Outline

- Characteristics of Modern Networks
  - Power law Distribution
  - Ubiquity of the Power Law
- Deriving the Power Law
  - How does network grow?
  - The theory of preferential attachment
  - Variations on the theme
- Properties of Scale Free Networks
  - Error, attack tolerance, and epidemics
  - Implications for modern distributed systems
  - Implications for everyday systems

Characteristics of Modern Networks

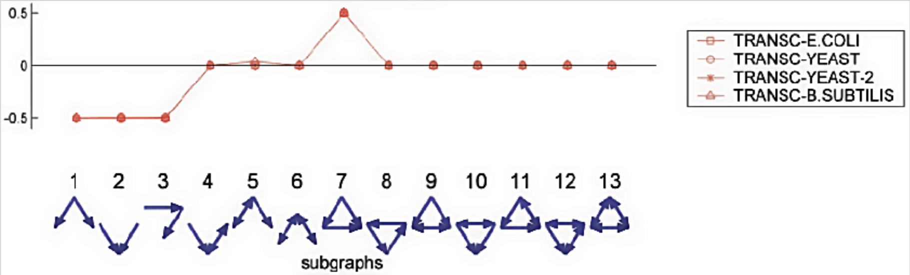
- Most networks
  - Social
  - Technological
  - Ecological
- Are characterized by being
  - Small world
  - Clustered
  - And SCALE FREE (Power law distribution)
- We now have to understand
  - What is the power law distribution
  - And how we can model it in networks

The Degree Distribution

- **What is the degree distribution?**
  - **It is the way the various edges of the network "distributes" across the vertices**
  - **How many edges connect the various vertices of the network**
- **For the previous types of networks**
- **In k-regular regular lattices, the distribution degree is constant**
  - $P(k_r)=1$ **for all nodes (all nodes have the same fixed k_r number of edges)**
- **In random networks, the distribution can be either constant or exponential**
  - $P(k_r)=1$ **for all nodes (is the randon network has been constructed as a k-regular network)**
  - $P(k_r)=\alpha e^{-\beta k}$ **, that is the normal "gaussian" distribution, as derived from the fact that edges are independently added at random**

Network Motifs

• When we looked for transitivity, we basically counted the number of subgraphs of a particular type (triangles and triples).



• We can generalize this approach to see which patterns are 'very frequent' in the network. Those patterns are called *network motifs*.

• For each subgraph, we measure its relative frequency in the network.

• To measure the frequency, we compare with how expected it is to see such patterns in a random network.

• As we are measuring for the 13 possible directed connected graphs of 3 vertices, it is called a *triad* significance profile (TSP).

• The significance profile (SP) of the network is a vector of those frequencies.

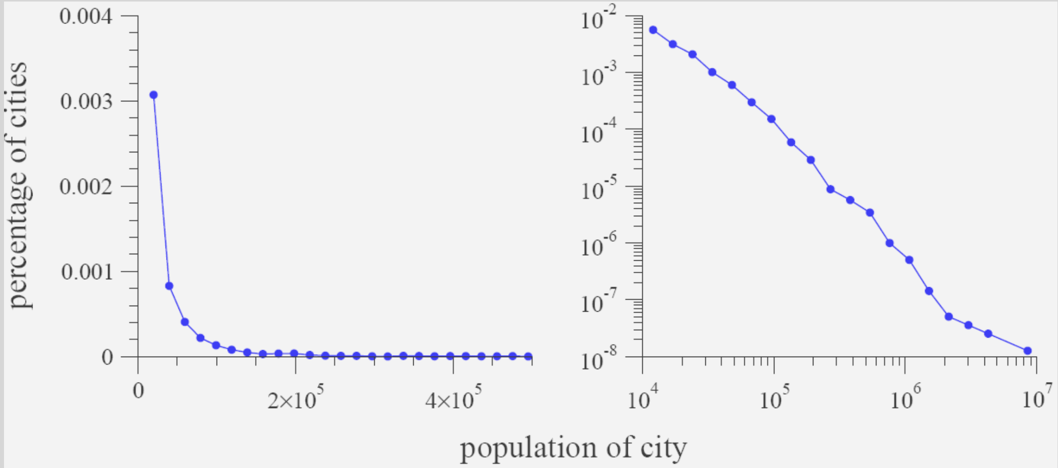
• 4 networks of different micro-organisms are shown to have very similar TSPs, and in particular the triad 7 called « feed-forward loop ».

