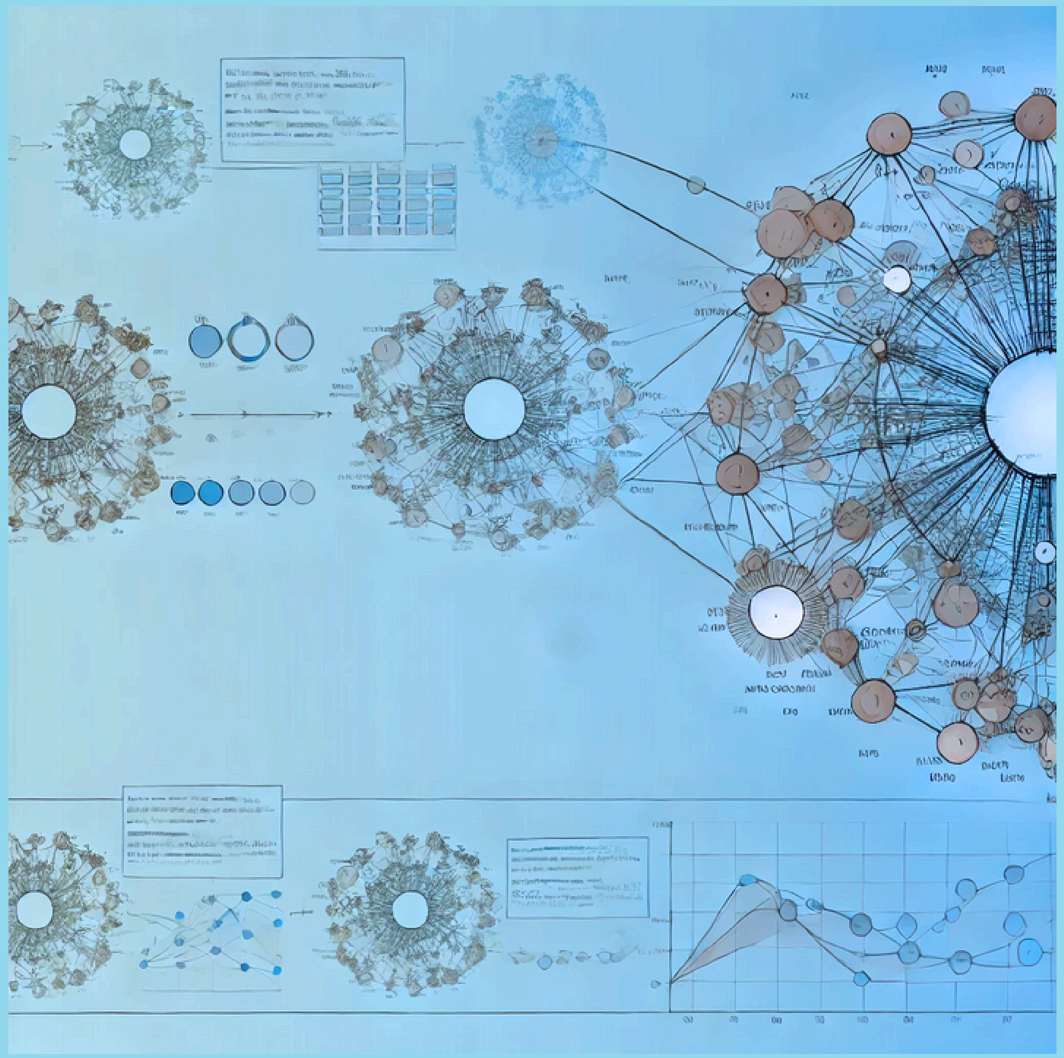


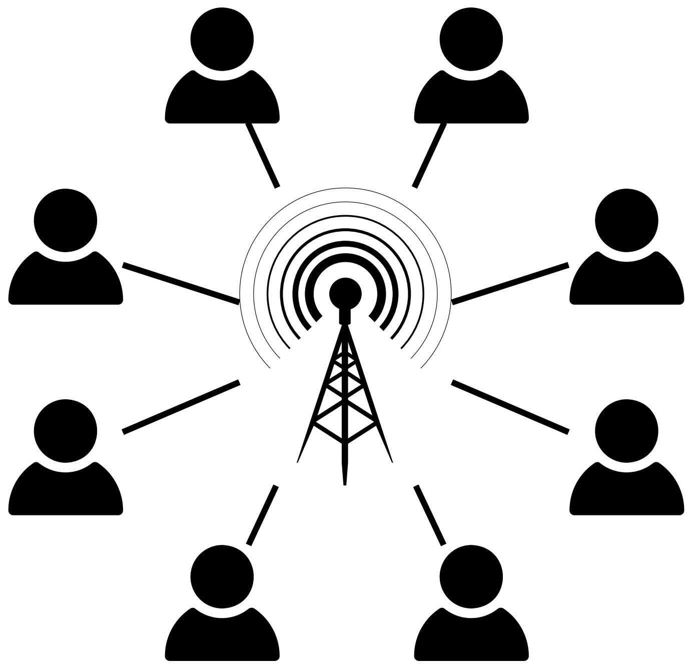
UNIVERSITÄT KONSTANZ

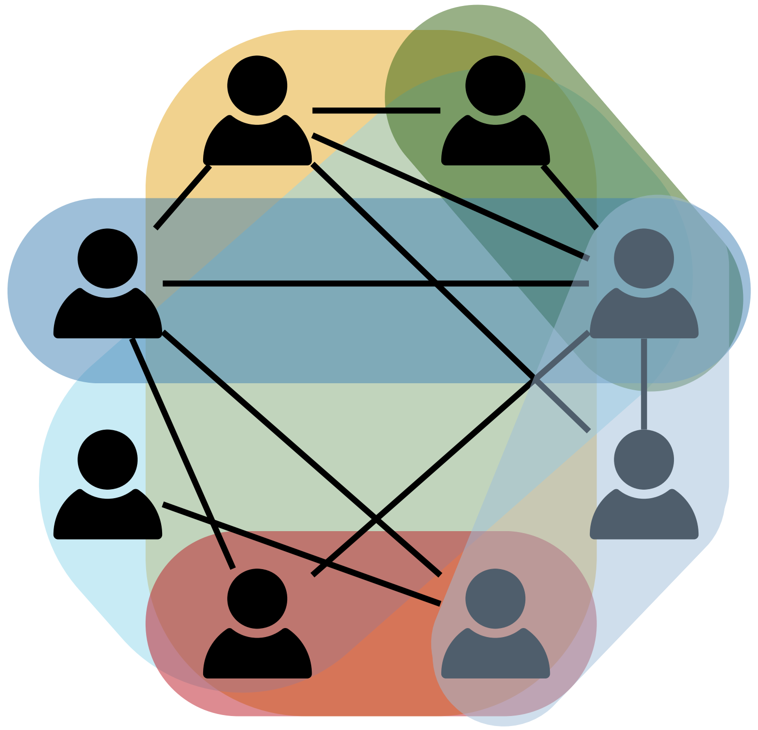
Random Graphs and Small World

Network Science of Socio-Economic Systems

Giordano De Marzo

Universität Konstanz





Recap

Networks Basics

We introduced the basics concepts of network science

Measuring Networks

Networks can be characterized in terms of diameter, clustering, degree distribution

Real World Networks

Real world networks are characterized by: small world, high clustering, scale free degree distribution, homophily, sparsity

The Value of Networks

We discussed different laws describing the value of networks

Outline

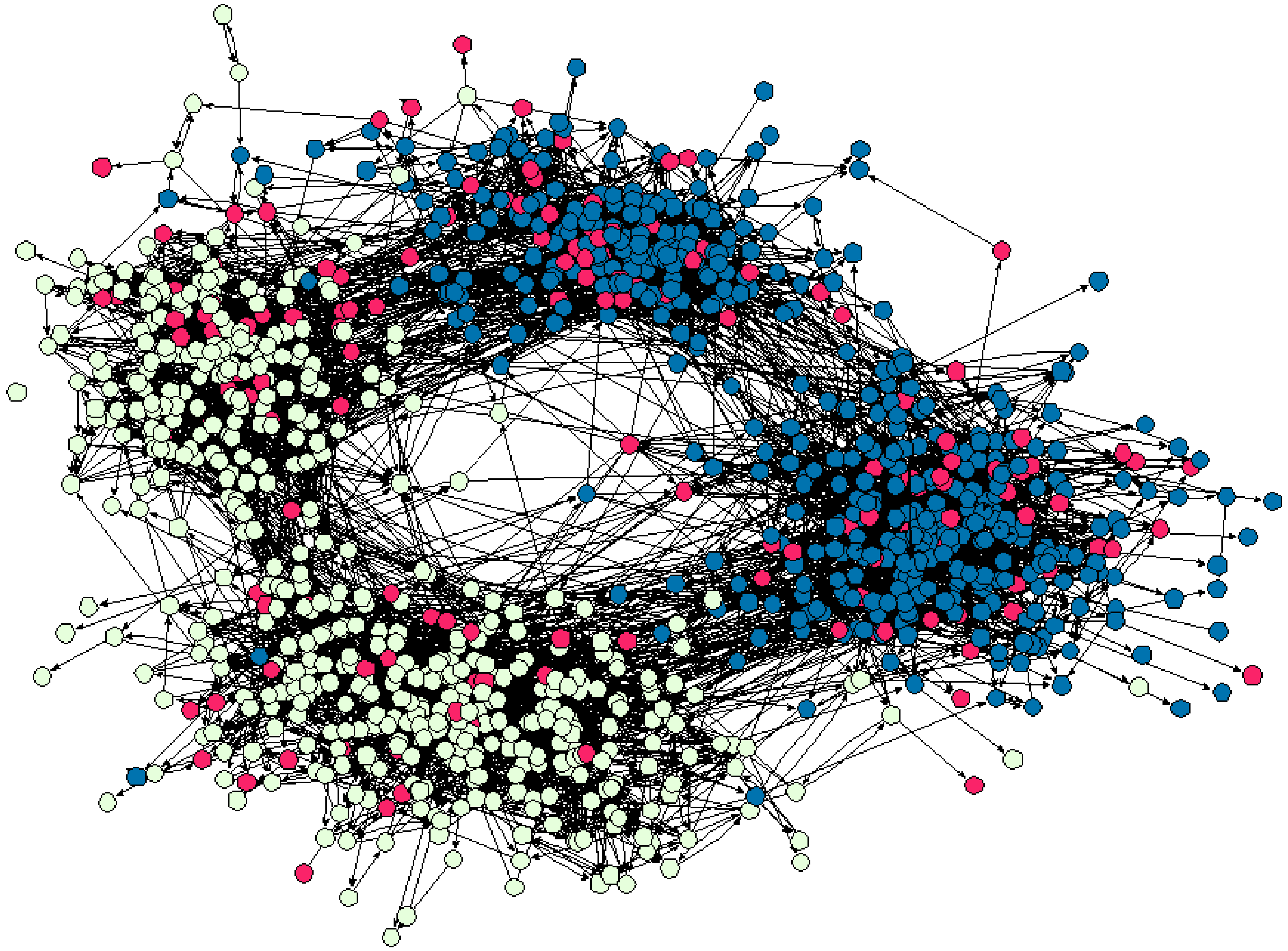
1. Random Graphs
2. Small World and Clustering
3. Watts-Strogatz Model
4. Network Robustness

Random Graphs

Why Random Graphs?

Real world networks have a structure and do not look being random. So why are we interested in random networks?

- in some cases real networks can be approximated as random
- if we want to understand which properties are significant, we need a null model
- random networks are the simplest possible null models for networks



The Erdős–Rényi Model

In the Erdős–Rényi model, a graph is generated by randomly connecting nodes with edges

- the model has two parameters
  - **N** total number of nodes
  - **p** probability of link creation
- it is generally denoted as **G(N, p)**

The model starts with N disconnected nodes

- for each pair of nodes a link is created with a probability p
- links are completely random and uncorrelated

Degree Distribution

We can easily compute the degree distribution for the Erdős–Rényi model given the random nature of the process

- Each node can connect to N-1 other nodes,
- Each connection forms independently with probability p.
- The degree k of a node is simply the count of successful connections (edges) out of N-1 possible trials.
- This is the definition of the binomial distribution

number of ways we can choose k links out on N-1 nodes

probability of k successful links

probability of N-1-k unsuccessful links

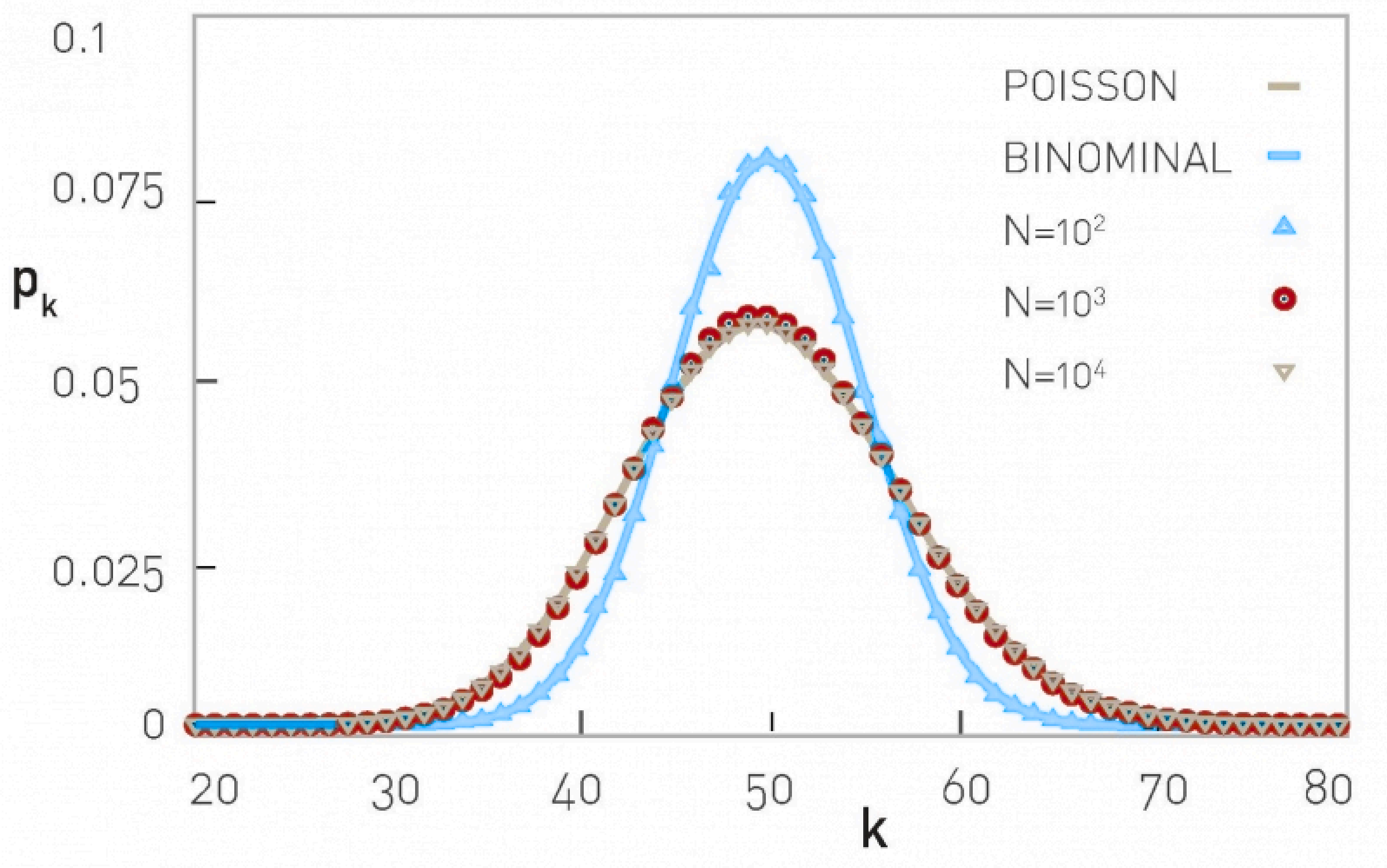
$$P(\text{degree} = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

Large Network **Limit**

The average degree in the G(N,p) is

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

- we take the limit of large networks N->∞
- we also consider small linking probability p->0
- in this way the average degree remains finite



