Networks:
Degree Distribution
Random Graphs
Small world

Elsa Arcaute

Outline

1. Definitions: node, link, degree, etc.
2. Undirected, weighted, directed networks
3. Adjacency matrix, paths and connectivity
4. Centrality measures: closeness, betweenness, etc.
5. Clustering, similarity and modularity
6. Community detection
7. Degree distribution
8. Scale-free networks
9. Preferential attachment
10. Random graphs
11. Small-world
12. Spatial networks

This lecture

Degree distribution

Degree distribution

Why?

- In my system, are all the agents equally important?
- Are my spaces homogeneous so that I can set up my business anywhere?
- Which airport should I fly from if I want to increase the chances of getting to my destination given possible flight restrictions, e.g. imposed by COVID?
- Is trading fair?
- Where should I place my mitigation strategy to reach more people?

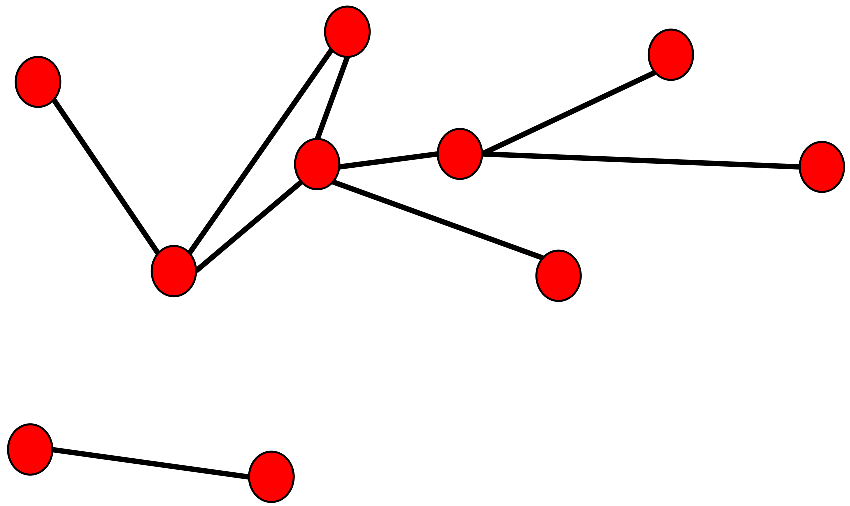
Degree distribution

What is it?

p_k: fraction of vertices with degree k
n_k: number of vertices with degree k

$n=10$

$n_0=0$ $\qquad p_1=6/10$
$n_1=6$ $\qquad p_2=1/10$
$n_2=1$ $\Rightarrow$ $\qquad p_3=2/10$
$n_3=2$ $\qquad p_4=1/10$
$n_4=1$ $\qquad p_k=0$, $k>4$ and $k=0$



$$\boxed{\sum_k p_k = 1}$$

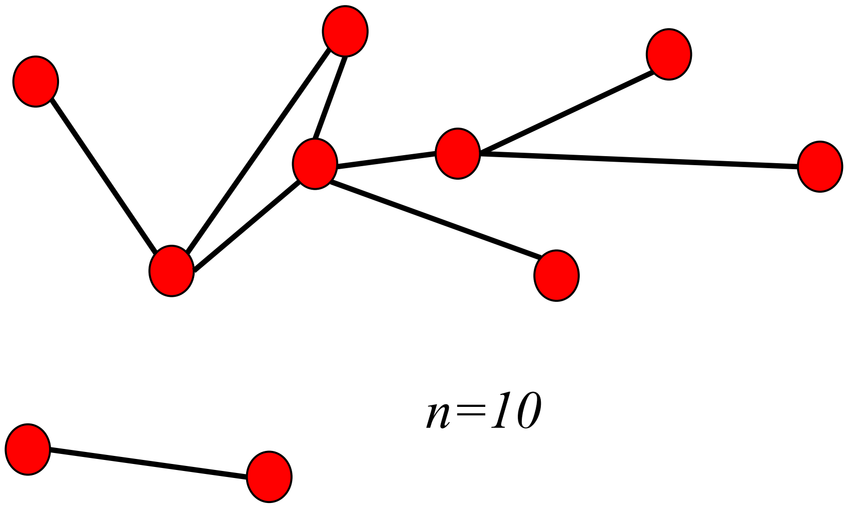
Slide from Barabási's lectures
https://www.barabasilab.com/course

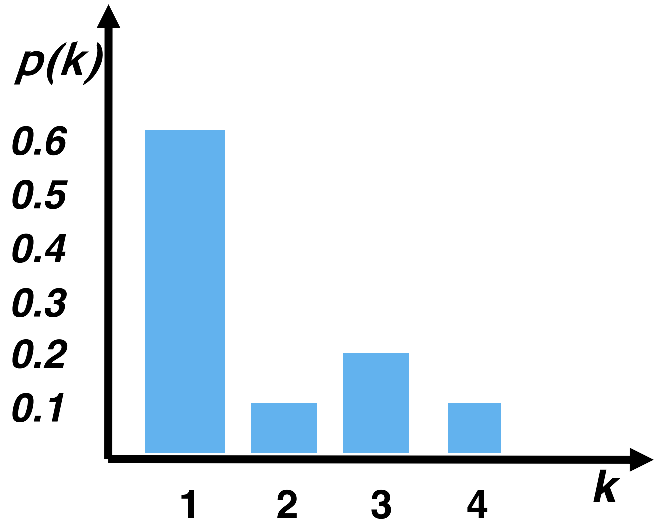
Degree distribution

p(k): probability that a randomly chosen vertex has degree k

n_k = # nodes with degree k

$p(k) = n_k / n$ ➔ plot





$n=10$

$n_0=0$
$n_1=6$
$n_2=1$
$n_3=2$
$n_4=1$

$p_1=6/10$
$p_2=1/10$
$p_3=2/10$
$p_4=1/10$
$p_k=0$, $k>4$ and $k=0$

Degree distribution: directed networks

Proceed to compute the in-degree AND the out-degree distribution for a directed network.

→ A joint distribution from both in- and out-degrees can be constructed

→ E.g. p_{jk}: fraction of vertices having simultaneously in-degree *j* and out-degree *k*.

→ get 2-d distribution that leads to a surface not a histogram

→ Can observe correlation between in- and out-degrees

A degree distribution can be represented through many different networks

→ We cannot infer the complete structure of a network from its degree distribution

Experiment

Consider a network that has n=8 nodes with degrees: {0,1,1,1,1,2,2,2}.

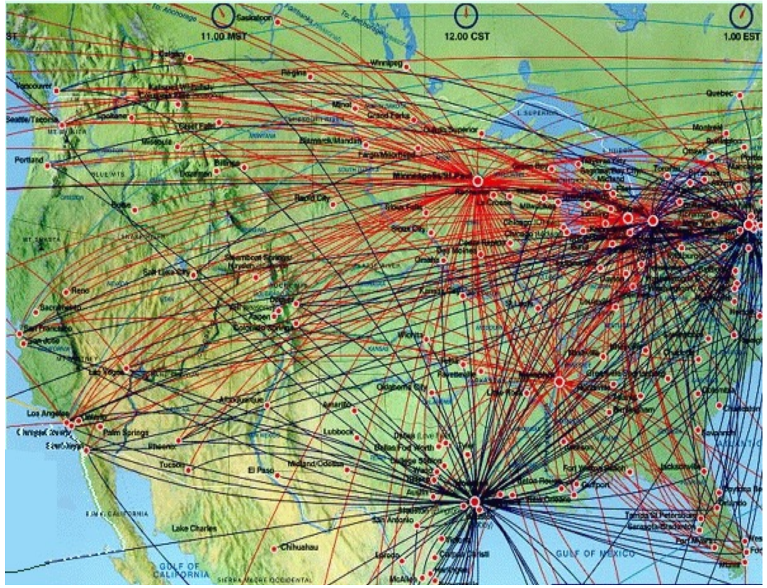
Take a piece of paper and draw the network, considering that it can be a disconnected network.

→ You will find many different solutions for these same properties.

