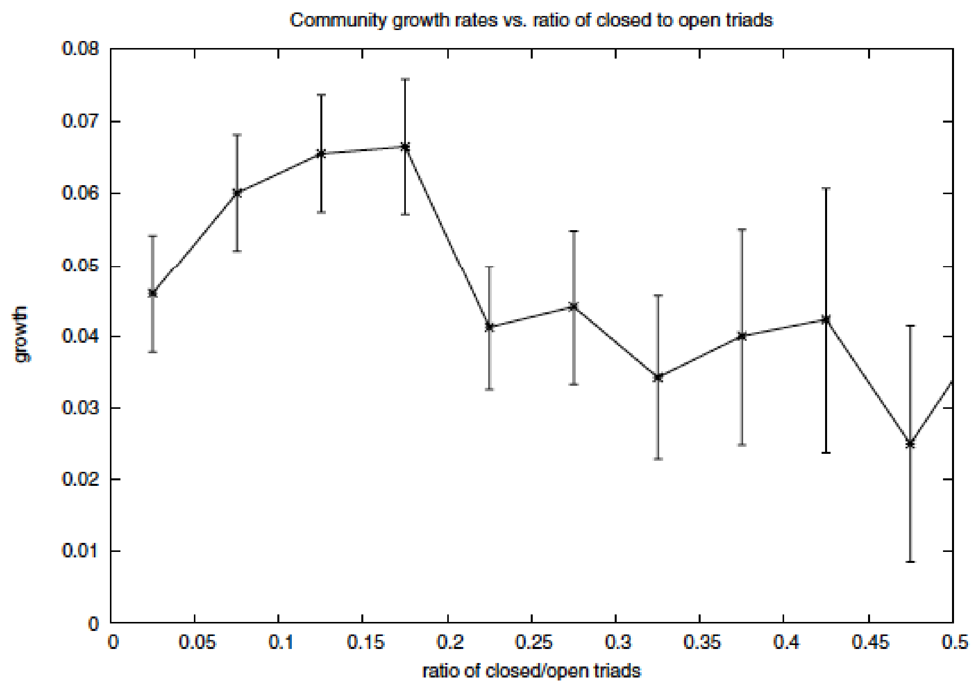


Figure 6: The rate of community growth as a function of the ratio of closed to open triads: having a large density of closed triads (triangles) is negatively related to growth. Error bars represent two standard errors.

Source: Backstrom, L., D. Huttenlocher, J. Kleinberg, and Et. Group formation in large social networks: Membership, growth, and evolution.

So,

- if your friends join, so will you
- if your friends who join know one another, you're even more likely to join
- cliquish communities grow more slowly

Evolution of aggregate network metrics

- as individual nodes and edges come and go, how do aggregate features change?
  - degree distribution?
  - clustering coefficient?
  - average shortest path?

An empirical puzzle of network evolution: Graph Densification

- Densification Power Law

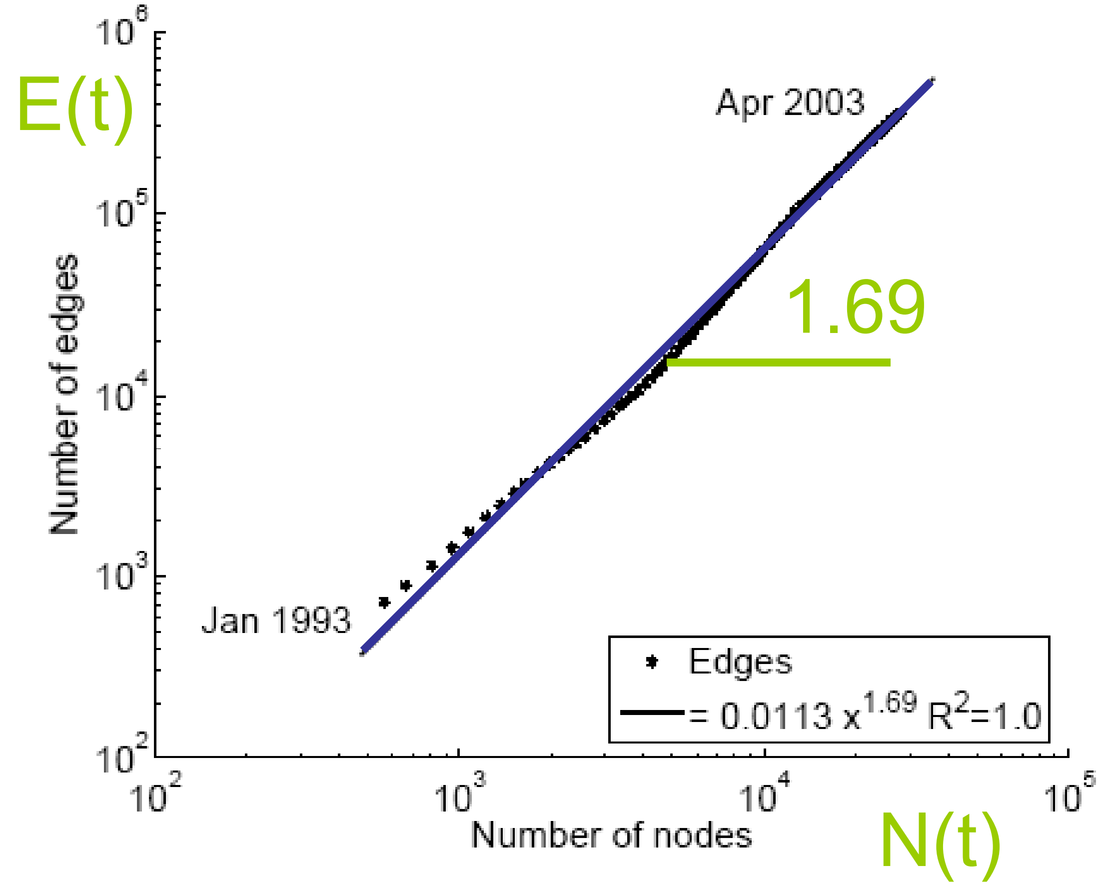
$$\boxed{E(t) \propto N(t)^a}$$

- E: Number of edges; N: Network size
- Densification exponent: $1 \leq a \leq 2$:
  - a=1: linear growth – constant out-degree (assumed in BA model)
  - a=2: quadratic growth – clique
- Let's see the real graphs!