The Scale-Free property

There are things that have an enormous variation in the distribution.

If we plot this histogram with logarithmic horizontal and vertical axis, a pattern will clearly emerge: a line.

In a normal histogram, this line is $p(x) = -\alpha x + c$. Here it's log-log, so:



apply exponent e

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

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

We say that this distribution follows a power-law, with exponent α.

A power law is the only distribution that is the same whatever scale we look at it on, *i.e.* $p(bx) = g(b)p(x)$. So, it's also called scale-free.

The Scale-Free property

It has been found that the human population has the scale-free property.

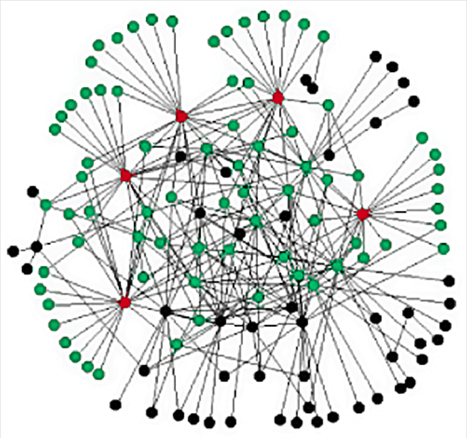
In 1955, Herbert Simon already showed that many systems follow a power law distribution, so that's neither new nor unique.

- Sizes of earthquakes
- Moon craters
- Solar flares
- Computer files
- Wars
- Number of citations received / paper
- Number of hits on web pages
- People's annual incomes

It has been found that the distribution of the degree of nodes follows a power-law in many networks, *i.e.* many networks are scale-free…

What is important is not so much to find a power-law as it's common, but to understand why and which other structural parameters can be there.

The Scale-Free property : *Myth and reality*



• **One mechanism was used to build scale-free networks, called preferential attachment, or « the rich get richer » paradigm.**

• Scaling distributions are a subset of a larger family of heavy-tailed distributions that exhibit high variability.

• One important claim of the literature for scale-free networks was the presence of highly connected central hubs.

• It was said that « the most highly connected nodes represent an Achilles' heel »: delete them and the graph breaks into pieces.

• **However, it only requires high variability and not strict scaling…**

• **It is only one of several, and not less than 7 other mechanisms give the same result, so preferential attachment gives little or no insight in the process.**

• **Recent research have shown that complex networks that claimed to be scale-free have a power-law but not this Achilles' heel.**

Other measurements

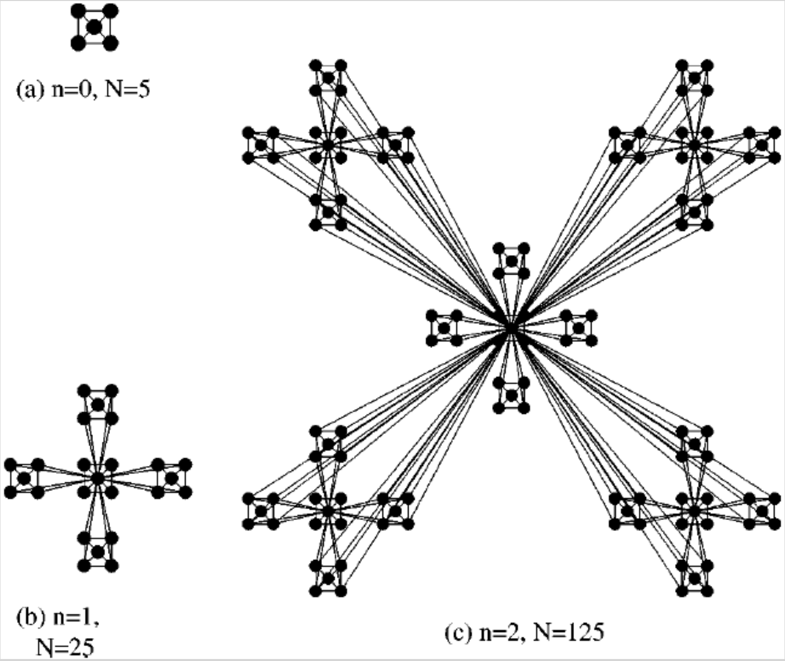
- **We already have the clustering, distribution of degree, etc. **
- Are there other global characteristics relevant to the performances of the network, in term of searchability or stability?
- Rozenfeld has proposed in his PhD thesis to study the cycles, with algorithms to approximate their counting (as it's exponential otherwise).
- Using cycles as a measure for complex networks has received attention:

Inhomogeneous evolution of subgraphs and cycles in complex networks (Vazquez, Oliveira, Barabasi. Phys. Rev E71, 2005).