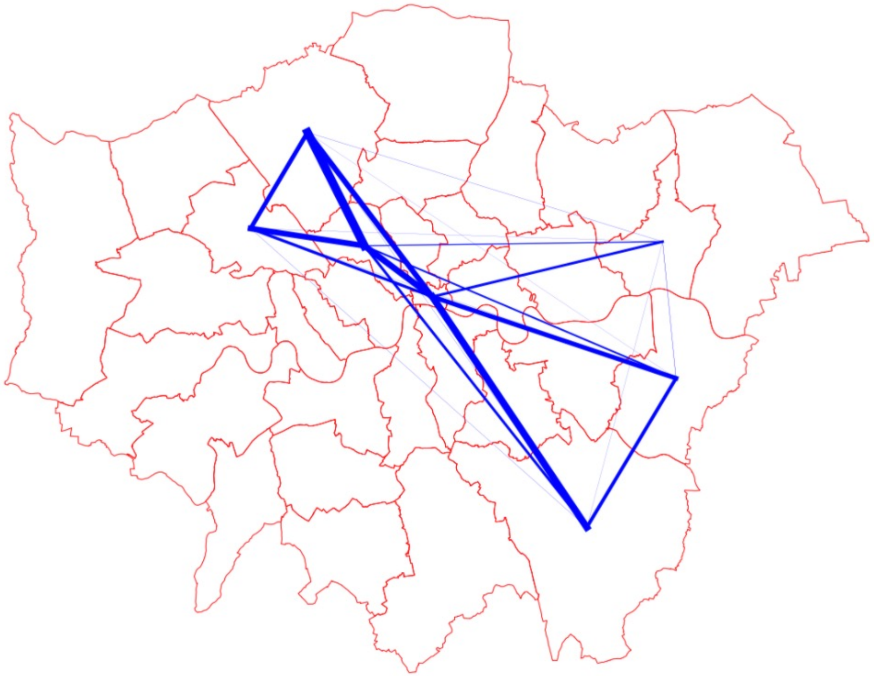
Real world networks

Road network

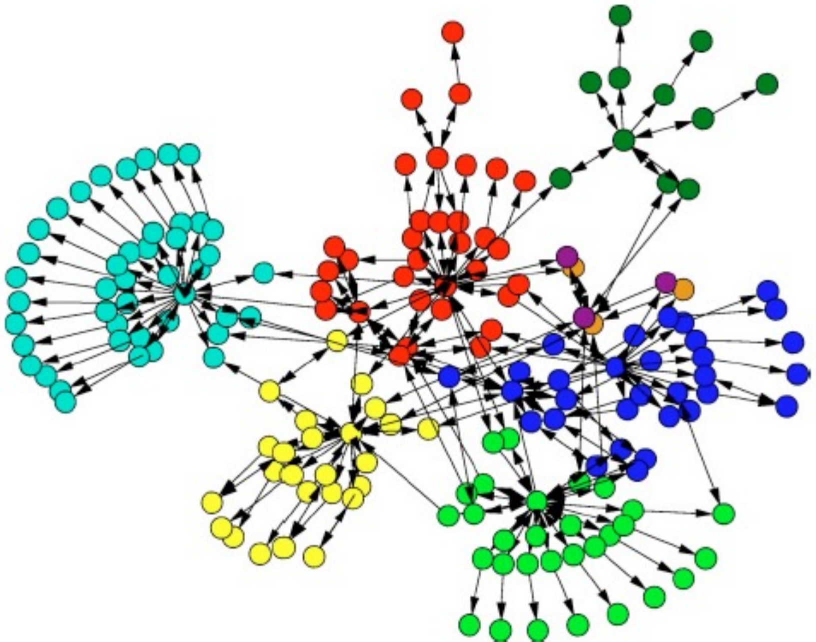
Flights network



Source pictures: Barábasi's lectures



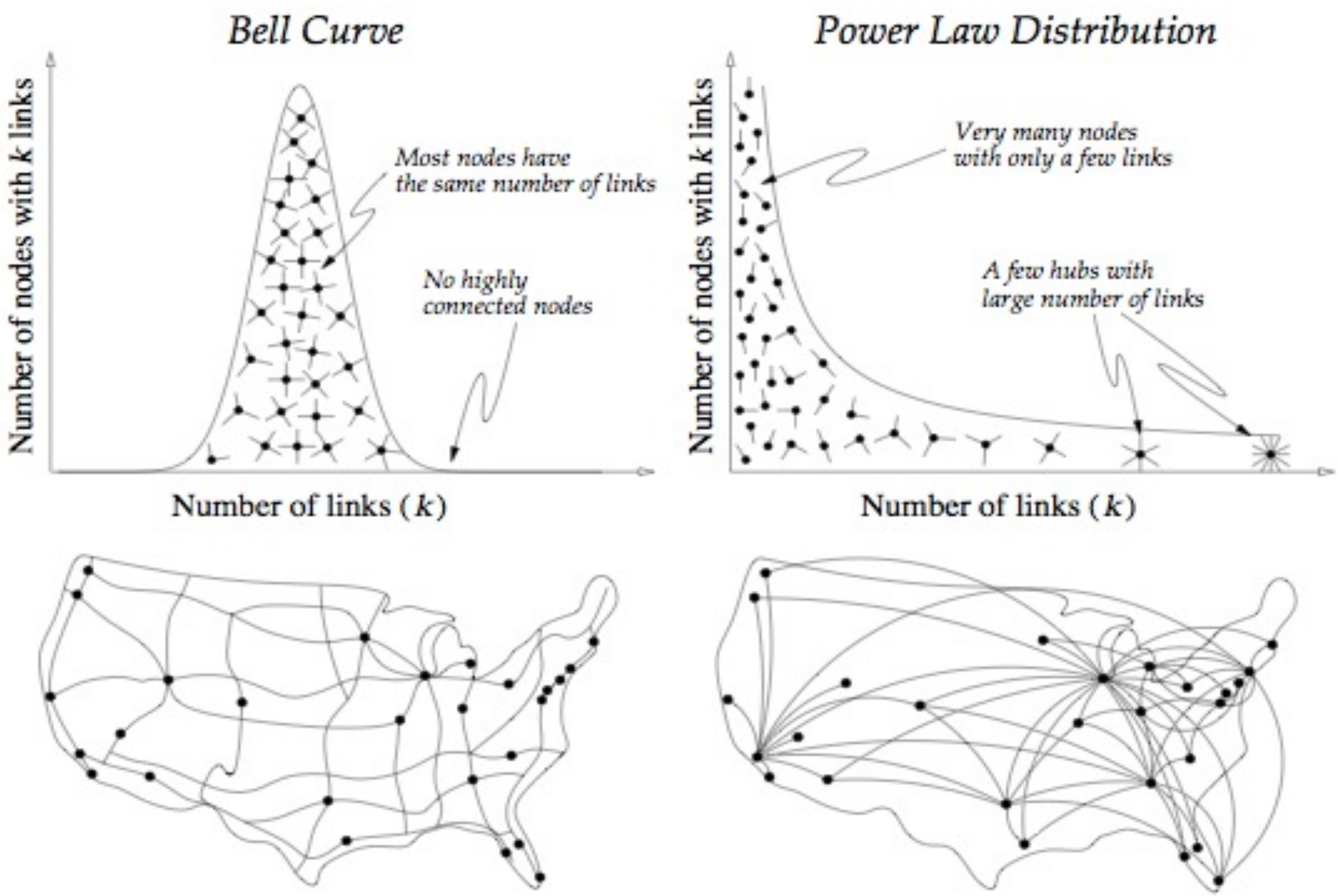
Some commuters in London



World Wide Web

M.E.J. Newman and M. Girvan, Phys. Rev. E 69, 026113, 2004

Slide from Barabási's lectures
https://www.barabasilab.com/course



Power laws and scale-free networks

If the degree distribution of a graph is a **power law**:

$$p_k \sim k^{-\gamma}$$

Or if directed:

$$p_{k_{in}} \sim k^{-\gamma_{in}}$$

$$p_{k_{out}} \sim k^{-\gamma_{out}}$$

, $\gamma_{in} \neq \gamma_{out}$ in general

→ **scale-free network**

A network whose degrees are randomly distributed:
→ Binomial or Poisson distribution instead (see later on)

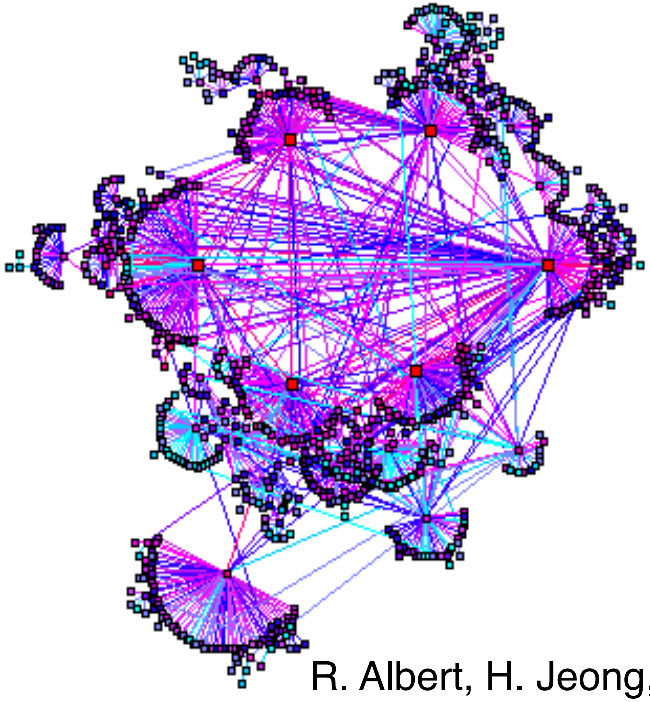
Slide from Barabási's lectures
https://www.barabasilab.com/course

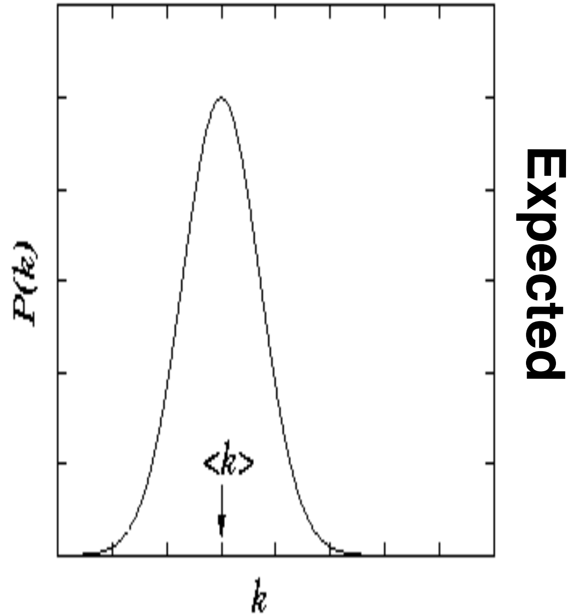
WORLD WIDE WEB

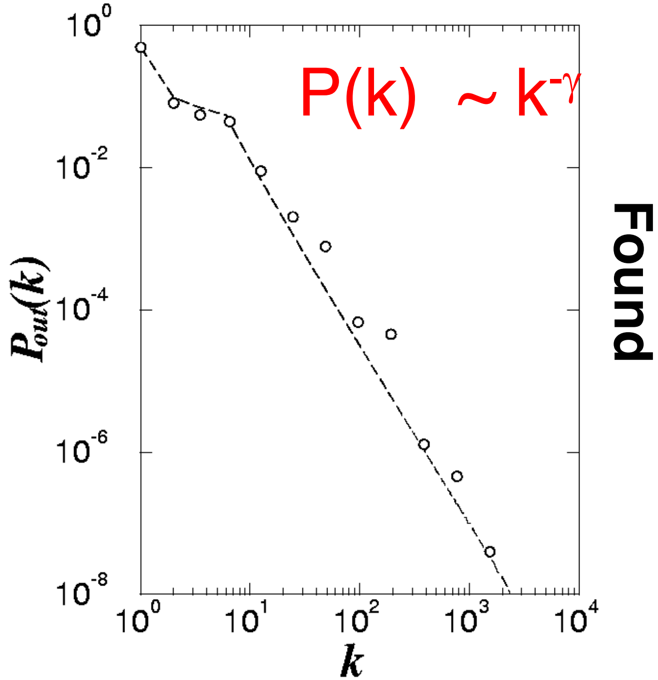
Nodes: **WWW documents**
Links: **URL links**