Source: *Leskovec et al. KDD 2005*

Densification – Physics Citations

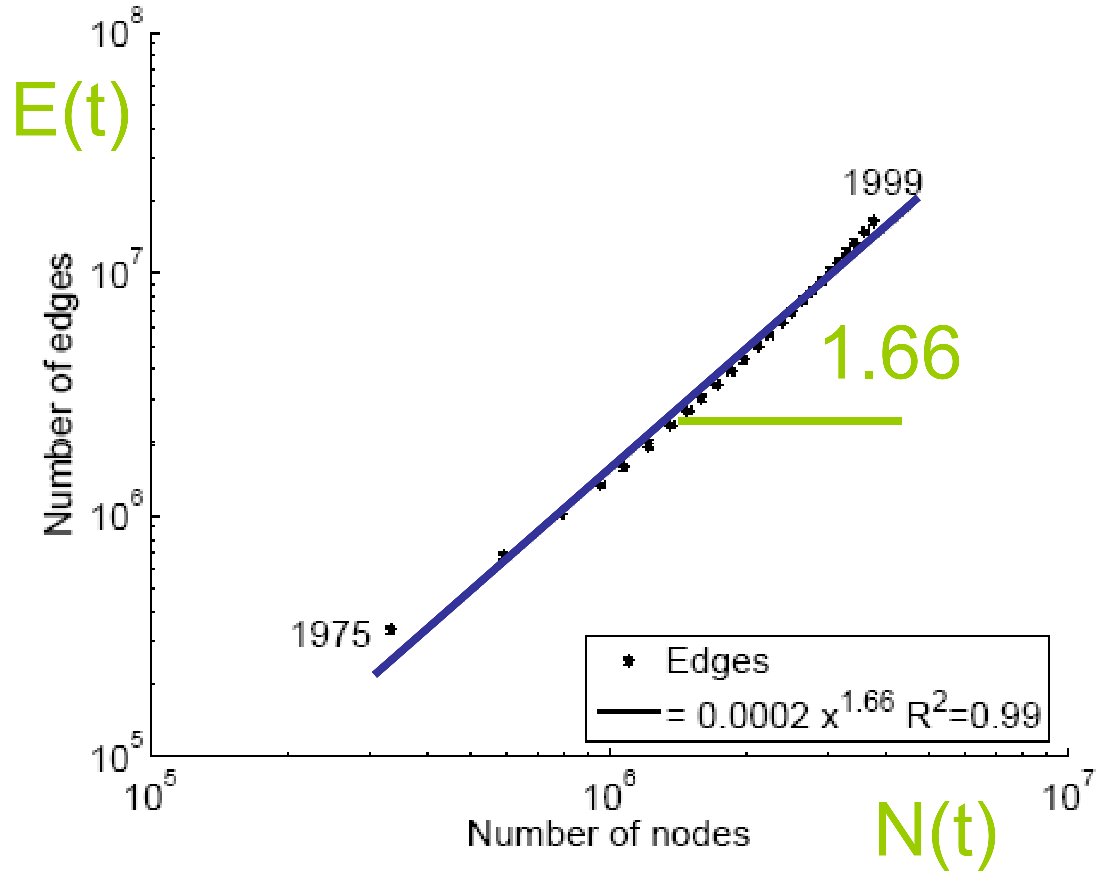
- Citations among physics papers
- 1992:
  - 1,293 papers, 2,717 citations
- 2003:
  - 29,555 papers, 352,807 citations
- For each month *M*, create a graph of all citations up to month *M*



Source: *Leskovec et al. KDD 2005*

Densification – Patent Citations

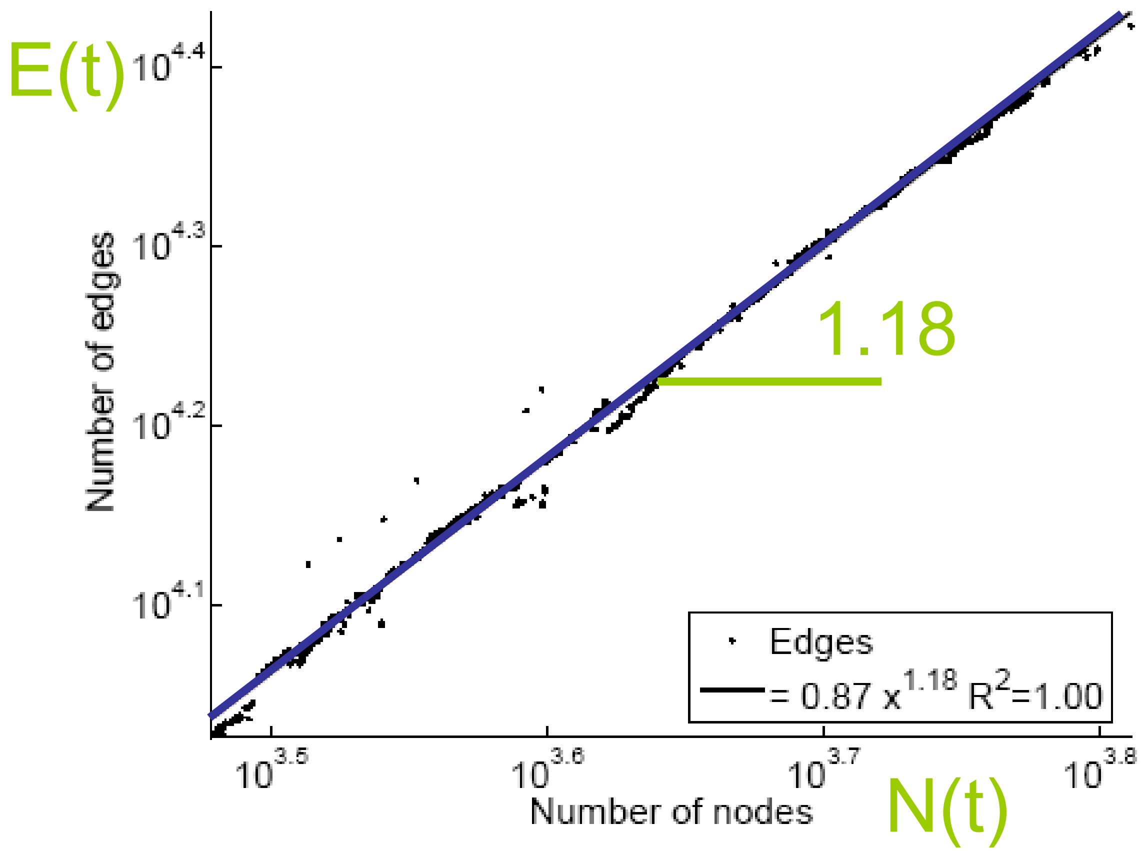
- Citations among patents granted
- 1975
  - 334,000 nodes
  - 676,000 edges
- 1999
  - 2.9 million nodes
  - 16.5 million edges
- Each year is a data point