This leads to an exponential random network with Poisson distribution

$$P(\text{degree} = k) \approx \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}$$

Evolution of Random Networks

The sparsity of a random network depends on the linking probability or equivalently on the average degree $\langle k \rangle$

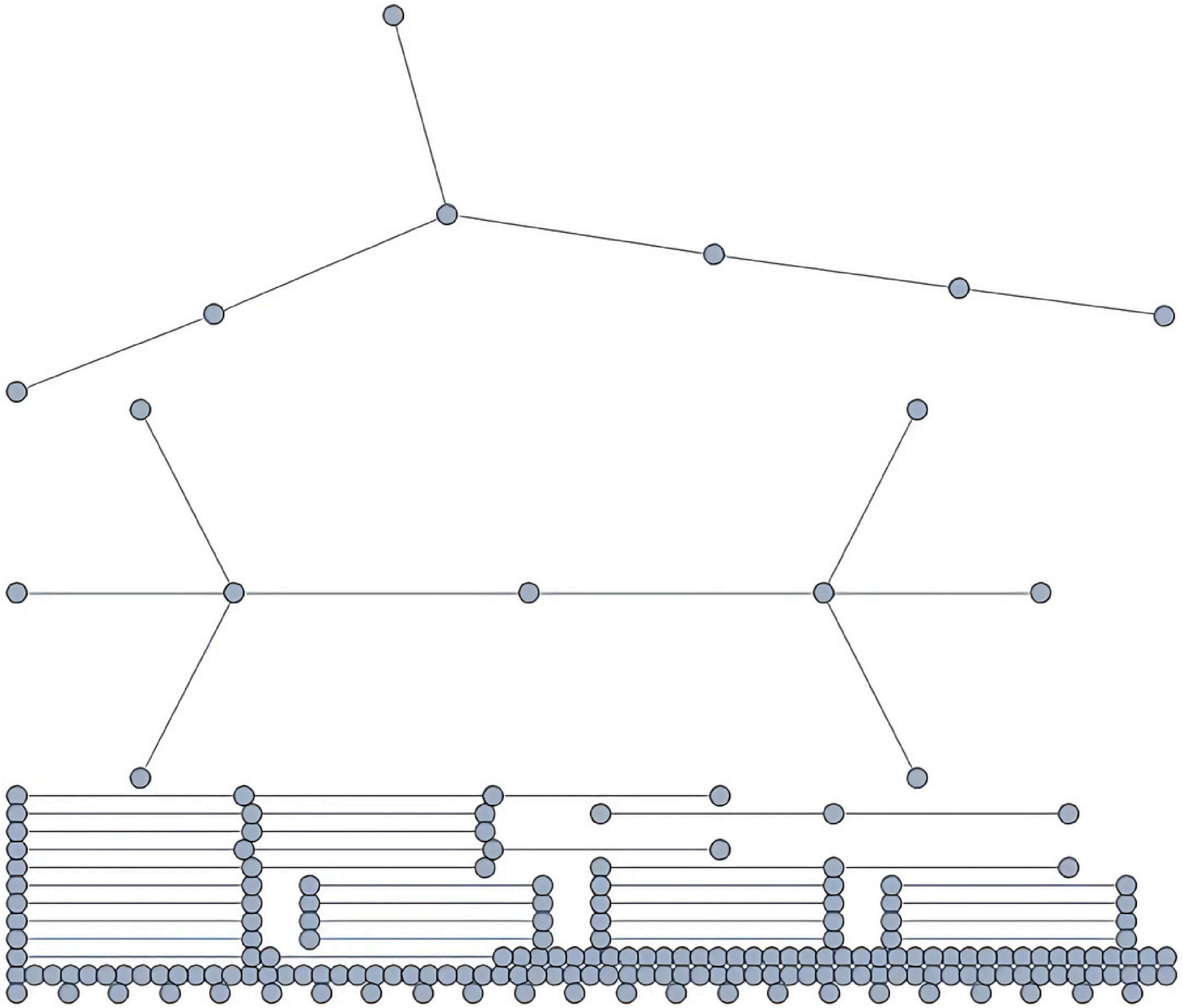
- we consider the largest connected component of the random graph
- we say that it is a **Giant Component** (GC) if it contains a non null fraction of the nodes in the network
- for large network size N, the size of the Giant Component N_G must scale as the network size

  $N_G = S \cdot N$ with $S > 0$

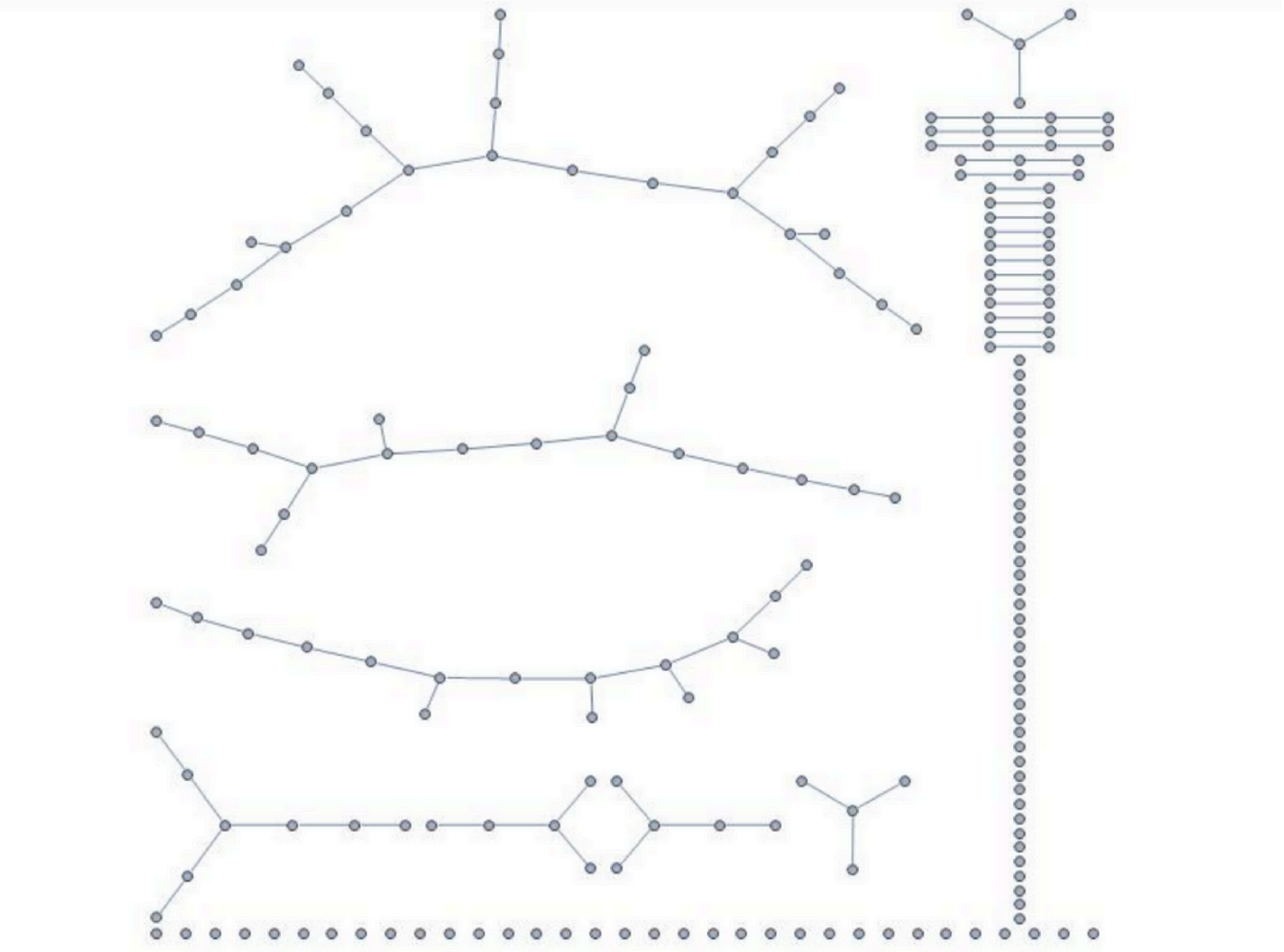
We want to understand the conditions under which the networks contains a GC

- when the network contains a GC, most of the nodes in the system are connected
- even if the network is sparse, still it is mostly connected