Over 3 billion documents

ROBOT: collects all URL's found in a document and follows them recursively



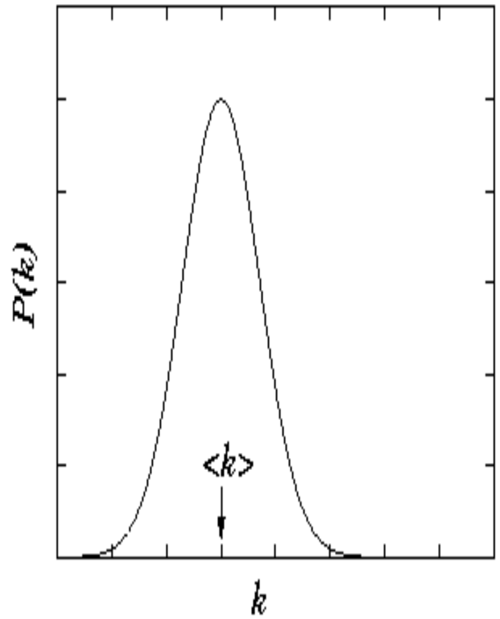




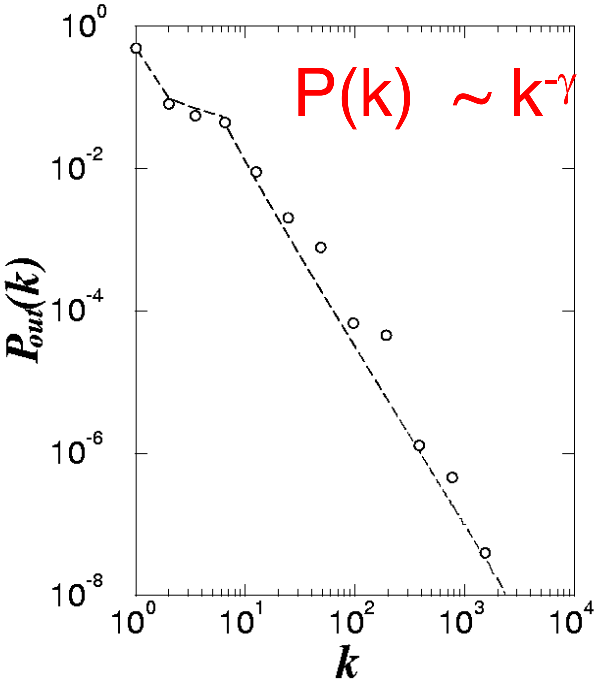
R. Albert, H. Jeong, A-L Barabasi, *Nature*, 401 130 (1999).

Slide from Barabási’s lectures
https://www.barabasilab.com/course

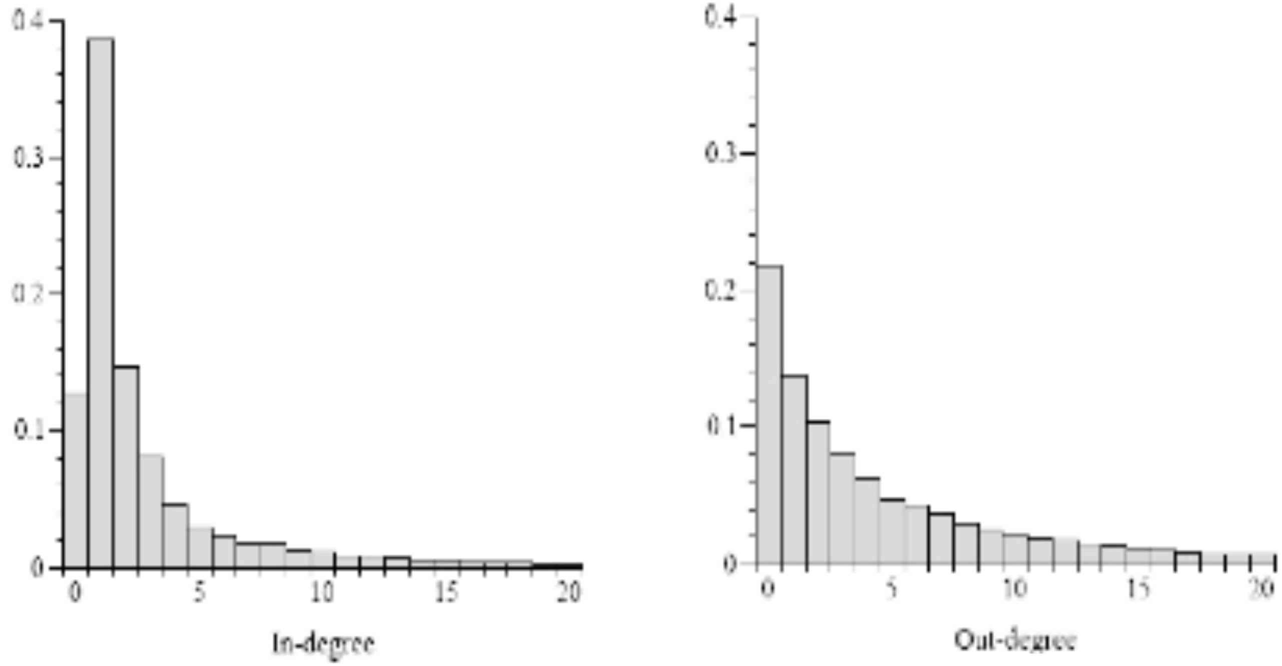
Degree distribution of the WWW



Expected



Found

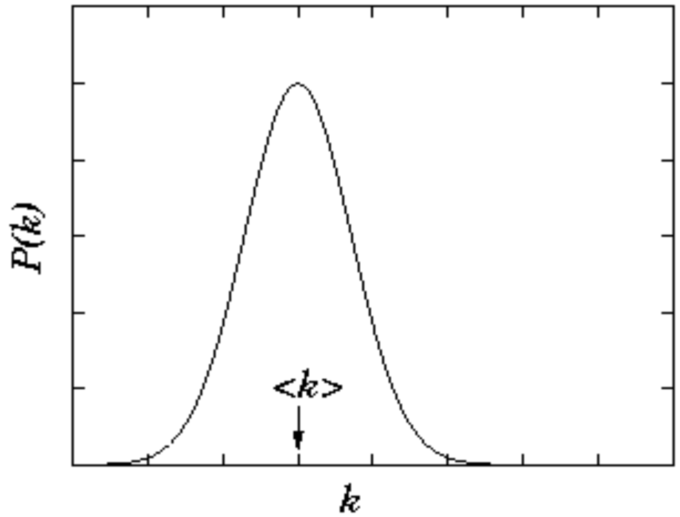


R. Albert, H. Jeong, A-L Barabasi, *Nature*, 401 130 (1999).

Slide from Barabási's lectures
https://www.barabasilab.com/course

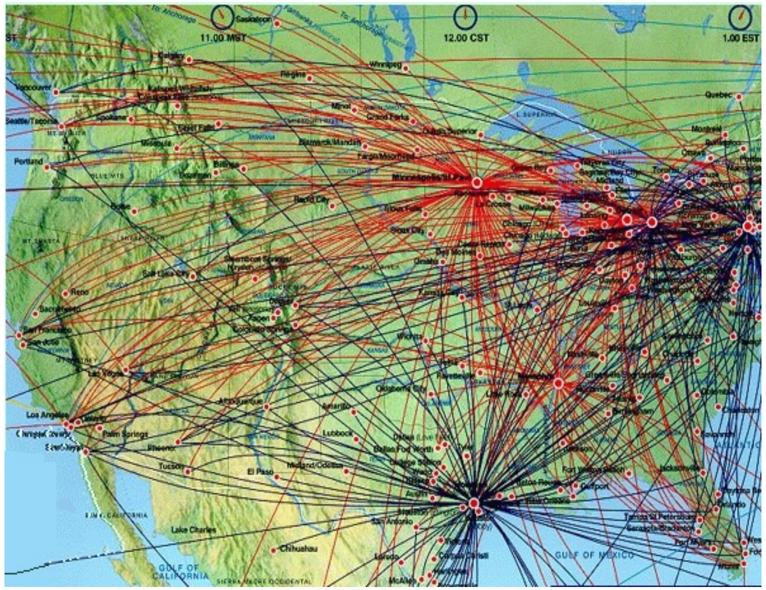
What does the difference mean? Visual representation.