Source: *Leskovec et al. KDD 2005*

Densification – Autonomous Systems

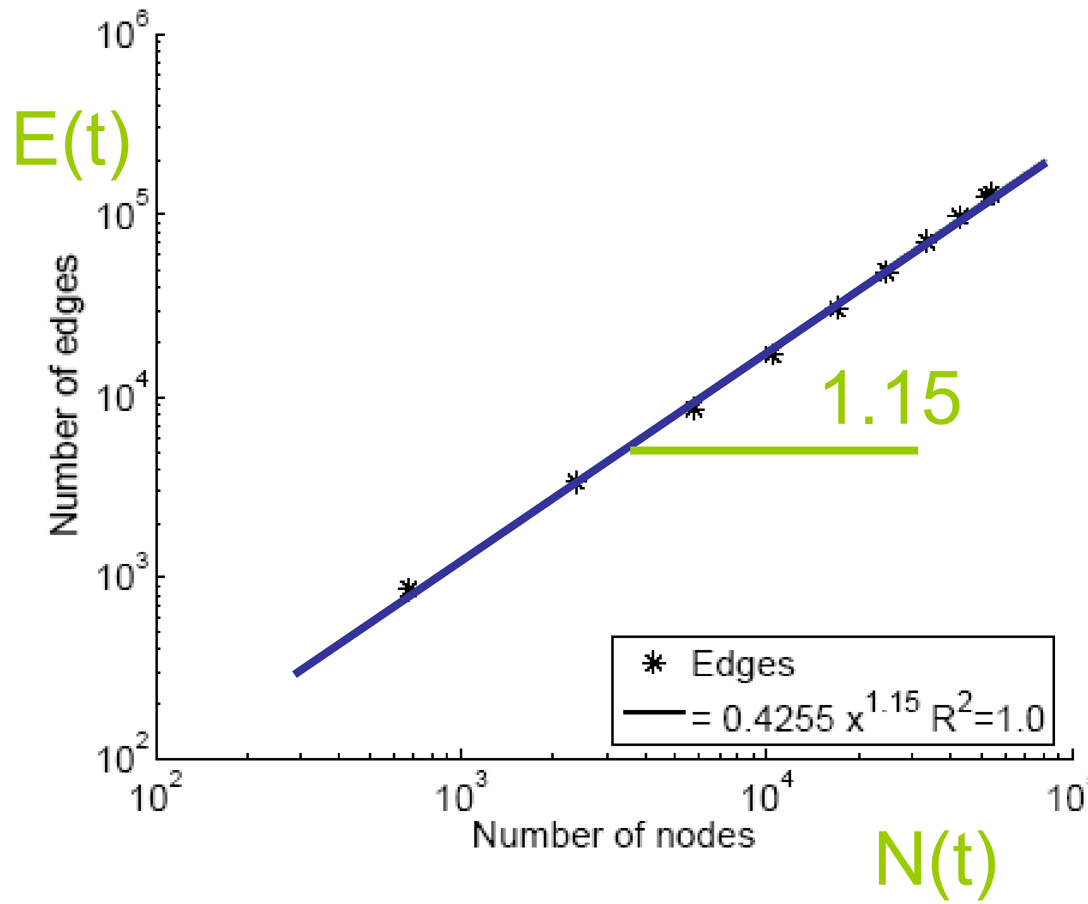
- Graph of the Internet
- 1997
  - 3,000 nodes
  - 10,000 edges
- 2000
  - 6,000 nodes
  - 26,000 edges
- One graph per day



Source: *Leskovec et al. KDD 2005*

Densification – Affiliation Network

- Authors linked to their publications
- 1992
  - 318 nodes
  - 272 edges
- 2002
  - 60,000 nodes
    - 20,000 authors
    - 38,000 papers
  - 133,000 edges



Source: *Leskovec et al. KDD 2005*

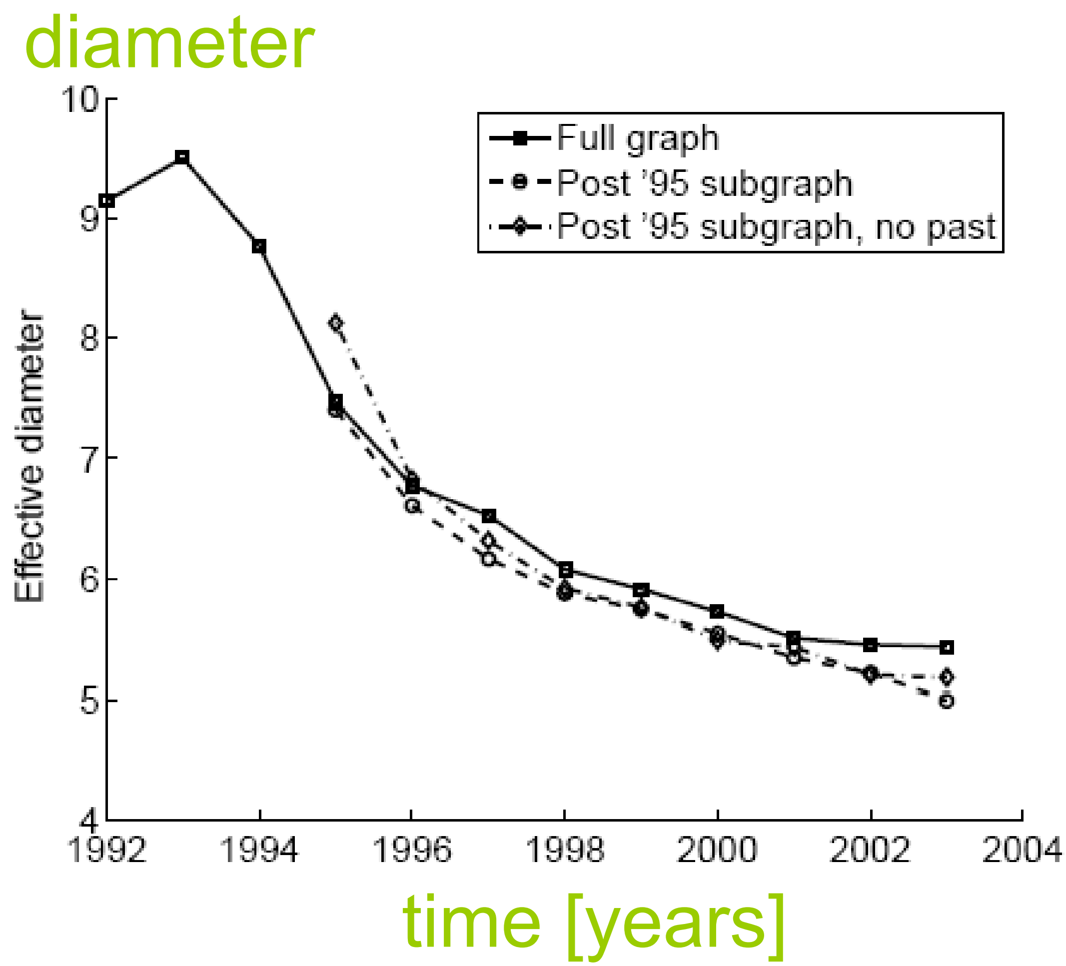
So,

- The traditional constant out-degree assumption does not hold
- Instead: $E(t) \propto N(t)^a$
- the number of edges grows faster than the number of nodes – average degree is increasing

