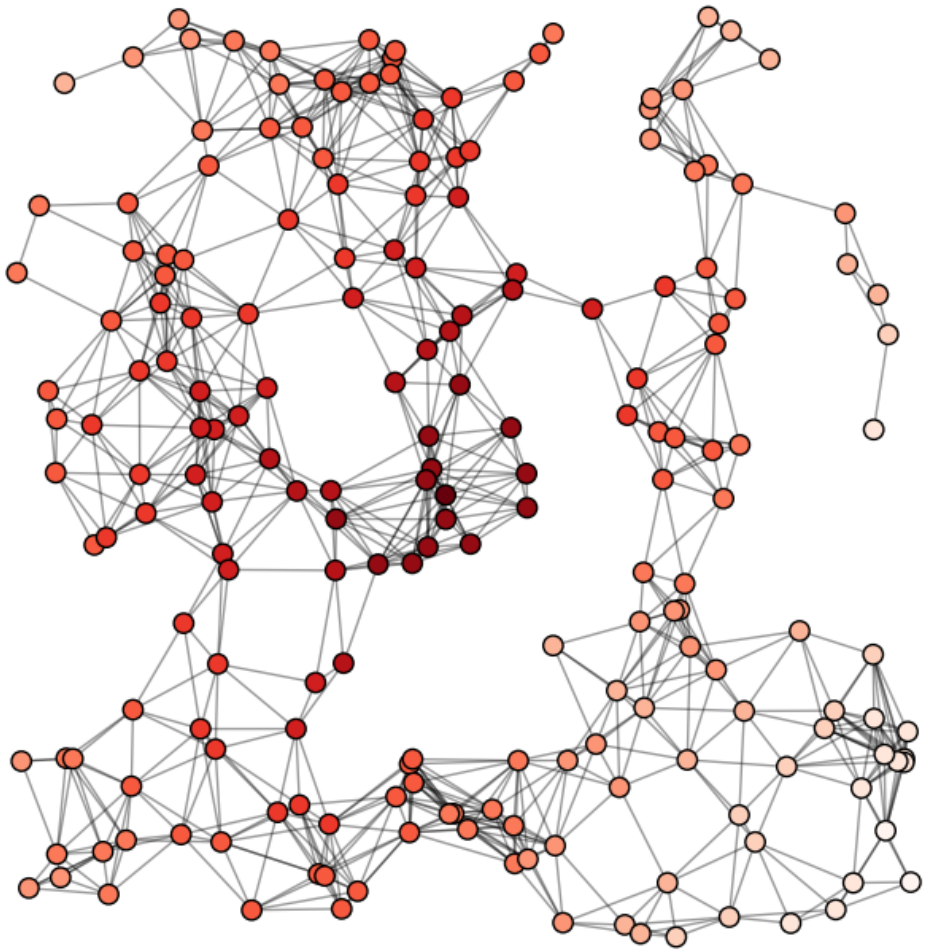
Advanced Algorithms and Computational Models (module A)

Random Networks

Giacomo Fiumara
gfiumara@unime.it

2021-2022



The Random Network Model

- Network science aims to build models that reproduce the properties of real networks
- Most networks do not have the regularity of crystal lattices or the radial architecture of a spider web
- Random network theory embraces the apparent randomness by constructing and characterizing networks that are truly random
- Although a network is a simple model, consisting of nodes and links, the challenge is to decide where to place the links between the nodes in order to reproduce the complexity of a real system
- The Random Network Model places the links randomly between the nodes

The Random Network Model

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

To construct a random network:

- Start with N isolated nodes
- Select a node pair and generate a random number between 0 and 1. If the number is smaller than p, connect the selected node pair with a link, otherwise leave them disconnected
- Repeat the previous step for each of the $N(N-1)/2$ pairs of nodes

The network obtained after this procedure is called a *random graph* or a *random network*. It is also called *Erdös-Rényi network* after the names of the mathematicians that studied the properties of these networks

Defining Random Networks

There are two definitions of a random network:

$G(N, L)$ Model

N labeled nodes are connected with L randomly placed links. Erdös and Rényi used this definition in their papers on random networks

$G(n, p)$ Model

Each pair of N labeled nodes is connected with probability p, a model introducted by Gilbert

- If a fair coin is tossed N times, tails and heads occur with the same probability $p = 1/2$
- The binomial distribution provides the probability $p_{x,N}$ that we obtain exactly x heads in a sequence of N throws
- In general, the binomial distribution describes the number of successes in N independent experiments with two possible outcomes in which the probability of one outcome is p and of the other is $1 - p$

The binomial distribution has the form

$$p_{x,N} = \binom{N}{x} p^x (1-p)^{N-x}$$

The binomial distribution has the form

$$p_{x,N} = \binom{N}{x} p^x (1-p)^{N-x}$$

The mean of the distribution (first moment) is

$$\langle x \rangle = \sum_{x=0}^{N} x p_{x,N} = Np$$

Its second moment is

$$\langle x^2 \rangle = \sum_{x=0}^{N} x^2 p_{x,N} = p(1-p)N + p^2 N^2$$

The mean of the distribution (first moment) is

$$\begin{aligned}\langle x \rangle &= \sum_{x=0}^{N} x p_{x,N} = \\ &= \sum_{x=0}^{N} x \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} = \end{aligned}$$

The mean of the distribution (first moment) is

$$\begin{aligned}
\langle x \rangle &= \sum_{x=0}^{N} x p_{x,N} = \\
&= \sum_{x=0}^{N} x \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} = \\
&= Np \sum_{x=1}^{N} \frac{(N-1)!}{[(N-1)-(x-1)]!(x-1)!} p^{x-1} (1-p)^{N-x} =
\end{aligned}$$

The mean of the distribution (first moment) is

$$
\begin{aligned}
\langle x \rangle &= \sum_{x=0}^{N} x p_{x,N} = \\
&= \sum_{x=0}^{N} x \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} = \\
&= Np \sum_{x=1}^{N} \frac{(N-1)!}{[(N-1)-(x-1)]!(x-1)!} p^{x-1} (1-p)^{N-x} = \\
&= Np \sum_{r=0}^{N-1} \frac{(N-1)!}{[(N-1)-r]!r!} p^r (1-p)^{N-r-1} =
\end{aligned}
$$

having replaced $x-1$ with r

The mean of the distribution (first moment) is

$$\begin{aligned}
\langle x \rangle &= \sum_{x=0}^{N} x p_{x,N} = \\
&= \sum_{x=0}^{N} x \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} = \\
&= Np \sum_{x=1}^{N} \frac{(N-1)!}{[(N-1)-(x-1)]!(x-1)!} p^{x-1} (1-p)^{N-x} = \\
&= Np \sum_{r=0}^{N-1} \frac{(N-1)!}{[(N-1)-r]!r!} p^r (1-p)^{N-r-1} = \\
&= Np(p+1-p)^N = Np
\end{aligned}$$

having replaced $x-1$ with r

The second moment of the distribution is

$$\begin{aligned}\langle x^2 \rangle &= \sum_{x=0}^{N} x^2 p_{x,N} = \\ &= \sum_{x=0}^{N} x^2 \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} = \end{aligned}$$

To be continued...

The second moment of the distribution is

$$\begin{aligned}
\langle x^2 \rangle &= \sum_{x=0}^{N} x^2 p_{x,N} = \\
&= \sum_{x=0}^{N} x^2 \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} = \\
&= Np \sum_{x=1}^{N} \frac{(N-1)!x}{(x-1)!(N-x)!} p^{x-1} (1-p)^{N-x} =
\end{aligned}$$

To be continued...

The second moment of the distribution is

$$\begin{aligned}
\langle x^2 \rangle &= \sum_{x=0}^{N} x^2 p_{x,N} = \\
&= \sum_{x=0}^{N} x^2 \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} = \\
&= Np \sum_{x=1}^{N} \frac{(N-1)!x}{(x-1)!(N-x)!} p^{x-1} (1-p)^{N-x} = \\
&= Np \sum_{r=0}^{N-1} (r+1) \frac{(N-1)!}{r!(N-r-1)!} p^r (1-p)^{N-r-1} =
\end{aligned}$$

having replaced $x-1$ with r

To be continued...

Continued

$$= \quad Np \sum_{r=0}^{N-1} (r+1) \frac{(N-1)!}{r!(N-r-1)!} p^r (1-p)^{N-r-1} =$$

Continued

$$= Np \sum_{r=0}^{N-1} (r+1) \frac{(N-1)!}{r!(N-r-1)!} p^r (1-p)^{N-r-1} =$$

$$= Np \sum_{r=0}^{N-1} (r+1) p_{r,N-1} =$$

Continued

$$
\begin{aligned}
&= Np \sum_{r=0}^{N-1} (r+1) \frac{(N-1)!}{r!(N-r-1)!} p^r (1-p)^{N-r-1} = \\
&= Np \sum_{r=0}^{N-1} (r+1) p_{r,N-1} = \\
&= Np \sum_{r=0}^{N-1} r p_{r,N-1} + Np \sum_{r=0}^{N-1} p_{r,N-1} =
\end{aligned}
$$

Continued

$$
\begin{aligned}
&= Np \sum_{r=0}^{N-1} (r+1) \frac{(N-1)!}{r!(N-r-1)!} p^r (1-p)^{N-r-1} = \\
&= Np \sum_{r=0}^{N-1} (r+1) p_{r,N-1} = \\
&= Np \sum_{r=0}^{N-1} r p_{r,N-1} + Np \sum_{r=0}^{N-1} p_{r,N-1} = \\
&= Np[(N-1)p+1] =
\end{aligned}
$$

Continued

$$
\begin{aligned}
&= Np \sum_{r=0}^{N-1} (r+1) \frac{(N-1)!}{r!(N-r-1)!} p^r (1-p)^{N-r-1} = \\
&= Np \sum_{r=0}^{N-1} (r+1) p_{r,N-1} = \\
&= Np \sum_{r=0}^{N-1} r p_{r,N-1} + Np \sum_{r=0}^{N-1} p_{r,N-1} = \\
&= Np[(N-1)p+1] = \\
&= N^2 p^2 - Np^2 + Np =
\end{aligned}
$$

Continued

$$
\begin{aligned}
&= Np \sum_{r=0}^{N-1} (r+1) \frac{(N-1)!}{r!(N-r-1)!} p^r (1-p)^{N-r-1} = \\
&= Np \sum_{r=0}^{N-1} (r+1) p_{r,N-1} = \\
&= Np \sum_{r=0}^{N-1} r p_{r,N-1} + Np \sum_{r=0}^{N-1} p_{r,N-1} = \\
&= Np[(N-1)p+1] = \\
&= N^2 p^2 - Np^2 + Np = \\
&= p(1-p)N + N^2 p^2
\end{aligned}
$$

Standard deviation

$$\sigma_x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} =$$

Standard deviation

$$\begin{aligned} \sigma_x &= \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = \\ &= \sqrt{p(1-p)Np + N^2p^2 - Np} = \end{aligned}$$