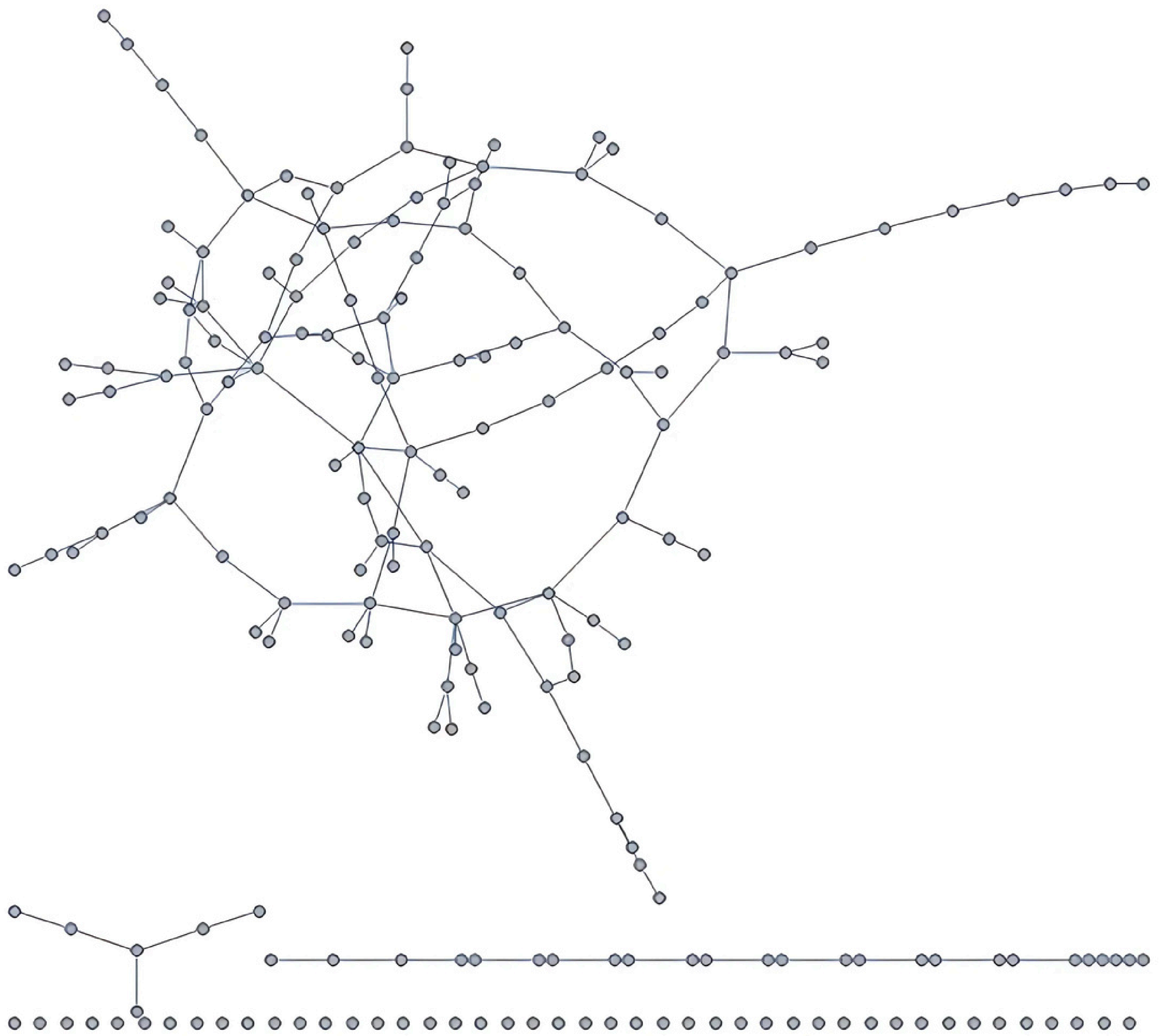
Example: N=200 ⟨k⟩=0.5



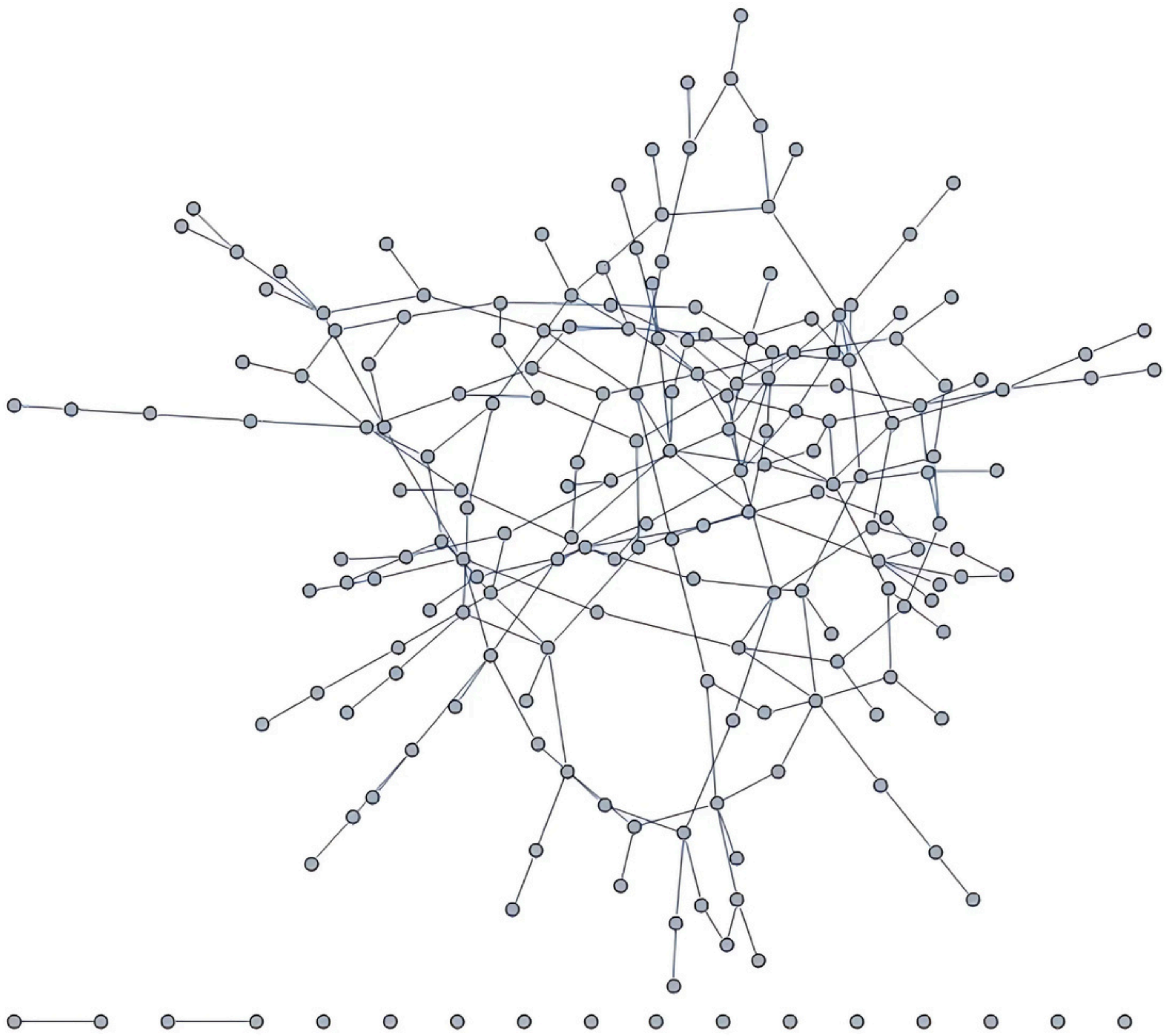
Example: N=200 ⟨k⟩=1



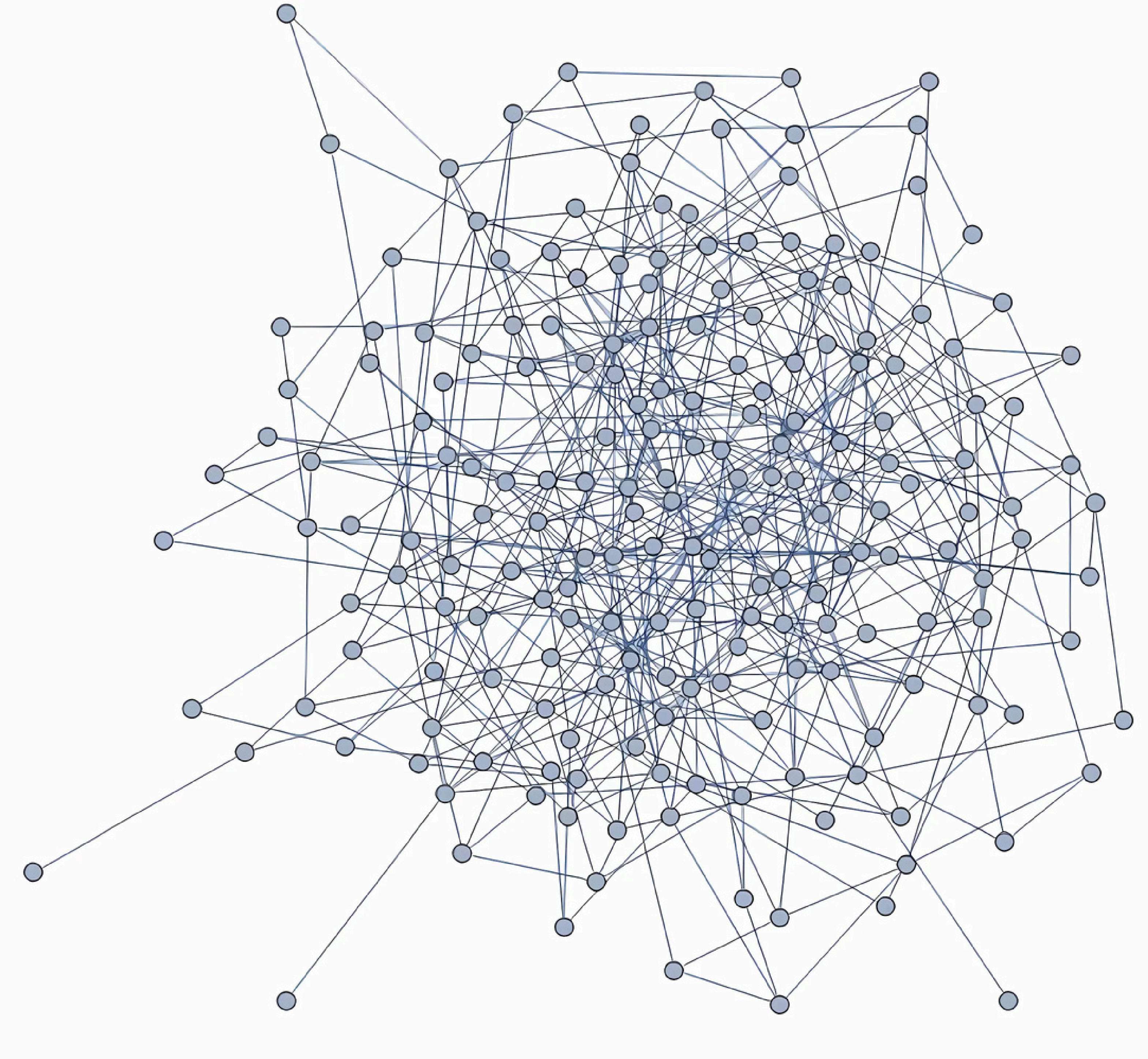
Example: N=200 $\langle k \rangle$=1.5



Example: N=200 ⟨k⟩=2



Example: N=200 $\langle k \rangle$=5



Computing the Critical Point

We denote by **Q** the **fraction of nodes not in the GC**.
Q gives the probability that a node i at random is not in the GC.
If we select a random node i, for each node j

- either i is not connected to j
  - This occur with a probability *1-p*
- or i is connected to j, but j is not in the giant component
  - This occur with a probability *pQ*
- the total probability for each node j is then *(1-p+pQ)*

There are (N-1) possible choices for j and so the total probability satisfies

$$Q = [(1-p) + p \cdot Q]^{N-1} = [1 - p(1-Q)]^{N-1}$$

Computing the Critical Point

We now take the logarithm of both sides and we expand for small p

$$\ln Q = (N-1)\ln[1-p(1-Q)] \approx -p(N-1)(1-Q)$$

The average degree satisfies ⟨k⟩=p(N-1) so we have

$$\ln Q \approx -\langle k\rangle(1-Q) \rightarrow Q \approx \mathrm{e}^{-\langle k\rangle(1-Q)}$$

Finally we denote by **S** the **fraction of nodes in the GC**, by the definition of Q it holds S=1-Q. In conclusion the fraction of nodes in the GC satisfies

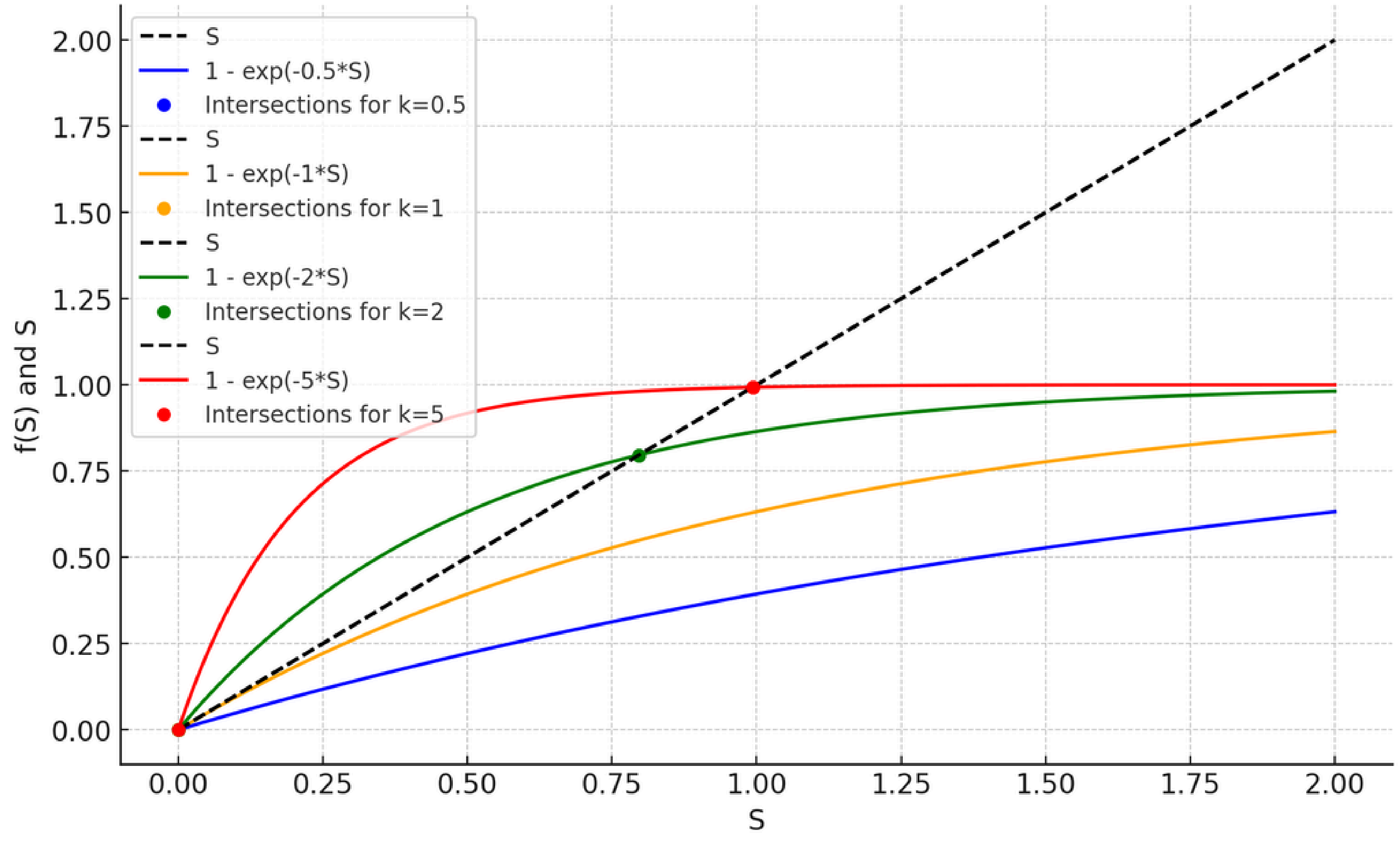
$$S \approx 1-\mathrm{e}^{-\langle k\rangle S}$$

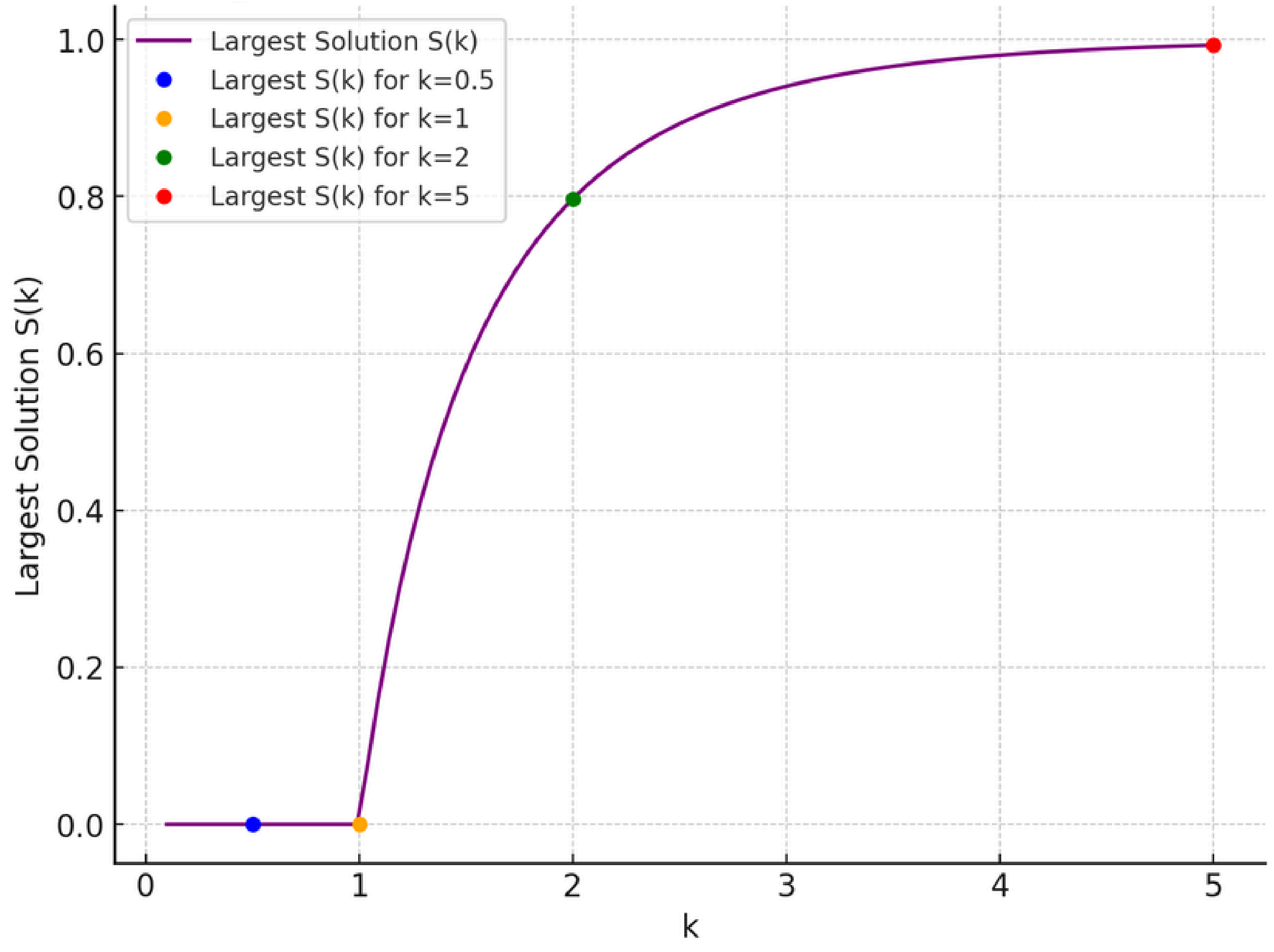
Depending on the average degree ⟨k⟩, this equation will have different solutions for S, the relative size of the giant component

Computing the Critical Point

In order to solve the equation we have to look at the intersection between y=S and y=1-exp(-⟨k⟩S)

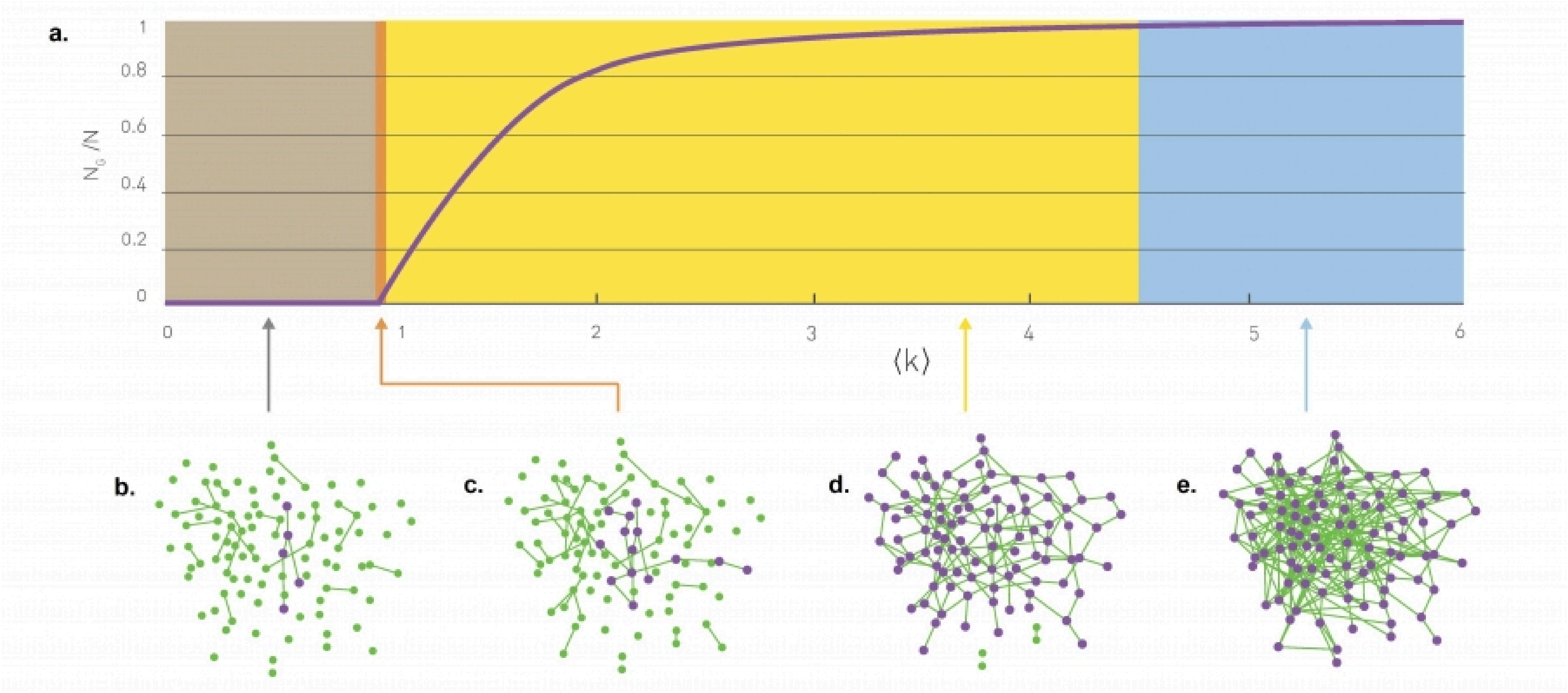
- when ⟨k⟩<1 there is only one intersection in S=0
- when ⟨k⟩>1 there is also an intersection for S>0