Degree-dependent intervertex separation in complex networks (Dorogovtsev, Mendes, Oliveira. Phys Rev. E73 2006)

- **One should study the correlation of degree (*i.e.* assortativity).**

Scale-free networks



Start with a complete graph of N vertices (a dense group).

Make N – 1 copies.

Link the root to all vertices but the copies of the root.

Repeat the process.

This construction was generalized in the family of graphs $H_{n,k}$.

Start with a group of n nodes and iterate k steps (*i.e.* build k levels).

- It has a scale-free distribution of degrees.
- **It has a high clustering: for a node with k links, its clustering is C(k) ~ 1/k.**
- **Its diameter is 2k – 1 (this network is <u>not</u> small-world).**

Scale-free and small-world

- This family $K_{n,k}$ is defined in a similar way as the previous one:

same
- Start at k = 0 with a complete graph K_n.
- Do k times the following:

change

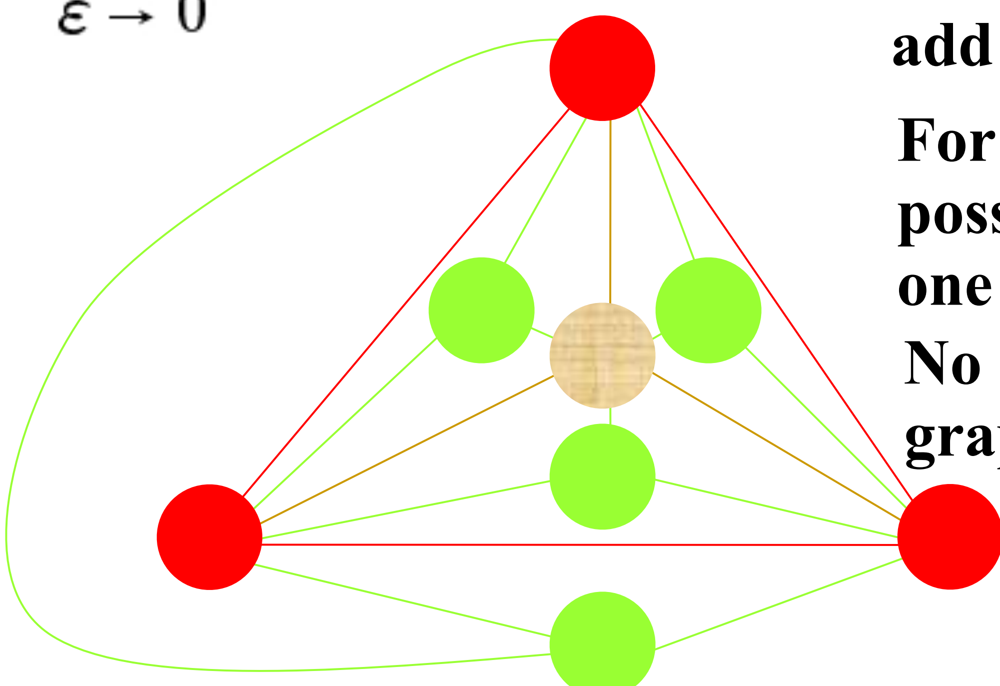
$$\frac{\int x \, dx}{\varepsilon \to 0}$$

Let's start our example with K_3.

There is one K_3 so we add one node.

For each of the 4 possible K_3, we add one node.

No other K_3 in the graph: next level.



Preferences of newcomers

Bob and Ted are really cool guys, everybody knows them, so we'll meet

Only one person knows Tracy, it's unlikely that we get in touch...

I'm new! Who should I get in touch with?

- Another very popular way to model networks is to reproduce the growth processes taking place in the real world: new nodes come in!
  - Herbert Simon showed in 1955 that power laws are encountered when the rich get richer: the more we already have, the more we get.
  - So, the most common way of generating a scale-free network is to use preferential attachment.