Standard deviation

$$\begin{aligned}\sigma_x &= \sqrt{\langle x^2\rangle - \langle x\rangle^2} = \\ &= \sqrt{p(1-p)Np + N^2p^2 - Np} = \\ &= \sqrt{N^2p^2 - Np^2 + Np - N^2p^2} =\end{aligned}$$

Standard deviation

$$\begin{aligned}
\sigma_x &= \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = \\
&= \sqrt{p(1-p)Np + N^2p^2 - Np} = \\
&= \sqrt{N^2p^2 - Np^2 + Np - N^2p^2} = \\
&= \sqrt{Np(1-p)}
\end{aligned}$$

Defining Random Networks

Number of links

- Each random network generated with the same parameters N and p looks slightly different
- Not only the detailed wiring diagram changes between replicas, but also the number of links L
- It is useful, therefore, to determine how many links are to be expected for a particular realization of a random network with fixed N and p

Defining Random Networks

Number of links

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

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

$$p^L$$

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

$$(1-p)^{N(N-1)/2-L}$$

- A combinatorial factor,

$$\binom{\frac{N(N-1)}{2}}{L}$$

counting the number of different ways we can place L links among $N(N-1)/2$ node pairs

The probability that a particular realization of a random network has exactly L links can be written as

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

The expected number of links in a random graph is

$$\langle L \rangle = \sum_{L=0}^{\frac{N(N-1)}{2}} L p_L = p \frac{N(N-1)}{2}$$

Note that $\langle L \rangle$ is the product of the probability p that two nodes are connected and the number of pairs we attempt to connect,

$$L_{max} = \frac{N(N-1)}{2}$$

From

$$\langle L \rangle = \sum_{L=0}^{\frac{N(N-1)}{2}} L p_L = p \frac{N(N-1)}{2}$$

it is possible to obtain the average degree of a random network

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N-1)$$

Note that $\langle k \rangle$ is the product of the probability p that two nodes are connected and $(N-1)$, which is the maximum number of links a node can have in a network of size N

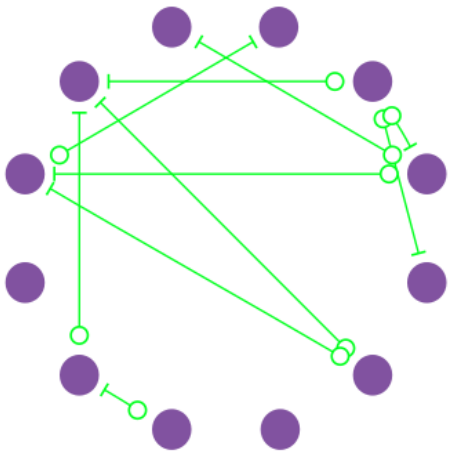
Defining Random Networks

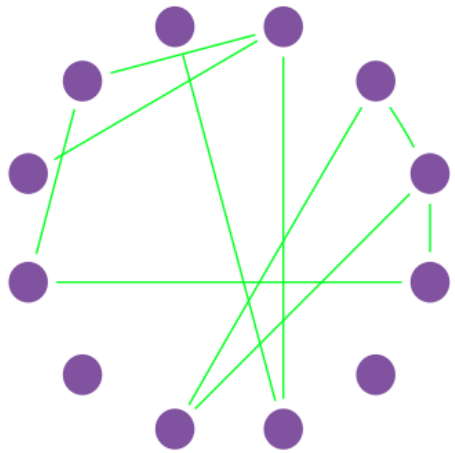
Summary

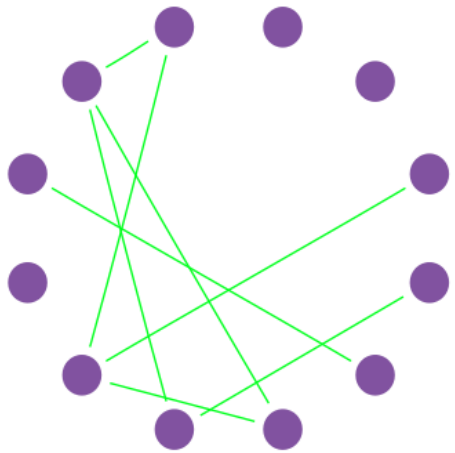
- The number of links in a random network varies between realizations
- Its expected value is determined by N and p
- If p is increased, a random network becomes denser
- The average number of links increases linearly from $\langle L \rangle = 0$ to L_{max}
- The average degree of a node increases from $\langle k \rangle = 0$ to $N - 1$

Defining Random Networks

$p = 1/6$ and $N = 12$

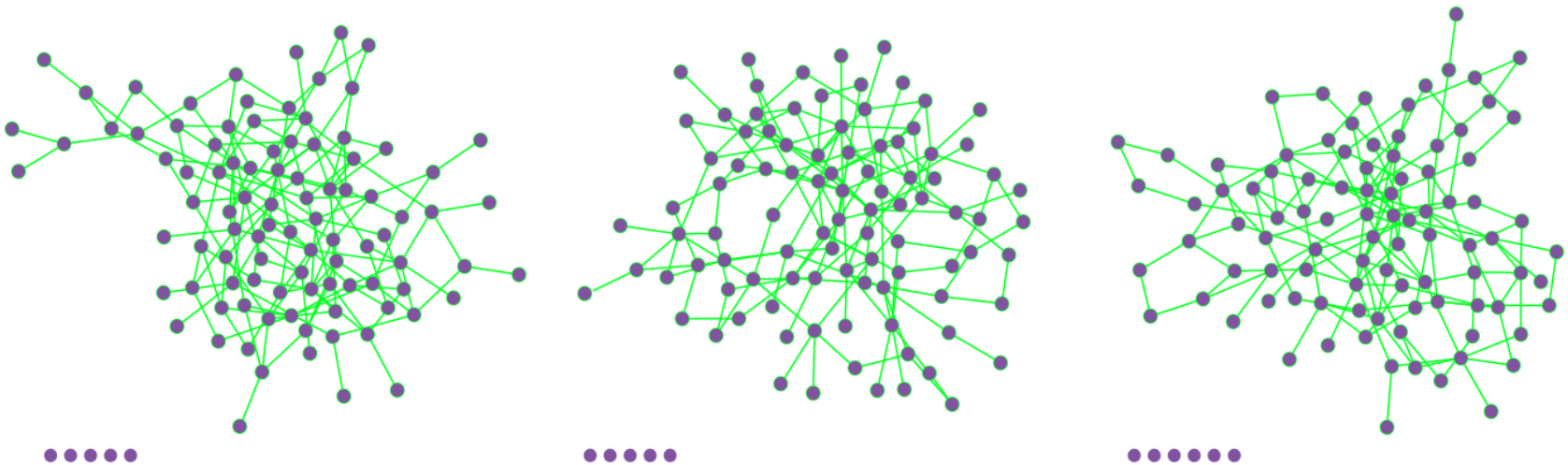


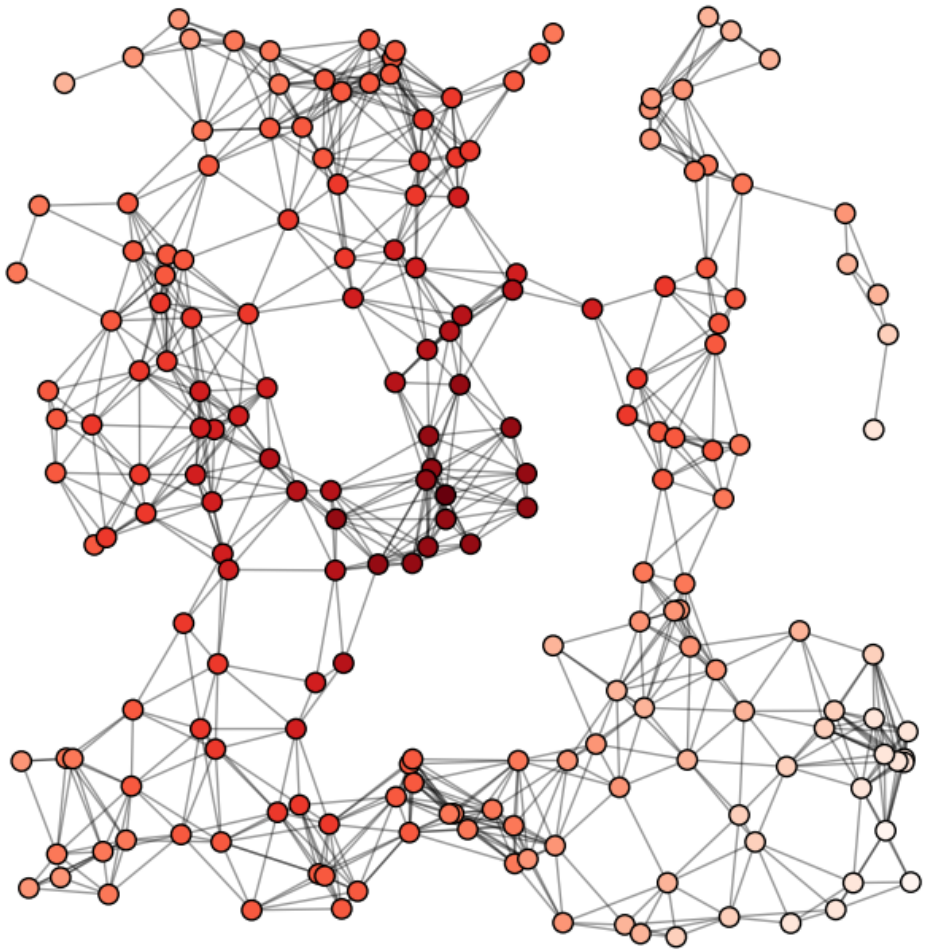




Defining Random Networks

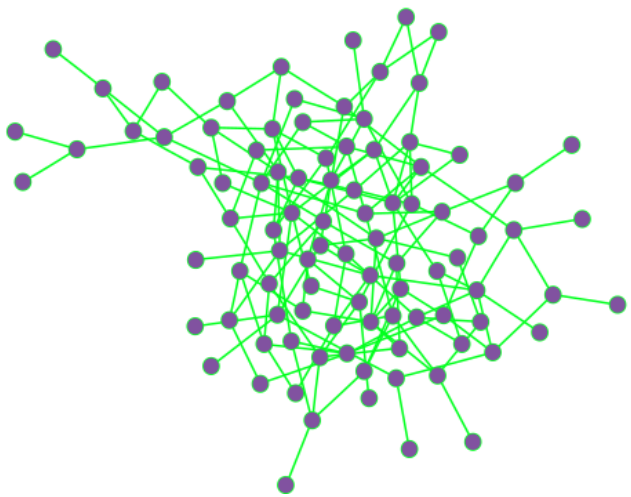
$p = 0.03$ and $N = 100$





Degree Distribution

- In a given realization of a random network, some nodes gain numerous links while others acquire only a few or no links
- These differences are captured by the degree distribution, p_k, which is the probability that a randomly chosen node has degree k