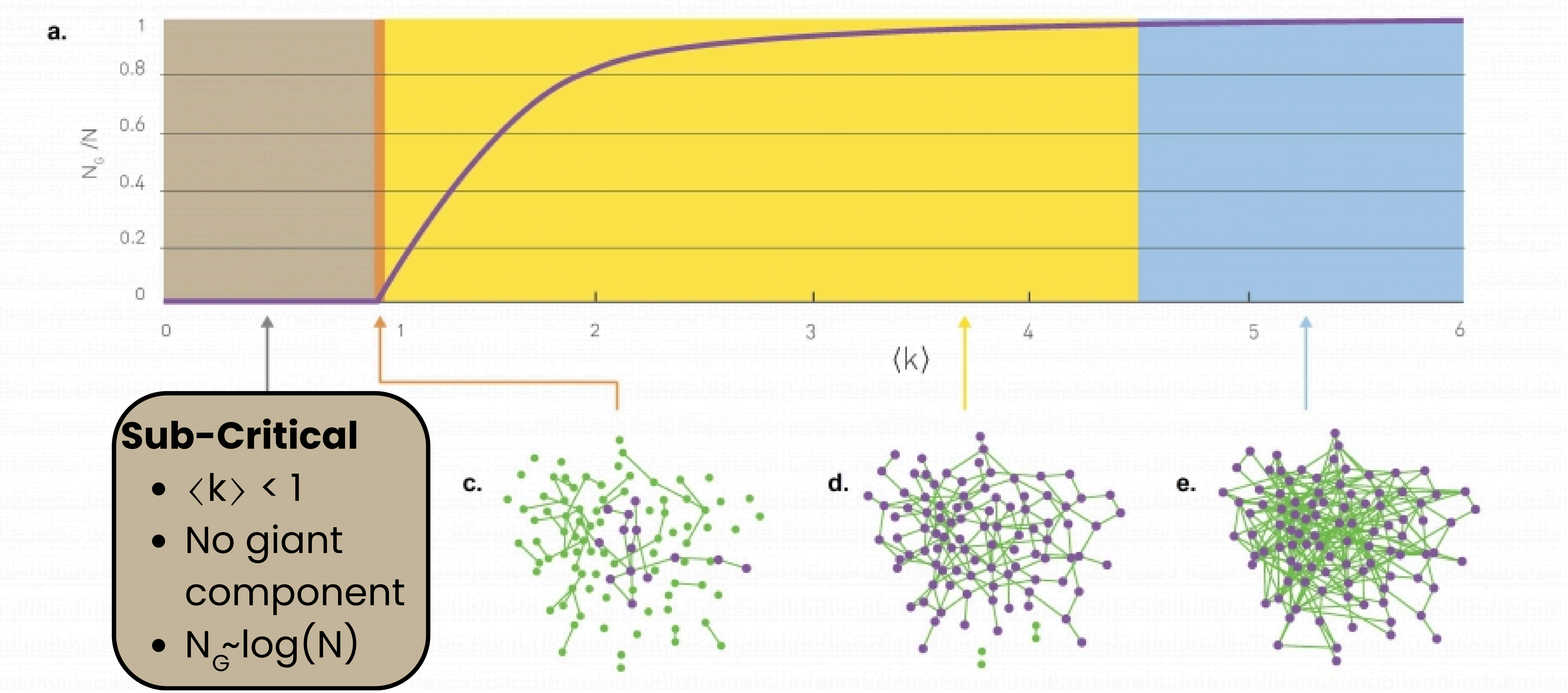
Phase Diagram

Random graphs show a second order phase transition in which a giant component emerges. The critical point is ⟨k⟩=1



Phase Diagram

For ⟨k⟩<1 the graph is subcritical and there are many disconnected components

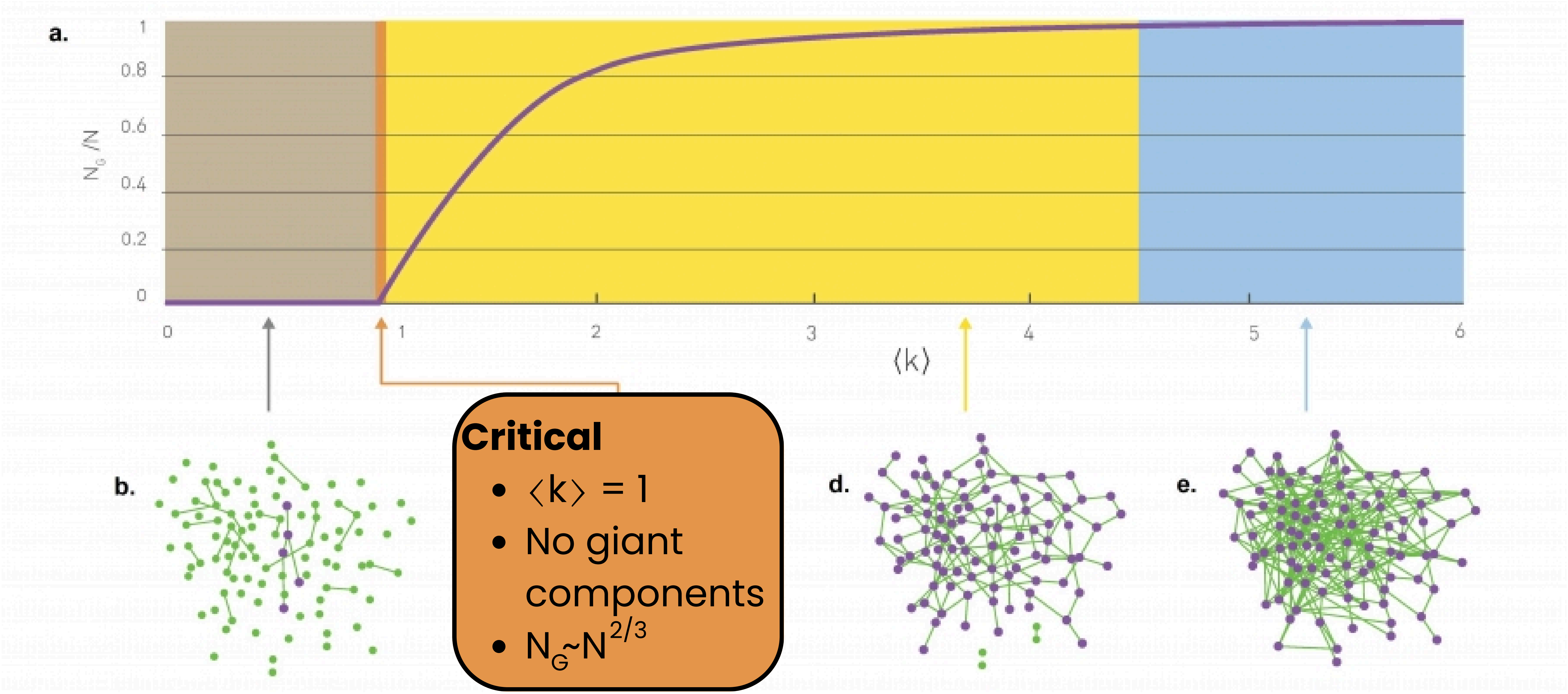


Sub-Critical

- ⟨k⟩ < 1
- No giant component
- N_G~log(N)

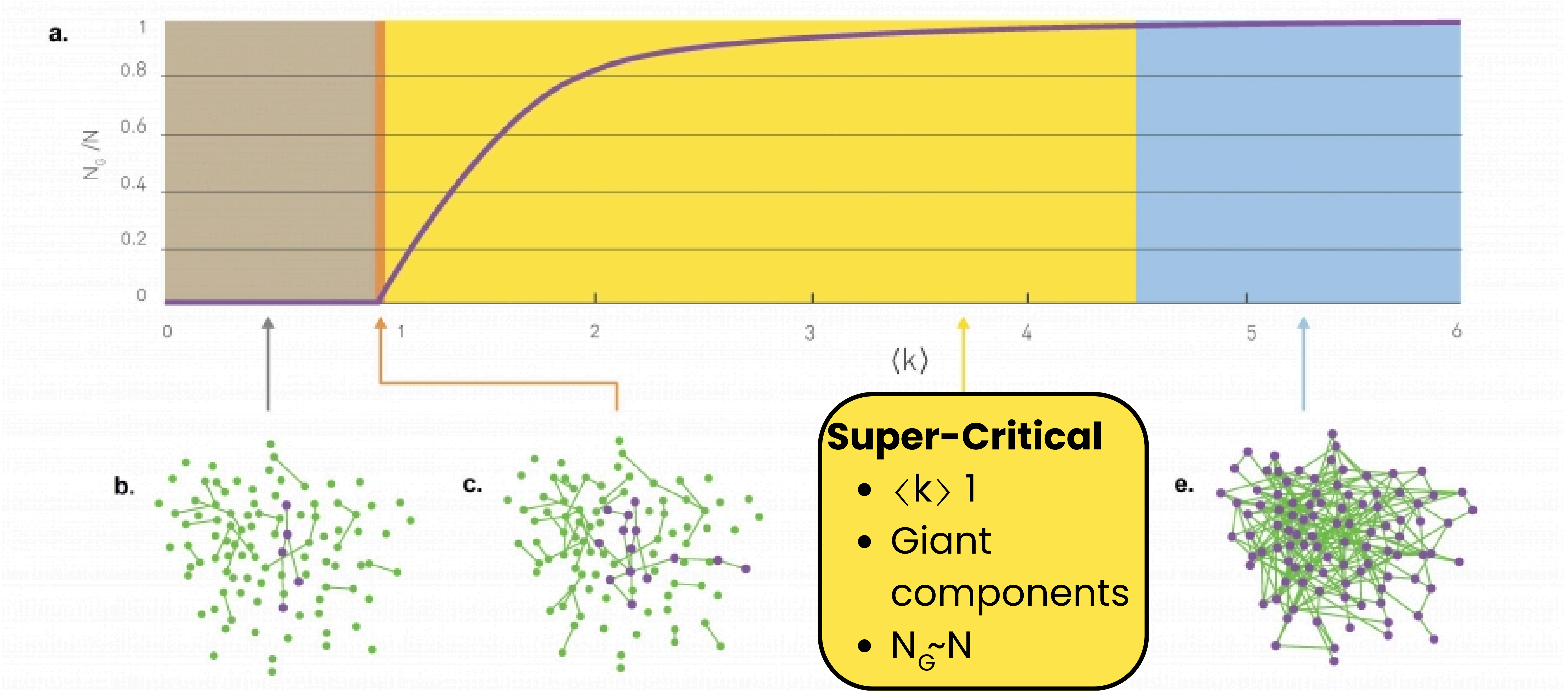
Phase Diagram

For ⟨k⟩=1 the graph is critical, there is no giant component, but larger connected components start to emerge



Phase Diagram

For ⟨k⟩>1 the graph is super-critical, there is a giant component containing a finite fraction of the nodes in the graph

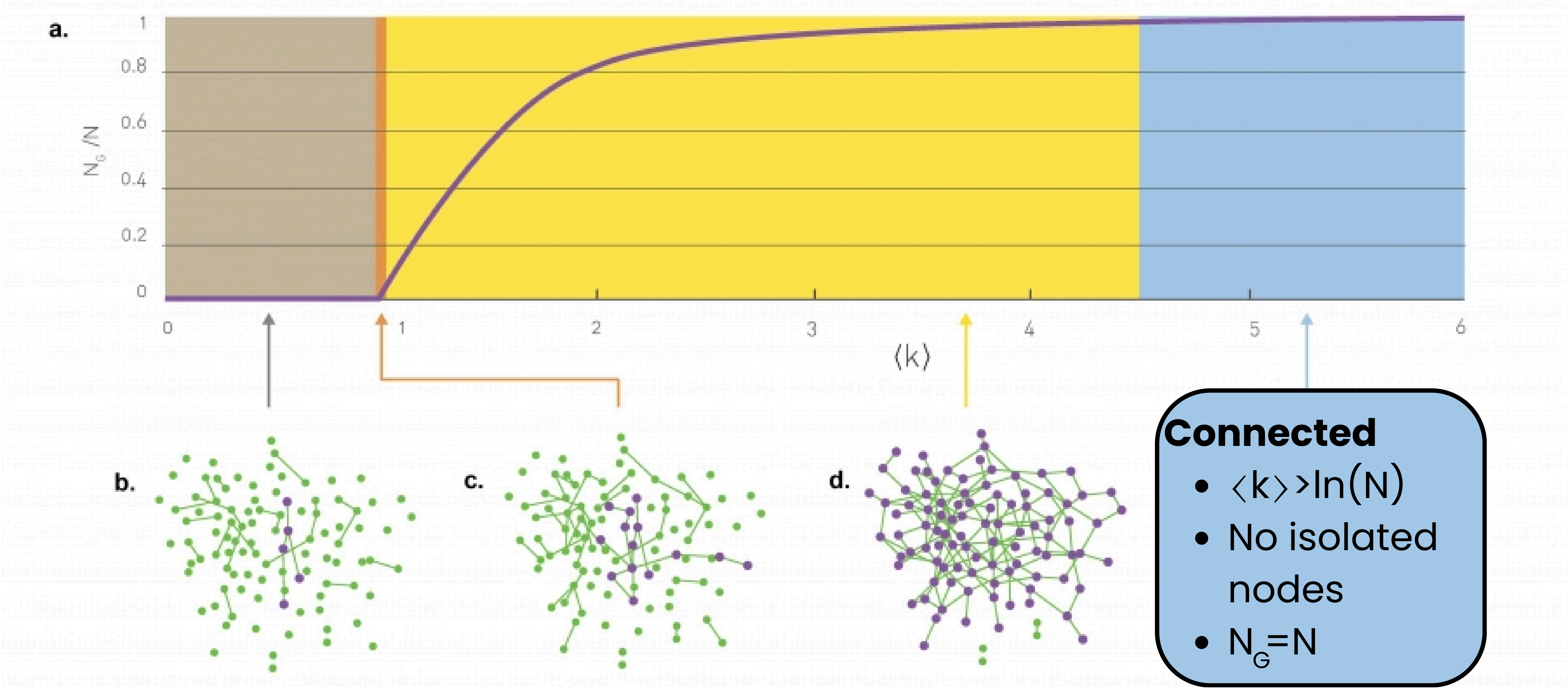


Super-Critical

- ⟨k⟩ 1
- Giant components
- N_G~N

Phase Diagram

For ⟨k⟩>ln(N) the graph is connected, all nodes belong to the same giant component that contains exactly N nodes