  → When a new node arrives, it prefers to link to the most popular nodes.

- In 1965, Derek de Solla Price set up a model where the probability that a new node links to another one is proportional to kin + 1, where kin is the incoming degree of the node.

The Barabasi-Albert (BA) model

- Price's model created a directed graph with variables number of edges added at each node. It gives the degree distribution $p_k \sim k^{-(2+1/m)}$.

- Thirty years after, in 1999, Barabasi and Albert came with their model: **undirected, constant number of edges, always gives $p_k \sim k^{-3}$** .

  - The BA model is the most famous and 'started' the field. Why?

    They gave an important situation where this model has a strong potential: the **Web**.

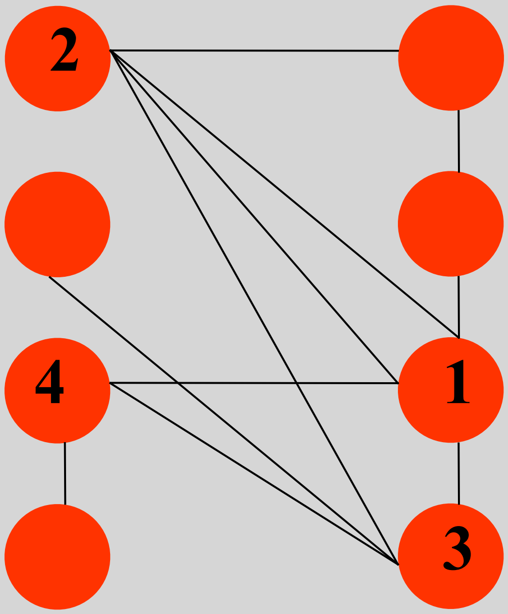
- You decided to start your website, and it's time to create a 'link' section.

  - Most likely, you will **link to popular websites, making them even more popular** (*i.e.* it creates a feedback loop system → rich get richer).

Including assortativity in a model

• Krapivsky and Redner have considered a directed version of the BA model with degree-degree correlations (which we don't have in BA).

• In general, Sokolov and Xulvi-Brunet have proposed a simple algorithm to make a network assortative with a parameter p:



1. Choose randomly two links.

2. Order their four end-nodes with respect to their degree.

3. Rewire to connect low degree vertices and high ones together with probability p, or 1 – p for random.

By repeating, one can make the network assortative and keep the degree distribution.

Generalization of the BA model

• The Dorogovtsev-Mendes-Samukhin (DMS) model adds a parameter k0 to the equations of preferential attachment. If k0 = 0, we have the BA.

• Krapivsky has shown that if we have a nonlinear attachment probability, then we don't have a power law but an exponential: a single nodes take all the newcomers.

• The Albert and Barabasi (AB) model uses two more parameters p and q to rewire changes on the connections after a node has been attached.

It produces power-law *but* unrealistic assortativity and clustering.

• Solé-Pastor Satorras-Smith-Kepler (SPSK) model uses 3 mechanisms.

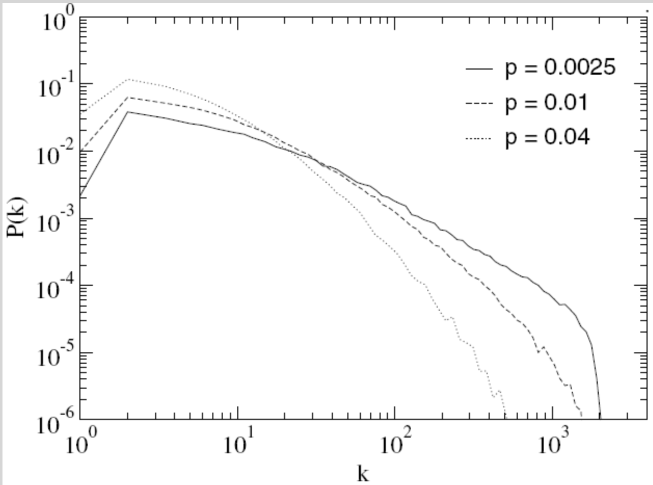
- *Duplicate* (copy a randomly selected node with its connection)
- *Divergence* (some connections of a duplicate are removed)

• A more realistic one with duplication and divergence is the Vazquez – Flammini – Maritan – Vespignani (VFMV) model.