Degree Distribution

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

- The probability that k of its links are present, or

$$p^k$$

- The probability that the remaining $N-1-k$ links are missing, or

$$(1-p)^{N-1-k}$$

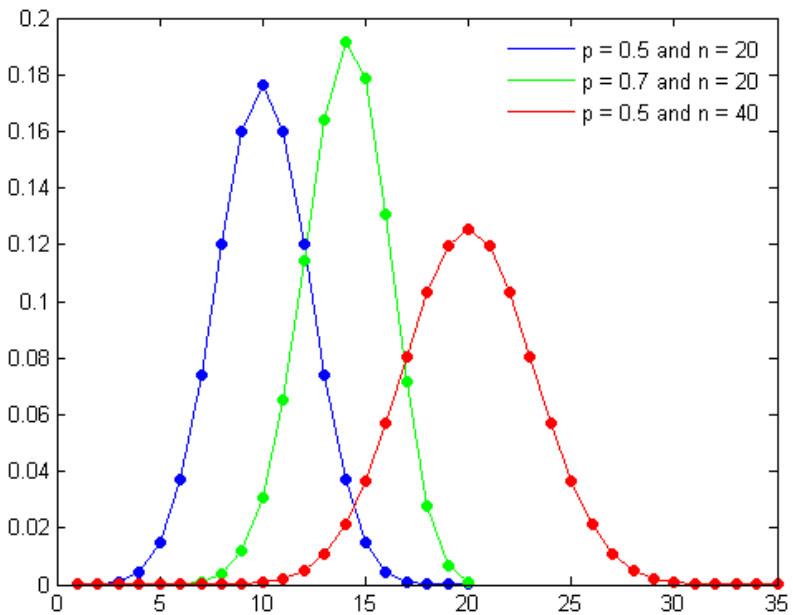
- The number of ways we can select k links from $N-1$ potential links a node can have, or

$$\binom{N-1}{k}$$

Degree Distribution

Therefore, the degree distribution of a random network follows the binomial distribution

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



The binomial distribution reads:

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

Under the hypothesis that $k \ll N$, the first term on the r.h.s. can be approximated as

$$\begin{aligned} \binom{N-1}{k} &= \frac{(N-1)(N-1-1)(N-1-2)\ldots(N-1-k+1)}{k!} \\ &\approx \frac{(N-1)^k}{k!} \end{aligned}$$

The term:

$$(1-p)^{N-1-k}$$

can be simplified as follows:

$$\begin{aligned} ln[(1-p)^{N-1-k}] &= (N-1-k)ln(1-p) = \\ &= (N-1-k)ln(1-\frac{\langle k \rangle}{N-1}) \end{aligned}$$

The series expansion

$$\begin{aligned} ln(1+x) &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n = \\ &= x - \frac{x^2}{2} + \frac{x^3}{3} - \dots \end{aligned}$$

Under the condition that $|x| \leq 1$

Having in mind the series expansion

$$ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$$

we can write

$$\begin{aligned} ln[(1-p)^{N-1-k}] &= (N-1-k)ln(1-\frac{\langle k \rangle}{N-1}) = \\ &= (N-1-k)(-\frac{\langle k \rangle}{N-1}) = \\ &= -\langle k \rangle (1-\frac{k}{N-1}) = \\ &\approx -\langle k \rangle \end{aligned}$$

Having in mind that $k \ll N$

Mathematical Digression

Deriving the Poisson Distribution

We therefore obtained that

$$ln[(1-p)^{N-1-k}] = -\langle k \rangle$$

which is equivalent to

$$(1-p)^{N-1-k} = e^{-\langle k \rangle}$$

Recall that

$$\binom{N-1}{k} \approx \frac{(N-1)^k}{k!}$$

Mathematical Digression

Deriving the Poisson Distribution

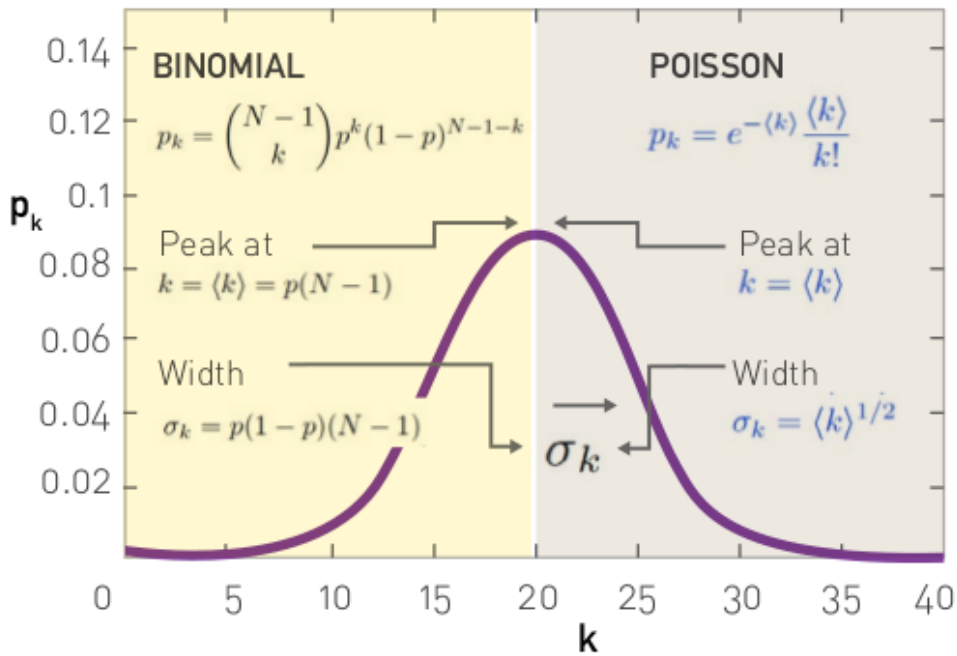
$$
\begin{aligned}
p_k &= \binom{N-1}{k} p^k (1-p)^{N-1-k} = \\
&= \frac{(N-1)^k}{k!} p^k e^{-\langle k \rangle} = \\
&= \frac{(N-1)^k}{k!} \left(\frac{\langle k \rangle}{N-1}\right)^k e^{-\langle k \rangle} = \\
&= e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}
\end{aligned}
$$

Mathematical Digression

Binomial and Poisson Distributions

Similar properties:

- Both distribution have a peak aroung $\langle k \rangle$. If p is increased, the network becomes denser, $\langle k \rangle$ increases and the peak moves to the right
- The width of the distribution is controlled by p or $\langle k \rangle$. The denser the network, the wider the distribution



When the Poisson form is used, it must be keep in mind that:

- The exact result for the degree distribution is the binomial form. The Poisson distribution represents an approximation valid in the $\langle k \rangle \approx N$ limit. As most networks of practical importance are sparse, this condition is typically satisfied
- The advantage of the Poisson form is that key network characteristics like $\langle k \rangle$, $\langle k^2 \rangle$, σ_k have a much simpler form, depending on a single parameter, $\langle k \rangle$.
- The Poisson distribution does not explicitly depend on the number of nodes N. The degree distribution of networks of different sizes but with the same average degree $\langle k \rangle$ are indistinguishable from each other

Real Networks are not Poisson

- The degree of a node in a random network can vary between 0 and $N - 1$
- How big are the differences between the node degrees in a particular realization of a random network?
- Can high degree nodes coexist with small degree nodes?

These questions can be addressed by estimating the size of the largest and the smallest node in a random network

Real Networks are not Poisson

- Let us assume that the world's social network is described by a random network model. This hypothesis may be sound in that there is significant randomness in whom we meet and whom we choose to be acquainted with
- Sociologists estimate that a typical person knows about 1000 individuals on a first name basis, therefore $\langle k \rangle \approx 1000$
- It is possible to demonstrate, using the Poisson distribution, that the most connected individual is expected to have $k_{max} = 1185$ acquaintances
- The degree of the least connected individual is $k_{min} = 816$, not that different from k_{max} or $\langle k \rangle$

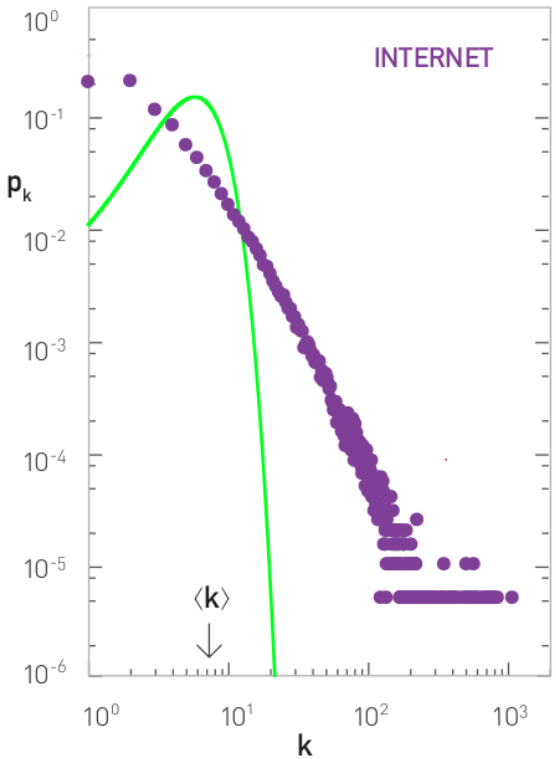
Real Networks are not Poisson

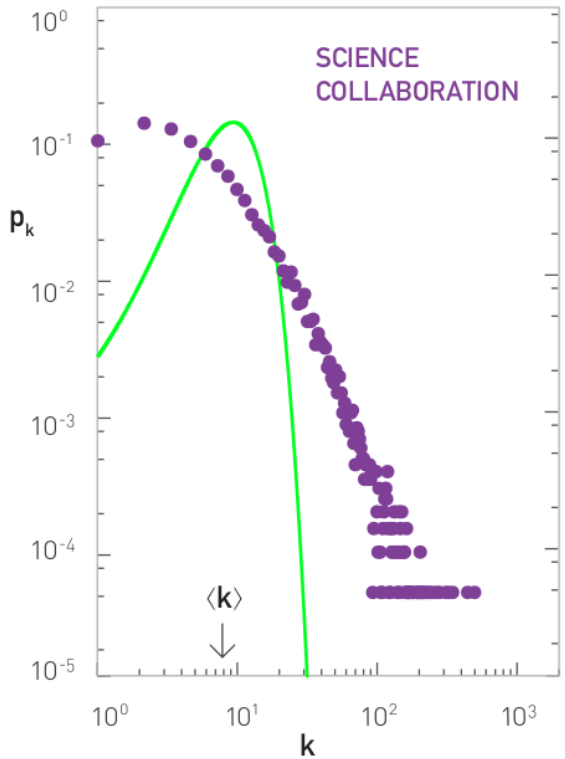
- The dispersion of a random network is $\sigma_k = \langle k \rangle^{1/2}$, which for $\langle k \rangle = 1000$ is $\sigma_k = 31.62$. This means that the number of friends a typical individual has is in the $\langle k \rangle \pm \sigma_k$ range, namely between 968 and 1032, a rather narrow window
- Therefore, in a random society all individuals are expected to have a comparable number of friends
- Hence, if people are randomly connected to each other, we lack outliers
- This surprising conclusion is a consequence of an important property of random networks: *in a large random network the degree of most nodes is in the narrow vicinity of* $\langle k \rangle$.

