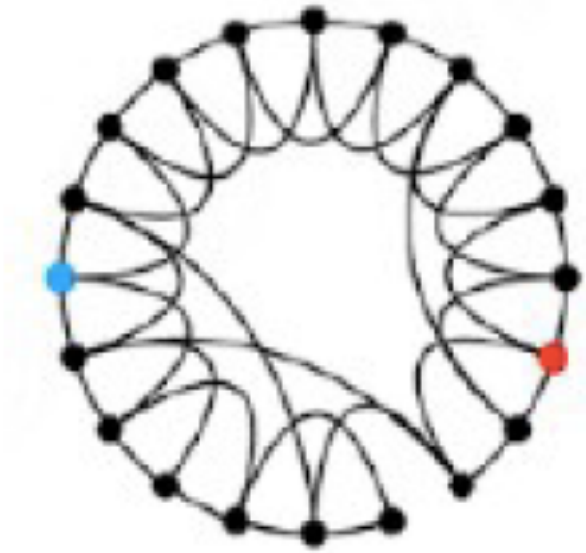
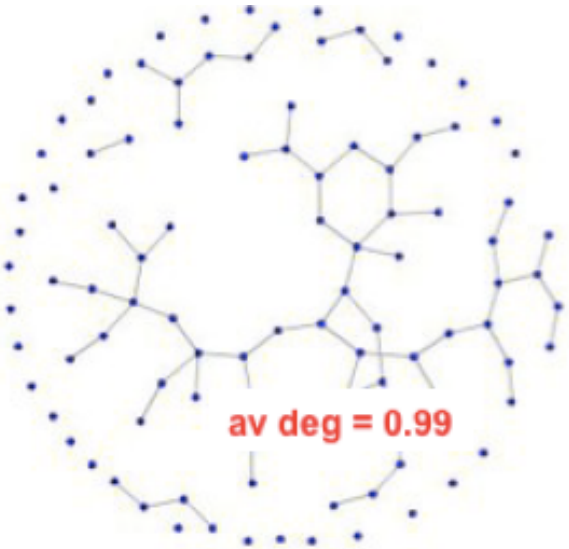
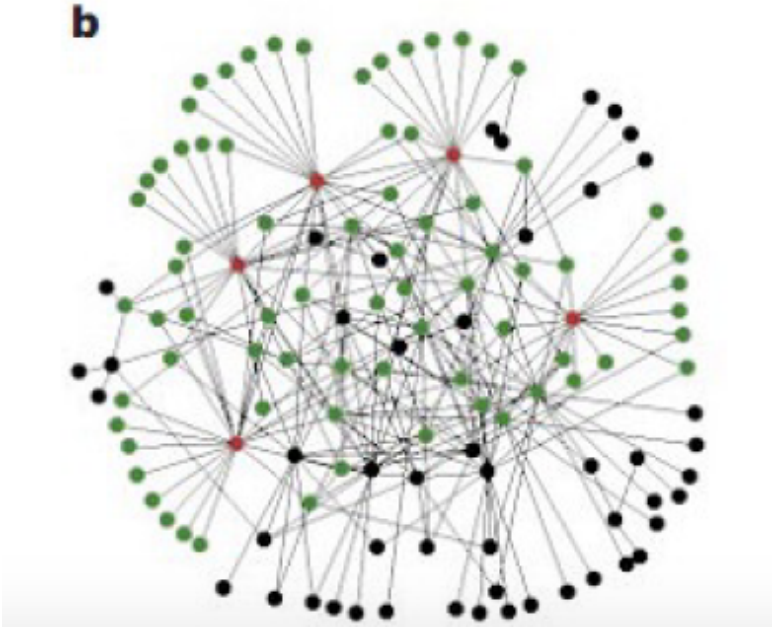


Complex Networks: Some Classical Models

Ana Paula Couto
Computer Science Department
Universidade Federal de Minas Gerais



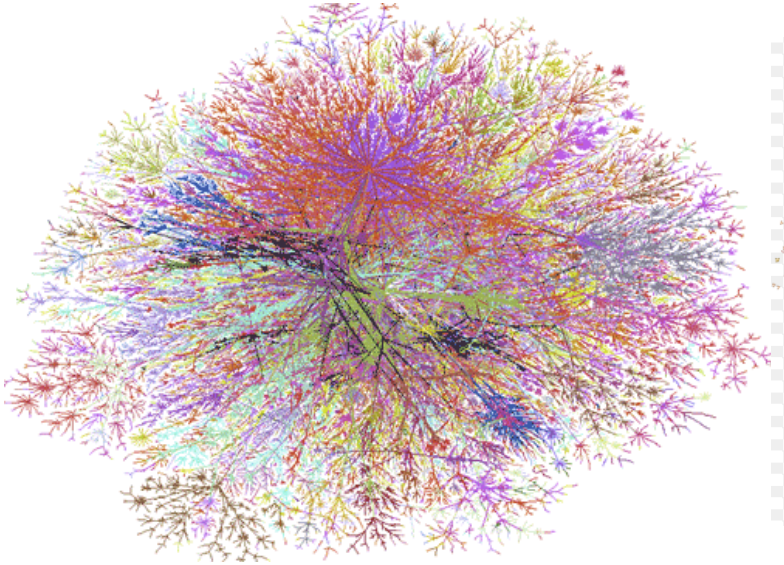


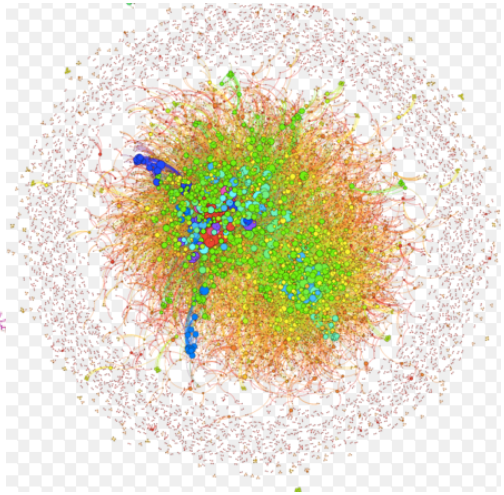
A little remainder…

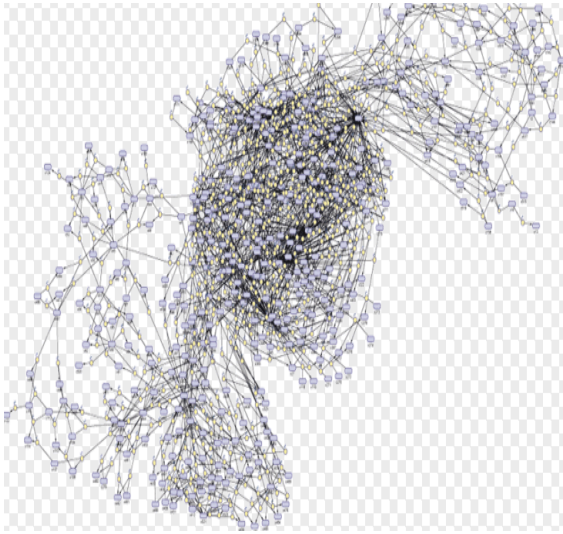
- Complex Networks
  - Try to understand how and why entities are connected with each other and the consequences
- Network structure is essential
  - Important for understanding the phenomena not yet explained

How to explore these networks?

- Graphic representation: A picture is worth a thousand words!
  - Not a good idea for huge networks!

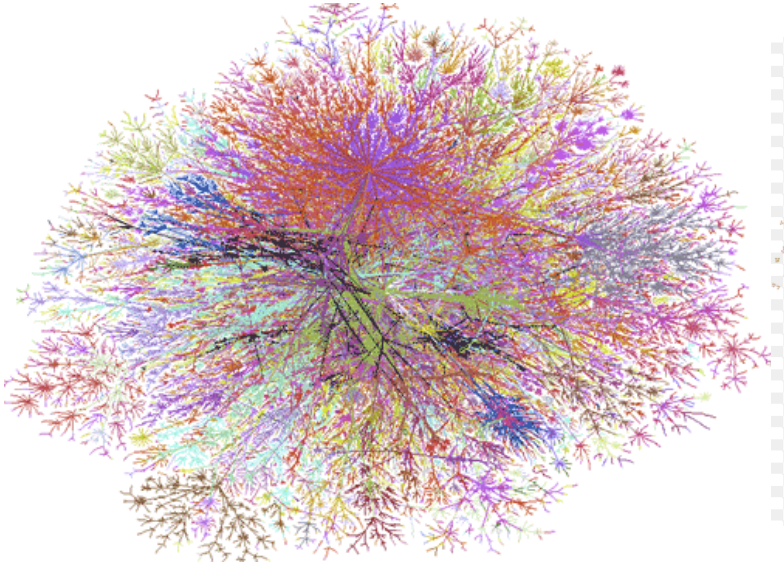


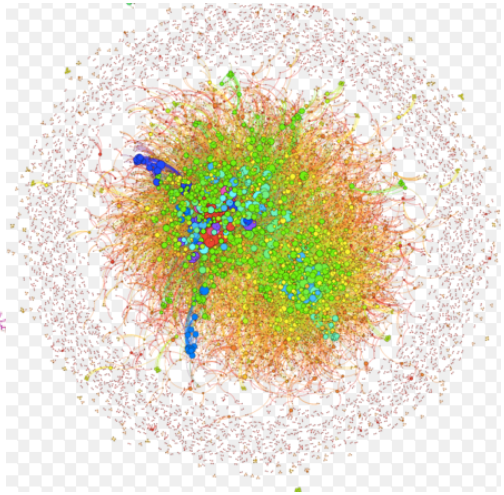


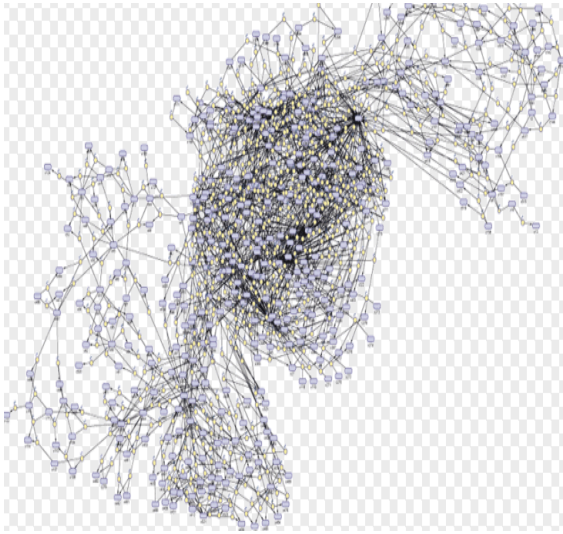


How to explore these networks?