- *Mutate* (connections are added to the duplicate)

Triangle-generating protocol

• In the same way as many sets of rules can guide a dynamic process toward to be scale-free, can it guide it to be small-world?



The distribution exhibits a power-law regime for small p, with an exponent of 1.35 for $p = 0.0025$.

Note that the distribution is largely insensitive to system size N, which here is $N = 7000$.

The exponential cutoff is a result of the finite age of nodes.

• Remember the example: when you marry somebody, it's quite likely that you will get to know his/her family (whether you want it or not).

• Nodes are dynamically introduced to each other by a common node:

- One random node chooses randomly two of its neighbours and link them. If the node has less than 2 neighbours, it links to a random node.
- With probability p, a random node is removed with its links, and replaced by a new node with a randomly chosen neighbour.

Proof

- Now let's remember that we add nodes at each time interval
- Therefore, the probability t_i 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

- **Now given the probability distribution:**
- **Which represents the probability that a node *i* has less than *k* link**
- **The probability that a node has exactly *k* link can be derived by the derivative of the probability distribution**

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

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

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

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 attach 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:**

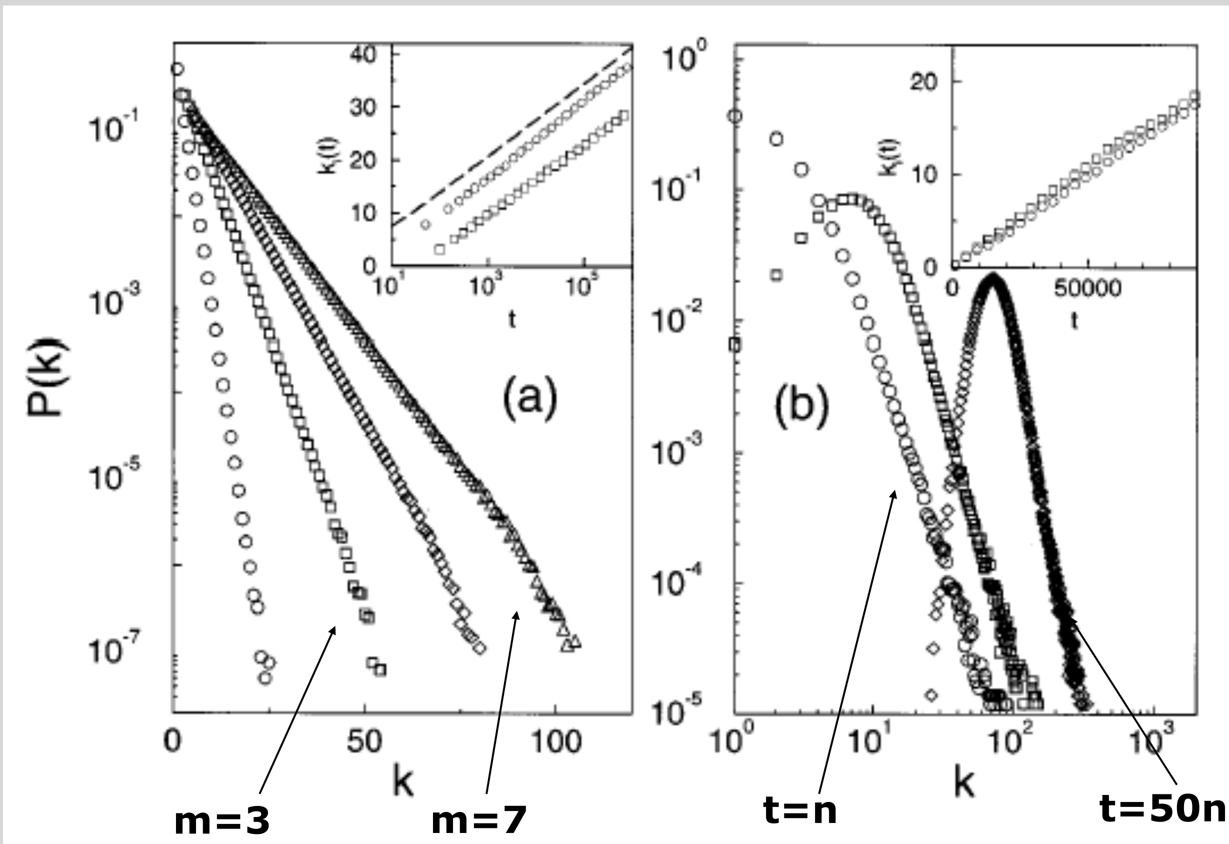
$$P(k) = \frac{1}{m} e^{-\frac{k}{m}}$$

Barabasi-Albert Model vs. Random Network Model

- **The difference in the evolution of the Barabasi-Albert model vs. the Random Network model (from Barabasi and Albert 2002)**

Barabasi-Albert Model n=800000

Simulations performed with various values of m



Random network model for n=10000

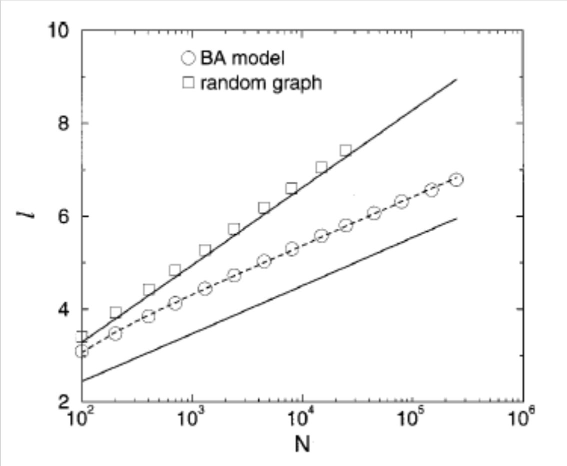
The degree distribution gradually becomes a normal one with passing time

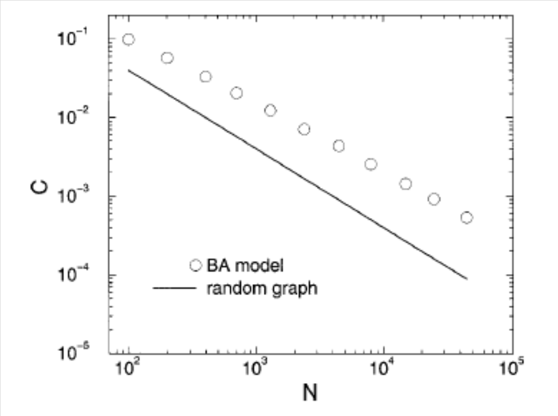
Generality of the Barabasi-Albert Model

- 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: 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 Barabasi-Albert 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!