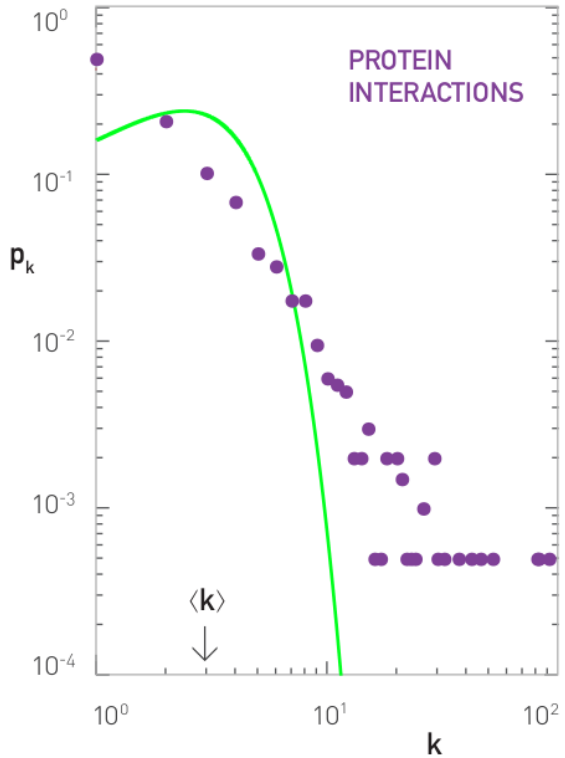
Real Networks are not Poisson

- This prediction clearly conflicts with reality
- There is extensive evidence of individuals who have considerably more than 1185 acquaintances
- For example, the social network behind Facebook documents numerous individuals with 5000 Facebook friends, the maximum allowed by the social network platform
- This is due to the presence of the term $\frac{1}{k!}$ in the Poisson distribution formula, which decreases the chances of observing large degree nodes

Real Networks are not Poisson







Real Networks are not Poisson

- The Poisson distribution significantly underestimates the number of high degree nodes. For example, according to the random network model the maximum degree of the Internet is expected to be around 20. In contrast the data indicates the existence of routers with degrees close to 10^3
- The spread in the degrees of real networks is much wider than expected in a random network. This difference is captured by the dispersion σ_k. If the Internet were to be random, we would expect $\sigma_k = 2.52$. The measurements indicate $\sigma = 14.14$, significantly higher than the random prediction.

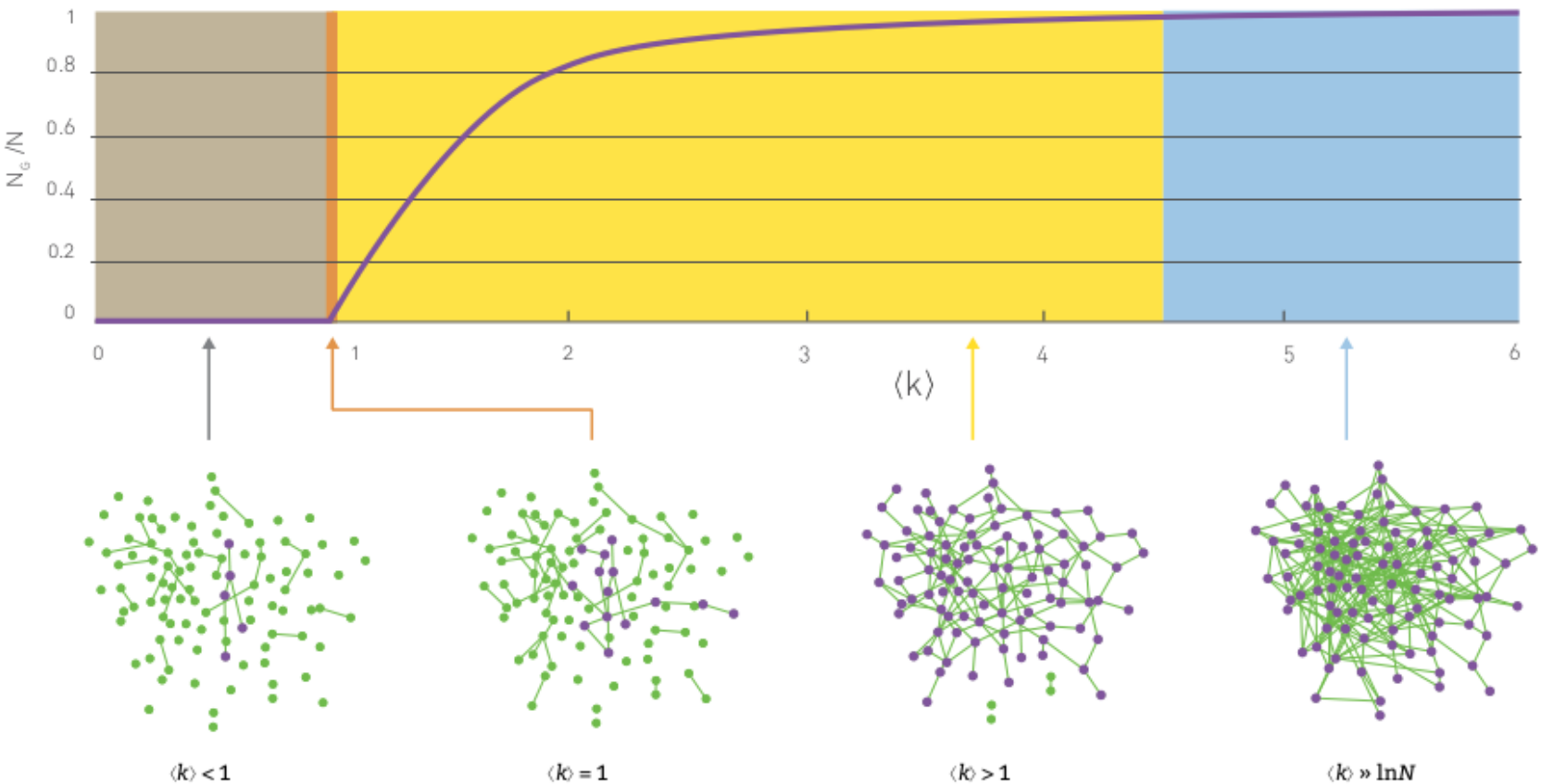
Real Networks are not Poisson

- The comparison with the real data indicates that the random network model does not capture the degree distribution of real networks. In a random network most nodes have comparable degrees, thus forbidding hubs. In contrast, in real networks we observe a significant number of highly connected nodes and there are large differences in node degrees

The Evolution of a Random Network

- The construction of a random network (start with N isolated nodes and randomly place links) is a dynamical process
- An analogous process takes place when the probability p is gradually increased, with interesting consequences on the network topology
- To quantify this process, we first inspect how the size of the largest connected cluster within the network, N_G, varies with $\langle k \rangle$. Two extreme cases:
  - For $p = 0$, we have $\langle k \rangle = 0$, hence all nodes are isolated. Therefore the largest component has size $N_G = 1$ and $N_G/N \to 0$ for large N
  - For $p = 1$ we have $\langle k \rangle = N$, hence the network is a complete graph and all nodes belong to a single component. Therefore $N_G = N$ and $N_G/N \to 1$ for large N

The Evolution of a Random Network



The Evolution of a Random Network

- One would expect that the largest component grows gradually from $N_G = 1$ to $N_G = N$ if $\langle k \rangle$ increases from 0 to $N - 1$.
- This is not the case: N_G/N remains zero for small $\langle k \rangle$, indicating the lack of a large cluster
- Once $\langle k \rangle$ exceeds a critical value, N_G/N increases, signaling the rapid emergence of a large cluster that is called *giant component*
- Erdös and Rényi predicted that the condition for the emergence of the giant component is

$$\langle k \rangle = 1$$

Giant Component

- Let us denote with $u = 1 - N_G/N$ the fraction of nodes that are not in the giant component (GC), whose size we take to be N_G
- If node i is part of the GC, it must link to another node j, which must also be part of the GC
- Hence, if i is not part of the GC, that could happen for two reasons:
  - There is no link between i and j (probability for this is $1 - p$)
  - There is a link between i and j, but j is not part of the GC (probability for this is pu)
- Therefore the total probability that i is not part of the GC via node j is $1 - p + pu$
- The probability that i is not linked to the GC via any other node is therefore $(1 - p + pu)^{N-1}$

Giant Component

As u is the fraction of nodes that do not belong to the GC, for any p and N, the solution of the equation

$$u = (1 - p + pu)^{N-1}$$

provides the size of the GC via $N_G = N(1 - u)$. Using

$$p = \langle k \rangle / (N - 1)$$

and taking the logarithm of both sides, for $\langle k \rangle = N$ we obtain

$$\begin{aligned} lnu &= (N-1) ln[1 - \frac{\langle k \rangle}{N-1}(1-u)] = \\ &= (N-1)[-\frac{\langle k \rangle}{N-1}(1-u)] = \\ &= -\langle k \rangle (1-u) \end{aligned}$$

Giant Component

Taking the exponentials of both sides leads to

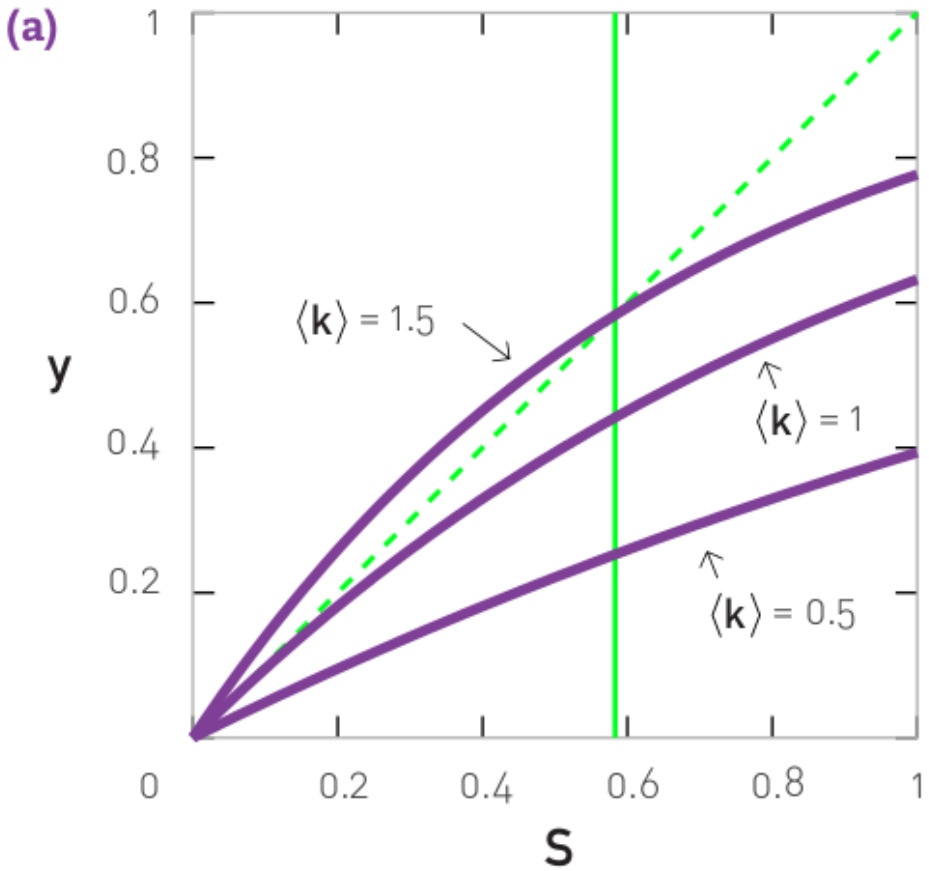
$$u = e^{-\langle k \rangle (1-u)}$$

If we denote with S the fraction of nodes in the GC, $S = N_G/N$, then $S = 1 - u$ and we have