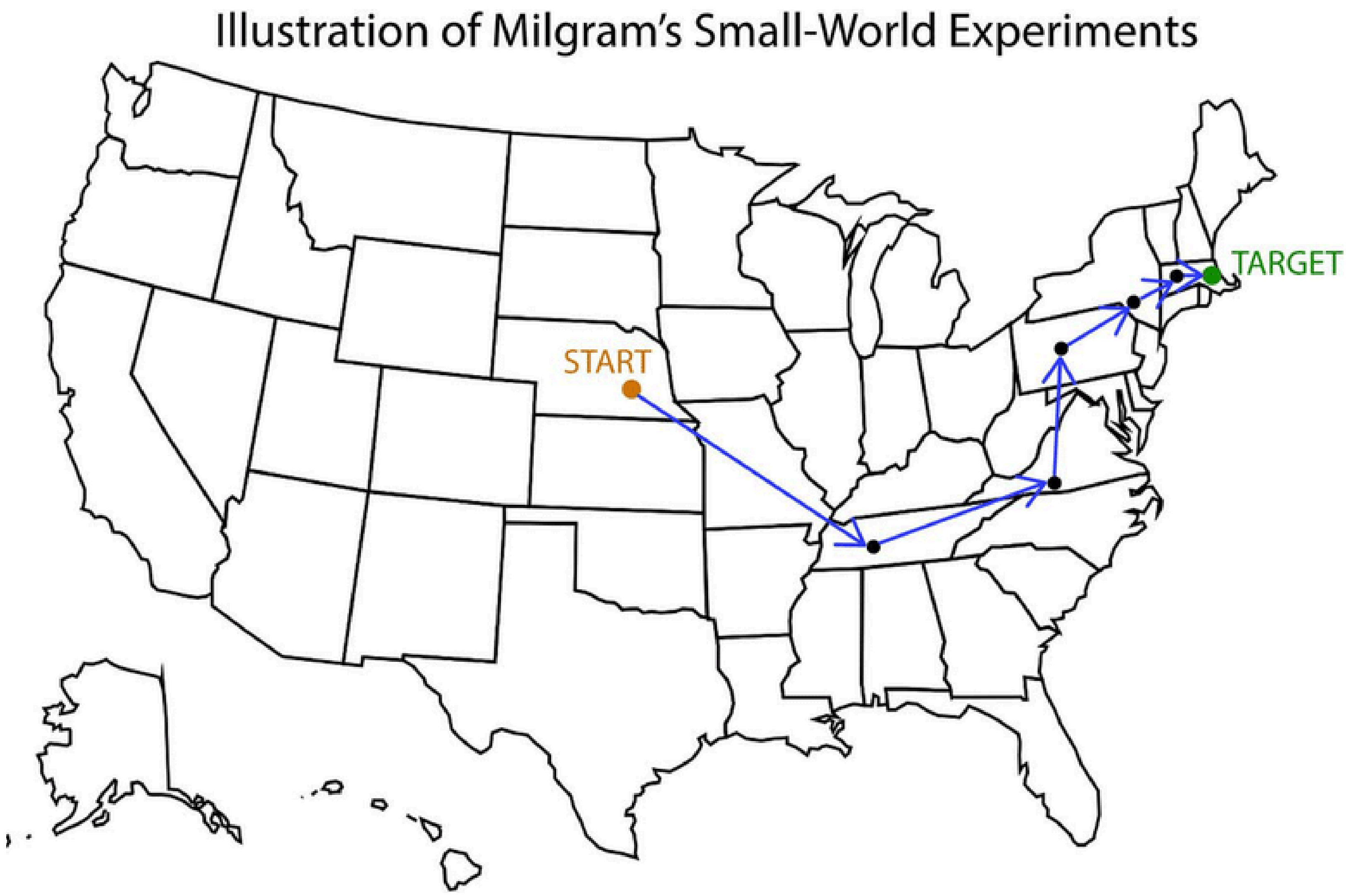


Connected

- ⟨k⟩>ln(N)
- No isolated nodes
- N_G=N

Small World and Clustering

Milgram's **Experiment**



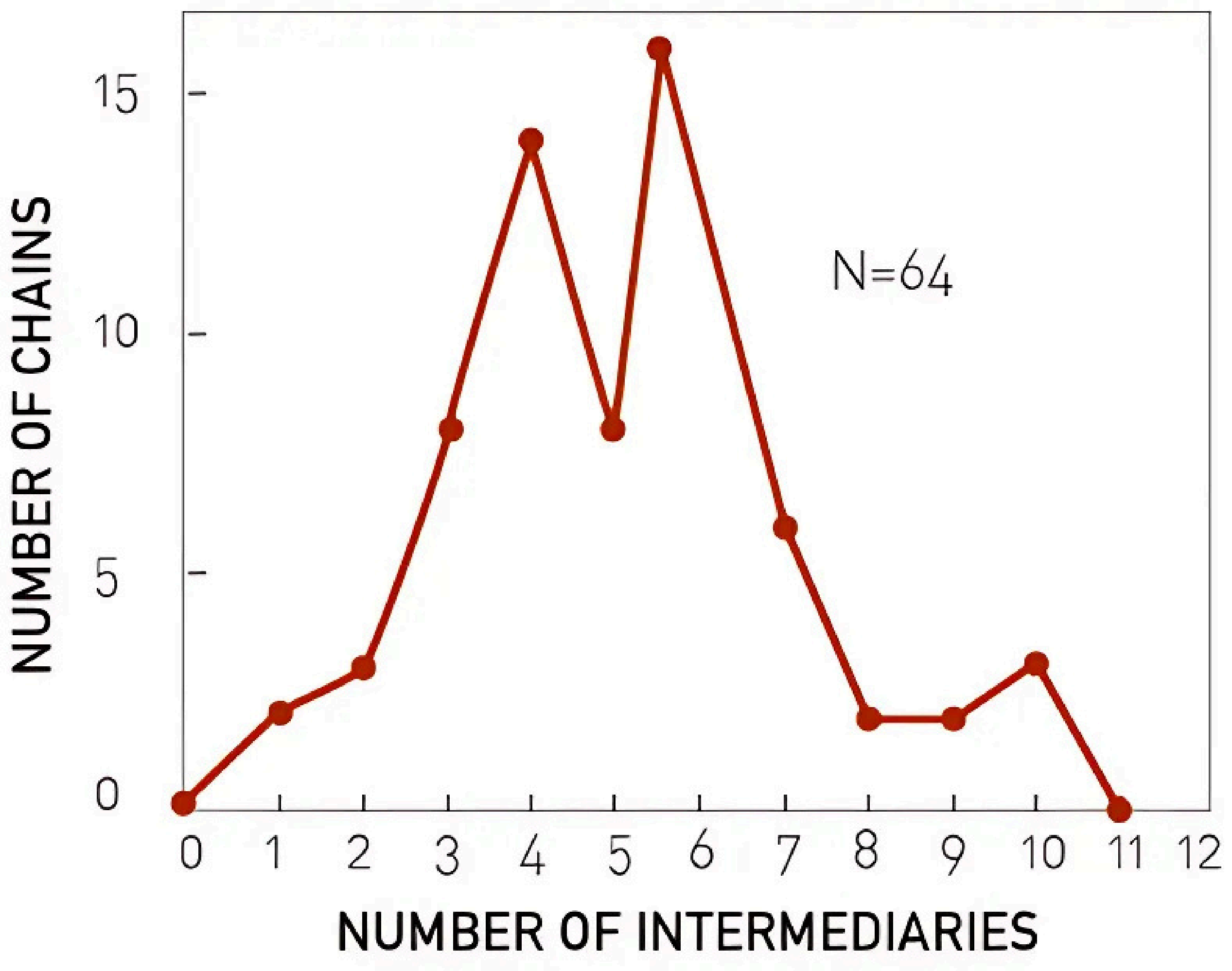
In 1967 Milgram measured the average path length in social networks

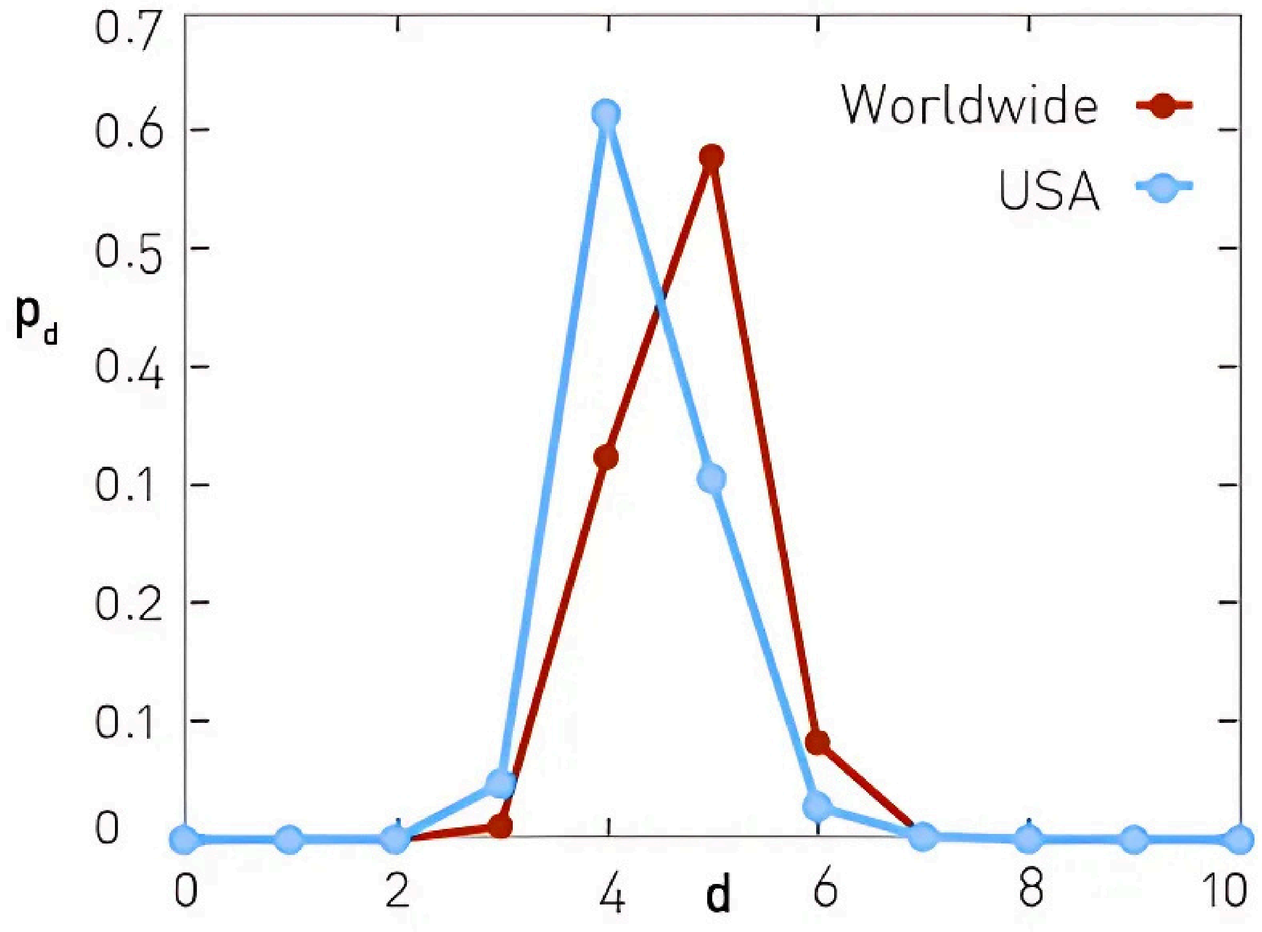
- **Participants:** Randomly selected individuals in Omaha, Nebraska
- **Task:** Send a package to a stockbroker in Boston
- **Method:** Each participant mailed a packet to a friend they thought was socially closer to the target. The process was repeated until the packet reached the stockbroker or the chain ended.

On average, 6 steps are needed to reach the target.

Milgram's Experiment on Facebook

The left plot shows the distribution of steps for Milgram's experiment, while on the right plot shows the distribution of distance among Facebook users. In this case the average path length is around 4





Erdős Number

The Small World property also applies to science collaboration networks. The famous mathematician Erdős is generally taken as a point of reference and used to compute the Erdős number

Author A

De Marzo, Giordano

Author B

Erdős, Paul[1]

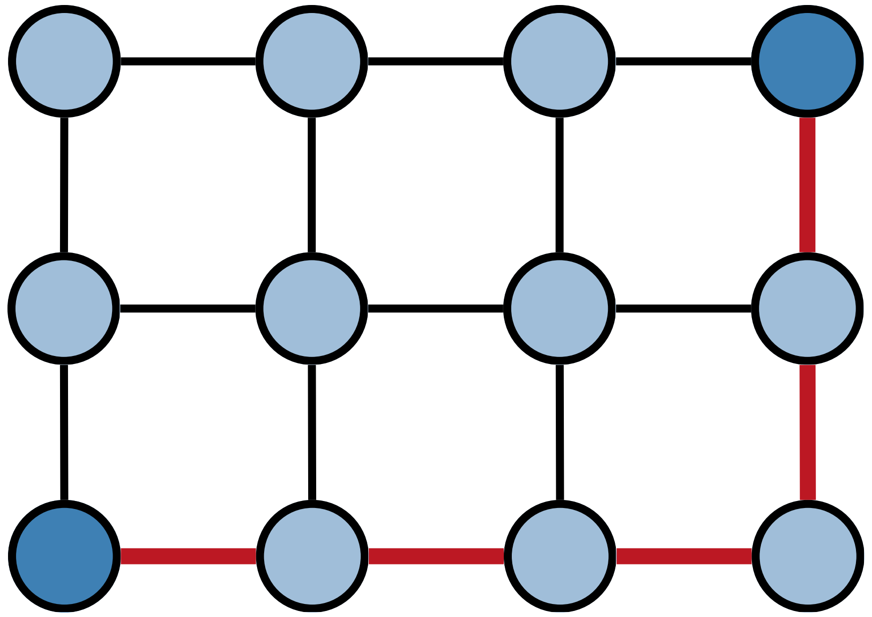
New Search

MR Collaboration Distance = 5

De Marzo, Giordano	coauthored with	Castellano, Claudio	**MR**4402863
Castellano, Claudio	coauthored with	Vespignani, Alessandro	**MR**3406040
Vespignani, Alessandro	coauthored with	Marsili, Matteo	**MR**1155944
Marsili, Matteo	coauthored with	Székely, László A.	**MR**3076123
Székely, László A.	coauthored with	Erdős, Paul[1]	**MR**0932227

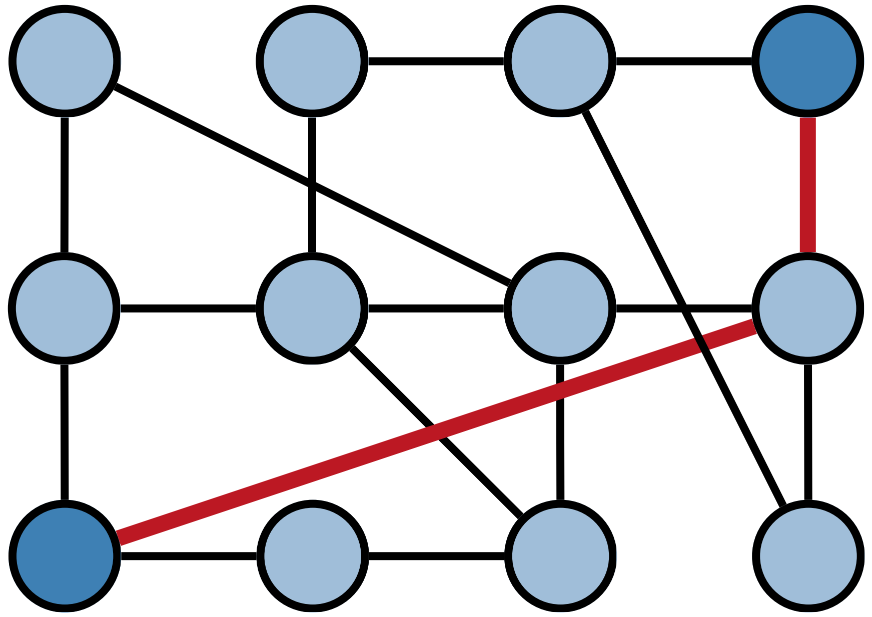
The Small World Property

Despite networks can be huge, often the path connecting any two elements in the network can be surprisingly short. This phenomenon is summarized by the popular notion of "**six degrees of separation**".