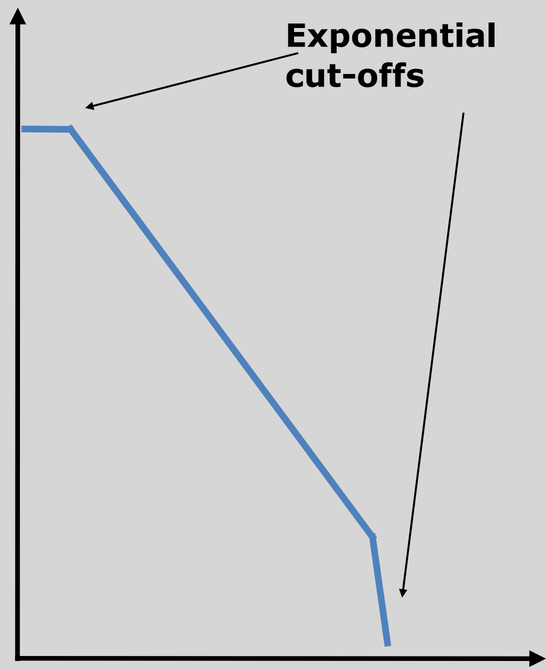


Problems of the Barabasi Albert Model

- 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

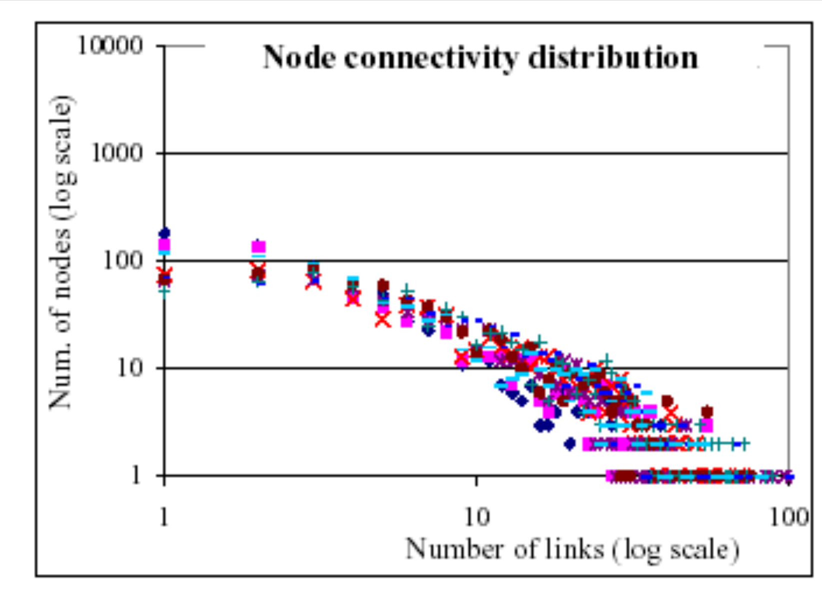
Problems of the Barabasi Albert Model

- As 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
  - The same sometimes happen for
- In general
  - The distribution has still a "heavy tailed" is 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
  - Viceversa, there could be a minimal amount of resources a node can have
- The Barabasi-Albert model 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 Barabasi-Albert 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 Barabasi-Albert 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 we start from the BA model and adapt it to somehow account for network evolution
  - And Obtain a bit more realistic model

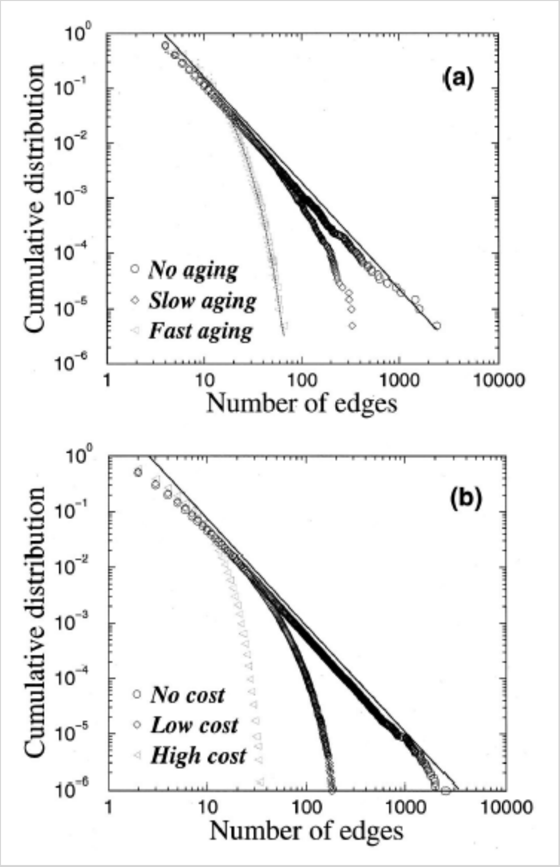
Variations on the Barabasi-Albert Model:
Edges Re-Wiring

- By coupling the model for node additions
  - Adding new nodes at new time interval
- One can consider also mechanisms for edge re-wiring
  - 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 Barabasi-Albert Model:

Aging and Cost

- One can consider that, in real networks (Amaral et al., 2000)
- 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
- Node Aging
  - The possibility of hosting new links decreased with the "age" of the node
  - E.g. nodes get tired or out-of-date
- These two models explain the "exponential cut-off" in power law networks



Variations on the Barabasi-Albert Model:

Fitness

- One can consider that, 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 pages 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

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

Summary

- The Barabasi-Albert model is very powerful to explain the structure of modern networks, but has some limitations
- With the proper extensions (re-wiring, node aging and link costs, fitness)
  - It can capture the structure of modern networks
  - The “rich get richer” phenomenon
  - As well as “the winner takes it all phenomena”
    - In the extreme case, when fitness and node re-wiring are allowed, it may happens that the network degenerates with a single node that attracts all link (monopolistic networks)
- Still, a proper unifying and sound model is missing