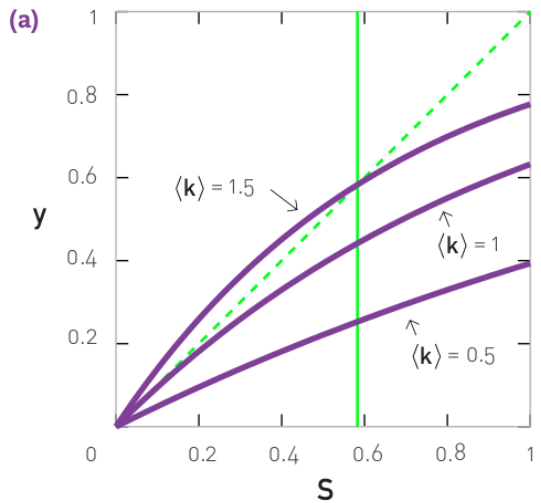
$$S = 1 - e^{-\langle k \rangle S}$$

This equation does not have a closed solution. It can be solved graphically by plotting the rhs as a function of S for various values of $\langle k \rangle$.
To have a nonzero solution, the obtained curve must intersect with the dotted diagonal representing the lhs of the equation.

Giant Component



Giant Component



- For small $\langle k \rangle$ the two curves intersect each other only at $S = 0$, indicating that for small $\langle k \rangle$ the size of the GC is zero
- Only when $\langle k \rangle$ exceeds a threshold value, does a non-zero solution emerge

Giant Component

To determine the value of $\langle k \rangle$ at which we start having a nonzero solution, we take a derivative of

$$S = 1 - e^{-\langle k \rangle S}$$

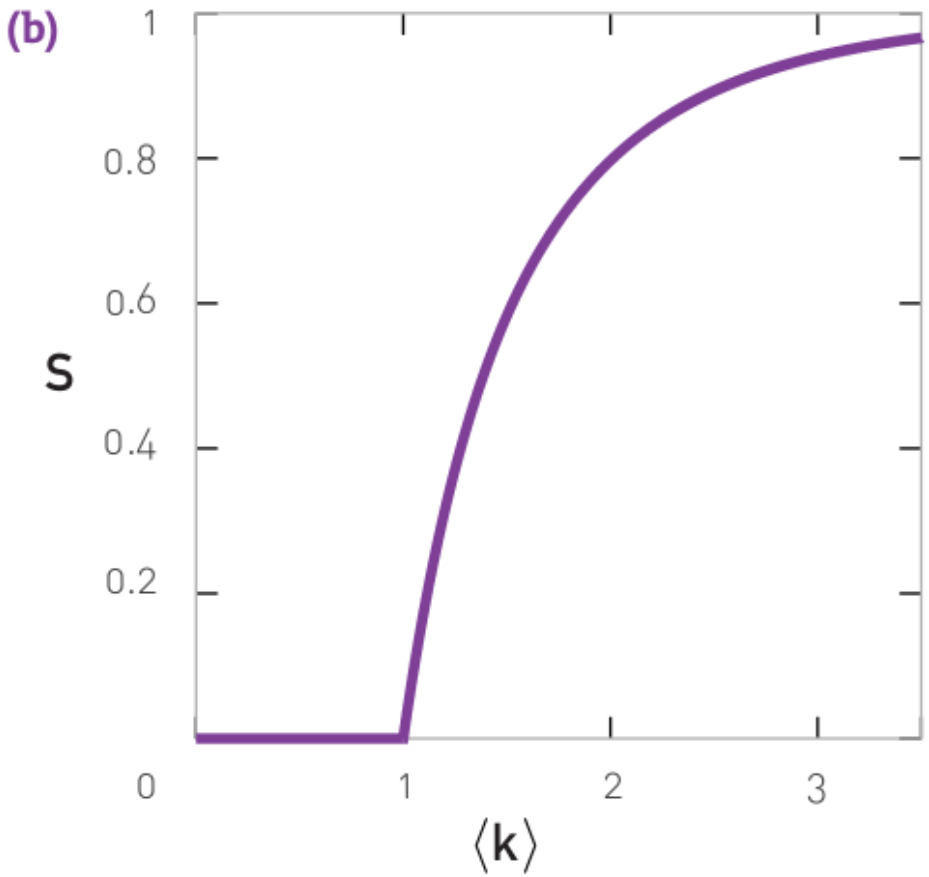
as the phase transition point is when the ths of the equation has the same derivative of the lhs, i.e. when

$$\frac{d}{dS}(1 - e^{-\langle k \rangle S}) = 1$$

$$-\langle k \rangle e^{-\langle k \rangle S} = 1$$

Giant Component

Size of the giant component in function of $\langle k \rangle$



Giant Component

The condition for the emergence of the giant component is

$$\langle k \rangle = 1$$

In other words, we have a giant component if and only if each node has on average more than one link

The condition

$$\langle k \rangle = 1$$

can be expressed in terms of p, obtaining

$$p_c = \frac{1}{N-1} \approx \frac{1}{N}$$

Therefore, the larger the network, the smaller p is sufficient for the giant component

Giant Component

The emergence of the giant component is only one of the transitions characterizing a random network as we change $\langle k \rangle$. Four topologically distinct regimes can be distinguished:

- Subcritical regime

$$0 < \langle k \rangle < 1 \quad (p < \frac{1}{N})$$

- Critical point

$$\langle k \rangle = 1 \quad (p = \frac{1}{N})$$

- Supercritical regime

$$\langle k \rangle > 1 \quad (p > \frac{1}{N})$$

- Connected regime:

$$\langle k \rangle > lnN \quad (p > \frac{lnN}{N})$$

Giant Component

Subcritical regime

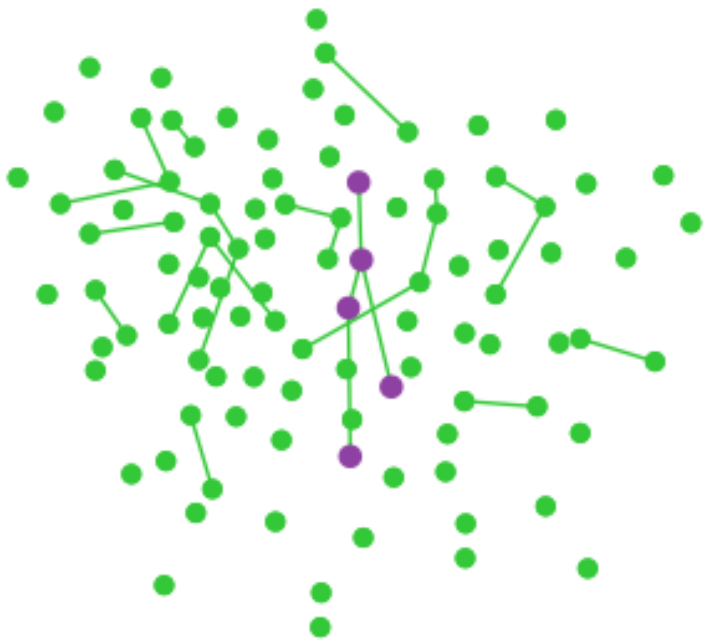
$$0 < \langle k \rangle < 1 \quad (p < \frac{1}{N})$$

- For $\langle k \rangle = 0$ the network consists of N isolated nodes
- Increasing $\langle k \rangle$ means that we are adding $N\langle k \rangle = pN(N-1)/2$ links to the network
- Nevertheless, given that $\langle k \rangle < 1$, we have only a small number of links in this regime, hence we mainly observe tiny clusters

Giant Component

Subcritical regime

$$0 < \langle k \rangle < 1 \quad (p < \frac{1}{N})$$



Giant Component

Critical point

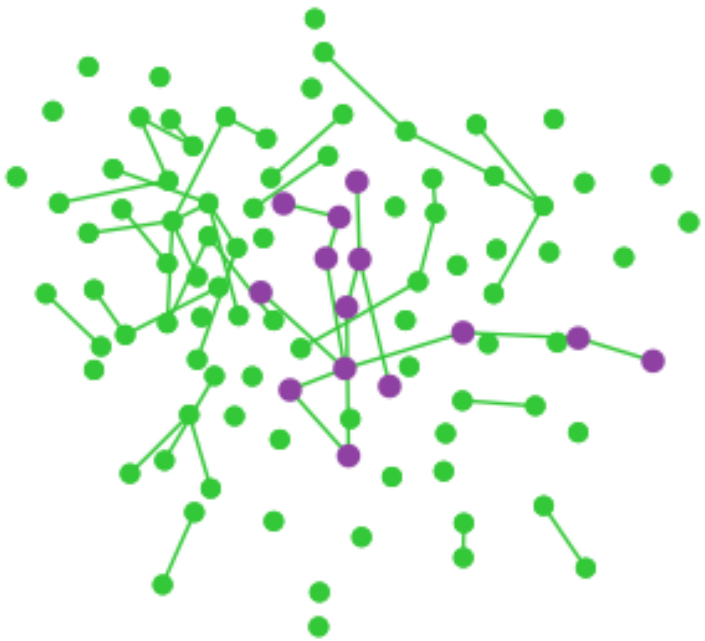
$$\langle k \rangle = 1 \quad (p = \frac{1}{N})$$

- The critical point separates the regime where there is not yet a giant component from the regime where there is one
- The relative size of the largest component is still zero
- At the critical point most nodes are located in numerous small components, and components of rather different sizes coexist
- Many properties of the network at the critical point resemble the properties of a physical system undergoing a phase transition