Source: *Leskovec et al. KDD 2005*

Diameter – ArXiv citation graph

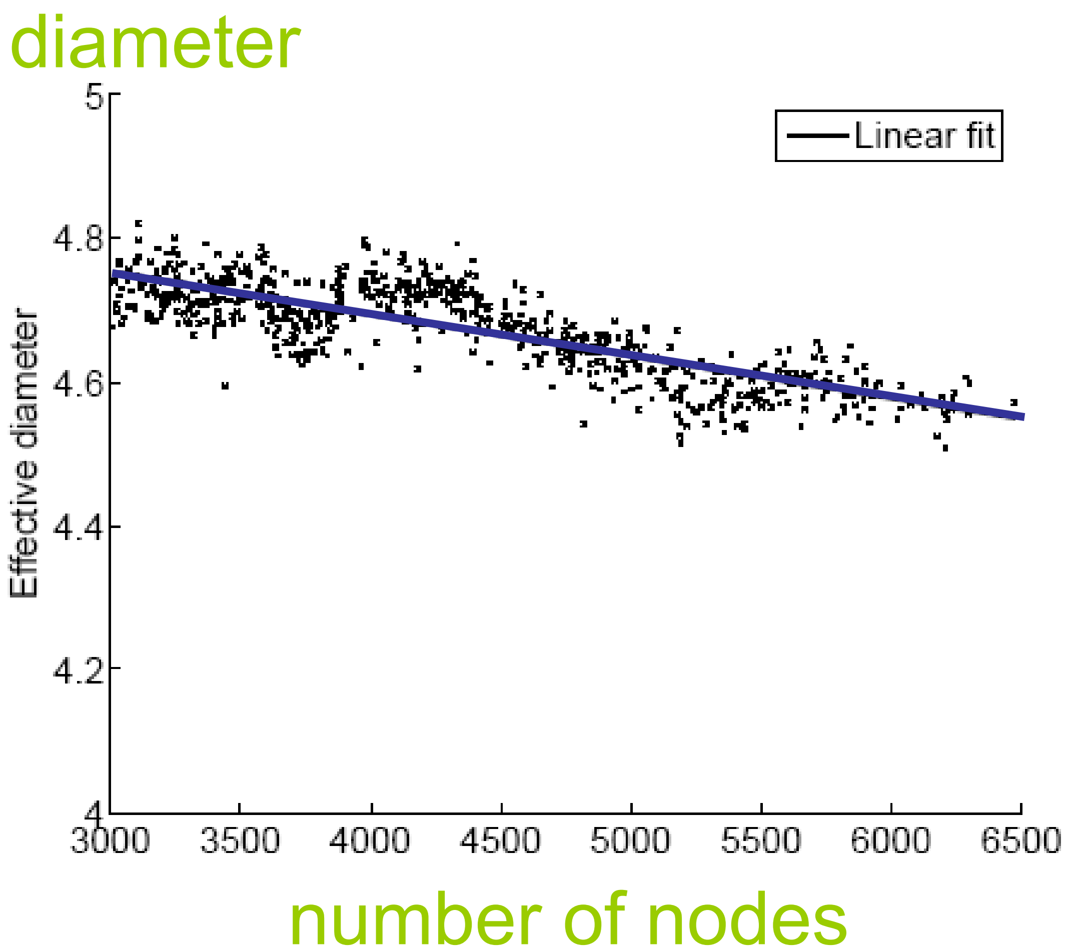
- Citations among physics papers
- 1992 –2003
- One graph per year



Source: *Leskovec et al. KDD 2005*

Diameter – "Autonomous Systems"

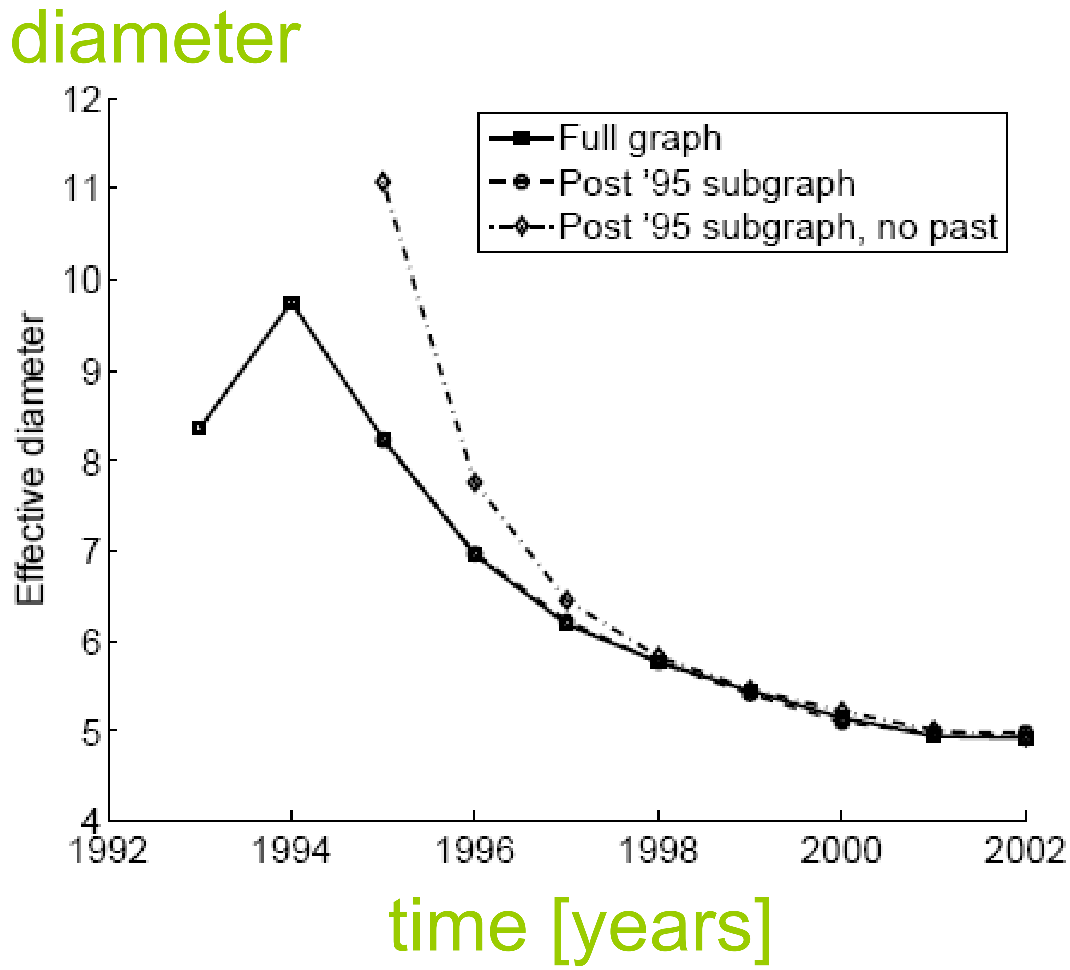
- Graph of the Internet
- One graph per day
- 1997 – 2000



Source: *Leskovec et al. KDD 2005*

Diameter – "Affiliation Network"