This property is expressed in terms of the average path length L and it is called **small world property**

$$L \sim \log(N)$$



Note that this is not true for lattices, for D=2

$$L \sim \sqrt{N}$$

Diameter of Random Networks

We approximate a random network as a tree. The number n(d) of nodes at distance d from a node is

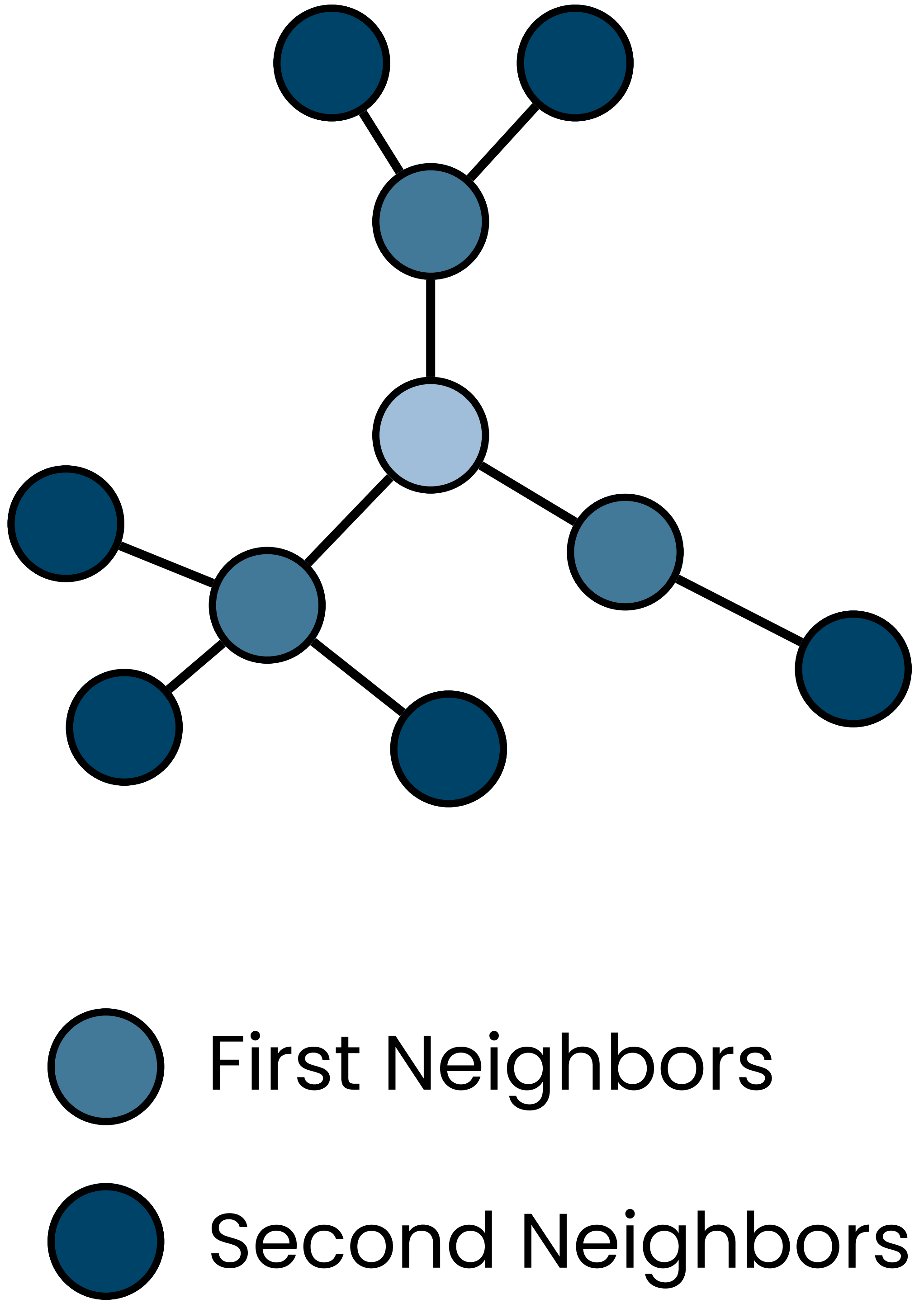
- $n(0) = 1$
- $n(1) = \langle k \rangle$
- $n(2) \approx \langle k \rangle^2$
- $n(d) \approx \langle k \rangle^d$

The total number of nodes N(d) up to distance d is

$$N(d) \approx 1 + \langle k \rangle + \langle k \rangle^2 + \cdots + \langle k \rangle^d \approx \langle k \rangle^d$$

When d is equal to the diameter d_{max} then $N(d)=N$

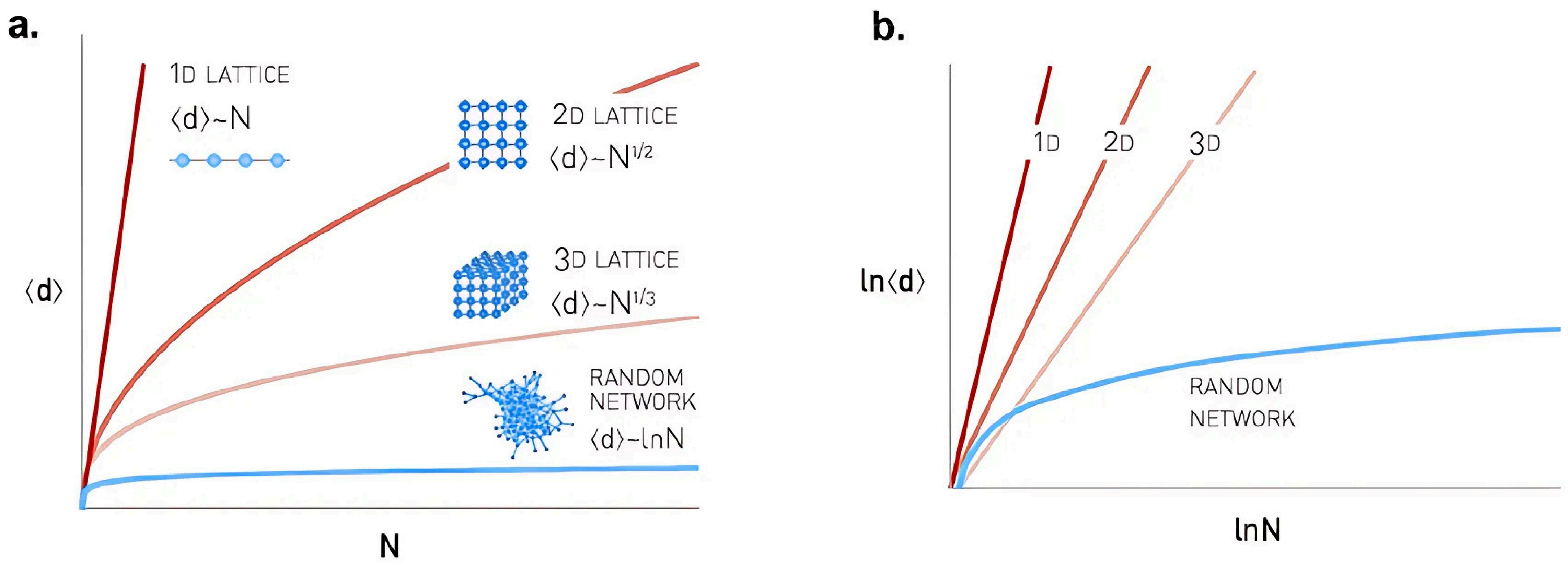
$$N = N(d_{max}) \approx \langle k \rangle^{d_{max}} \rightarrow d_{max} \approx \frac{\log N}{\log \langle k \rangle}$$



Random Networks vs Lattices

The result we obtained implies that random networks are **Small World**. The approximation actually works better for the average path length

$$\langle d \rangle \approx \frac{\log N}{\log \langle k \rangle}$$



Clustering in Random Networks

To compute the clustering coefficient we need the **number of triangles**

- we denote by L_i the number of links among the connections of node i
- the number of triangles t_i will be given by L_i
- the probability for 2 nodes to link is p
- the number of possible links among the k_i connection of i is $k_i(k_i-1)/2$

As a consequence the number of connections L_i is

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

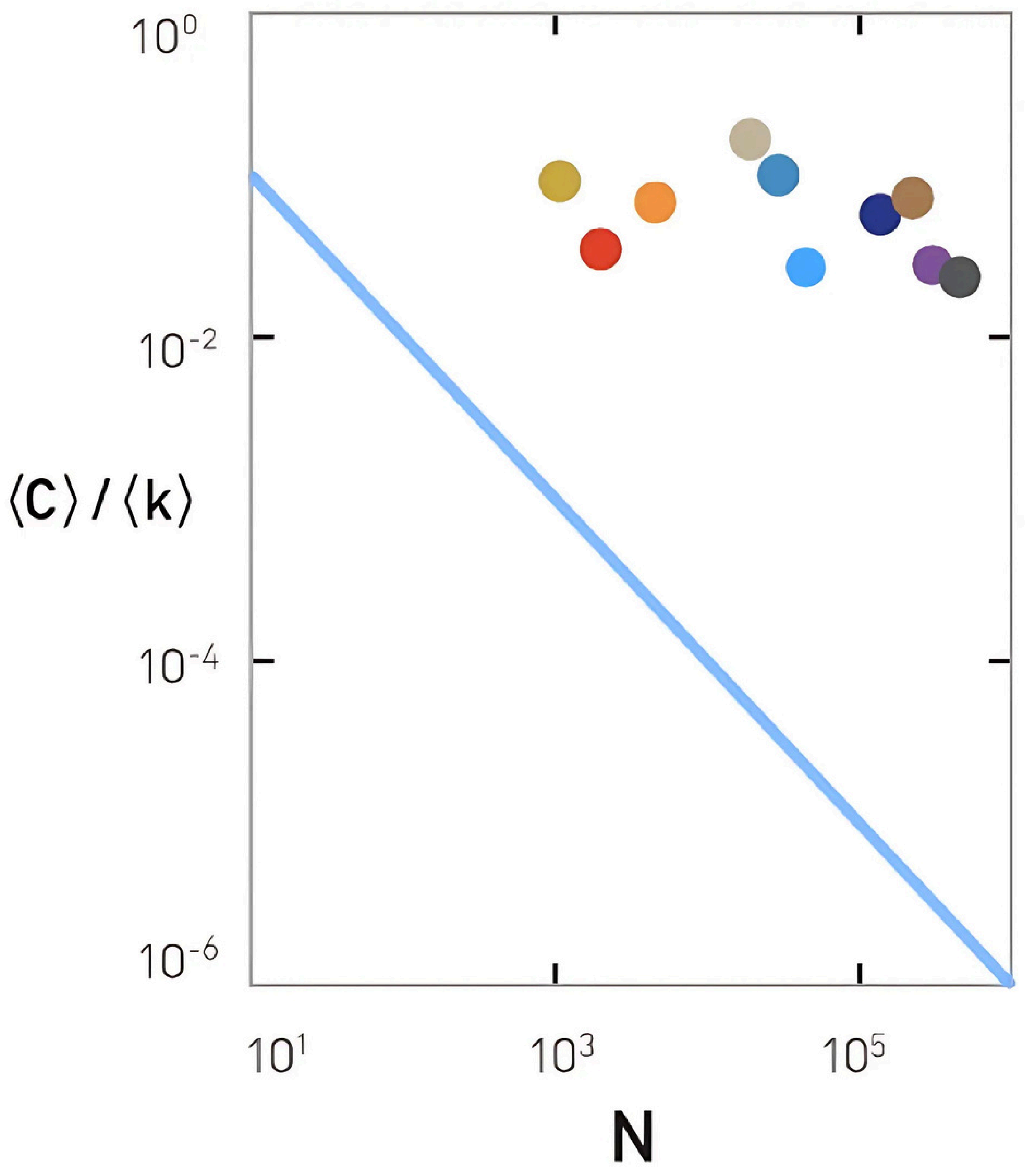
The local and global **clustering coefficients** are then

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

Random vs Real Networks

Since C= ⟨k⟩/N the global clustering coefficient in random networks goes to zero as the network size increases

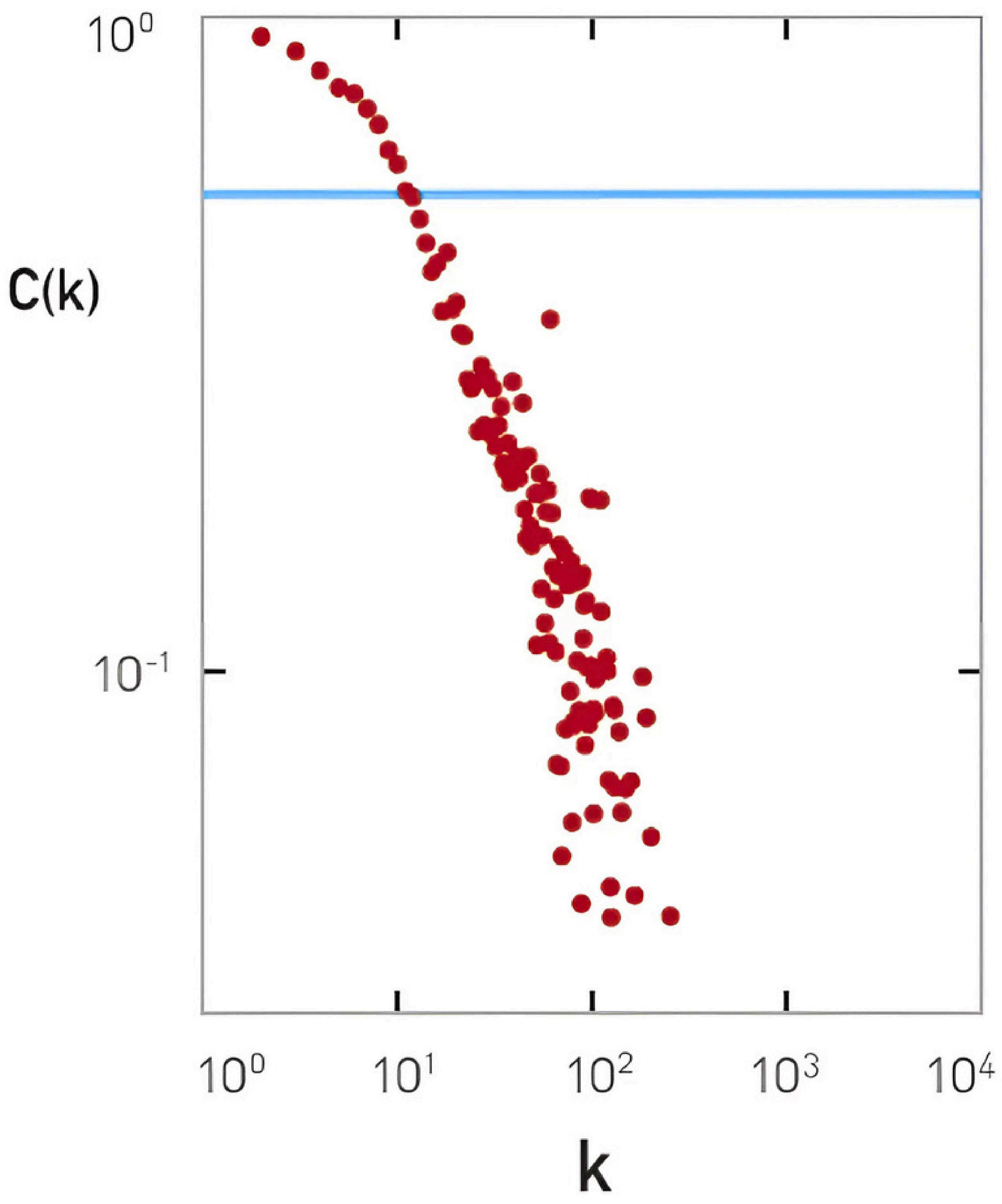
- the figure shows several real world network and the prediction for random networks
- the clustering of the real networks is well above the random network scenario



Random vs Real **Networks**

In random networks also the local clustering coefficient is $C_i = \langle k \rangle / N$

- the clustering coefficient is the same for all nodes
- it does not depends on the node's degree
- real world networks show a different behavior
- the figure shows the case of a science collaboration network