Giant Component

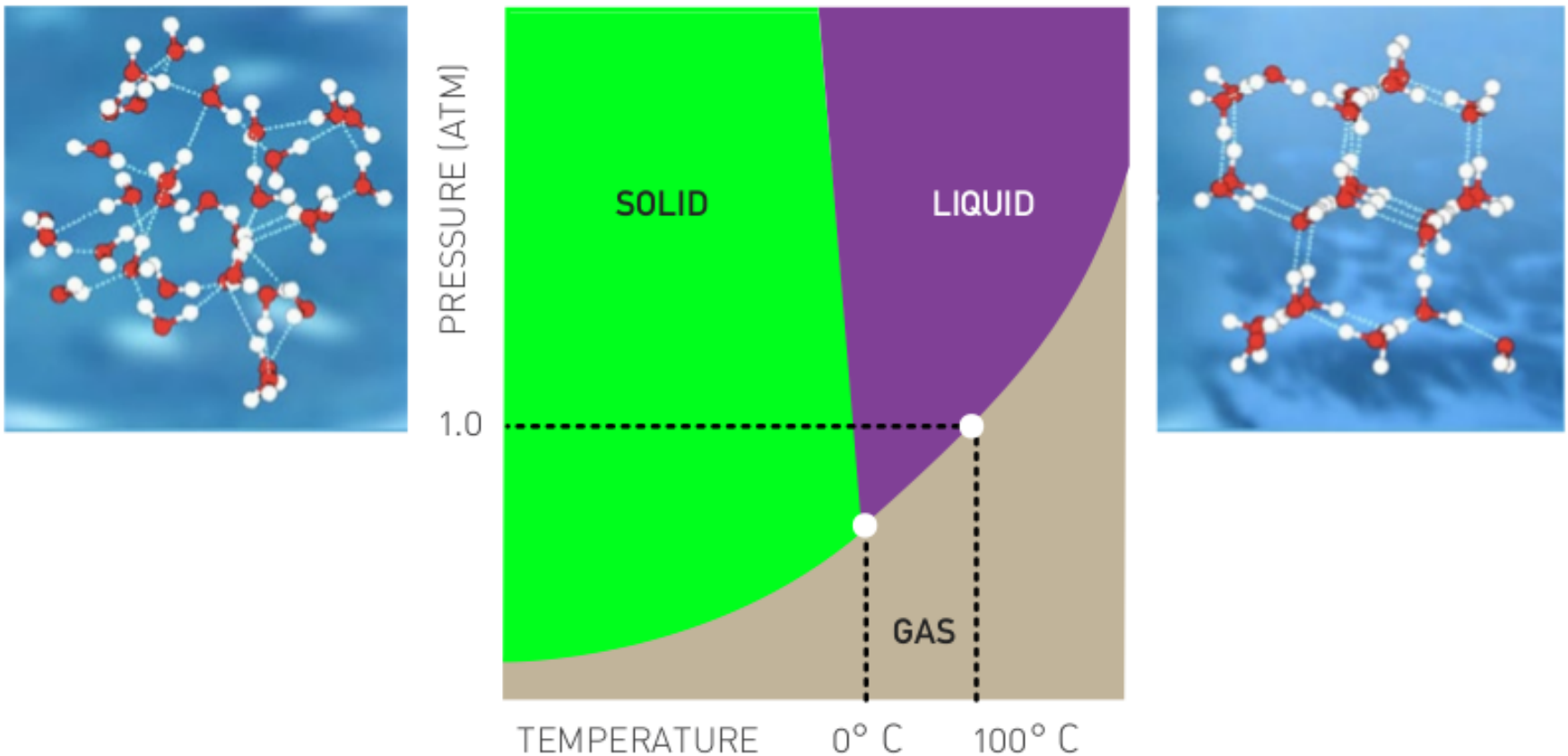
Critical point

$$\langle k \rangle = 1 \quad (p = \frac{1}{N})$$



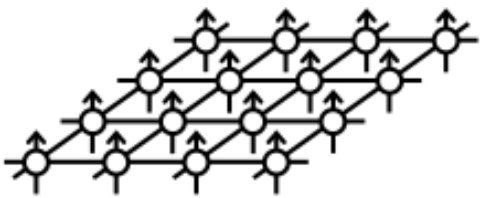
Giant Component

Critical point: water-ice phase transition

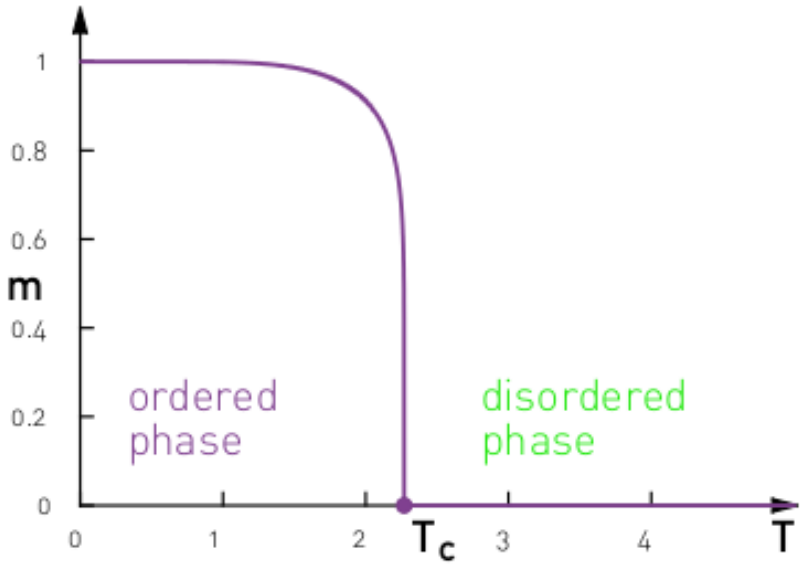


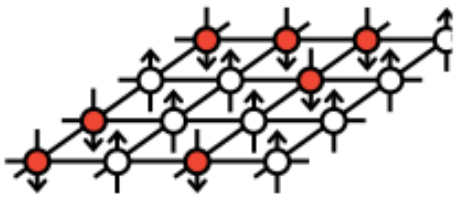
Giant Component

Critical point: magnetic phase transition



ordered phase





disordered phase

Giant Component

Supercritical regime

$$\langle k \rangle > 1 \quad (p > \frac{1}{N})$$

- This regime is relevant for most systems, as for the first time we have a giant component that looks like a network
- It can be demonstrated that in the vicinity of the critical point, the size of the giant component varies as

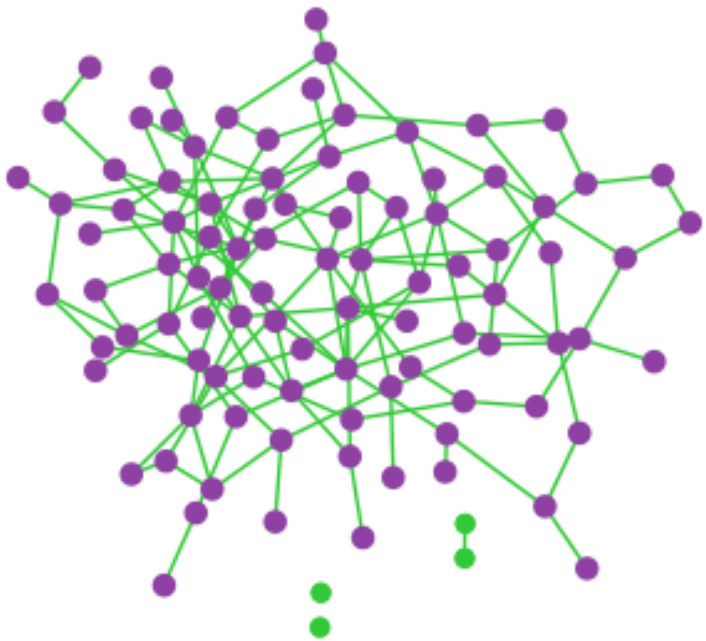
$$N_G \approx (p - p_c)N$$

- In other words, the giant component contains a finite fraction of the nodes. The further we move from the critical point, a larger fraction of nodes will belong to it
- In the supercritical regime numerous isolated components coexist with the giant component
- The supercritical regime lasts until all nodes are a bsorbed by the giant component

Giant Component

Supercritical regime

$$\langle k \rangle > 1 \quad (p > \frac{1}{N})$$



Giant Component

Connected regime

$$\langle k \rangle > lnN \quad (p > \frac{lnN}{N})$$

- For sufficiently large p, the giant component absorbs all nodes and components, hence $N_G \approx N$
- In the absence of isolated nodes the network becomes connected and the average degree depends on N as

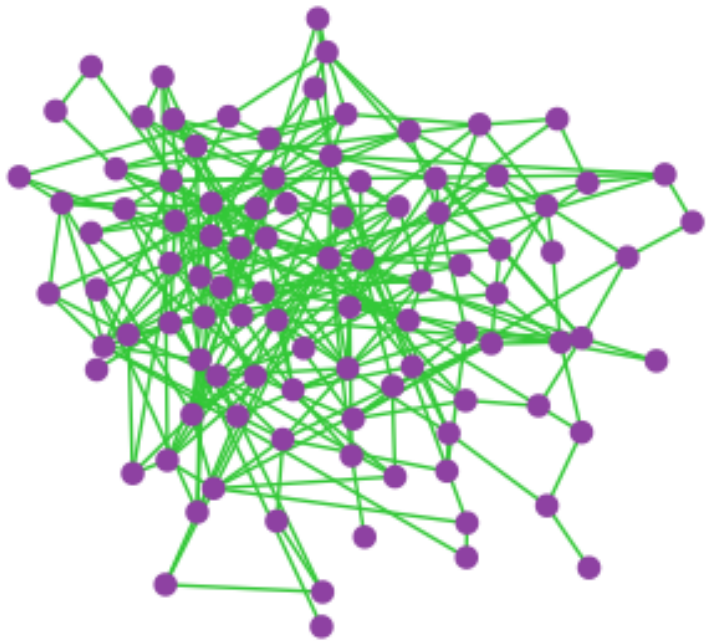
$$\langle k \rangle = lnN$$

- When we enter the connected regime, the network is still relatively sparse, as $lnn/N \to 0$ for large N. The network turns into a complete graph only at $\langle k \rangle = N - 1$

Giant Component

Connected regime

$$\langle k \rangle > lnN \quad (p > \frac{lnN}{N})$$



Real Networks are Supercritical

Two predictions of random network theory are important for real networks:

- Once the average degree exceeds $\langle k \rangle = 1$, a giant component should emerge that contains a finite fraction of all nodes. Hence only for $\langle k \rangle > 1$ the nodes organize themselves into a recognizable network
- For $\langle k \rangle > lnN$ all components are absorbed by the giant component, resulting in a single connected network

Real Networks are Supercritical

The measurements indicate that real networks exceed the $\langle k \rangle = 1$ threshold

Network	N	L	$\langle k \rangle$	lnN
Internet	192244	609066	6.34	12.17
Power Grid	4941	6594	2.67	8.51
Science Collaboration	23133	94439	8.08	10.05
Actor Network	702388	29397908	83.71	13.46
Protein Interactions	2018	2930	2.90	7.61