- Describing their structures!
- But, how???







Some important features

- Size, density, degree distribution, diameter, centrality
  - General ideas of the network structures

- What features we should evaluate?
- What are the most important ones?
  - Depends on what we want to study and understand!

Important feature

- What makes a feature to be important?
  - Predicts the behavior of some process?
  - Influences many different processes?

- Example: degree distribution
  - Predicts the random walk behavior
  - Influences the epidemical process

Important feature

Many room for discovering what are the really important features

Why models?

- Simple representation of complex systems
- Proprieties can be written using mathematical stuff
- What-if analysis

- Some questions:
  - What are the differences between the real system and its model?
  - What are the main insights we can get from them?

Main characteristics

- Observed in various real networks
  - beginning of modern empirical studies (98, 99)

- "Small world" effect

- "My friends are also friends" effect

- "Down with democracy" effect

Deterministic models

- Network topology is deterministic
- Proprieties and characteristics are deterministic
  - Diameter can be previously known

- Examples:
  - Complete graph, Hypercube

Probabilistic models

- Random Topology
- Random process defines the graph topology
- Sample space are all possible graphs
- Examples:
  - Random Graphs
    - By Erdos and Reyni, 1959: N labeled nodes are connected with L randomly placed links: G(N,L)
    - By Gilbert, 1959: Each pair of N labeled nodes is connected with probability p: G(N,p)

Random Networks

Random networks: G(N,p)

A random network consists of N labeled nodes where each node pair is connected with the same probability p

How to build a G(N,p)?

1) Start with N isolated nodes
2) Select a node pair, and generate a random number between 0 and 1. With probability p we connect the nodes, with probability (1-p) the nodes are not connected.
3) Repeat step (2) for each of the N(N-1)/2 node pairs

Random networks: Number of links is variable

Each random network we generate with the same parameters *(N, p)* will look slightly different

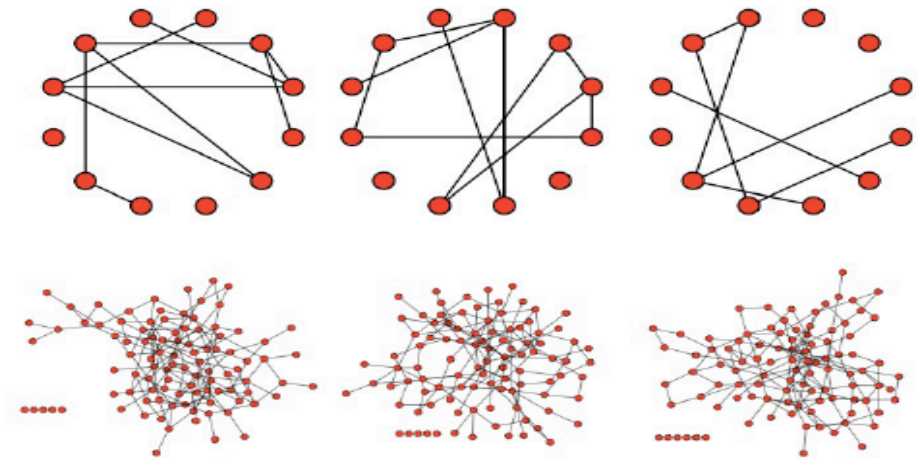


Image 3.3
Random networks are truly random.

Top row: Three realizations of a random network generated with the same parameters $N = 12$ and $p = 1/6$. Despite the identical parameters, the networks not only look different, but they differ in the number of links they have ($L = 8, 10, 7$) and in the degree of the individual nodes.
Bottom row: Three realizations of a random network with $N = 100$ and $p = 1/6$.

Network Science Book – Albert Barabasi

Number of links is variable

The probability that a random network has exactly L links is the product of three terms:

1. The probability that L of the attempts to connect the $N(N-1)/2$ pairs of nodes have resulted in a link, which is p^L.

2. The probability that the remaining $N(N-1)/2 - L$ attempts have not resulted in a link, which is $(1-p)^{N(N-1)/2-L}$

3. A combinational factor, $\binom{\binom{N}{2}}{L}$ counting the number of different ways we can place L links among $N(N-1)/2$ node pairs.

Hence the probability that a particular realization of a random graph has exactly L links is

$$p_L = \binom{\binom{N}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2}-L}. \qquad (1)$$

Network Science Book – Albert Barabasi

Random networks

- Number of links is not fixed, but varies between realizations (simulations)
- Expected value is given by N and p
- From p=0 to p=1, the random network becomes denser and the average number of links increase from $L = 0$ to L_{max} and the average degree of node increases from k=0 to k=N-1

Degree distribution

In a random network the probability that node i has exactly k links is the product of three terms [5]:

- The probability that k of its links are present, or p^k.
- The probability that the remaining $(N - 1 - k)$ links are missing, or $(1-p)^{N-1-k}$.
- The number of ways we can select k links from $N - 1$ potential links a node can have, or $\binom{N-1}{k}$.

Hence the degree distribution of a random network follows the binomial distribution

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}. \qquad (7)$$

Network Science Book – Albert Barabasi

Degree distribution

- The shape of degree distribution depends on the system size N and the probability p
- Mean degree and its variance can be determined
- Most real networks are sparse, and ***<k> << N***
  - Binomial distribution is approximated by the Poisson distribution, with N➔∞

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!},$$

Degree distribution

- Both distributions have a peak around <*k*>. If we keep N constant and increase p, the network becomes denser: peak moves to the right
- The dispersion of the distribution is controlled by p or <*k*>. The denser the network, the wider is the distribution

Degree distribution

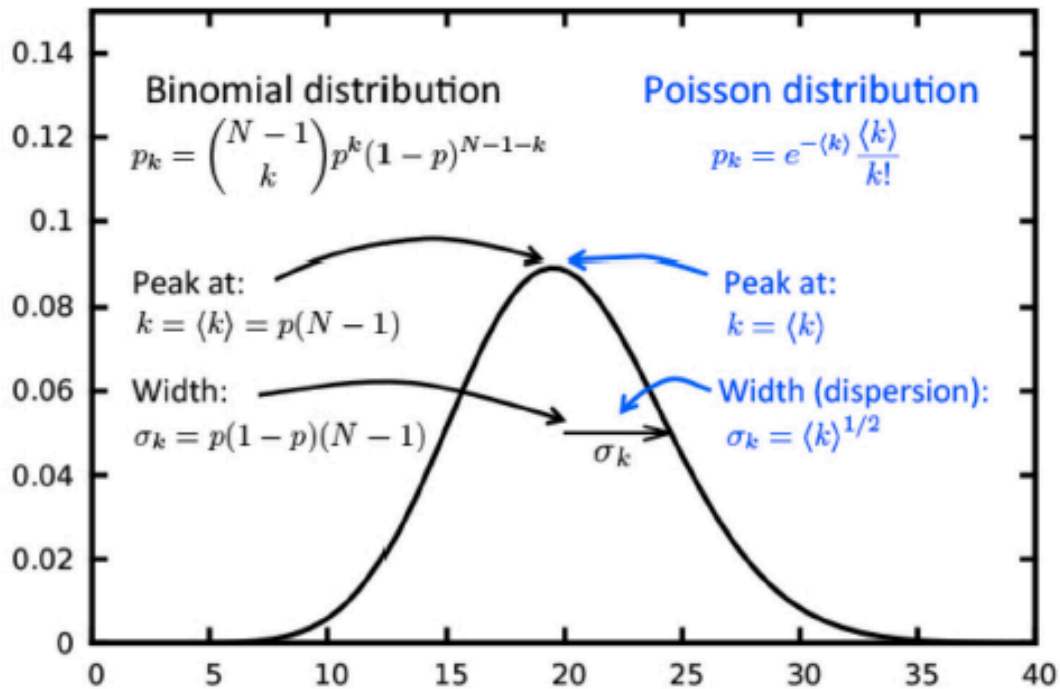


Image 3.4a

Anatomy of a binomial and a Poisson degree distribution.

The exact form of the degree distribution of a random network is the binomial distribution (left). For $N \gg \langle k \rangle$, the binomial can be well approximated by a Poisson distribution (right). As both distributions describe the same quantity, they have the same properties, which are expressed in terms of different parameters: the binomial distribution uses p and N as its fundamental parameters, while the Poisson distribution has only one parameter, $\langle k \rangle$.