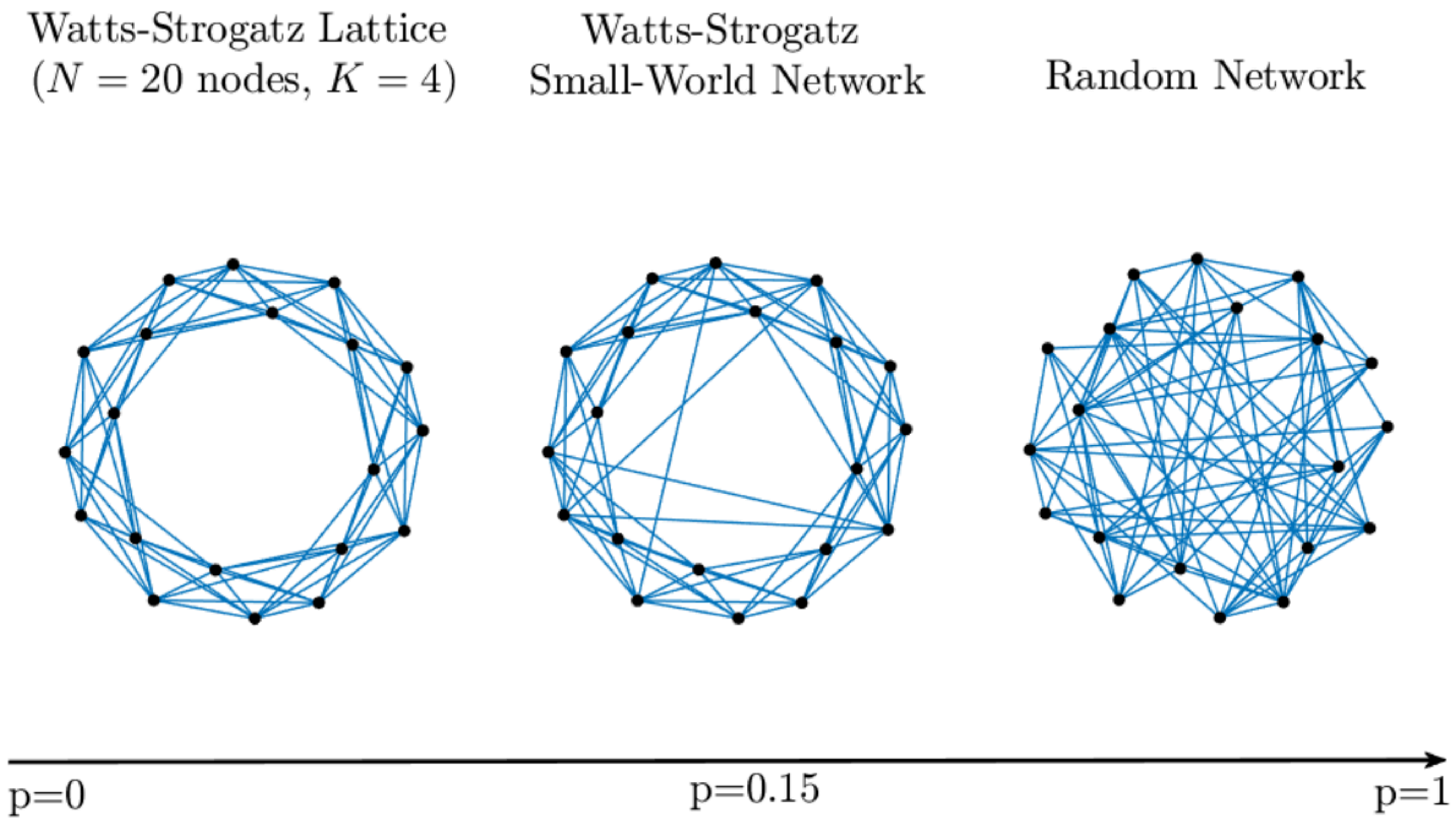
Watts-Strogatz Model

Real Networks are not Random!

Random networks have similar average path length compared to real networks, but the clustering and the degree distribution are very different

	Small World	Clustering	Degree Distribution
Random Networks	Yes	Small	Poisson
Real Networks	Yes	Large	Scale Free

The Watts-Strogatz Model



The **Watts-Strogatz** Model is one of the most simple models

- start with a ring with connections only to near nodes (on both sides)
- rewire each link with probability p

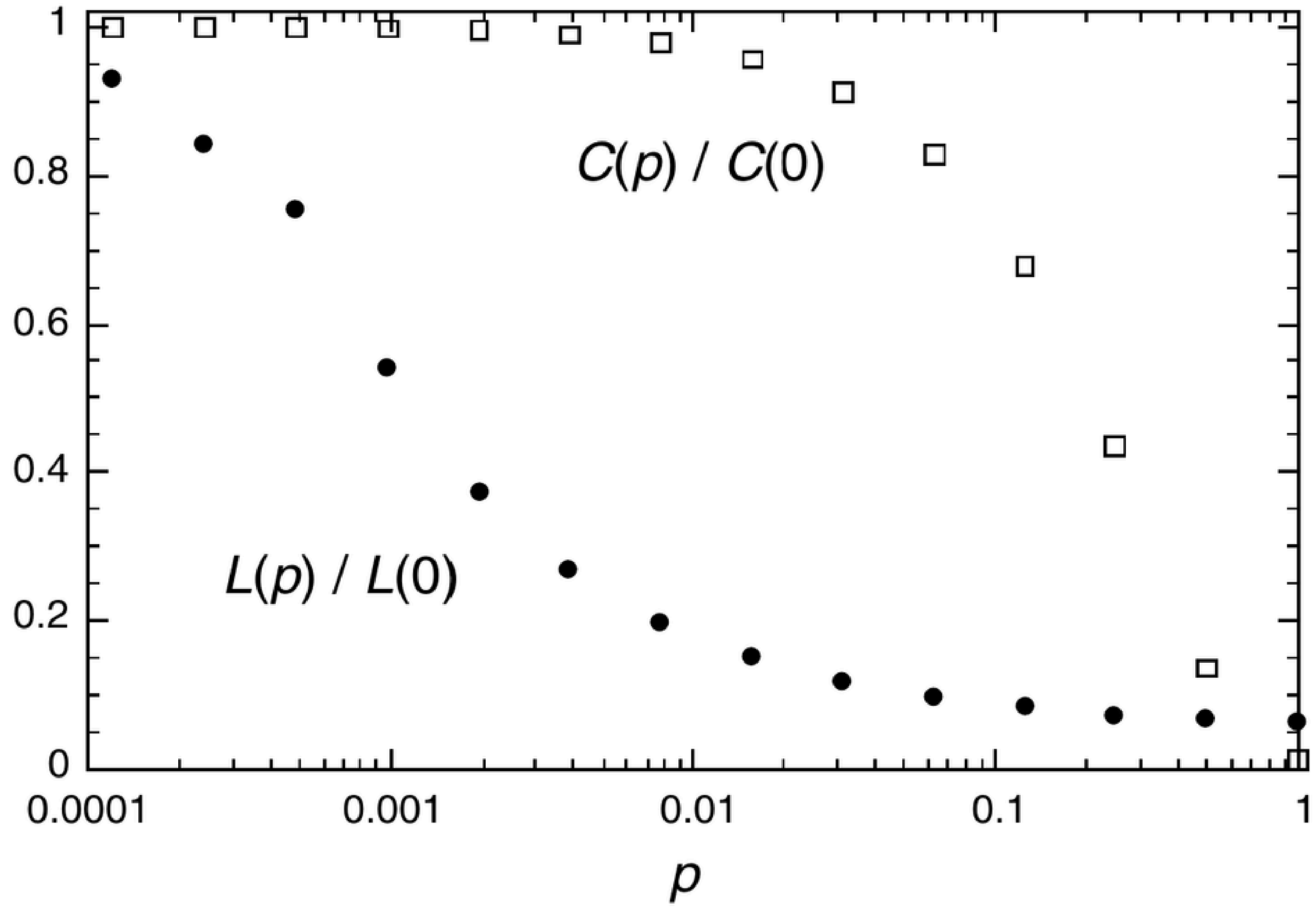
For p=0 we have a regular ring network (similar to a lattice), while for p=1 we have a random network. What happens in between?

Properties of the Model

The Watts-Strogatz Model interpolates between a regular graph and a random graph. For intermediate values of p we observe:

- high clustering (inherited from the initial regular graph)
- low average path length (deriving from the rewiring)

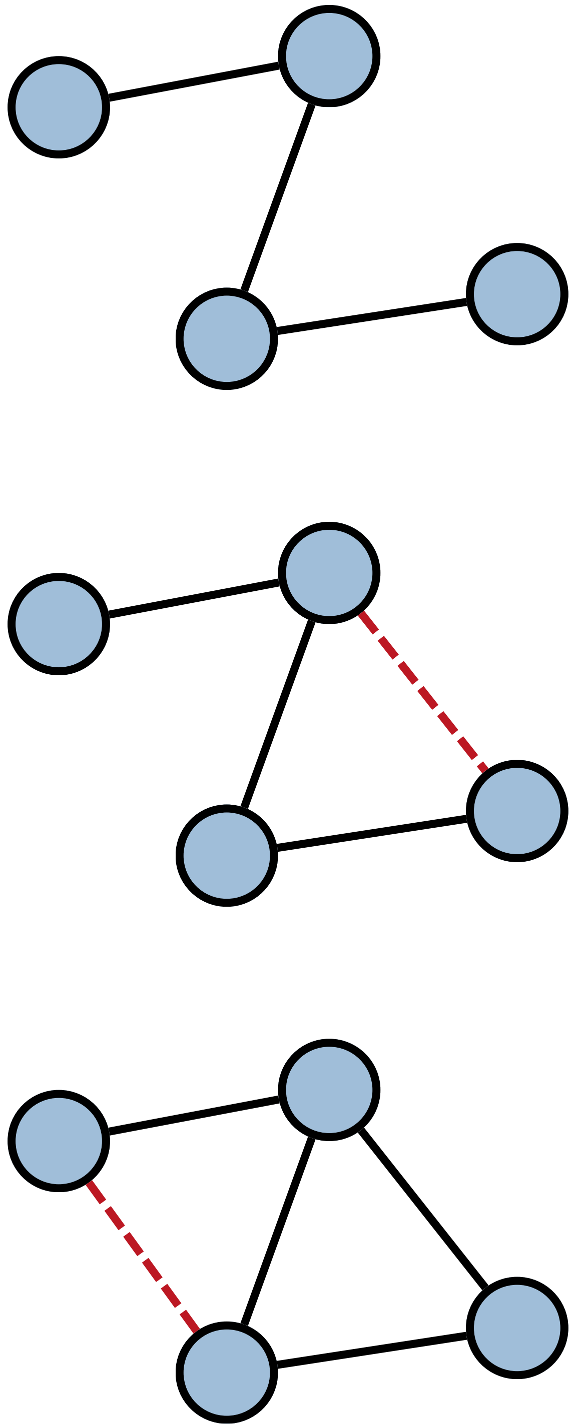
In practice the few random connection we are adding make it much easier to move around the network.



Triadic Closure **Mechanism**

The Watts-Strogatz Model reproduces real networks properties, however it is not very realistic:

- in real life we don't know much about the full network, we tend to link more with close people

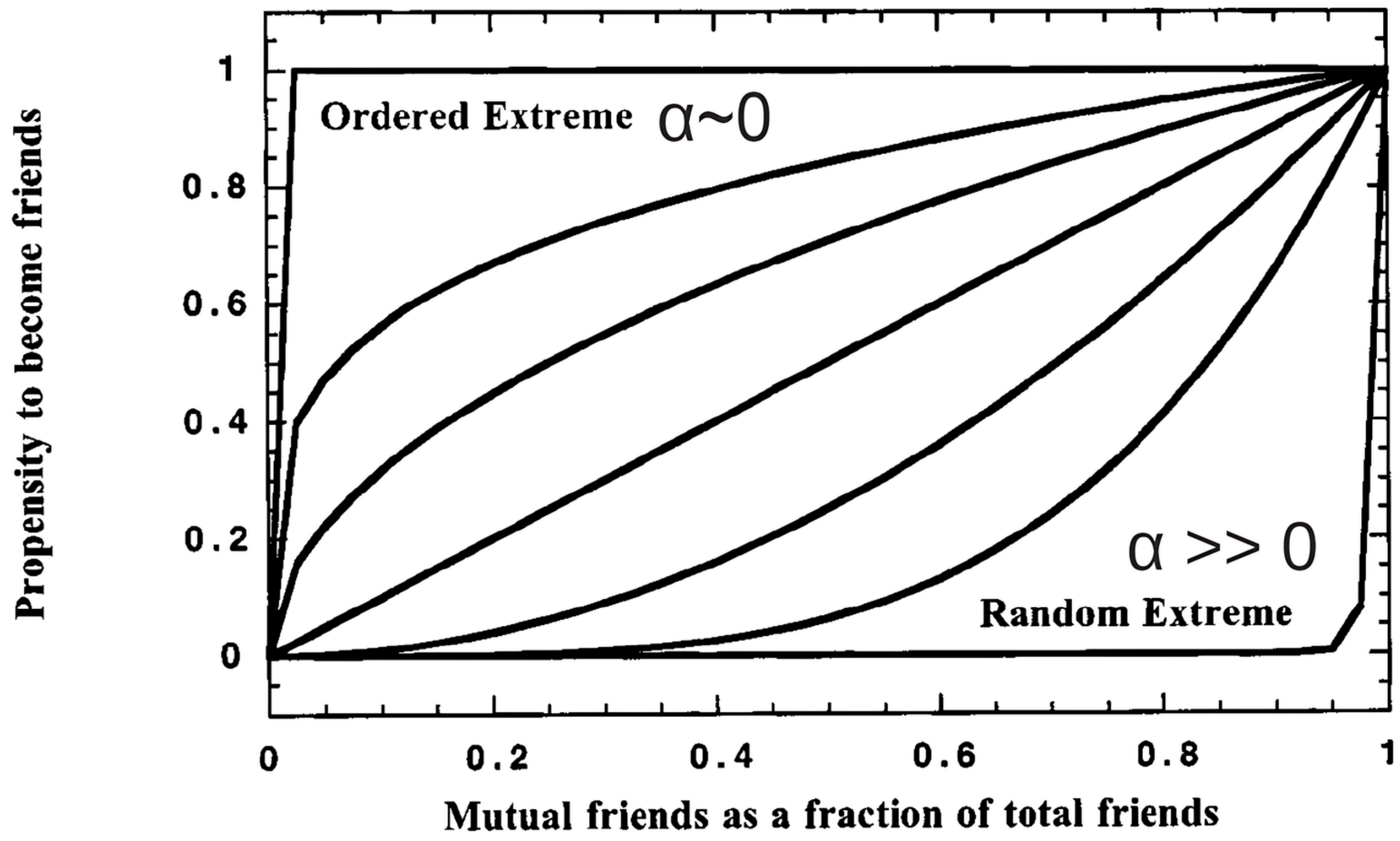


We can obtain similar networks performing rewiring based on triadic closure instead of random

- the idea is that nodes having a "common friend" are more likely to link
- we always start with a regular ring
- we add new links with a probability that depends on the number of shared friends

Propensity to Triadic Closure

$$R_{i,j} = \begin{cases} 1 & m_{i,j} \geq k \\ \left[\frac{m_{i,j}}{k}\right]^{\alpha} (1-p) + p & k > m_{i,j} > 0 \\ p & m_{i,j} = 0 \end{cases}$$