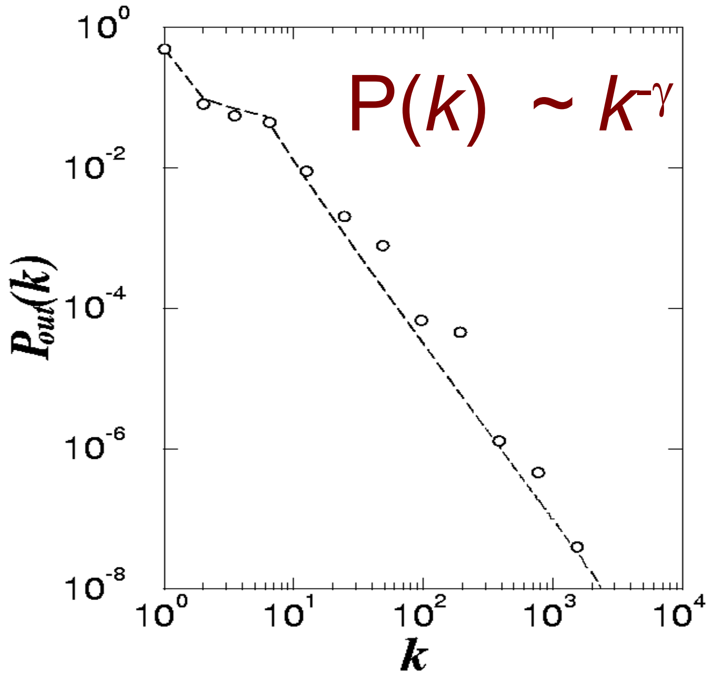
Exponential Network



Expected

Scale-free Network



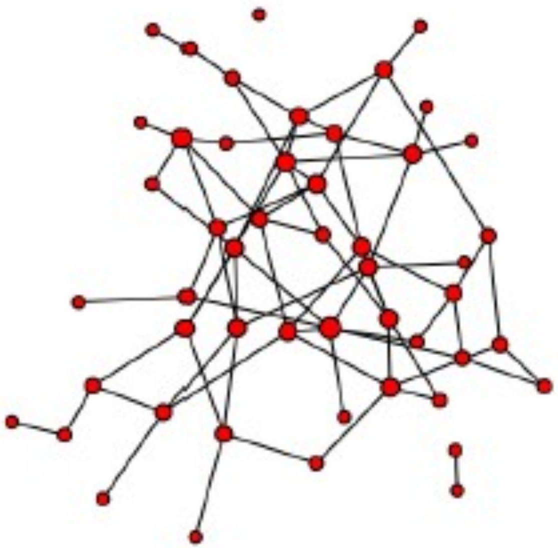


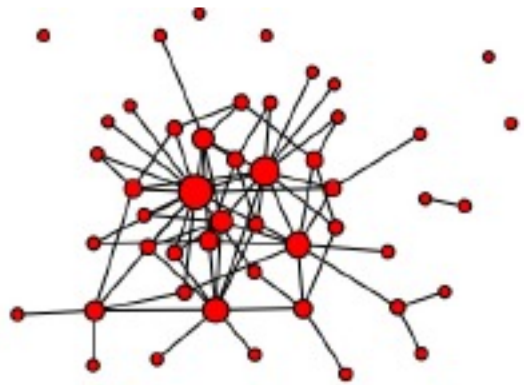
Found

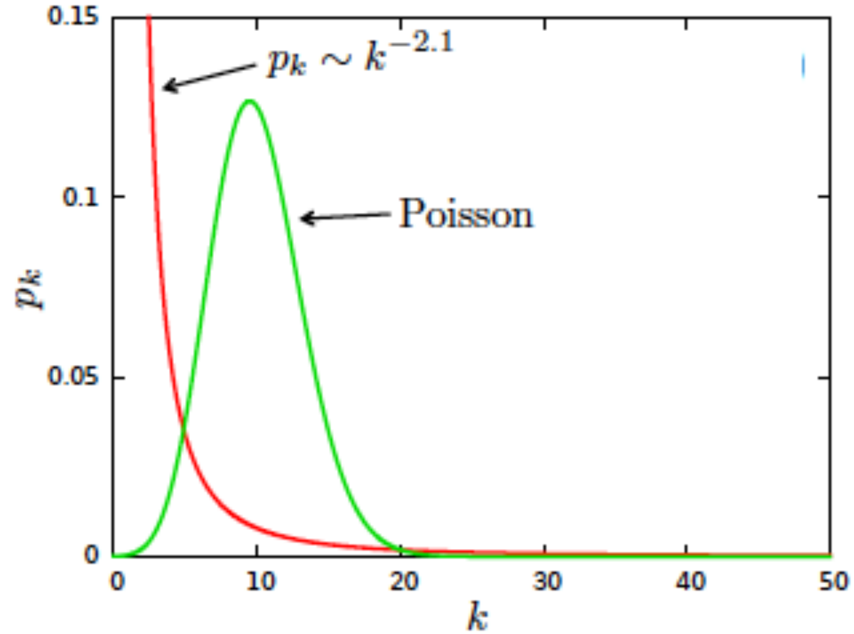
R. Albert, H. Jeong, A-L Barabasi, *Nature*, 401 130 (1999).

Slide from Barabási's book

<k>=3

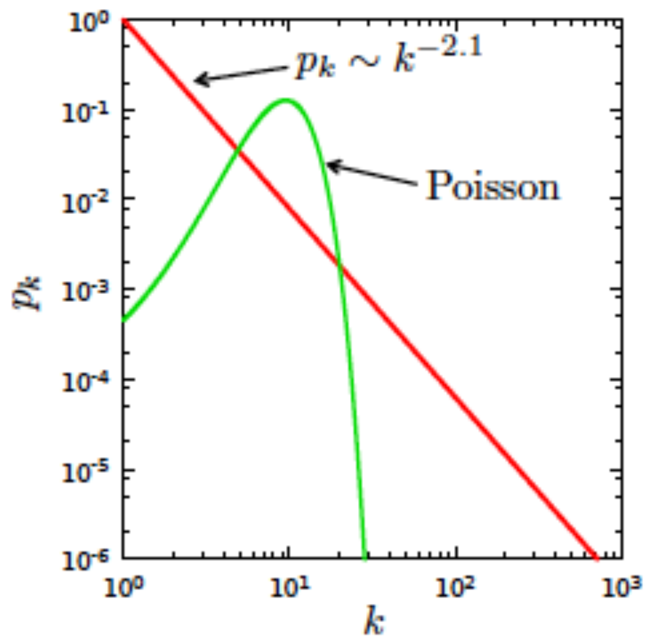






random

Scale-free



Most nodes k~<k>

numerous small-degree nodes coexist with a few highly connected hubs

Moments of the degree distribution

$$\langle k^n \rangle = \sum_{k_{min}}^{\infty} k^n p_k = \int_{k_{min}}^{\infty} k^n p(k) dk.$$

The lower moments have important interpretation:

- n=1: the first moment is the average degree, $\langle k \rangle$.
- n=2: the second moment, $\langle k^2 \rangle$, provides the variance $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$, measuring the spread in the degrees. Its square root, σ, is the standard deviation.
- n=3: the third moment, $\langle k^3 \rangle$, determines the skewness of a distribution, telling us how symmetric is p_k around the average $\langle k \rangle$. Symmetric distributions have zero skewness.

n^{th} moment for a scale-free network

$$\langle k^n \rangle = \int_{k_{min}}^{k_{max}} k^n p(k) dk = C \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n-\gamma+1}.$$

For very large networks, consider asymptotic limit: $k_{max} \to \infty$

It has been observed that scale-free networks have values: $2 \leq \gamma \leq 3$

All moments for $n \geq \gamma - 1$ will diverge

- For a network with a power-law degree distribution and $\gamma < 3$ the first moment is finite but the second moment is infinite. The divergence of $\langle k^2 \rangle$, and hence of σ_k for large N indicates that the fluctuations around the average could be arbitrary large. That is, when we randomly choose a node, we do not know what to expect, as the chosen node's degree could be tiny or arbitrarily large. Hence networks with $\gamma < 3$ do not have a meaningful internal scale. They are "scale-free"

- If the degrees follow a normal distribution, then the degree of a randomly chosen node is

$$k = \langle k \rangle \pm \sigma_k \tag{4.21}$$

For a random network with a Poisson degree distribution $\sigma_k = \sqrt{\langle k \rangle}$, which is always smaller than $\langle k \rangle$. Hence 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 $\langle k \rangle$ serves as the "scale" of a random network.

Preferential attachment

What is the origin of scale-free networks?
Why very different systems have the same architecture?

→ e.g. living organisms vs the www
→ evolution over billions of years vs tens?

SOLUTION

In summary, the random network model differs from real networks in two important characteristics:

GROWTH

While the random network model assumes that the number of nodes, *N*, is fixed (time invariant), real networks are the result of a growth process that continuously increases *N*.

PREFERENTIAL ATTACHMENT

While nodes in random networks randomly choose their interaction partner, in real networks new nodes prefer to link to the more connected nodes.

Slide from Barabási's book

PREFERENTIAL ATTACHMENT: A BRIEF HISTORY

Preferential attachment has emerged repeatedly in mathematics and social sciences. Consequently today we can encounter it under different names in the scientific literature:

- It made its first appearance in 1923 in the celebrated urn model of the Hungarian mathematician György Pólya (1887-1985) [2], proposed to explain the nature of certain distributions. Hence, in mathematics preferential attachment is often called a *Pólya process*.

- George Udmy Yule (1871-1951) in 1925 used preferential attachment to explain the power-law distribution of the number of species per genus of flowering plants [3]. Hence, in statistics preferential attachment is often called a *Yule process*.

- Rober Gibrat (1904-1980) in 1931 proposed that the size and the growth rate of a firm are independent. Hence, larger firms grow faster [4]. Called *proportional growth*, this is a form of preferential attachment.

Slide from Barabási's book

- George Kinsley Zipf (1902-1950) in 1941 used preferential attachment to explain the fat tailed distribution of wealth in the society [5].

- Modern analytical treatments of preferential attachment use of the master equation approach pioneered by the economist Herbert Alexander Simon (1916-2001). Simon used preferential attachment in 1955 to explain the fat-tailed nature of the distributions describing city sizes, word frequencies in a text, or the number of papers published by scientists [6].

- Building on Simon's work, Derek de Solla Price (1922-1983) used preferential attachment to explain the citation statistics of scientific publications, referring to it as *cumulative advantage* [7].

- In sociology preferential attachment is often called the *Matthew effect*, named by Robert Merton (1910-2003) [8] after a passage in the Gospel of Matthew: "For everyone who has will be given more, and he will have an abundance. Whoever does not have, even what he has will be taken from him."

- The term *preferential attachment* was introduced in the 1999 paper by Barabási and Albert [1] to explain the ubiquity of power laws in networks.

Preferential attachment can be seen as a mechanism to generate power laws!

The Barábasi-Albert model

Emergence of Scaling in Random Networks

Albert-László Barabási*, **Réka Albert**

+ See all authors and affiliations

Science 15 Oct 1999:
Vol. 286, Issue 5439, pp. 509-512
DOI: 10.1126/science.286.5439.509

The Barábasi-Albert model

We start with m_0 nodes, the links between which are chosen arbitrarily, as long as each node has at least one link. The network develops following two steps Fig. 5.3:

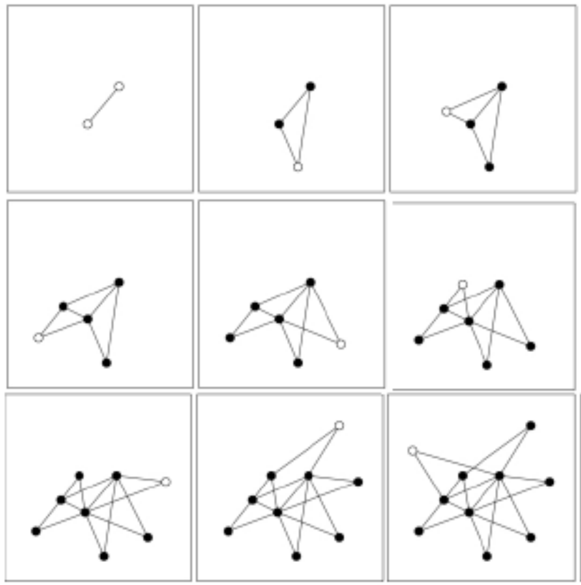
(A) GROWTH

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.