Network Science Book – Albert Barabasi

Degree distribution

- Poisson distribution represents only an approximation to the binomial distribution in the case k<<N limit. For most networks of practical importance this condition is easily satisfied
- The advantage of the Poisson form is that key network characteristics have a much simpler form, depending on a single parameter, $\langle k \rangle$

Degree distribution

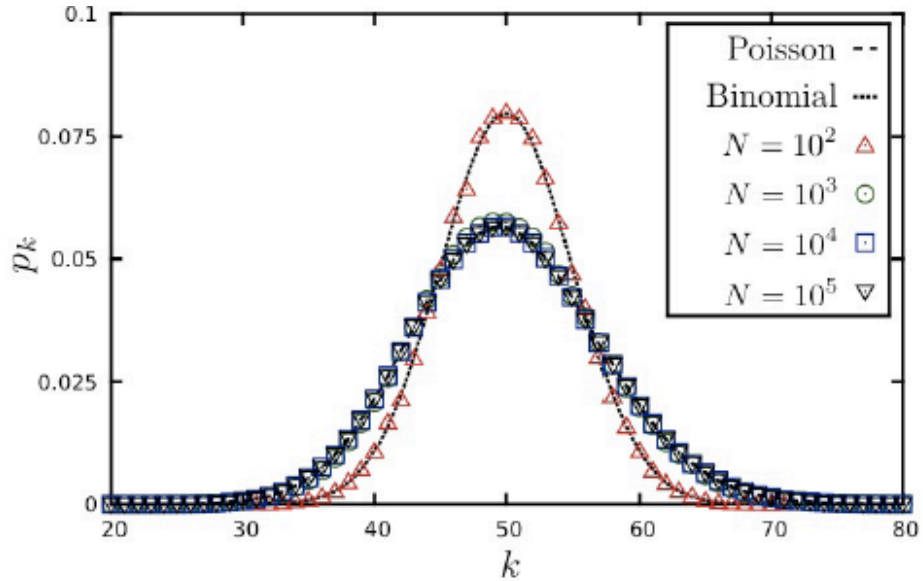


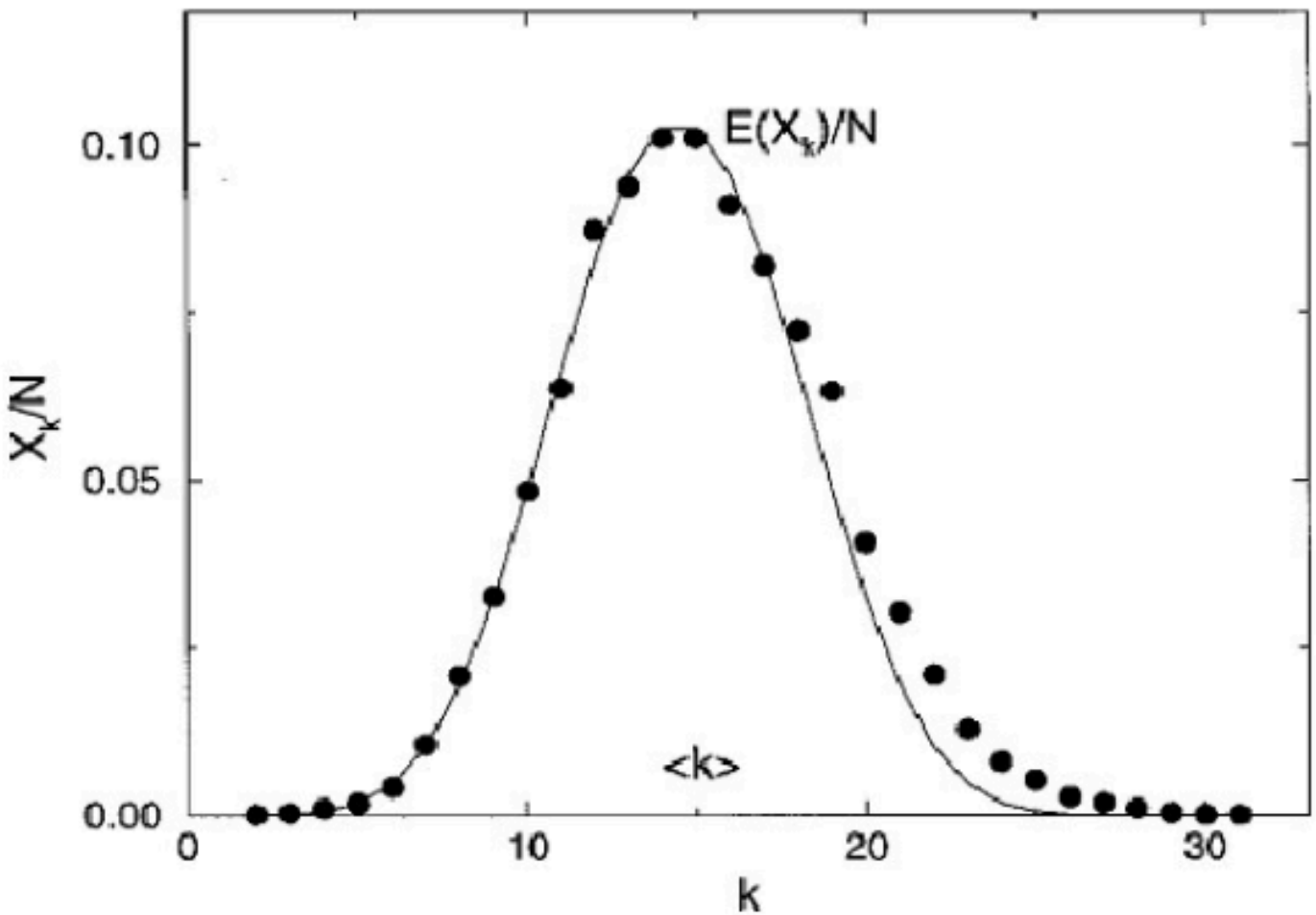
Image 3.4b

Degree distribution is independent of the network size.

The degree distribution of a random network with average degree $\langle k \rangle = 50$ and sizes $N = 10^2$, 10^3, 10^4. For $N = 10^2$ the degree distribution deviates significantly from the Poisson prediction (8), as the condition for the Poisson approximation, $N \gg \langle k \rangle$, is not satisfied. Hence for small networks one needs to use the exact binomial form of Eq. (7) (dotted line). For $N = 10^3$ and larger networks the degree distribution becomes indistinguishable from the Poisson prediction, (8), shown as a continuous line, illustrating that for large N the degree distribution is independent of the network size. In the figure we averaged over *1,000* independently generated random networks to decrease the noise in the degree distribution.

Network Science Book – Albert Barabasi

Degree distribution



R. Albert and A.-L. Barabási: Statistical mechanics of complex networks

Diameter

- For the random networks:

$$d_{max} \propto \frac{\log N}{\log \langle k \rangle}$$

Represents the quantitative formulation of the small world phenomenon!

Diameter

- Small world implies that average path length or the diameter depends logarithmically on the system size

- Small means that <d> is proportional to log N, rather than N or some power of N

Diameter

Network Name	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\frac{\log N}{\log \langle k \rangle}$
Internet	192,244	609,066	6.34	6.98	26	6.59
WWW	325,729	1,497,134	4.60	11.27	93	8.32
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	186,936	8.08	5.35	15	4.81
Actor Network	212,250	3,054,278	28.78	-	-	-
Citation Network	449,673	4,707,958	10.47	11.21	42	5.55
E Coli Metabolism	1,039	5,802	5.84	2.98	8	4.04
Yeast Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

Table 3.2

Six degrees of separation.

The average distance $\langle d \rangle$ and the maximum distance d_{max} of the ten networks explored in this book. The last column provides $\langle d \rangle$ predicted by Eq. (19), indicating that it offers a reasonable approximation to $\langle d \rangle$. Yet, the agreement is not perfect – we will see in the next chapter that for many real networks Eq. (19) needs to be adjusted. For directed networks we list the average out-degree $\langle k_{out} \rangle$ and the path lengths are measured only along the direction of the links.

Clustering coefficient

- Local clustering coefficient C_i captures the density of links in node i's neighborhood:
  - C = 0: no links between i's neighbors
  - C = 1: each of the i's neighbors link to each other

Clustering coefficient

- To calculate C_i for a node in a random network
  - Estimate the expected number of links L_i between the node's k_i neighbors
  - The probability that two of i's neighbors link to each other is p
  - So:

$$\langle L_i \rangle = p\frac{k_i(k_i - 1)}{2}.$$

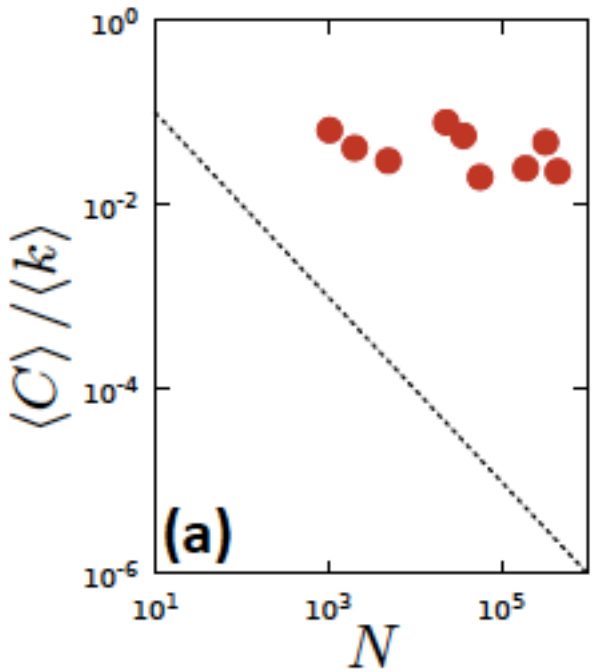
  - The local clustering coefficient of a random graph:

$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

Clustering coefficient

- For fixed $\langle k \rangle$, the larger the network, the smaller is a node's clustering coefficient
- The local clustering coefficient of a node is independent of the node's degree

Clustering coefficient



Clustering in real networks.