Real Networks are Supercritical

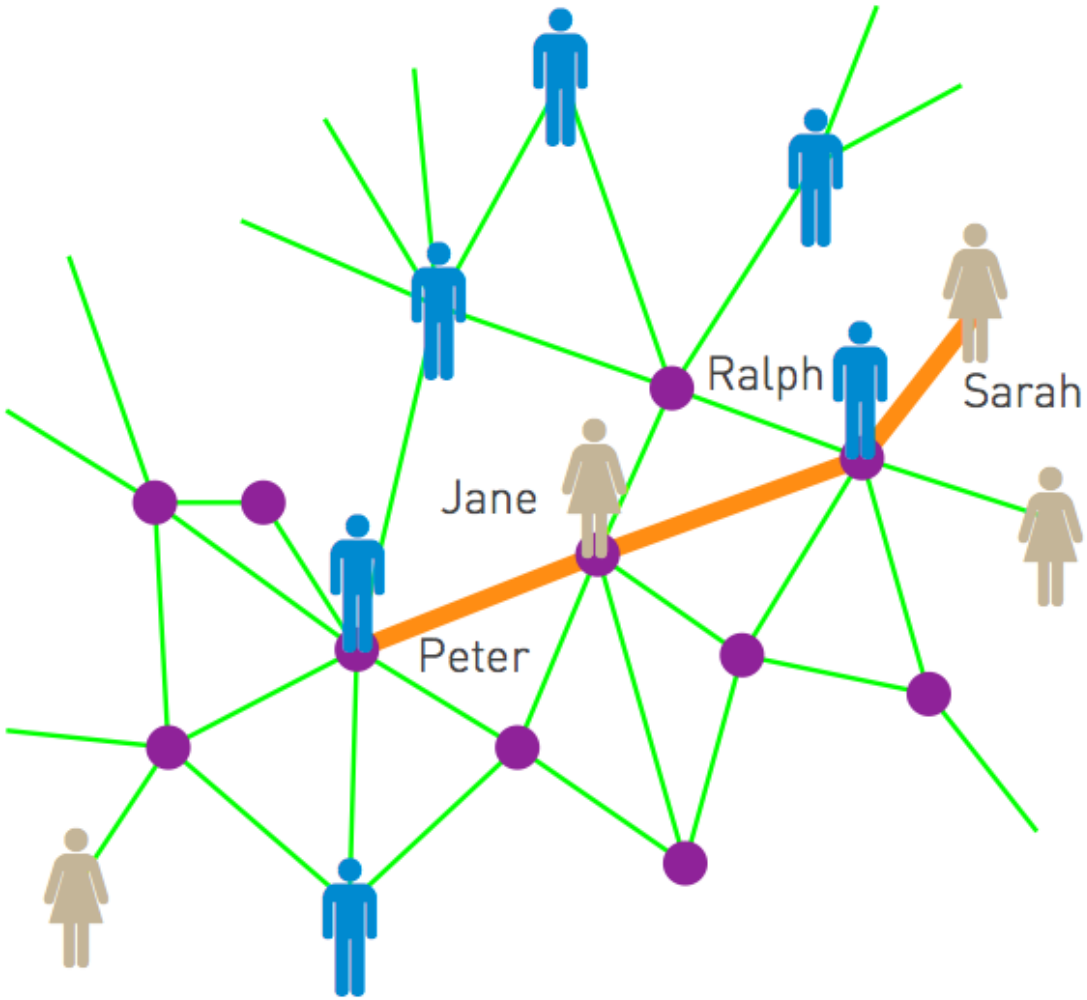
Some considerations

- As for the first prediction, the average degree of the networks shown in Table is well beyond the threshold value $\langle k \rangle > 1$ and therefore they all have a giant component
- As for the second prediction ($\langle k \rangle > ln N$), for social networks the transition between the supercritical and the fully connected regime should be at $\langle k \rangle > ln(7 \times 10^9) \approx 22.7$
- That is, if the average individual has more that two dozens acquaintances, then a random society must have a single component, leaving no individual disconnected.
- According to random network model, most real networks are in the supercritical regime. Therefore these networks are expected to have a giant component which should coexist with many disconnected components

Small World

- The *small world phenomenon* states that if you choose any two individuals anywhere on Earth, you will find a path of at most six acquaintances between them
- The small world concept states that even individuals who are on the opposite side of the world can be connected to us via a few acquaintances
- In the language of network science, the small world phenomenon implies that *the distance between two randomly chosen nodes in a network is short*
- What does short mean?
- How do we explain the existence of these short distances?

Small World



Small World

To answer these questions, consider a random network with average degree $\langle k \rangle$. A node in this network has in average

- $\langle k \rangle$ nodes at distance 1 ($d = 1$)
- $\langle k \rangle^2$ nodes at distance 2 ($d = 2$)
- $\langle k \rangle^3$ nodes at distance 3 ($d = 3$)
- ...
- $\langle k \rangle^d$ nodes at distance d

For example, if $\langle k \rangle \approx 1000$, we expect 10^6 individuals at distance two and about a billion at distance three

Small World

To be more precise, the expected number of nodes up to distance d from our starting node is

$$N(d) \approx 1 + \langle k \rangle + \langle k \rangle^2 + \cdots + \langle k \rangle^d = \frac{\langle k \rangle^{d+1} - 1}{\langle k \rangle - 1}$$

Of course $N(d) < N$, therefore the maximum distance d_{max} is

$$N(d_{max}) \approx N$$

Assuming that $\langle k \rangle \gg 1$ we can write

$$\begin{aligned} N(d_{max}) &\approx \frac{\langle k \rangle^{d_{max}+1} - 1}{\langle k \rangle - 1} \\ &\approx \frac{\langle k \rangle^{d_{max}+1}}{\langle k \rangle} \\ &= \langle k \rangle^{d_{max}} \end{aligned}$$

Small World

Therefore we can write

$$\langle k \rangle^{d_{max}} \approx N$$

from which, taking the logarithms of both sides

$$d_{max} \approx \frac{lnN}{ln\langle k \rangle}$$

which is the mathematical formulation of the small world phenomenon

d_{max} is often dominated by a few extreme paths. For this reason it is better to use $\langle k \rangle$ which is averaged over all node pairs. Hence the small world phenomenon property is defined by

$$\langle d \rangle \approx \frac{lnN}{ln\langle k \rangle}$$

Small World

Some considerations

$$\langle d \rangle \approx \frac{lnN}{ln\langle k \rangle}$$

- In general $lnN \ll N$, hence the dependence of $\langle d \rangle$ on lnN implies that the distances in a random network are *orders of magnitude smaller than the size of the network*. As a consequence, by *small* in the "small world phenomenon" we mean that the *average path length or the diameter dependes logarithmically on the system size*
- The $1/ln\langle k \rangle$ implies that the denser the network, the smaller is the distance between the nodes

Small World

Some considerations

$$\langle d \rangle \approx \frac{lnN}{ln\langle k \rangle}$$

- In social networks we have $N \approx 7 \times 10^9$ and $\langle k \rangle \approx 10^3$
- We obtain

$$\langle d \rangle \approx \frac{ln(7 \times 10^9)}{ln(10^3)} = 3.28$$

- Therefore, all individuals on Earth should be within three to four handshakes of each other. This estimate is probably closer to the real value that the frequently cited six degrees

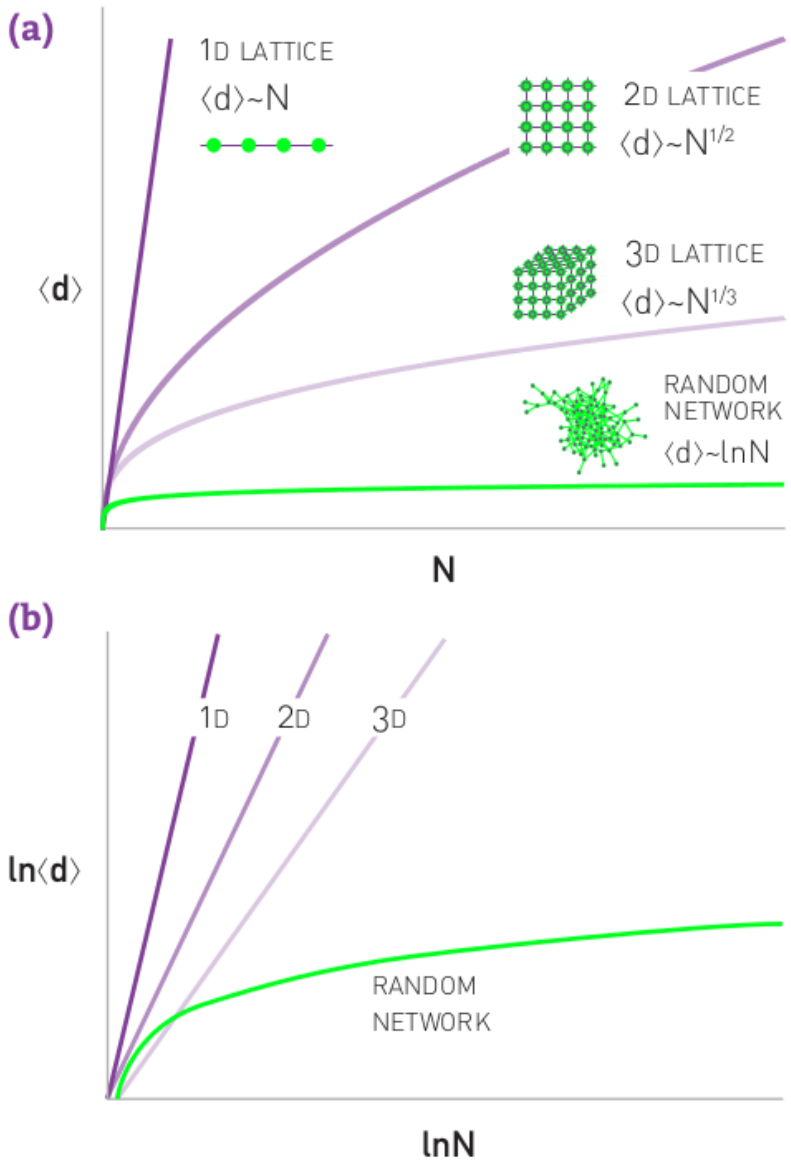
Small World

Some considerations

NETWORK	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\frac{\ln N}{\ln \langle k \rangle}$
Internet	192,244	609,066	6.34	6.98	26	6.58
WWW	325,729	1,497,134	4.60	11.27	93	8.31
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	93,439	8.08	5.35	15	4.81
Actor Network	702,388	29,397,908	83,71	3,91	14	3,04
Citation Network	449,673	4,707,958	10.43	11,21	42	5.55
E. Coli Metabolism	1,039	5,802	5.58	2.98	8	4.04
Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