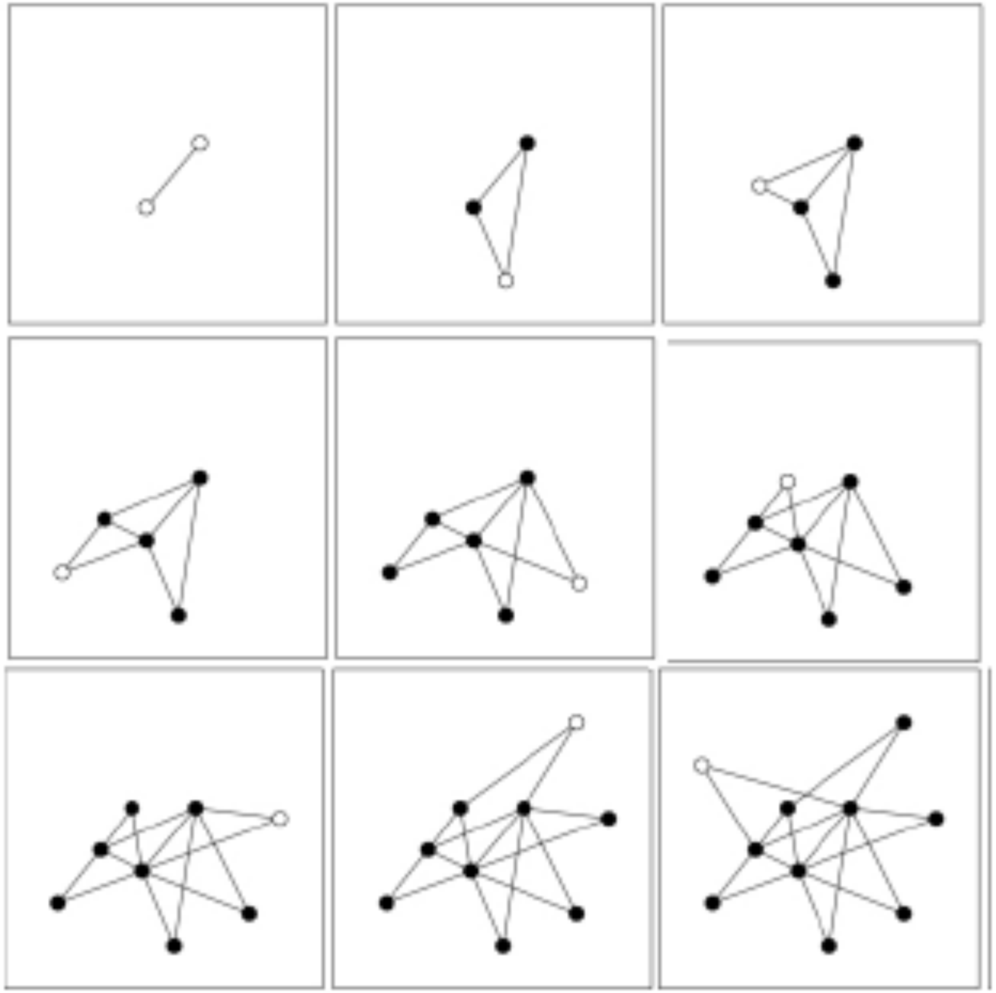
(B) PREFERENTIAL ATTACHMENT

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

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



Slide from Barabási's book



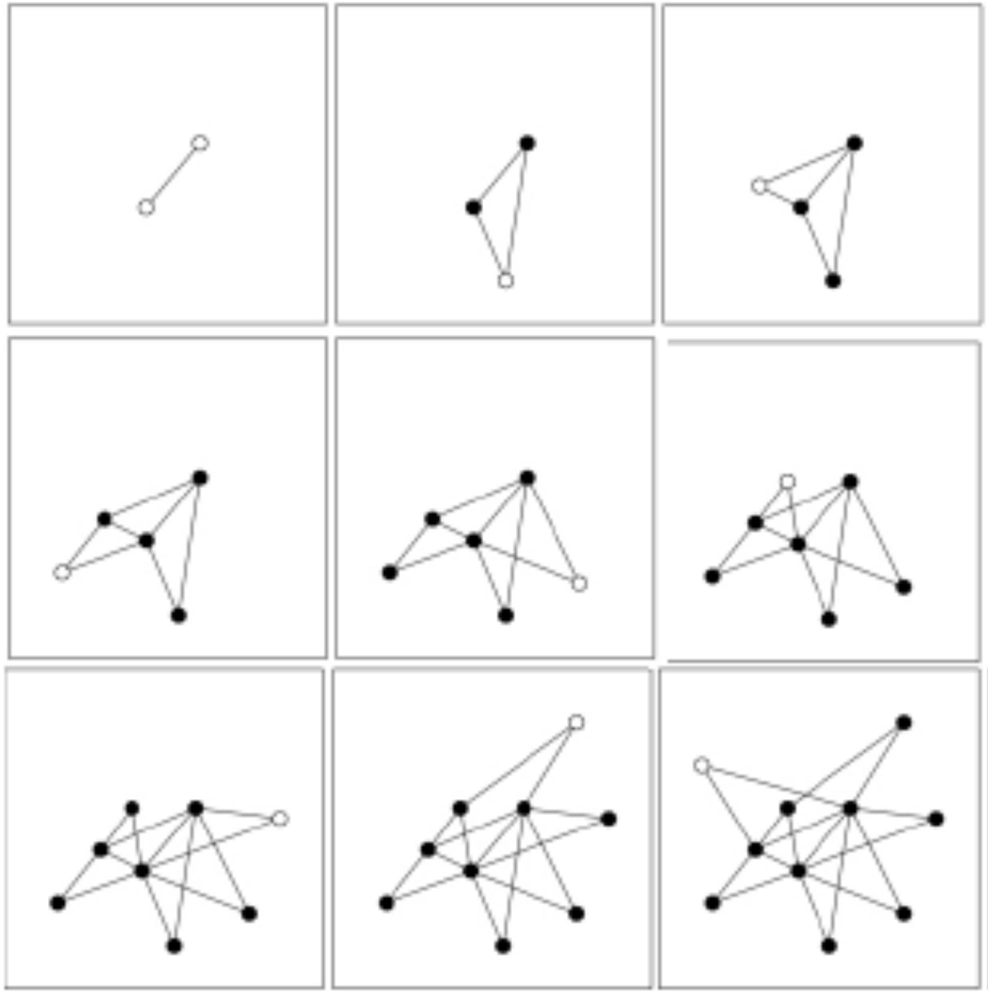
The sequence of images shows the gradual emergence of a few highly connected nodes, or hubs, through growth and preferential attachment. White circles mark the newly added node to the network, which decides where to connect its two links (m=2) through preferential attachment Eq. 5.1. After [9].

A. Add a new node with links $m \le m_0$

B. Probability to connect to node *i*:

$$\Pi_i = \frac{k_i}{\sum_j k_j}$$

Slide from Barabási's book



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].

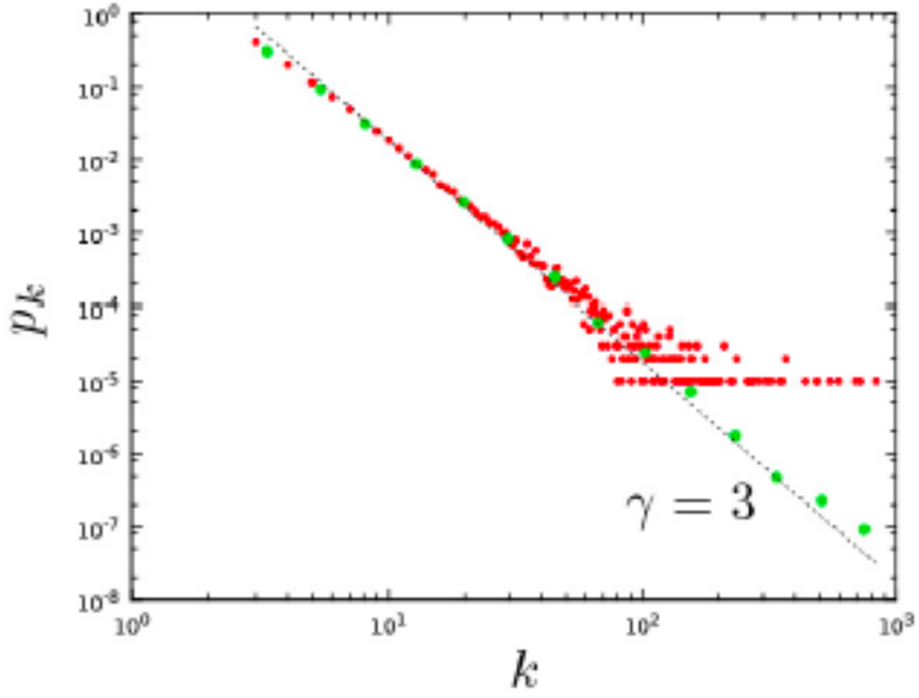


Figure 5.4

The degree distribution

The degree distribution of a network generated by the Barabási-Albert model. The plot shows p_k for a single network of size N=100,000 and m=3. It shows both the linearly-binned (red symbols) as well as the log-binned version (green symbols) of p_k. The straight line is added to guide the eye and has slope γ=3, corresponding to the resulting network's degree distribution.

BUT!!!!!!! Recent paper: PL everywhere, really???

Article | Open Access | Published: 04 March 2019

Scale-free networks are rare

Anna D. Broido & Aaron Clauset

Nature Communications **10**, Article number: 1017 (2019) | Cite this article

Abstract

Real-world networks are often claimed to be scale free, meaning that the fraction of nodes with degree k follows a power law $k^{-\alpha}$, a pattern with broad implications for the structure and dynamics of complex systems. However, the universality of scale-free networks remains controversial. Here, we organize different definitions of scale-free networks and construct a severe test of their empirical prevalence using state-of-the-art statistical tools applied to nearly 1000 social, biological, technological, transportation, and information networks. Across these networks, we find robust evidence that strongly scale-free structure is empirically rare, while for most networks, log-normal distributions fit the data as well or better than power laws. Furthermore, social networks are at best weakly scale free, while a handful of technological and biological networks appear strongly scale free. These findings highlight the structural diversity of real-world networks and the need for new theoretical explanations of these non-scale-free patterns.

Testing for Power Laws

SIAM REVIEW
Vol. 51, No. 4, pp. 661–703

Power-Law Distributions in Empirical Data*

Aaron Clauset†
Cosma Rohilla Shalizi‡
M. E. J. Newman§

http://tuvalu.santafe.edu/~aaronc/slides/Broido_Clauset_2017_ScaleFreeNetsAreRare_NetSci.pdf

Conclusions

Genuine scale-free networks are rare:

- only 4% of network datasets are "strongly scale-free"
- only 33% of network datasets are "weakly scale-free"
- Of remaining 77%, *the majority* favor a non-PL distribution over the power law.

Future directions

Are any structural patterns "universal"? Maybe not.