(a) Comparison between the average clustering coefficient of real networks and the prediction Eq. (21) for random networks. Each circle corresponds to a network from Table 3.2. Directed network were made undirected to calculate C. The dashed line corresponds to Eq. (21), predicting that for random networks the average clustering coefficient should decrease as N^{-1}. In contrast, for real networks $\langle C \rangle$ has only a weak dependence on N.

Random networks

The model suggests that if a network is not as regular as a square lattice: it is random

Do we really believe that real networks are random?

Real networks are not random

- There is a deep order behind most complex systems
  - Interactions between our proteins are governed by the strict laws of biochemistry
- The order must be reflected in the structure of the network that describes their architecture, resulting in systematic deviations from a pure random configuration

How random networks fail to describe real systems?

- Degree distribution
  - In real systems, there are more highly connected nodes than the random network model could account for
- Clustering coefficient
  - In real systems, clustering coefficient is largely independent of system size and it decreases with the node degrees.

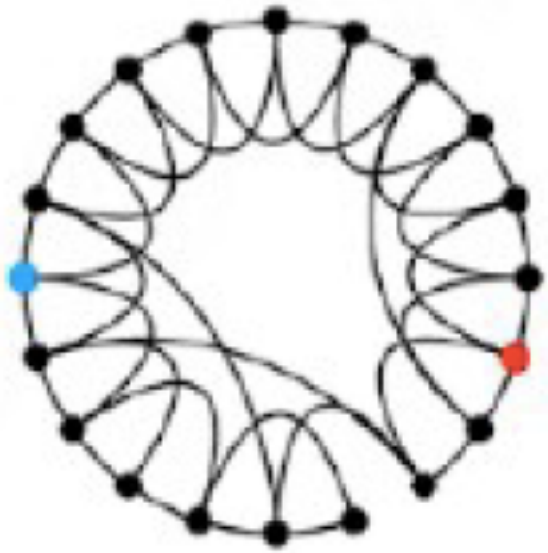
"Down with democracy" effect

"My friends are also friends" effect

How random networks fail to describe real systems?

- The small world phenomena is the only property reasonably explained by the random network model
- Random models serves as a fundamental reference as we try to understand the properties of real networks

Small-World Networks



Small world networks

- Based on Milgram's experiment: small diameter
- High clustering coefficient: local triangles
  - High probability of node's neighbors being connected each other
- Proposed by Watts and Strogatz (1998)

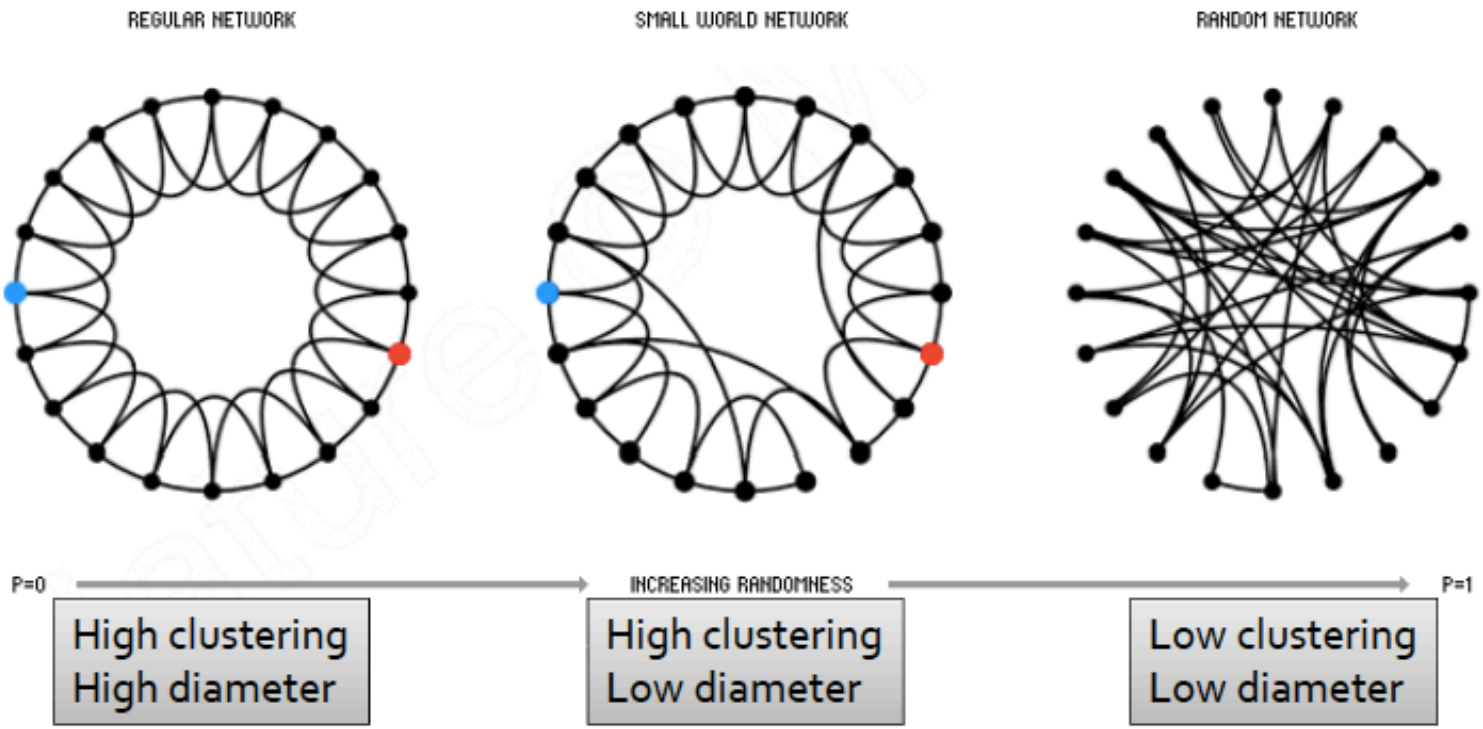
Small world networks

Small world and high clustering coefficient properties can co-exist

How to build a Small world?

1) Begin with a regular lattice
2) Rewire
   - For each link, with probability p, change one of the vertices

Small world networks

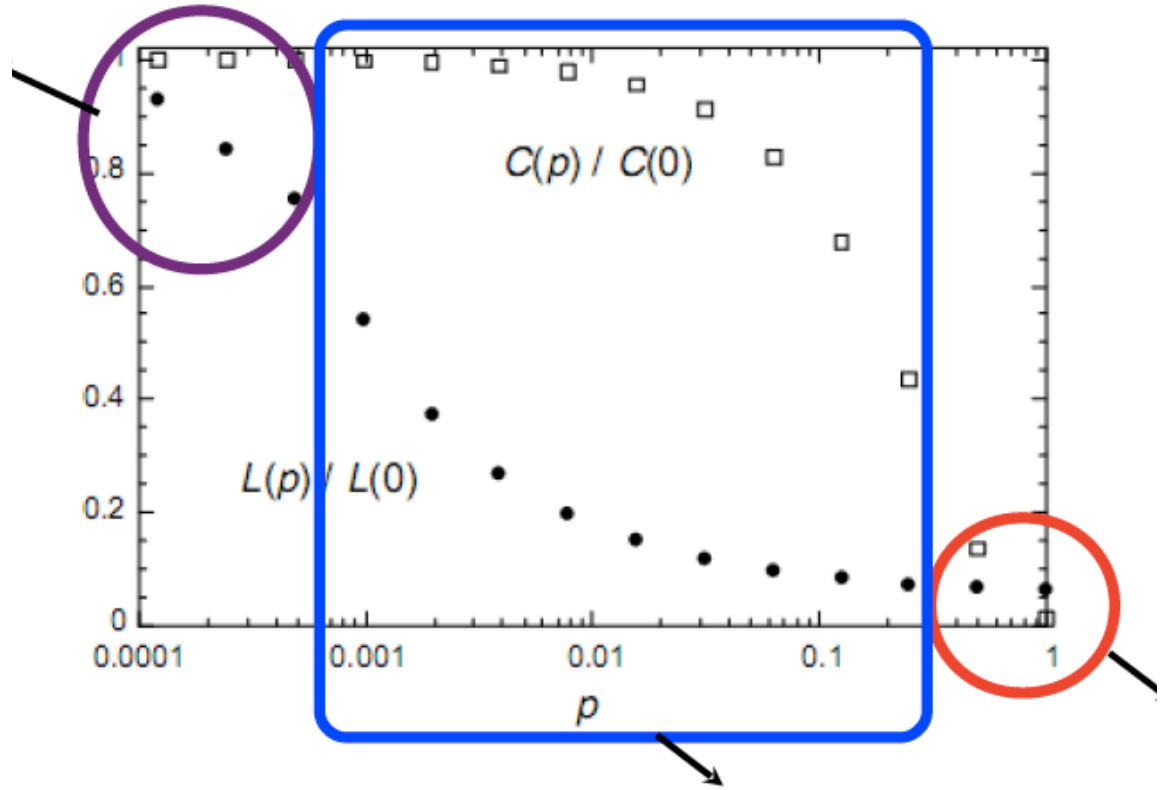


Rewiring allows to interpolate between regular lattice and a random graph

What means to rewire links? (Social networks)