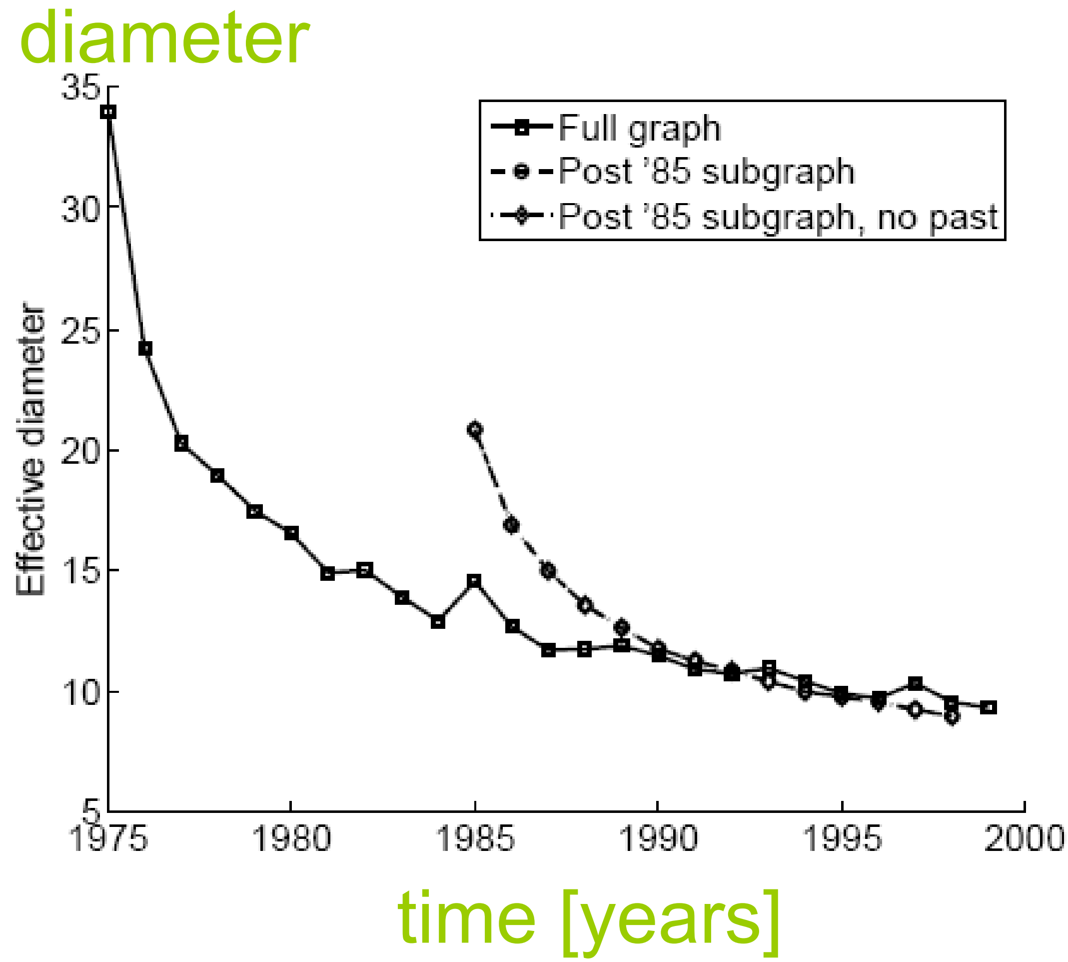
- Graph of collaborations in physics – authors linked to papers
- 10 years of data



Source: *Leskovec et al. KDD 2005*

Diameter – "Patents"

- Patent citation network
- 25 years of data



Source: *Leskovec et al. KDD 2005*

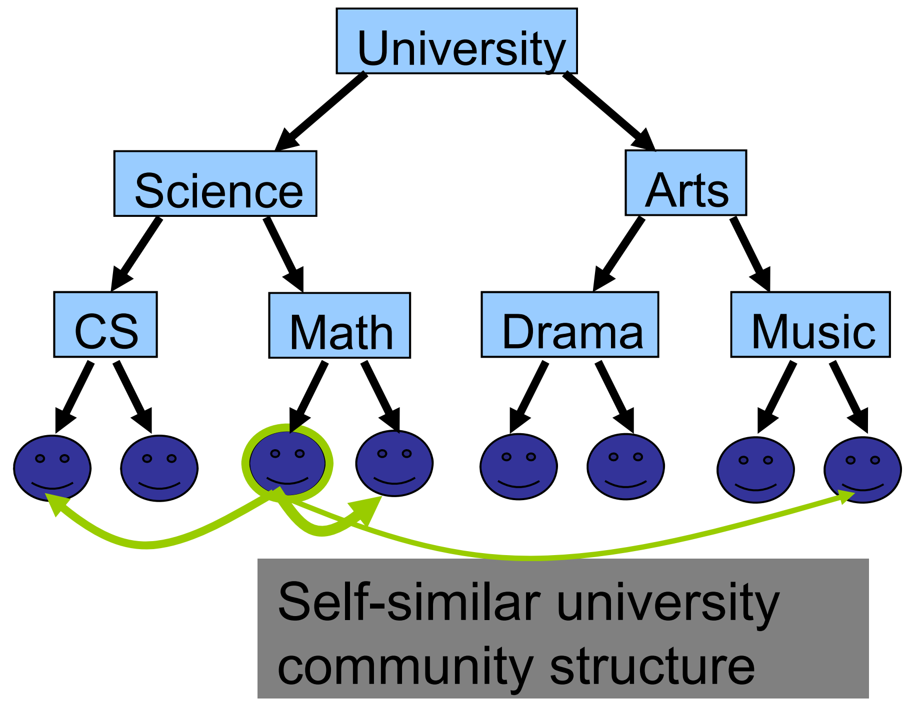
Densification – Possible Explanation

- BA model does not capture the Densification Power Law and Shrinking diameters
- Can we find a simple model of local behavior, which naturally leads to observed phenomena?
- Yes! 2 models have been presented:
  - Community Guided Attachment – obeys Densification
  - Forest Fire model – obeys Densification, Shrinking diameter (and Power Law degree distribution)

Source: *Leskovec et al. KDD 2005*

Community structure

- Let's assume the community structure
- One expects many within-group friendships and fewer cross-group ones
- How hard is it to cross communities?



- If the cross-community linking probability of nodes at tree-distance *h* is scale-free
- cross-community linking probability: $\boxed{f(h) = c^{-h}}$

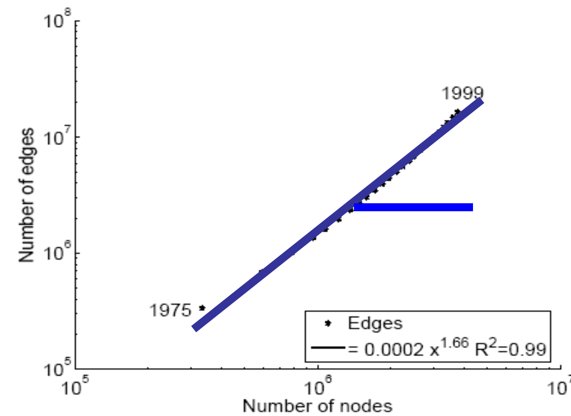
where: c ≥ 1 … is the Difficulty constant and h is the tree-distance

Source: *Leskovec et al. KDD 2005*

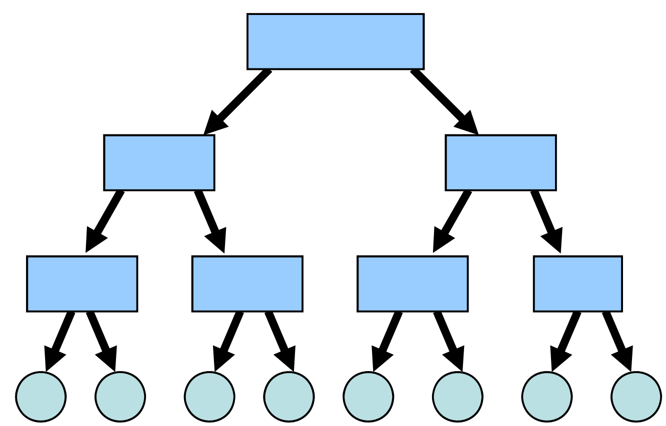
Densification Power Law

- <u>Theorem:</u> The Community Guided Attachment leads to Densification Power Law with exponent

$$\boxed{a = 2 - \log_b(c)}$$



- a: densification exponent
- b: community structure branching factor
- c: difficulty constant