Some domains have more scale-free networks than others

- e.g., Biological and Technological networks [good theories for why]
- Social networks at best weakly scale-free
- *we need new mechanistic models of general structural patterns*

http://tuvalu.santafe.edu/~aaronc/slides/Broido_Clauset_2017_ScaleFreeNetsAreRare_NetSci.pdf

Random graphs

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

Two definitions of random networks.

There are two equivalent ways of defining a random network:

- **G(N,L) model:** *N* labeled nodes are connected with *L* randomly placed links. Erdős and Rényi (Erdős & Rényi, 1959) used this definition in their string of articles on random networks.

- **G(N,p) model:** Each pair of *N* labeled nodes is connected with probability *p*, a model introduced by Gilbert (Gilbert, 1959).

Hence the *G(N,p)* model fixes the probability *p* that two nodes are connected and the *G(N,L)* model fixes the total number of links *L*.

Random graphs

→mostly known as Erdős-Rényi random graphs

To construct a random network, denoted with $G(N, p)$ (Box 3.1):

1. Start with N isolated nodes.

2. Select a node pair, and generate a random number between 0 and 1. If the random number exceeds p, connect the selected node pair with a link, otherwise leave them disconnected.

3. Repeat step (2) for each of the $N(N-1)/2$ node pairs.

Random graphs

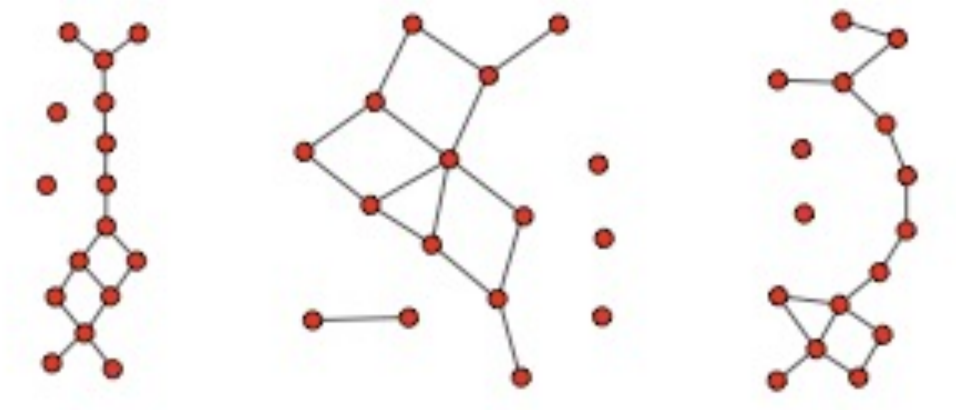


Fig. 4.1. Three different realizations of model A, with $N = 16$ and $K = 15$

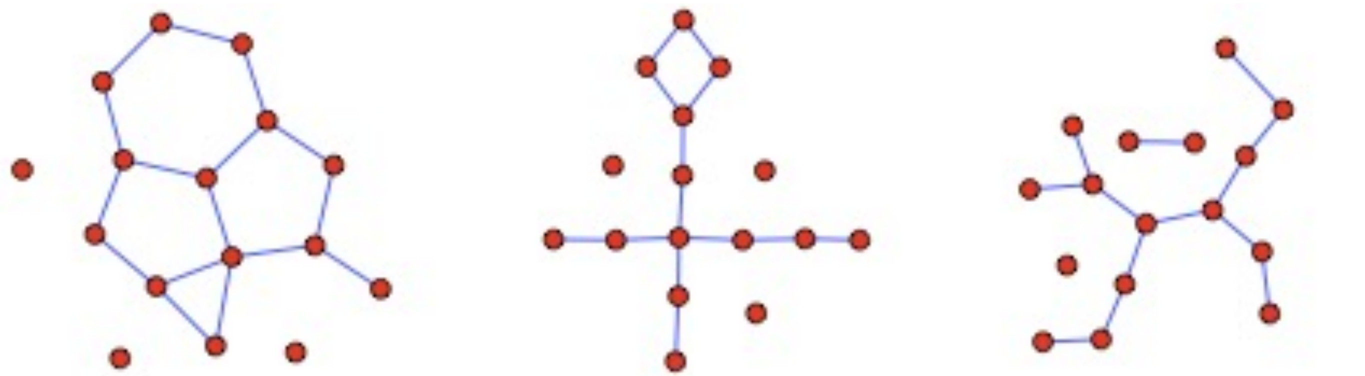


Fig. 4.2. Three different realizations of model B, with $N = 16$ and $p = 0.125$

Figures from Vito Latora Lectures

Random graphs

A random graph model is NOT a single generated random graph!!!

→ Need to consider an ensemble of networks and derive properties of random graphs as the average properties of the ensemble.

→ E.g. observable *D* (could be any measure, such as the diameter)

$$\boxed{< D >= \sum_G P(G)D(G)}$$

Probability distribution P and value D of graph G

→ Need to generate a statistically significant measure: need many different samples

Random graphs

Degree distribution

The next slides contain a derivation of how to compute the probability distribution for a random graph and its degree distribution. These slides are taken from the lectures by Barabási, have a look there for more detail.
I will not look at these in class, but I will leave them for the curious student.

In the limit of N large and $<k>$ small (most of the time the network will be sparse), the derivation gives:

$$\boxed{P(k) = e^{-<k>} \frac{<k>^k}{k!}}$$

→ **Poisson distribution**

Clustering coefficient:

$$\boxed{C = \frac{<k>}{N-1}}$$

Random graphs

To look at home if interested

Computing the probability distribution

→ Let P(L) be the total probability of getting a network with N nodes and exactly L links for a prob. p from our ensemble

The maximum number of links in a network of N nodes.

Prob. that the rest of the possible links don't exist

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

Binomial distribution...

Prob. of having L links

Number of different ways we can choose L links among all potential links.

From Barabási's Lectures

Slide from Barabási's lectures
https://www.barabasilab.com/course

RANDOM NETWORK MODEL

To look at home if interested

P(L): the probability to have a network of exactly *L* links

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

- The average number of links *<L>* in a random graph

$$< L >= \sum_{L=0}^{\frac{N(N-1)}{2}} LP(L) = p\frac{N(N-1)}{2}$$

$$< k >= 2L/N = p(N-1)$$

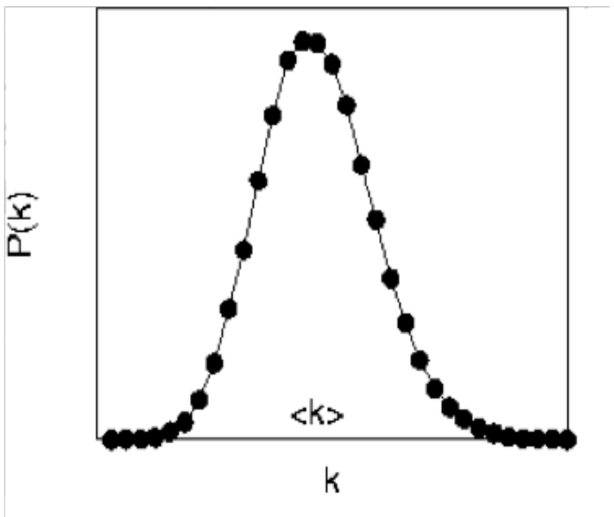
- The standard deviation

$$\sigma^2 = p(1-p)\frac{N(N-1)}{2}$$

Slide from Barabási's lectures
https://www.barabasilab.com/course

Degree distribution of random graph

To look at home if interested



$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

Select k
nodes from N-1

probability of
having *k*
edges

probability of
missing N-1-
k
edges

$$< k >= p(N-1)$$

$$\sigma_k^2 = p(1-p)(N-1)$$

$$\frac{\sigma_k}{< k >} = \left[\frac{1-p}{p}\frac{1}{(N-1)}\right]^{1/2} \approx \frac{1}{(N-1)^{1/2}}$$

As the network size increases, the distribution becomes increasingly narrow—we are increasingly confident that the degree of a node is in the vicinity of <k>.

Random graphs: limits for degree distribution

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

$$<k> = p(N-1)$$

$$p = \frac{<k>}{(N-1)}$$

If *N* large and *k* small:

$$\lim_{\substack{k<< \\ N>>}} P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k} = e^{-<k>} \frac{<k>^k}{k!}$$

→ **Poisson distribution**

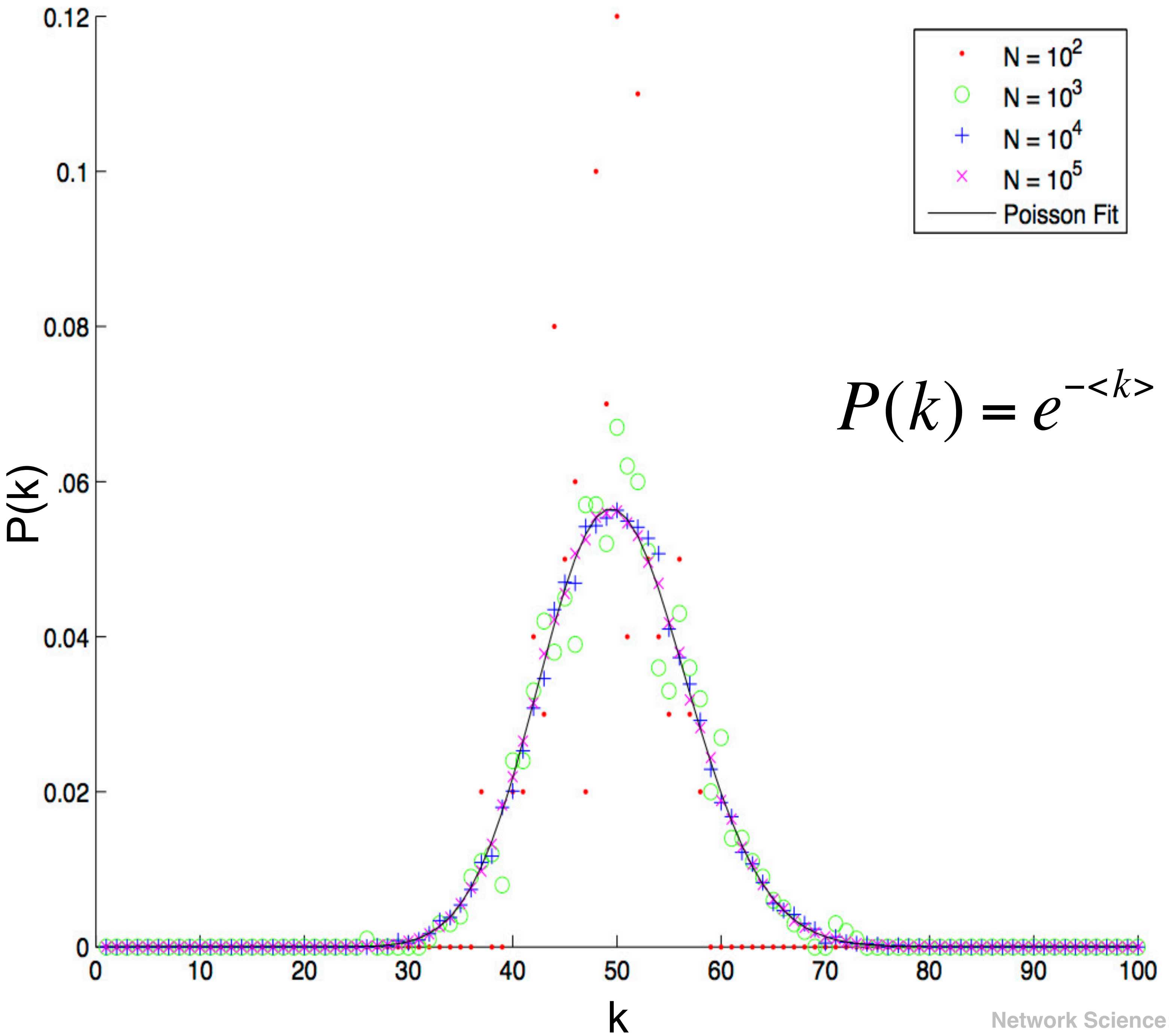
Clustering coefficient:

$$C = \frac{<k>}{N-1}$$

To look at home if interested

Degree distribution of random graph

Slide from Barabási's lectures
https://www.barabasilab.com/course



$$P(k) = e^{-<k>} \frac{<k>^k}{k!}$$

Slide from Barabási's lectures
https://www.barabasilab.com/course

NO OUTLIERS IN A RANDOM SOCIETY

$$P(k) = e^{-<k>} \frac{<k>^k}{k!}$$

According to sociological research, for a typical individual k ~1,000

The probability to find an individual with degree k > 2,000 is 10^{-27}.

Given that N ~10^9, (7.9 billion people) the chance of finding an individual with 2,000 acquaintances is so tiny that such nodes are virtually inexistent in a random society.

→ a random society would consist of mainly average individuals, with everyone with roughly the same number of friends.

→ It would lack outliers, individuals that are either highly popular or recluse.

Small-world model: Watts and Strogatz (1998)

→ We notice that real networks have clustering coefficients much larger than those found for random networks

Recall for a random network:

Clustering coefficient: $$C = \frac{<k>}{N-1}$$

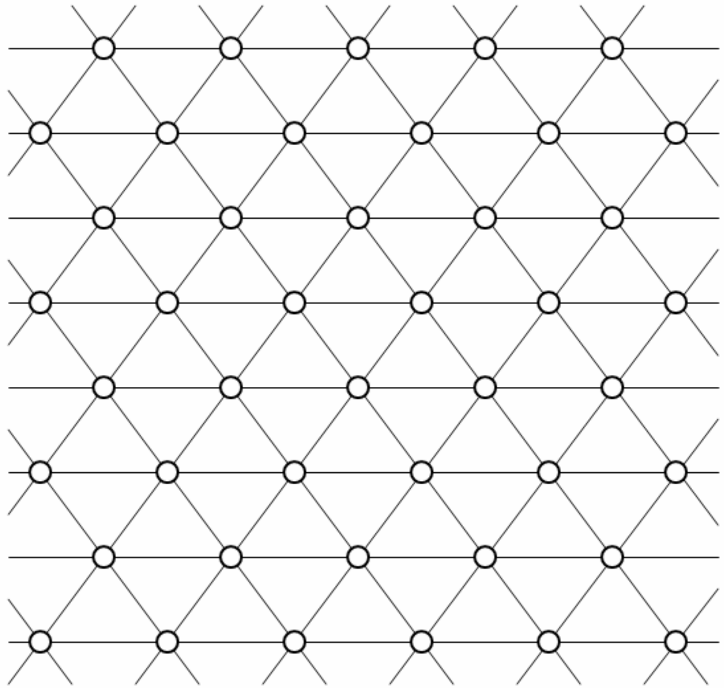
Since in a random network <k> can be seen as a constant as N increases C→0

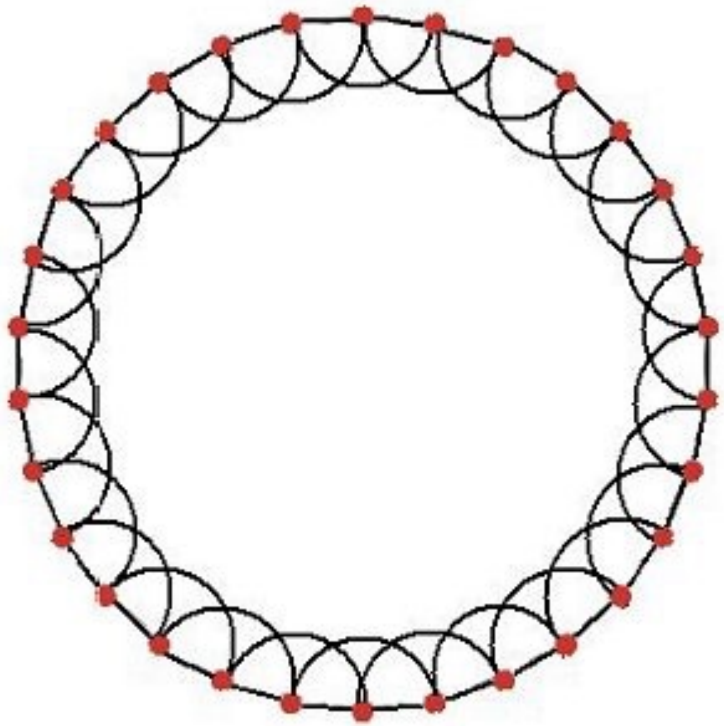
→ A model is needed where the n. of triangles are increased: increased transitivity
→ Model also needs to produce short paths, since these are observed

→ **Small-world model**

Initial solutions for clustering coefficient:

→ Regular lattices





C=0.4

$$C = \frac{3(\langle k \rangle - 2)}{4(\langle k \rangle - 1)}$$

→ C independent of size of network

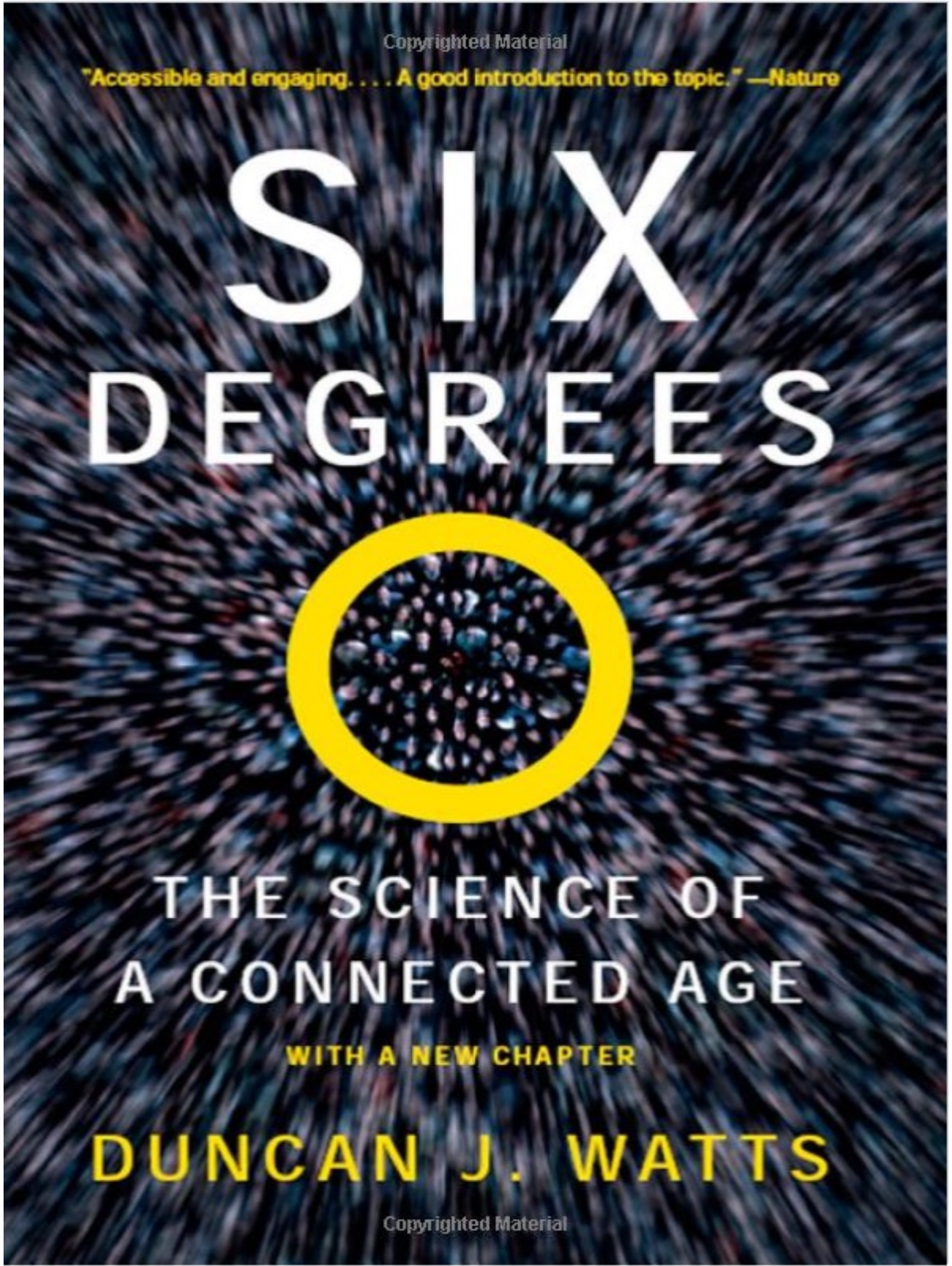
Initial solutions for clustering coefficient:

→ Regular lattices have a constant clustering coefficient regardless of their size. So we can make the lattice such that C has the value measured in social networks.