- People have many local connections that are also connected each other (neighbors, co-workers, friends, family): high clustering coefficient
- People also have global connections (other countries, cities, communities of interest), established by rewiring links: small diameter

Small world networks



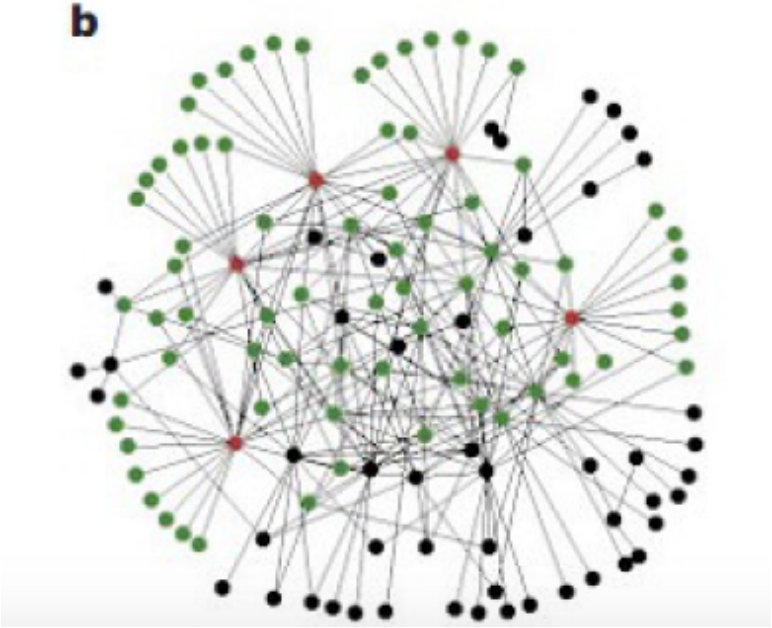
Collective dynamics of 'Small-World' networks,
D. Watts & S. Strogatz, Nature 393, 1998

How small-world networks fail to describe real systems?

"Down with democracy" effect

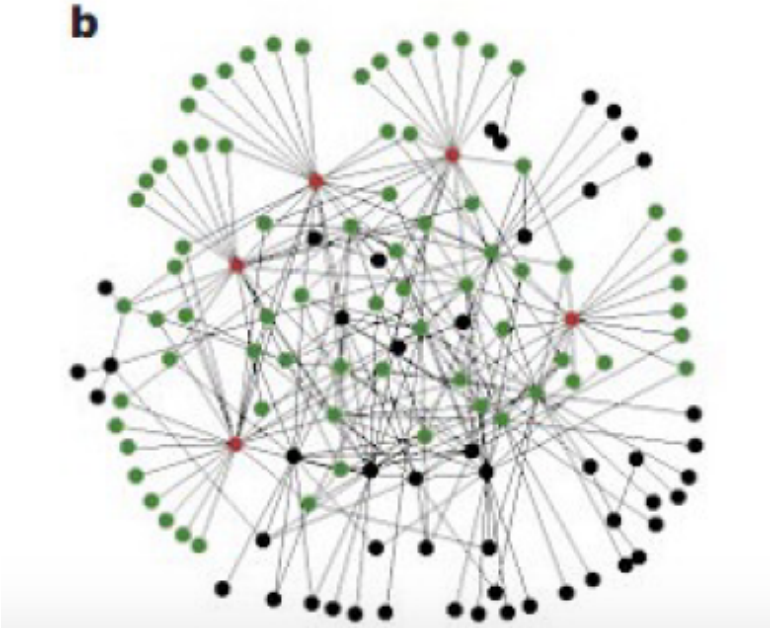
- Degree distribution
  - In real systems, there are more highly connected nodes than the random network model could account for
- Dynamics
  - Evolution over time is not considered: total number of nodes and links are static

Barabási-Albert Networks



Real networks and hubs

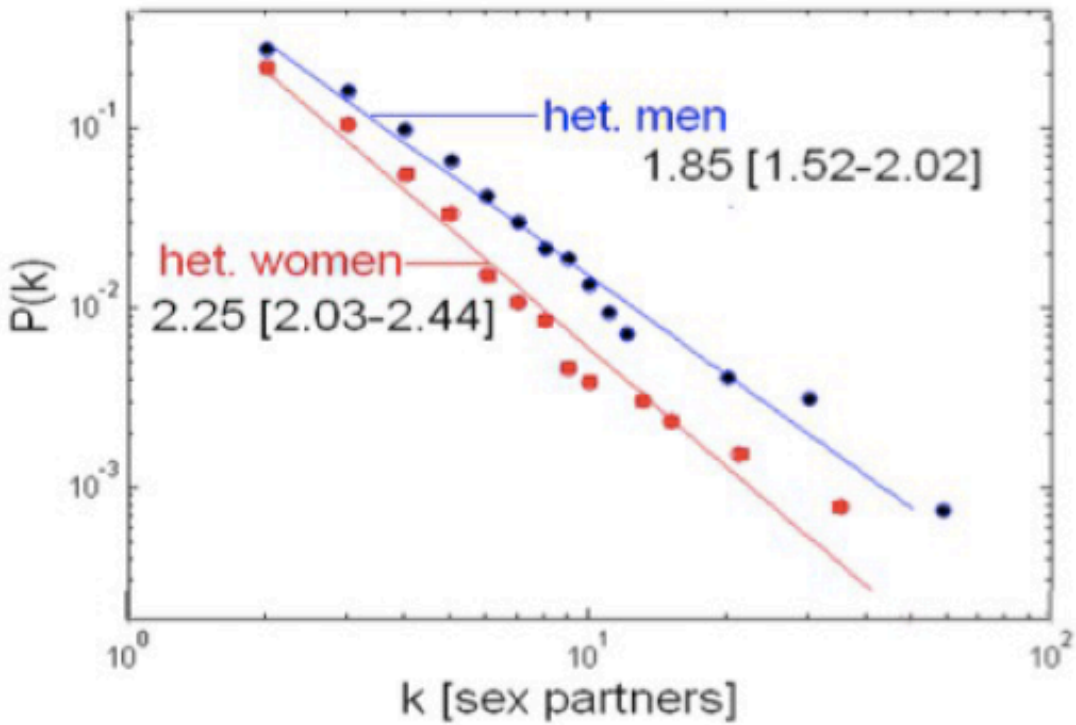
- Real networks have a small number of nodes with many connections and a large number of nodes with few connections



Red nodes have many neighbors

Example

- Sexual networks
- Great variation in contact numbers

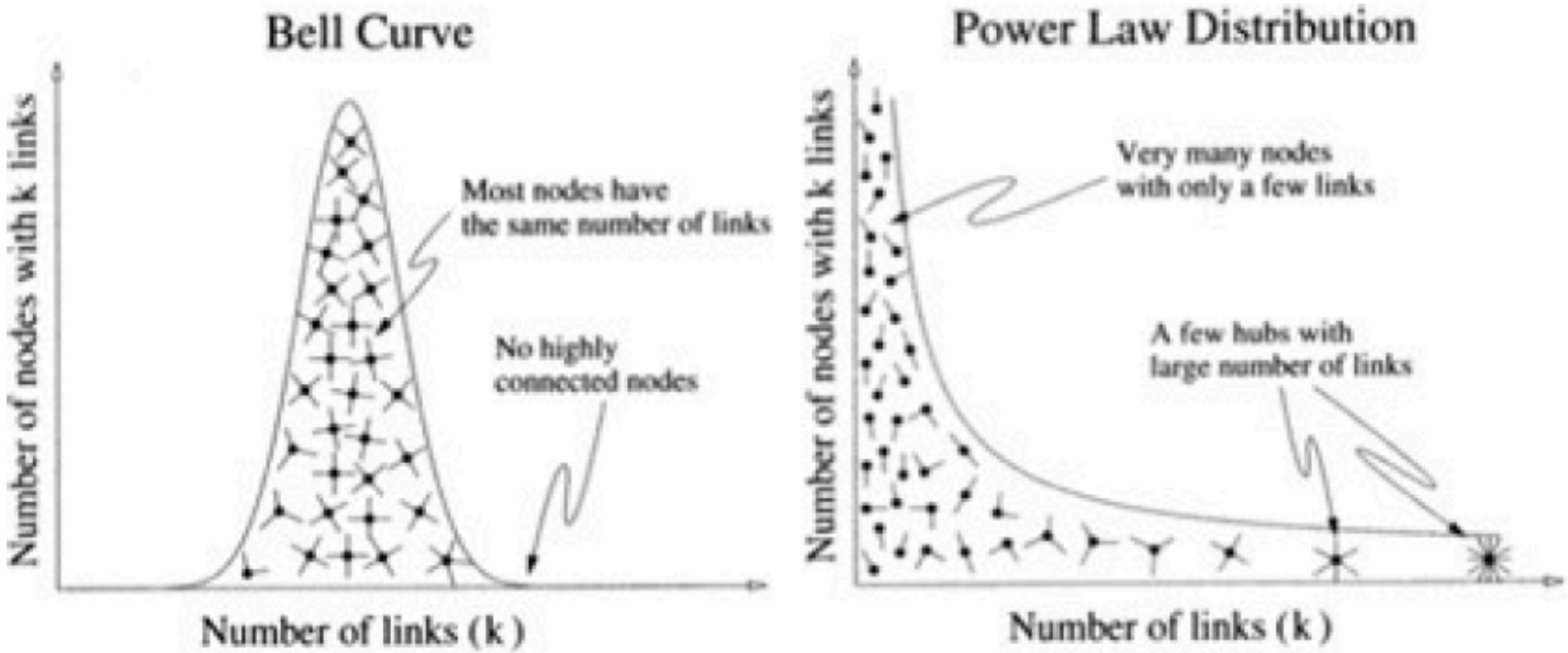


Real networks and hubs

- Degree distribution follows a power-law distribution (scale-free networks)