Source: *Leskovec et al. KDD 2005*

Difficulty Constant

- Theorem: $\boxed{a = 2 - \log_b(c)}$
- Gives any non-integer Densification exponent
- If c = 1: easy to cross communities
  - Then; a=2, quadratic growth of edges – near clique
- If c = b: hard to cross communities
  - Then; a=1, linear growth of edges – constant out-degree
- Room for improvement:
  - Community Guided Attachment explains Densification Power Law
  - Issues:
    - Requires explicit Community structure
    - Does not obey Shrinking Diameters

Source: *Leskovec et al. KDD 2005*

“Forest Fire” model – Wish List

- Want no explicit Community structure
- Shrinking diameters
- and:
  - “Rich get richer” attachment process, to get heavy-tailed in-degrees
  - “Copying” model, to lead to communities
  - Community Guided Attachment, to produce Densification Power Law

Source: *Leskovec et al. KDD 2005*

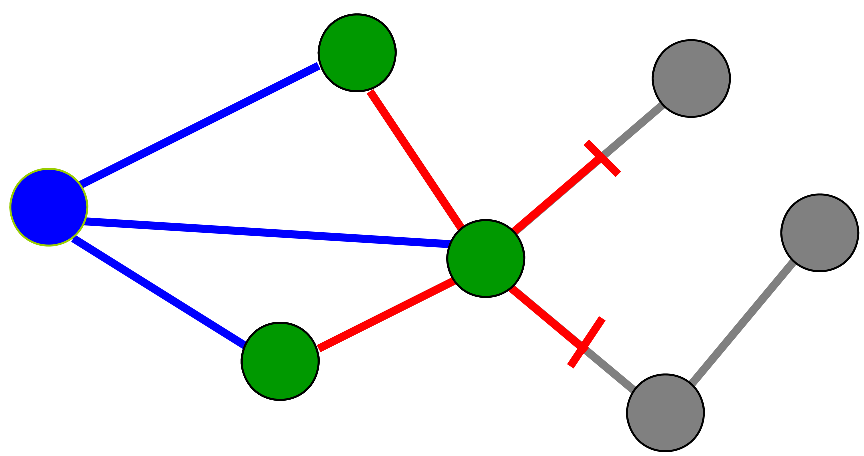
“Forest Fire” model – Intuition

- How do authors identify references?
  1. Find first paper and cite it
  2. Follow a few citations, make citations
  3. Continue recursively
  4. From time to time use bibliographic tools (e.g. CiteSeer) and chase back-links
- How do people make friends in a new environment?
  1. Find first a person and make friends
  2. Follow of his friends
  3. Continue recursively
  4. From time to time get introduced to his friends
- Forest Fire model imitates exactly this process

Source: *Leskovec et al. KDD 2005*

“Forest Fire” – the Model

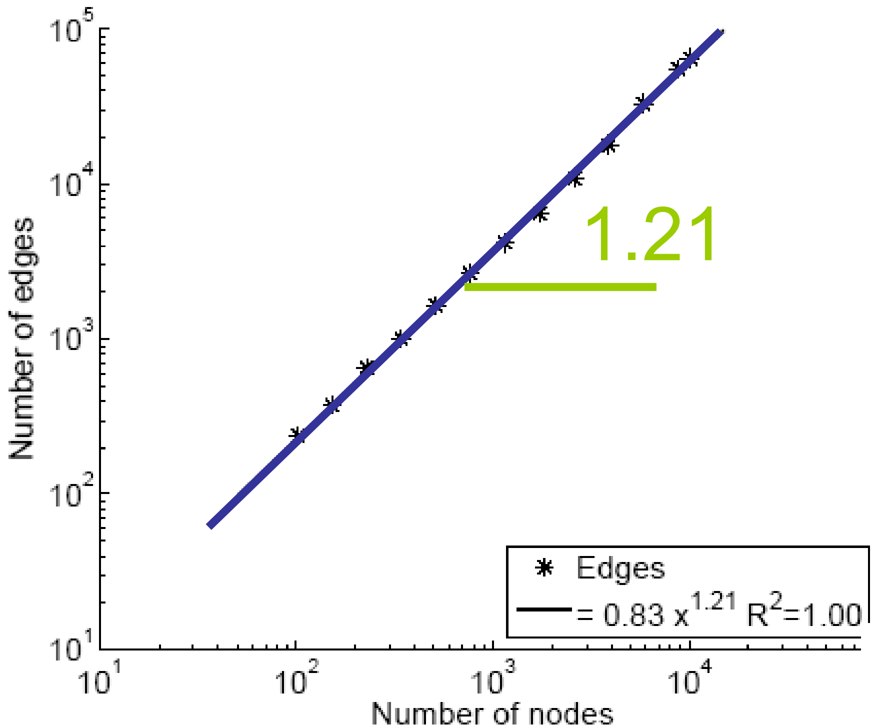
- A node arrives
- Randomly chooses an “ambassador”
- Starts burning nodes (with probability p) and adds links to burned nodes
- “Fire” spreads recursively



Forest Fire in Action

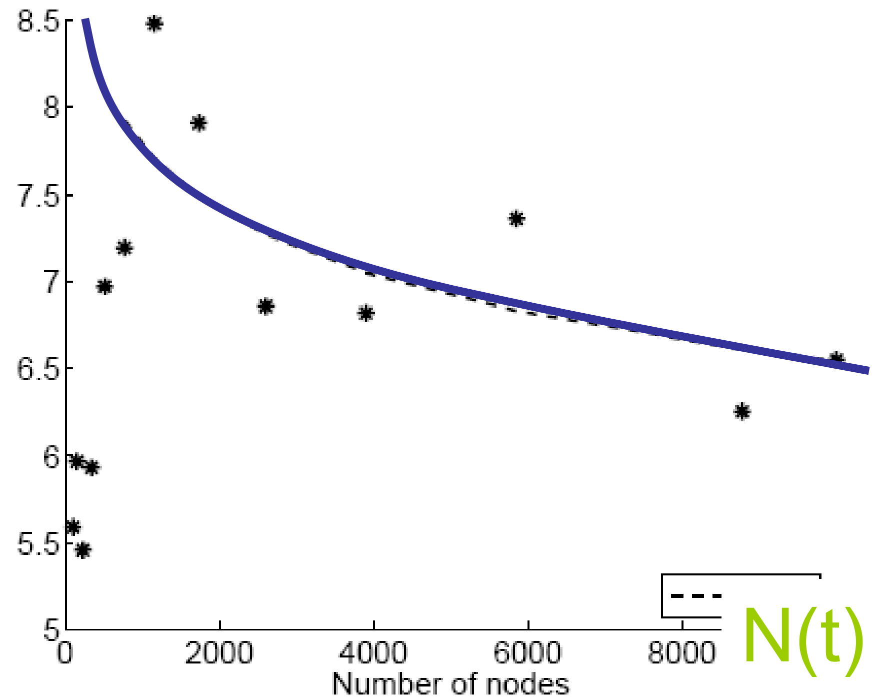
Forest Fire generates graphs that Densify and have Shrinking Diameter