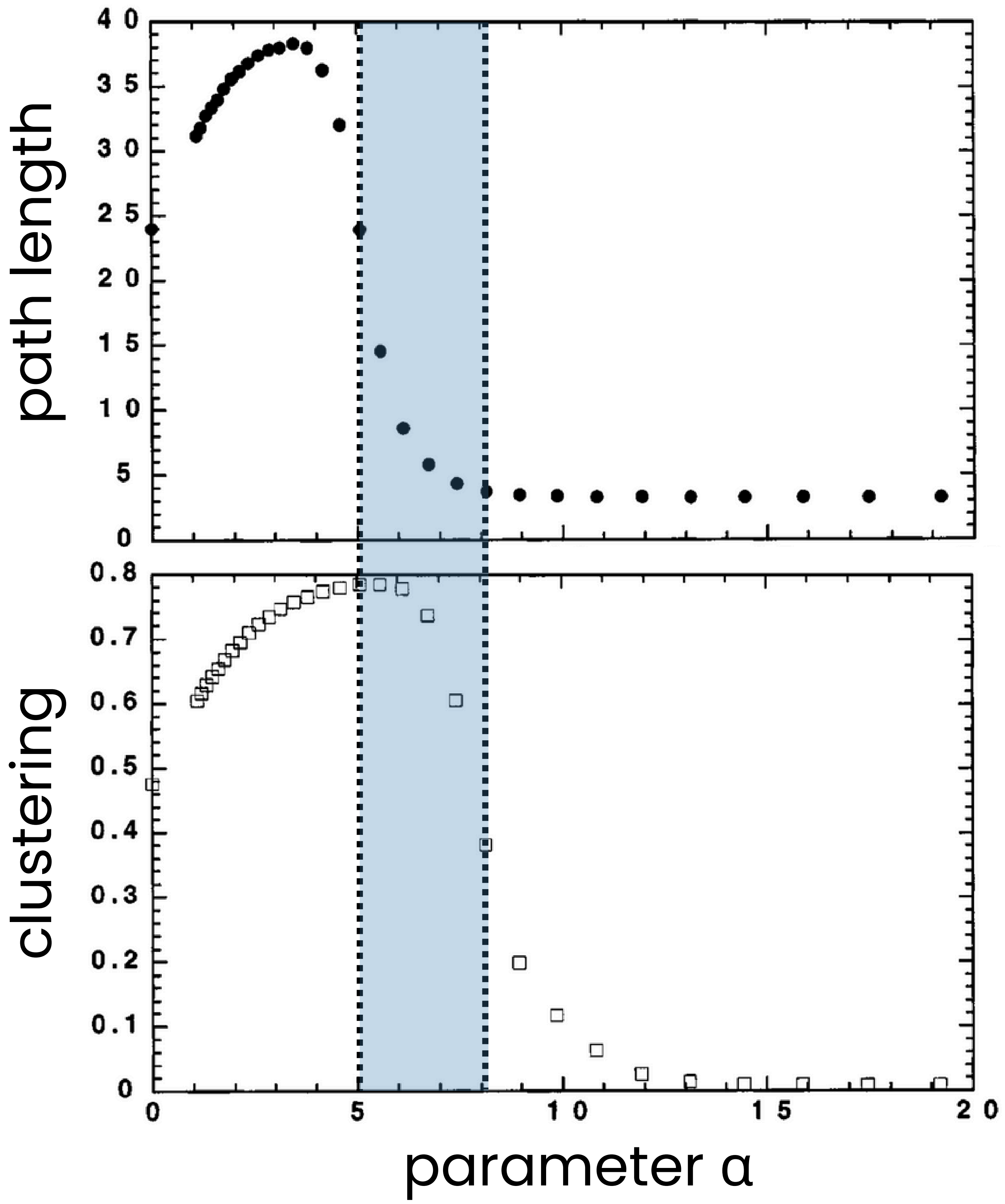
The model works as it follows:

- Start with a ring of n nodes
- For each pair of nodes:
    - Calculate number of shared friends $m_{i,j}$
    - Calculate probability to connect $R_{i,j}$ based on $m_{i,j}$
    - Connect them with prob. $R_{i,j}$
- p gives the probability to connect even in absence of mutual friends
- α sets the relevance of the common friend mechanism

Properties of the Model

Similarly to the Watts-Strogatz Model, we observe a sweet spot (in α) for which the model produces networks with both low path length and high clustering

- this is much more realistic than the Watts-Strogatz model
- the rewiring process is based on local characteristics of the network
- the process resembles what we humans tend to do in real life



Network Robustness

Network **Failures**

Nodes or links within networks may fail and cause chain reactions through the whole structure. Disruption are significant when the giant component breaks, leading to a fragmented network

- delays in trains or flights
- power outages
- supply chain shortages

Rail Networks

Power Grids

Supply Chains

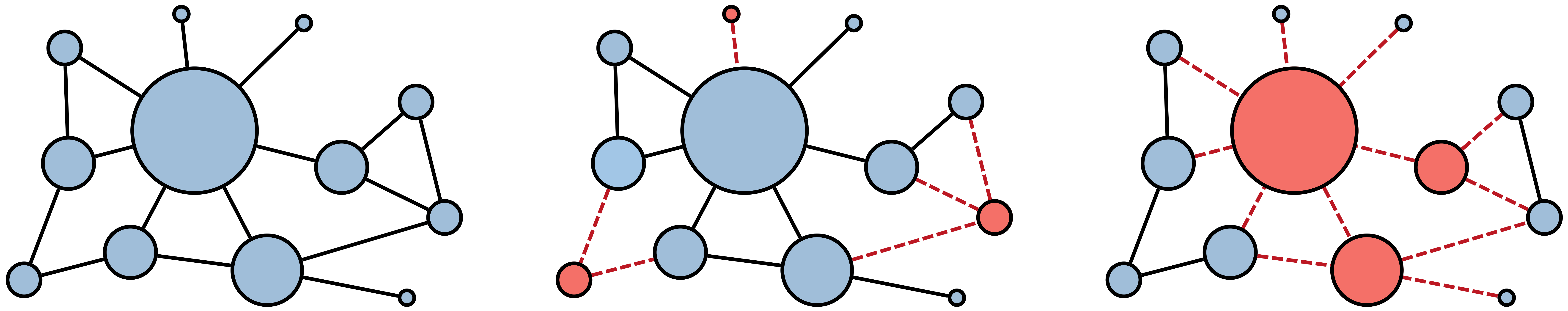
Random Failures vs Targeted Attacks

We focus on nodes and we distinguish between two possible scenarios

- **Random Failure.** Random nodes in the network fail (are removed)
- **Targeted Attacks.** Nodes with the largest degree are removed

A network may be very robust against random failures but very susceptible to targeted attacks

Random Failure **Targeted Attack**



Your Friends have more Friends!?

Networks have strage properties, one of them is the **Friendship paradox**

On average, an individual's friends have more friends than that individual.

To explain this we denote by

- N(k) the number of nodes with degree k
- $P_n(k)$ the probability that the neighbor of a node has degree k
- $k \cdot N(k)$ is the number of nodes connected to nodes with degree k
- $\langle k \rangle \cdot N$ is the total number of links in the network

The probability $P_n(k)$ to follow a random link and reach a node with degree k is

$$P_n(k) = \frac{k \cdot N(k)}{\langle k \rangle \cdot N} = \frac{k}{\langle k \rangle} P(k)$$

And we can compute the average degree of a random neighbor using $P_n(k)$

$$\langle k \rangle_n = \sum_k P_n(k) k = \sum_k k \frac{k}{\langle k \rangle} P(k) = \frac{\langle k^2 \rangle}{\langle k \rangle} > \langle k \rangle$$

Molloy-Reed Criterion

Molloy-Reed Criterion states that:

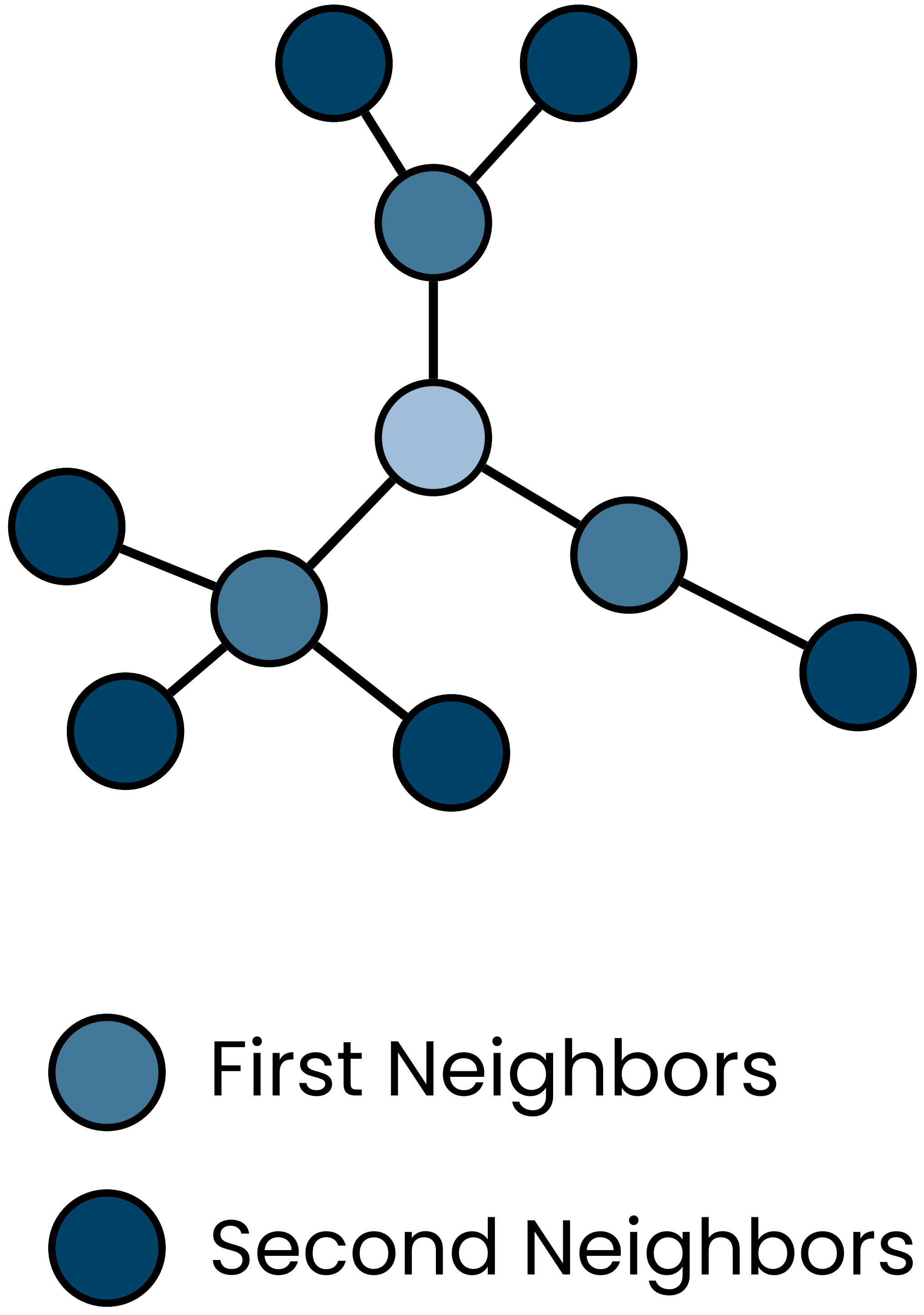
A giant component can exists only if the average number of second neighbors is larger than the average number of first neighbors

Starting from a node, the network must expand

- the degree of the neighbors is $\langle k \rangle_n = \langle k^2 \rangle / \langle k \rangle$
- we have to subtract 1 (remove starting node)
- the total number of second neighbors is $Z_2 = \langle k \rangle [\langle k \rangle_n - 1] = \langle k^2 \rangle - \langle k \rangle$

The criterion then reads

$Z_2 > \langle k \rangle$ -> $\langle k^2 \rangle - \langle k \rangle > \langle k \rangle$



Percolation Threshold

Using Molloy-Reed criterion we can obtain the giant component phase transition of random graphs

$$\langle k^2\rangle - \langle k\rangle > \langle k\rangle \rightarrow \mathrm{Var}[k] + \langle k\rangle^2 - \langle k\rangle > \langle k\rangle \rightarrow \langle k\rangle^2 > \langle k\rangle \rightarrow \langle k\rangle > 1$$

We can also use Molloy-Reed criterion to asses the robustness of networks.

We consider a random failure involving a fraction f of the nodes (percolation)

- the number of first neighbors is reduced by (1-f)
- the number of second neighbors is reduced by $(1-f)^2$

Molloy-Reed criterion becomes

$$(1-f)^2(\langle k^2\rangle - \langle k\rangle) > (1-f)\langle k\rangle$$

From which we get the critical percolation point f_c

$$f_c = 1 - \frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}$$

Random Graphs vs
Real Networks

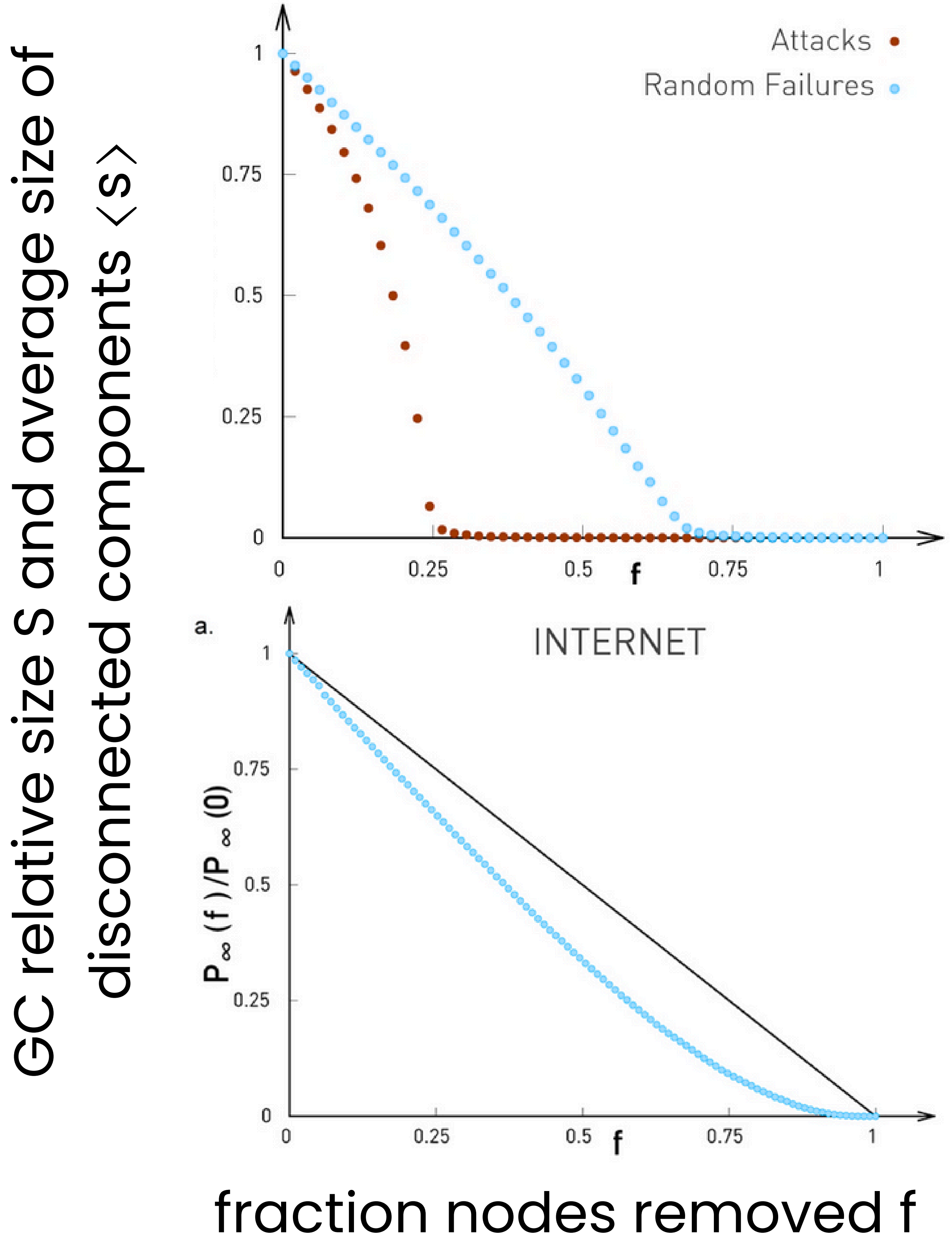
In the case of a random network ⟨k²⟩=⟨k⟩²+⟨k⟩ and the critical fraction is

$$f_c = 1 - \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} = 1 - \frac{1}{\langle k \rangle} = \frac{\langle k \rangle - 1}{\langle k \rangle}$$

For large ⟨k⟩, a very large fraction of nodes can fail
The figures shows the behavior for

- a random network with ⟨k⟩=3
- an internet network with ⟨k⟩ ≈ 6

The real network is much more tolerant to random failures. The GC exists up to f≈1 and therefore the robustness is extremely high.



Conclusions

Random Networks

Random networks are characterized by a Poisson degree distribution and present a phase transition leading to the emergence of a Giant Component

Small World and Clustering

Random networks present the small world property like real networks, but differently from them are characterized by a small clustering

Watts and Strogatz Model

We can get both small world and high clustering using the Watts-Strogatz model or the more realistic triadic closure model

Network Robustness

The robustness of networks to failures can be computed using the Molloy-Reed criterion. For random graphs this leads to a percolation transition.

Quiz

- What are some networks that can be schematized as random?
- Do you have any real life example of the small world property?
- Is there any flaw in Milgram’s experiment?
- Why there is no clustering in random graphs?
- What are the implausible assumptions of the Watts-Strogatz model?
- Which networks are expected to be more tolerant to attacks, random or real?
- Which characteristics make a network more or less tolerant to attacks and failures?