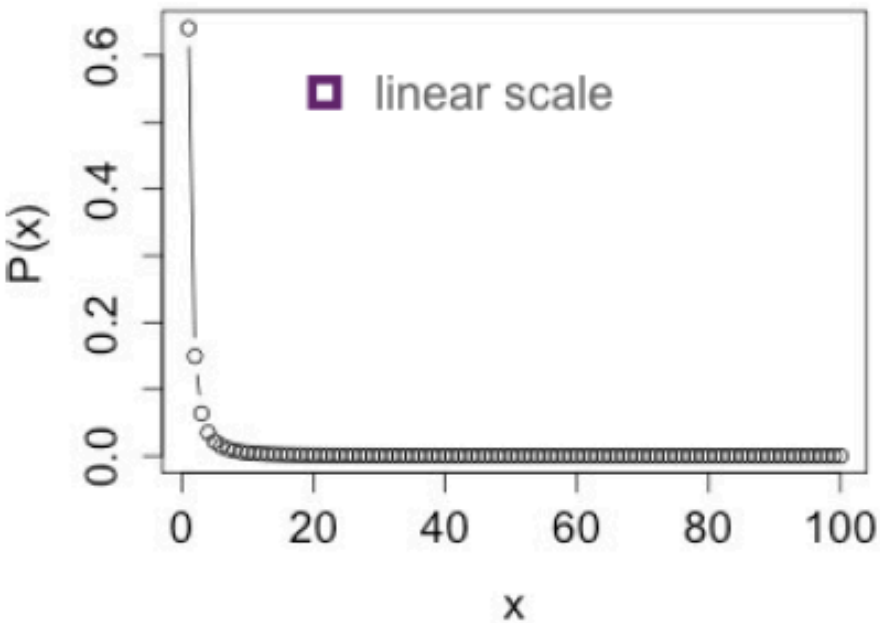
$$P[D = k] \sim k^{-\gamma}$$

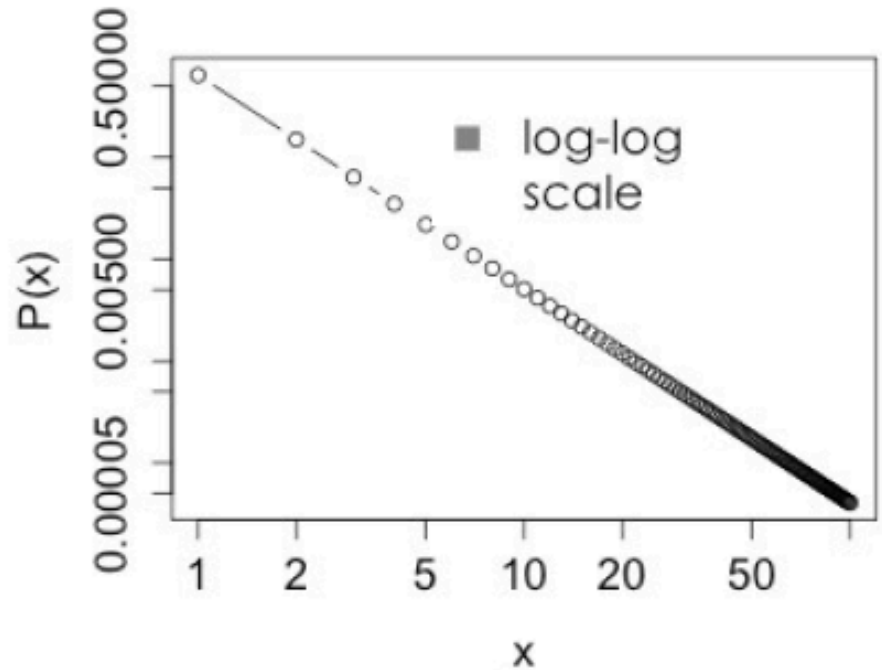
Power law distribution



Fonte: *Linked: The New Science of Networks*, A. Barabási, 2002

Power law distribution





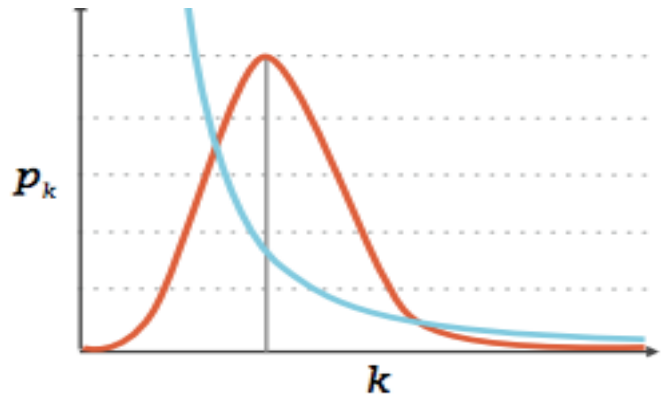
Why scale-free?

- For a random network with a Poisson degree distribution, the degree variance is always smaller than $\langle k \rangle$
- The degrees are in the range $k = \langle k \rangle \pm \langle k \rangle^{1/2}$, indicating that nodes in a random network have comparable degrees
- Therefore the average degree $\langle k \rangle$ servers as the "scale" of a random network

Why scale-free?

- The average degree of the WWW sample is $\langle k \rangle = 4.6$. However, the second moment diverges
- This means that our expectation for the in-degree of a randomly chosen WWW document is $\langle k \rangle = 4.60 \pm \infty$
- A randomly chosen webpage could easily yield a document of degree one or two, as 74.02% of nodes have in-degree less than $\langle k \rangle$
- It could also yield a node with hundreds of millions of links, like google.com or facebook.com

Why scale-free?



Random network
Randomly chosen node: $k = \langle k \rangle \pm \langle k \rangle^{1/2}$
Scale: $\langle k \rangle$

Scale-free network
Randomly chosen node: $k = \langle k \rangle \pm \infty$
$\langle k \rangle$ is meaningless as 'scale'

Figure 4.7
Scale-free networks lack an internal scale

For any bounded distribution (e.g. a Poisson or a Gaussian distribution) the degree of a randomly chosen node will be in the vicinity of $\langle k \rangle$. Hence $\langle k \rangle$ serves as the network's scale. In a scale-free network the second moment diverges, hence the degree of a randomly chosen node can be arbitrarily different from $\langle k \rangle$. As a scale-free network lacks an intrinsic scale, is it scale-free.

Network Science Book – Albert Barabasi

Why scale-free?

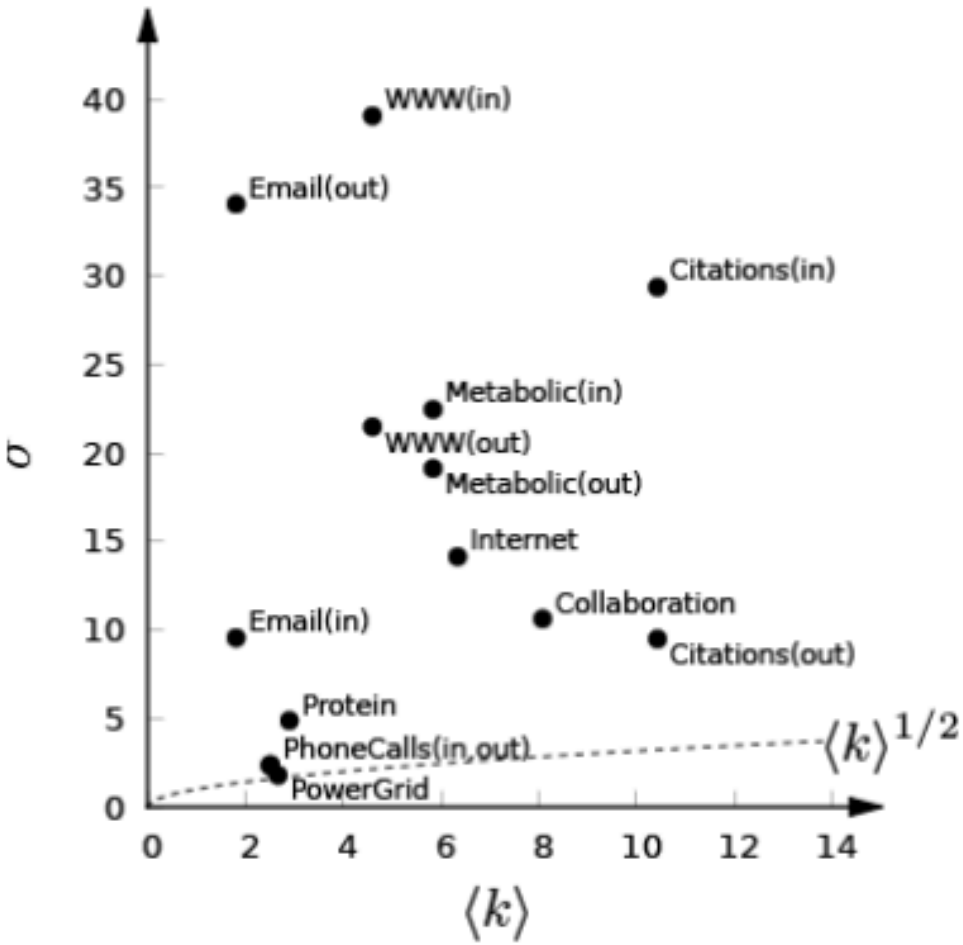


Figure 4.8
Standard deviation is large in real networks

For a random network the standard deviation follows $\sigma_k = \sqrt{\langle k \rangle}$ shown as a dashed line on the figure. The symbols show σ for ten reference networks Table 4.1, indicating that for each σ is larger than expected for a random network with similar $\langle k \rangle$. The only exception is the power grid, which is not scale-free. While the phone call network is scale-free, it has a large γ, hence it behaves like a random network.