E(t) densification



N(t)

diameter



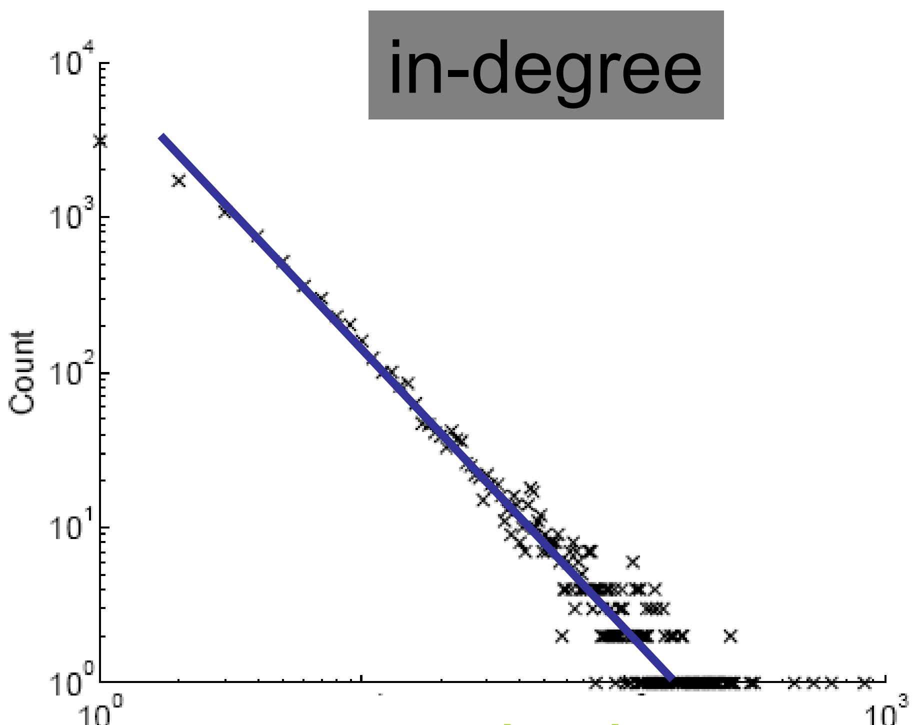
diameter

N(t)

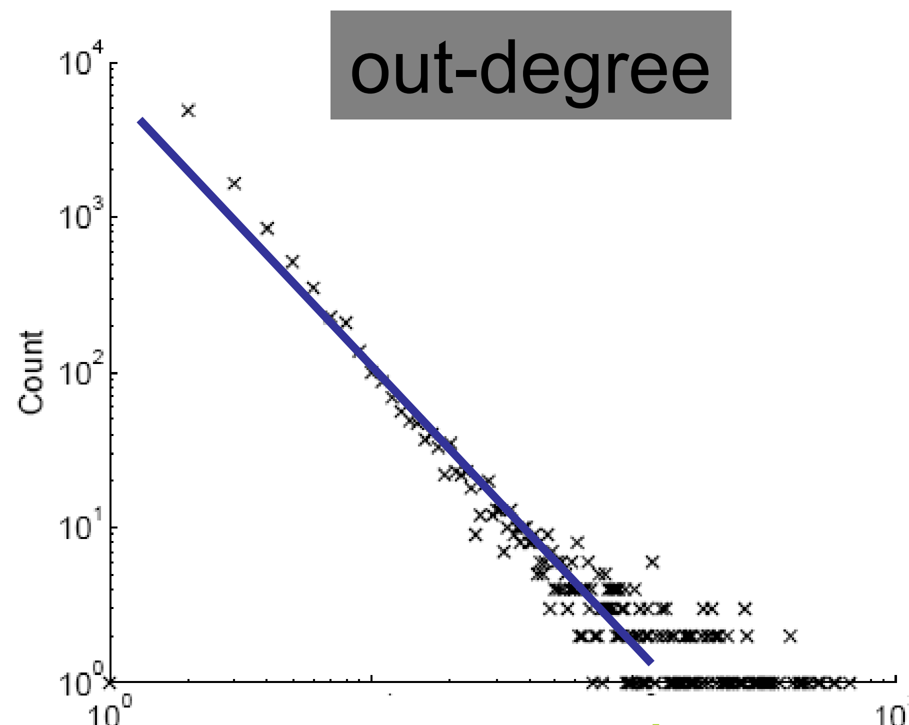
Source: *Leskovec et al. KDD 2005*

Forest Fire in Action

Forest Fire also generates graphs with heavy-tailed degree distribution



count vs. in-degree



count vs. out-degree

Source: *Leskovec et al. KDD 2005*

Forest Fire model – Justification

- Densification Power Law:
  - Similar to Community Guided Attachment
  - The probability of linking decays exponentially with the distance – Densification Power Law
- Power law out-degrees:
  - From time to time we get large fires
- Power law in-degrees:
  - The fire is more likely to burn hubs
- Communities:
  - Newcomer copies neighbors' links
- Shrinking diameter

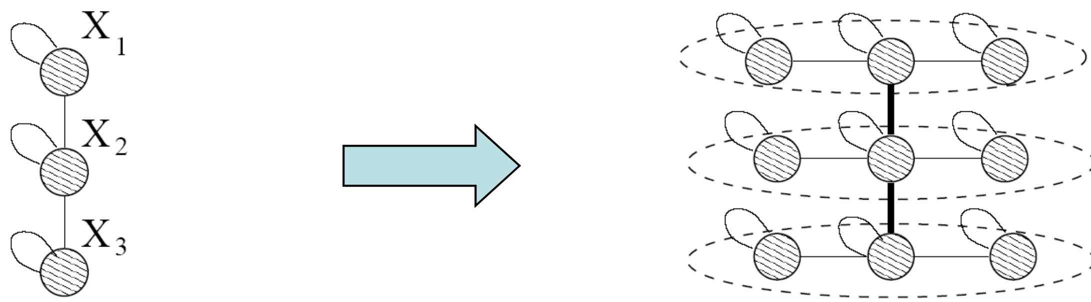
Source: *Leskovec et al. KDD 2005*

Kronecker graphs

- But, want to have a model that can generate a realistic graph with realistic growth:
  - Static Patterns
    - Power Law Degree Distribution
    - Small Diameter
    - Power Law Eigenvalue and Eigenvector Distribution
  - Temporal Patterns
    - Densification Power Law
    - Shrinking/Constant Diameter
- For Kronecker graphs [Leskovec et al, PKDD05] all these properties can actually be proven

Idea: Recursive graph generation

- Starting with our intuitions from densification
- Try to mimic recursive graph/community growth because self similarity leads to power-laws
- There are many obvious (but wrong) ways:

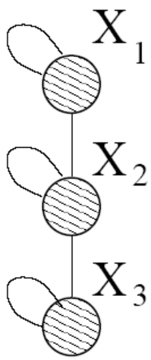


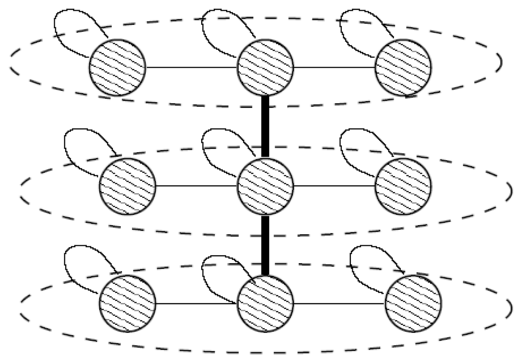
Initial graph

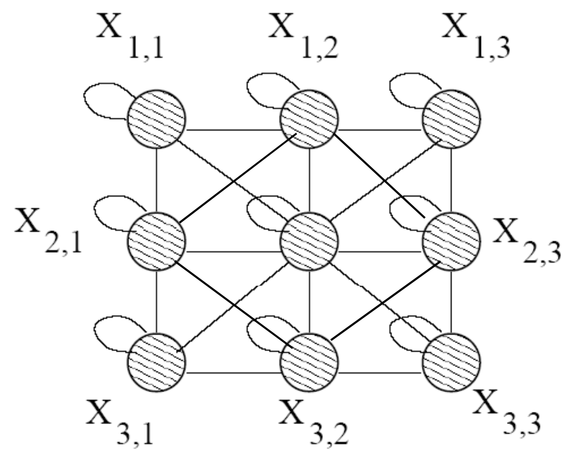
Recursive expansion

  - Does not densify, has increasing diameter
- Kronecker Product is a way of generating self-similar matrices