Small World

Some considerations



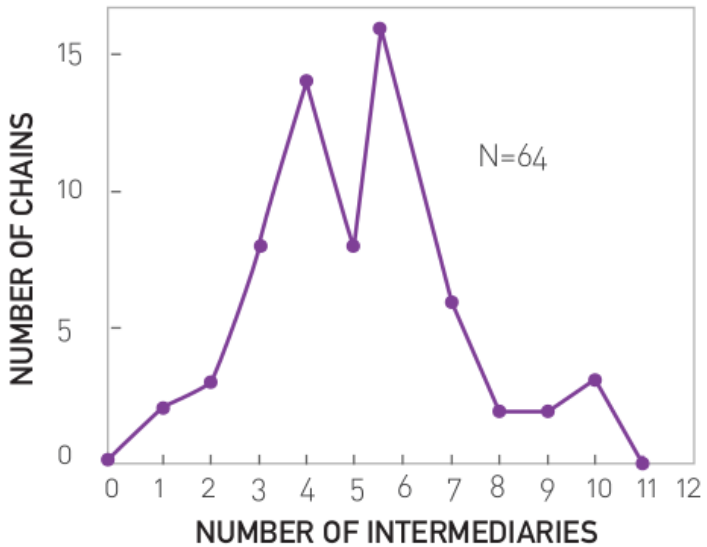
Small World

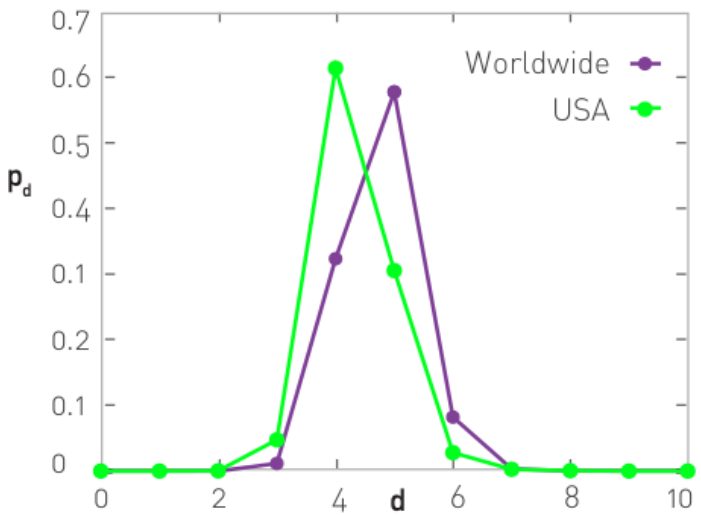
Six degrees of separation

- In 1967 Stanley Milgram designed an experiment to measure the distances in social networks
- He chose a stock broker in Boston and a student in Sharon, Massachussets as *targets*
- He randomly selected residents of Wichita and Omaha sending them a letter containing a short summary of the purpose of the study, the name, address and information about the target person
- They were asked to forward the letter to a friend, relative or acquaintance who is most likely to know the target person
- Within a few days the first letter arrived, passing through only two links
- Eventually 64 of the 296 letters arrived

Small World

Six degrees of separation





Clustering Coefficient

- The degree of a node contains no information about the relationship among a node's neighbours
- Information about this relationship is provided by the local clustering coefficient C_i, that measures the density of links in the immediate neighbourhood of node i
- $C_i = 0$ means that there are no links among the neighbours
- $C_i = 1$ means that each of the neighbours of i link to each other

Clustering Coefficient

- To calculate C_i for a node in a random network we need to estimate the expected number of links L_i between the k_i neighbours of the node
- In a random network the probability that two of the neighbours of i link to each other is p
- There are $k_i(k_i-1)/2$ possible links between the k_i neighbours of node i
- The expected value of L_i is

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

The local clustering coefficient of a random network is

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

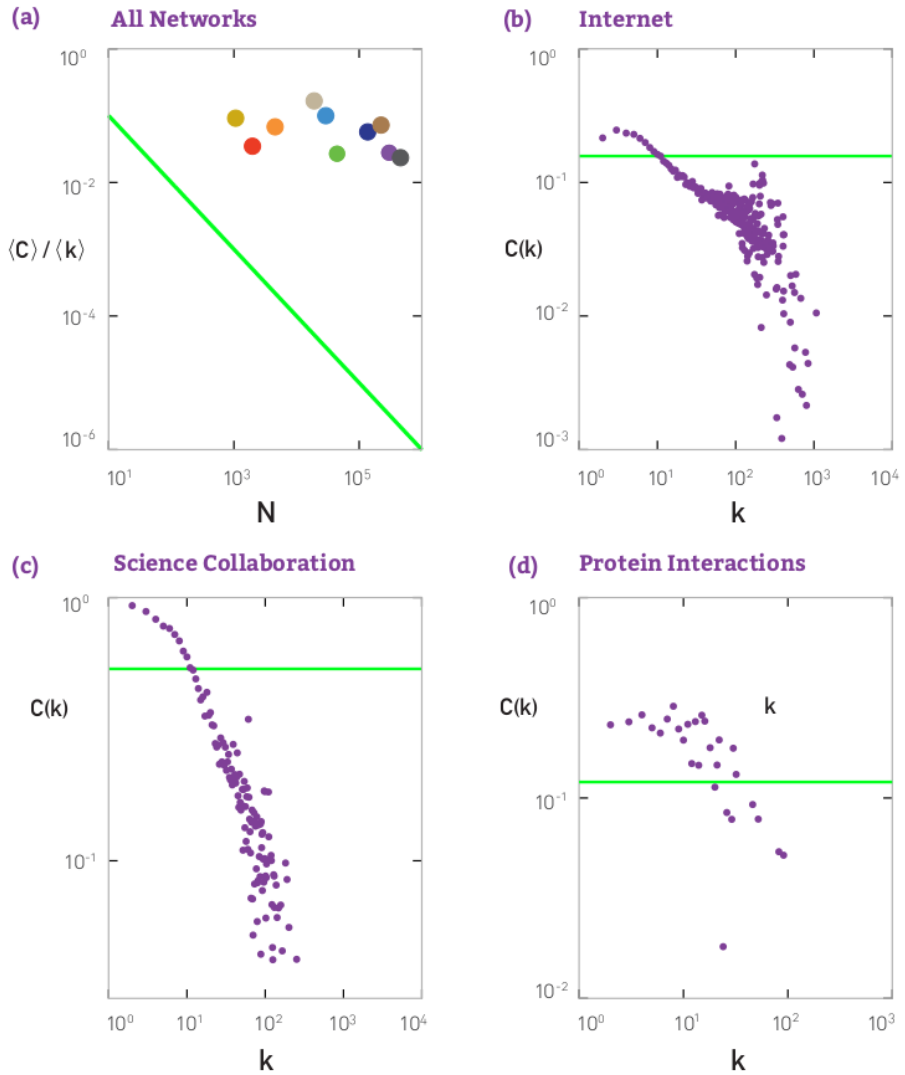
Clustering Coefficient

The local clustering coefficient of a random network is

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

- For fixed $\langle k \rangle$, the larger the network, the smaller is the clustering coefficient of a node. As a consequence, C_i is expected to decrease as $1/N$
- The local clustering coefficient of a node is independent of the node's degree

Clustering Coefficient



Clustering Coefficient

- The random network model does not capture the clustering of real networks
- Instead real networks have a much higher clustering coefficient than expected for a random network of similar N and L
- An extension of the random network model has been proposed by Watts and Strogatz

Watts-Strogatz Model

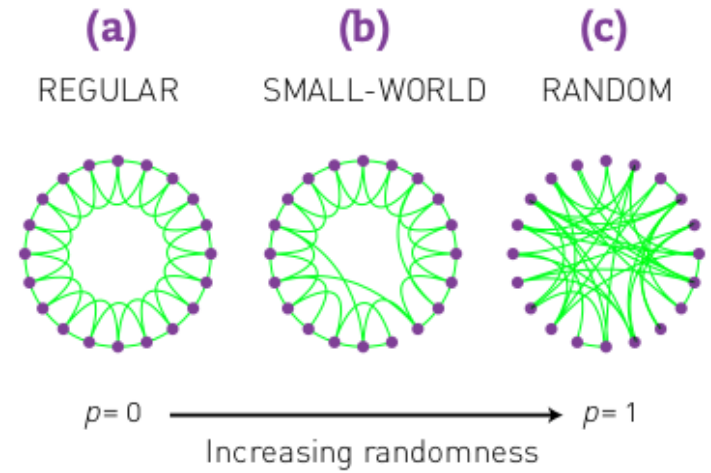
Duncan Watts and Steven Strogatz proposed an extension of the random network model motivated by two considerations:

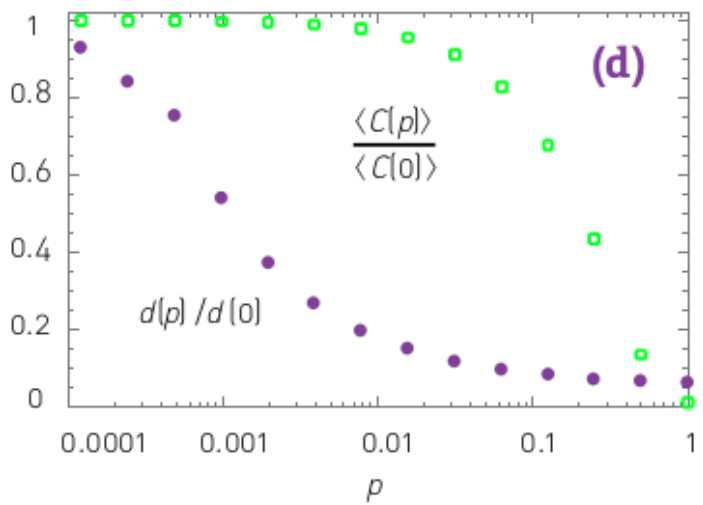
- In real networks the average distance between two nodes depends logarithmically on N rather than following a polynomial expected for regular lattices

$$d_{max} \approx \frac{lnN}{ln\langle k \rangle}$$

- The average clustering coefficient of real networks is much higher than expected for a random network of similar N and L

Watts-Strogatz Model





Watts-Strogatz Model

A Watts-Strogatz network can be constructed as follows:

- We start from a ring of nodes, each node being connected to their immediate and next neighbours. Hence initially each node has $\langle C \rangle = 3/4$
- With probability p each link is rewired to a randomly chosen node. For small p the network maintains high clustering but the random long-range links can drastically decrease the distances between the nodes
- For $p = 1$ all links have been rewired, so the network turns into a random network

Watts-Strogatz Model

- The Watts-Strogatz model interpolates between a *regular lattice*, which has high clustering but lacks the small-world phenomenon, and a *random network*, which has low clustering but displays the small-world property
- Numerical simulations show that for a range of rewiring parameters the average path length is low but the clustering coefficient is high, hence reproducing the coexistence of high clustering and small-world phenomena
- The Watts-Strogatz model predicts a Poisson-like bounded degree distribution. Consequently high degree nodes are absent from it
- Furthermore it predicts a k-independent $C(k)$