Network Science Book – Albert Barabasi

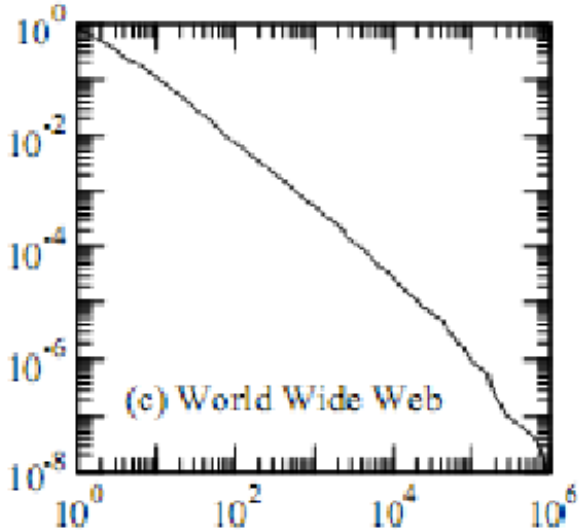
Examples of Scale-Free Networks

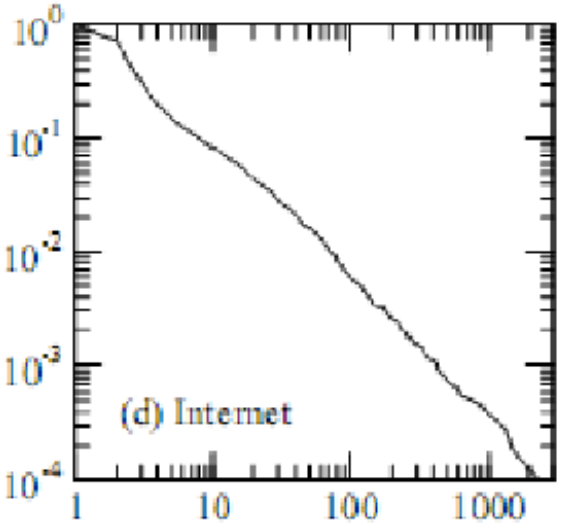
NETWORK	NODES	LINKS
Cellular metabolism	Molecules involved in burning food for energy	Participation in the same biochemical reaction
Hollywood	Actors	Appearance in the same movie
Internet	Routers	Optical and other physical connections
Protein regulatory network	Proteins that help to regulate a cell's activities	Interactions among proteins
Research collaborations	Scientists	Co-authorship of papers
Sexual relationships	People	Sexual contact
World Wide Web	Web pages	URLs

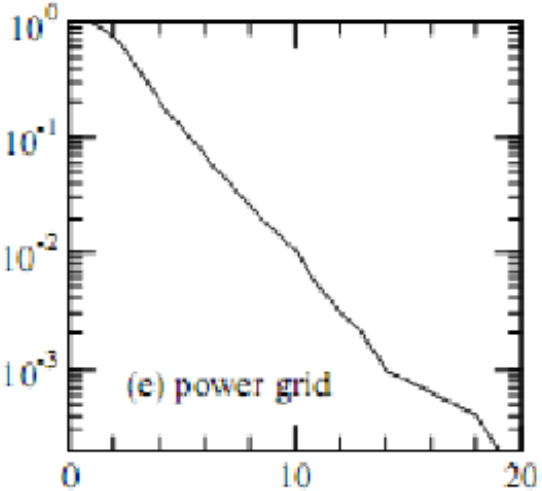
ALBERT-LÁSZLÓ BARABÁSI AND ERIC BONABEAU

Scale-free networks

Degree distribution







Real networks have $2<\gamma<3$

Barabási-Albert model

- Proposed by Barabási & Albert in 1999:
  - Emergence of Scaling in Random Networks, Science
- Dynamics
  - Growth: Total number of nodes is a result of a growth process, that continuously increases N
  - Preferential attachment: New nodes prefer to link to more connected nodes

How to build a Barabási-Albert model?

(A) GROWTH

At each timestep we add a new node with m ($\leq m_0$) links that connect the new node to m nodes already in the network.

(B) PREFERENTIAL ATTACHMENT

The probability $\Pi(k)$ that one of the links of the new node connects to node i depends on the degree k_i of node i as

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \tag{5.1}$$

Barabási-Albert model

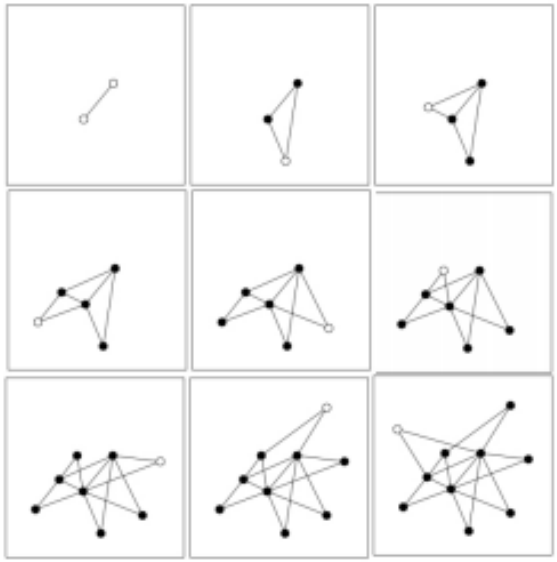
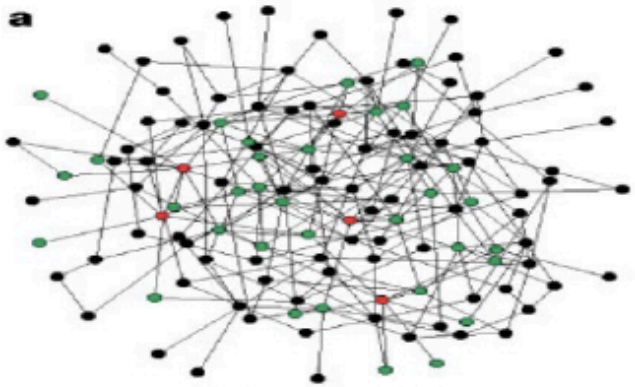


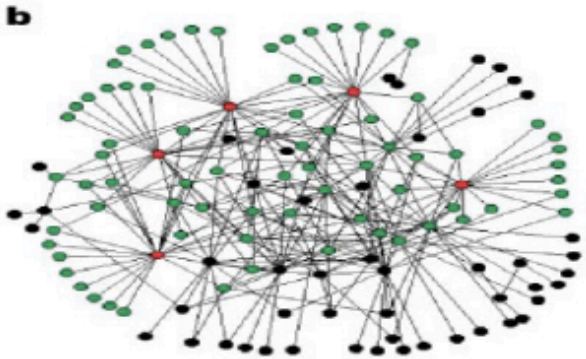
Figure 5.3
Time evolution of the Barabási-Albert model

The sequence of images shows the gradual emergence of a few highly connected nodes, or hubs, through growth and preferential attachment. White circles mark the newly added node to the network, which decides where to connect its two links (m=2) through preferential attachment Eq. 5.1. After [9].

Random versus Barabási-Albert



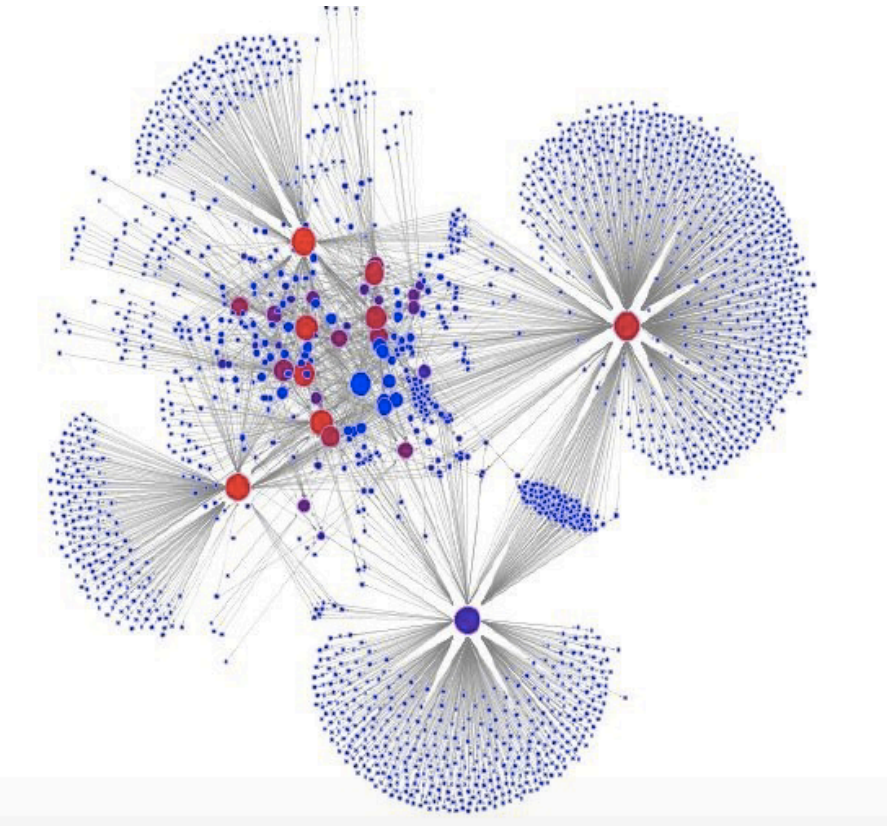
- Random network
- 130 nodes, 215 edges
- Homogeneous: most nodes have approximately the same number of links
- Five red nodes with the highest number of links reach 27% of the nodes



- Scale-free network
- 130 nodes, 215 edges
- Heterogeneous: the majority of the nodes have one or two links but a few nodes have a large number of links
- Five red nodes with the highest degrees reach 60% of the nodes (**hubs**)

Albert et al., Nature, 406:378, 2000

Preferential attachment



Rich-gets-richer phenomenon: due to preferential attachment new nodes are more likely to connect to the more connected nodes than to the smaller degree nodes.