Kronecker product: Graph







Intermediate stage

1	1	0
1	1	1
0	1	1

(3x3)

G_1

Adjacency matrix

G_1	G_1	0
G_1	G_1	G_1
0	G_1	G_1

(9x9)

$$G_2 = G_1 \otimes G_1$$

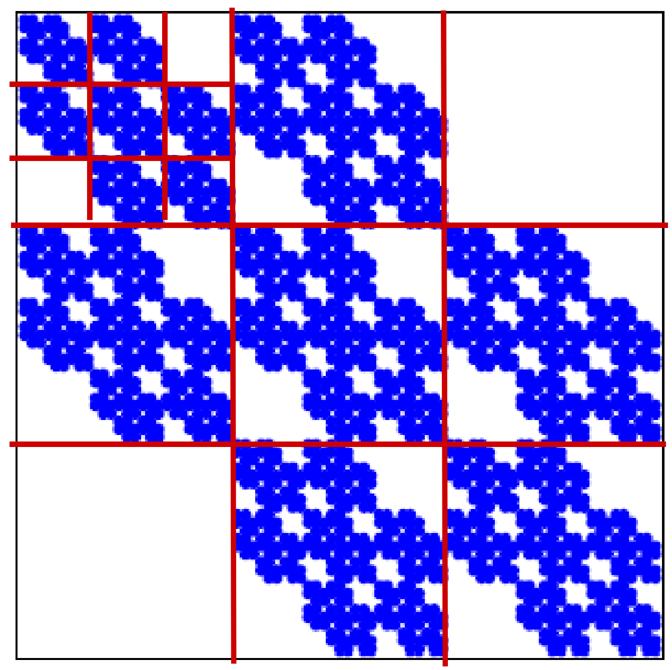
Adjacency matrix

Kronecker product: Graph

- Continuing multypling with G_1 we obtain G_4 and so on …

1	1	0
1	1	1
0	1	1

G_1



G_4 adjacency matrix

Kronecker product: Definition

- The Kronecker product of matrices A and B is given by

$$\underset{N\,x\,M}{\mathbf{C} = \mathbf{A}} \otimes \underset{K\,x\,L}{\mathbf{B}} \doteq \begin{pmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \dots & a_{1,m}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \dots & a_{2,m}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}\mathbf{B} & a_{n,2}\mathbf{B} & \dots & a_{n,m}\mathbf{B} \end{pmatrix}$$

$$N{*}K \; x \; M{*}L$$

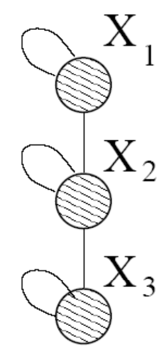
- We define a Kronecker product of two graphs as a Kronecker product of their adjacency matrices

Kronecker graphs

- We propose a growing sequence of graphs by iterating the Kronecker product

$$G_k = \underbrace{G_1 \otimes G_1 \otimes \ldots G_1}_{k\ times}$$

- Each Kronecker multiplication exponentially increases the size of the graph
- G_k has N_1^k nodes and E_1^k edges, so we get densification



G_1

Stochastic Kronecker graphs

- But, want a randomized version of Kronecker graphs
- Possible strategies:
  - Randomly add/delete some edges
  - Threshold the matrix, e.g. use only the strongest edges
- Wrong, will destroy the structure of the graph, e.g. diameter, clustering

Stochastic Kronecker graphs

- Create $N_1 \times N_1$ probability matrix P_1
- Compute the k^{th} Kronecker power P_k
- For each entry p_{uv} of P_k include an edge (u,v) with probability p_{uv}

0.5	0.2
0.1	0.3

P_1

Kronecker multiplication

Probability of edge p_{ij}

0.25	0.10	0.10	0.04
0.05	0.15	0.02	0.06
0.05	0.02	0.15	0.06
0.01	0.03	0.03	0.09

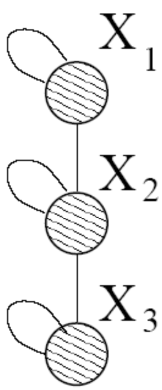
$P_2 = P_1 \otimes P_1$

Instance matrix K_2

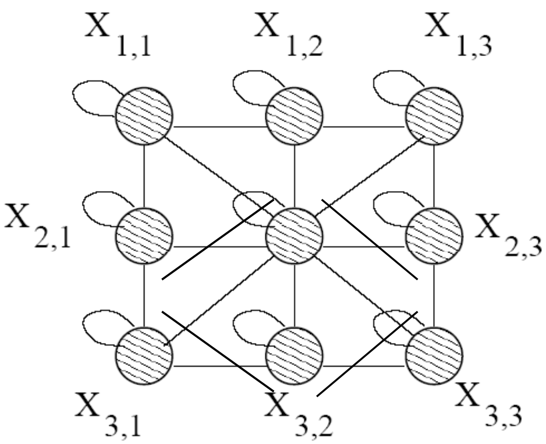
flip biased coins

Kronecker graphs: Intuition

- Intuition:
  - Recursive growth of graph communities
  - Nodes get expanded to micro communities
  - Nodes in sub-community link among themselves and to nodes from different communities



G_1



$G_2 = G_1 \otimes G_1$

Kronecker graphs: Intuition

- Node attribute representation
  - Nodes are described by (binary) features [likes ice cream, likes chocolate]
  - *E.g.*, u=[1,0], v=[1, 1]
  - Parameter matrix gives linking probability: p(u,v) = 0.1 * 0.5 = 0.15

	1	0
1	0.5	0.2
0	0.1	0.3

Kronecker multiplication

	11	10	01	00
11	0.25	0.10	0.10	0.04
10	0.05	0.15	0.02	0.06
01	0.05	0.02	0.15	0.06
00	0.01	0.03	0.03	0.09

Properties of Kronecker graphs

- One can show that Kronecker multiplication generates graphs that have:
  - Properties of static networks
    - ✓ Power Law Degree Distribution
    - ✓ Power Law eigenvalue and eigenvector distribution
    - ✓ Small Diameter
  - Properties of dynamic networks
    - ✓ Densification Power Law
    - ✓ Shrinking/Stabilizing Diameter

Readings

- "Networks, Crowds, and Markets" by Easley and Kleinberg (Chapter 18)
- Barabasi A-L, Albert R (1999) Emergence of scaling in random networks. Science 286: 5009-5012
- Leskovec L, Kleinberg J, Faloutsos, (2007) Evolution: Densification and shrinking diameters, ACM Transactions on Knowledge Discovery from Data, 1 (1), 1-41