→ BUT!!!!! We also observe that we are all within 6 degrees of separation of Kevin Bacon!! In this lattice it doesn't work….

- In *Everything is Different (story Chains)* in 1929 Karinthy proposed this.
- Milgram's 1967 "small world experiment": packages were sent and tracked: 6 degrees separation on average

Need to include this short path phenomenology, lattices have very large paths!


"Accessible and engaging. . . . A good introduction to the topic." —Nature
SIX
DEGREES
THE SCIENCE OF
A CONNECTED AGE
WITH A NEW CHAPTER
DUNCAN J. WATTS

Watts and Strogatz solution: small world network

- → Mathematically showed this in 1998
  - → Power grids, neural network of C. elegans, connections between Hollywood actors

- → Watts repeated the experiment in the 2000s with 60,000 participants and 18 different targets in 13 countries: <connection>~5-7 steps

- → Facebook study in 2011 confirmed 4.57 degrees (Ugander et al 2011, arxiv:1111.4503 and Backstrom et al 2012, Proc 4th Ann ACM Web Sc Conf, 33-42)

- → Granovetter's "The strength of weak ties" pointed out that although most people could be reached within a few steps, the most efficient way through weak and not strong ties, in particular for diffusion of information.

The Strength of Weak Ties

Mark S. Granovetter

American Journal of Sociology, Volume 78, Issue 6 (May, 1973), 1360-1380.

Published: 04 June 1998

Collective dynamics of ‘small-world’ networks

Duncan J. Watts & Steven H. Strogatz

Nature **393**, 440–442(1998) | Cite this article

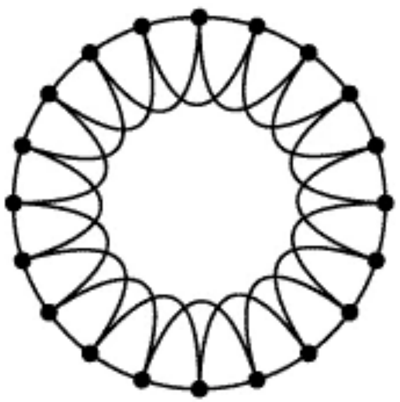
52k Accesses | **22556** Citations | **381** Altmetric | Metrics

Need a model where the network has 2 main properties:

1) A small diameter: people can be connected to other people within a few steps.

2) A small clustering coefficient.

Regular



Clustering coefficient for regular lattices:

Good clustering coefficient

$$C = \frac{3(\langle k \rangle - 2)}{4(\langle k \rangle - 1)}$$

For the smallest number of neighbours $\langle k \rangle = 2$, $C = 0$

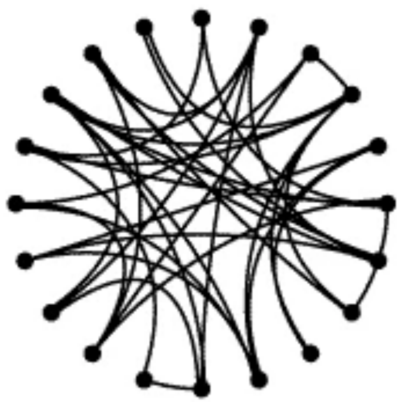
For the denser lattice $\langle k \rangle \to \infty$, $C_{\text{max}} = 3/4$.

Diameter for regular lattice win N nodes:

Very large diameter

$$d = \frac{N}{2k}$$

Random



Clustering coefficient for random networks:

Very small clustering coefficient

$$C = \frac{\langle k \rangle}{N - 1}$$

Diameter for random networks:

Smaller diameter

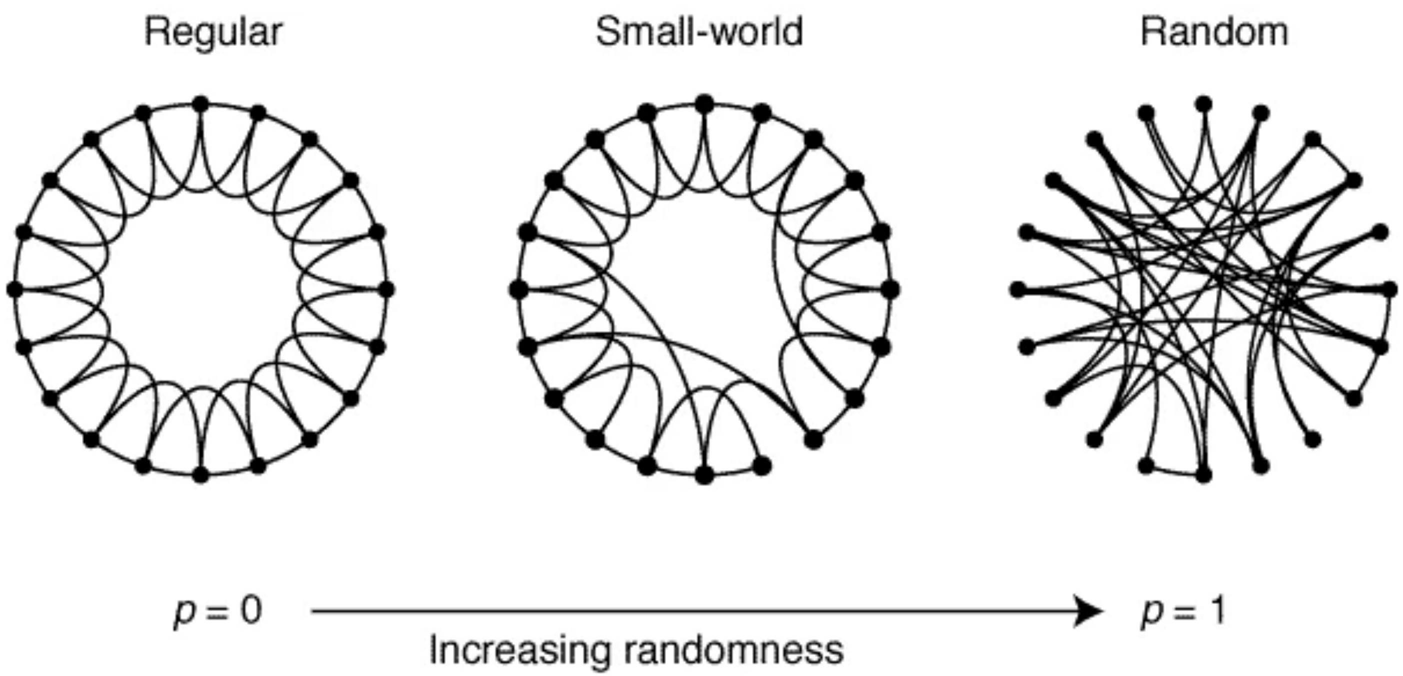
$$d = \frac{\ln N}{\ln k}$$

Small world: Watts Strogatz model

Watts and Strogatz, Nature **393**,409 (1998)

1. Start with a regular lattice

2. Rewire with prob p



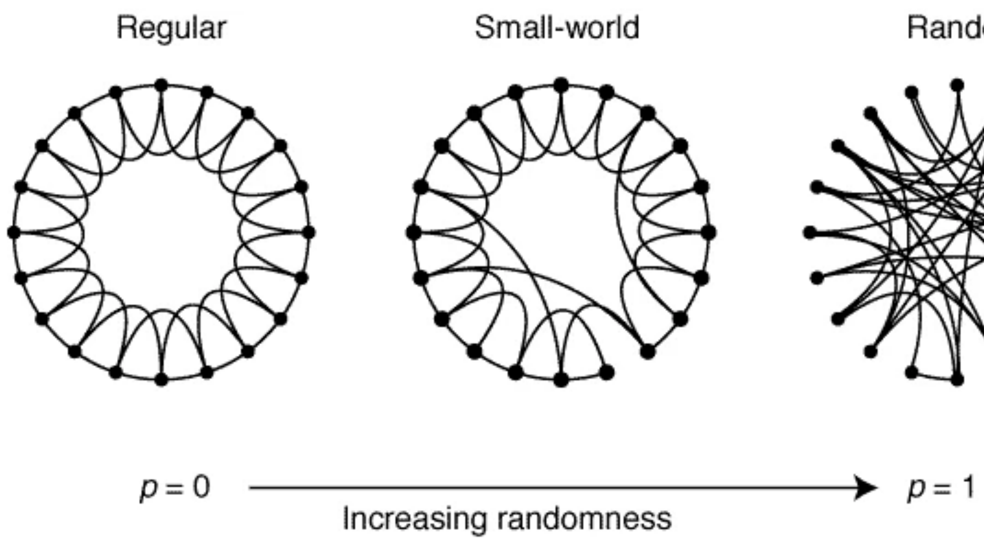
Randomly replacing edges

shortcuts

Network with clustering coefficient larger than random networks and diameter smaller than regular lattices

Watts and Strogatz, Nature **393**,409 (1998)

Figure 1: Random rewiring procedure for interpolating between a regular ring lattice and a random network, without altering the number of vertices or edges in the graph.



We start with a ring of n vertices, each connected to its k nearest neighbours by undirected edges. (For clarity, $n = 20$ and $k = 4$ in the schematic examples shown here, but much larger n and k are used in the rest of this Letter.) We choose a vertex and the edge that connects it to its nearest neighbour in a clockwise sense. With probability p, we reconnect this edge to a vertex chosen uniformly at random over the entire ring, with duplicate edges forbidden; otherwise we leave the edge in place. We repeat this process by moving clockwise around the ring, considering each vertex in turn until one lap is completed. Next, we consider the edges that connect vertices to their second-nearest neighbours clockwise. As before, we randomly rewire each of these edges with probability p, and continue this process, circulating around the ring and proceeding outward to more distant neighbours after each lap, until each edge in the original lattice has been considered once. (As there are $nk/2$ edges in the entire graph, the rewiring process stops after $k/2$ laps.) Three realizations of this process are shown, for different values of p. For $p = 0$, the original ring is unchanged; as p increases, the graph becomes increasingly disordered until for $p = 1$, all edges are rewired randomly. One of our main results is that for intermediate values of p, the graph is a small-world network: highly clustered like a regular graph, yet with small characteristic path length, like a random graph. (See Fig. 2.)

PNAS | February 21, 2012 | vol. 109 | no. 8 | 2825–2830

A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks

Lazaros K. Gallos[a], Hernán A. Makse[a,b,1], and Mariano Sigman[b]

- Multiplicity of percolation transitions
- Hierarchy of clusters
- Correlated percolation process