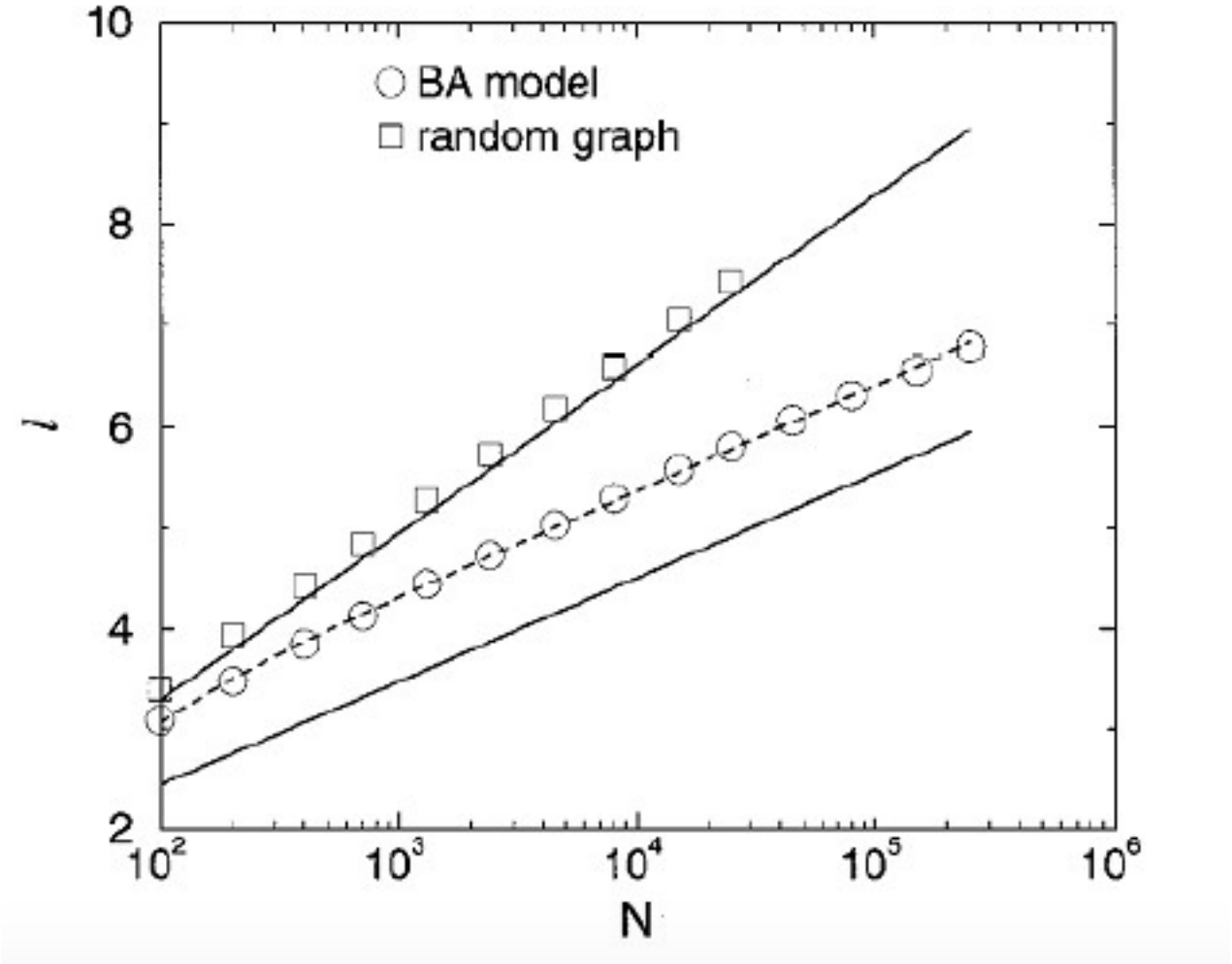
Diameter

- Diameter value is smaller than that of random network

$$\langle l \rangle \sim \frac{\log N}{\log \log N}$$

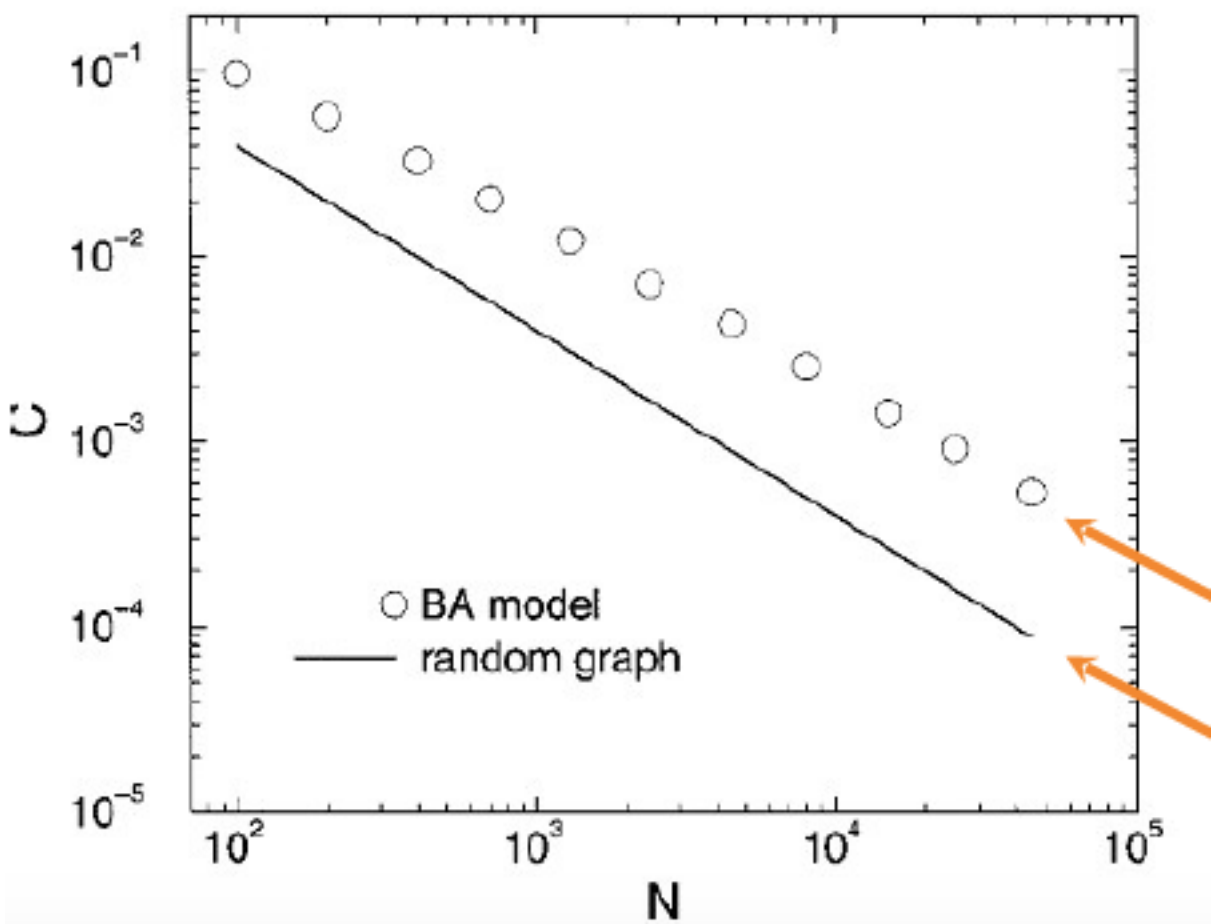
- Hubs: nodes get closer!
- Information diffusion is more efficient

Diameter



Clustering coefficient

- Larger than clustering coefficient of random network



Popular network features measured

1. Degree distribution: what fraction of nodes have degree k?
   power-law degree distribution

2. Network distances: what is the average shortest path distance between two nodes (in the largest connected component)?
   logarithmic in the number of nodes

3. Clustering: does the network exhibit community structure?
   high clustering coefficient

4. Connected components: are nodes disconnected from each other?
   one giant connected component, many smaller components

http://www.cs.cmu.edu/~02317/slides/lec_17.pdf

Model	Power-law?	Short paths?	Clustering?
Erdos-Renyi	No	Yes	No
Watts-Strogatz	No	Yes	Yes
LPA	Yes	Yes	No

Barabasi-Albert