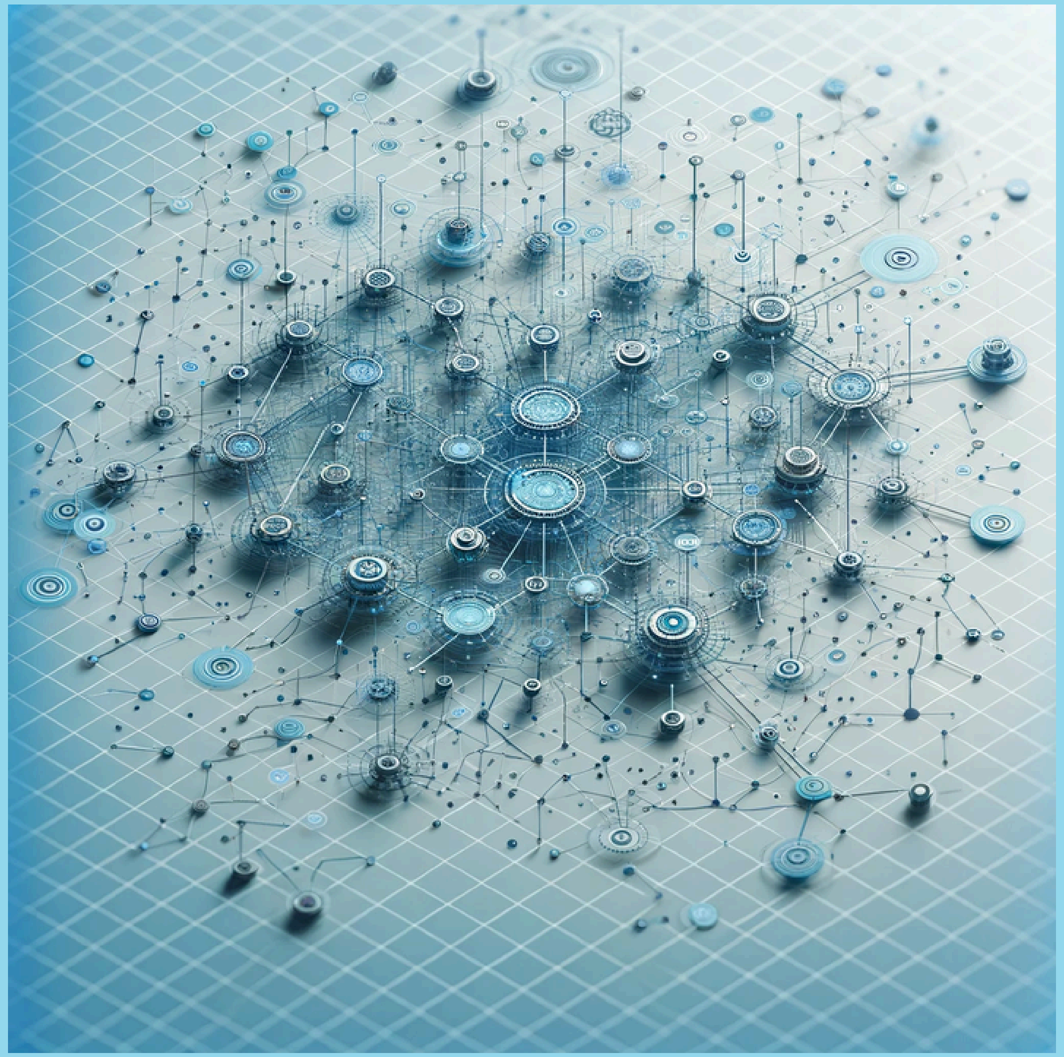


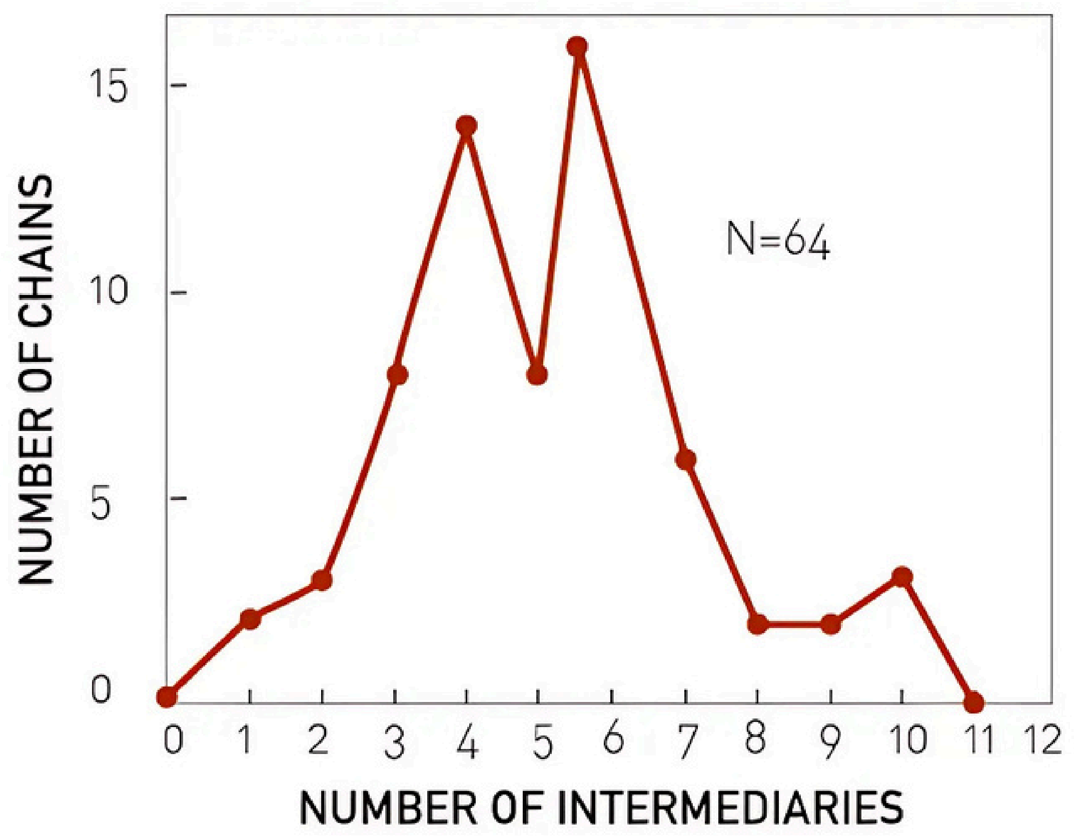
UNIVERSITÄT KONSTANZ

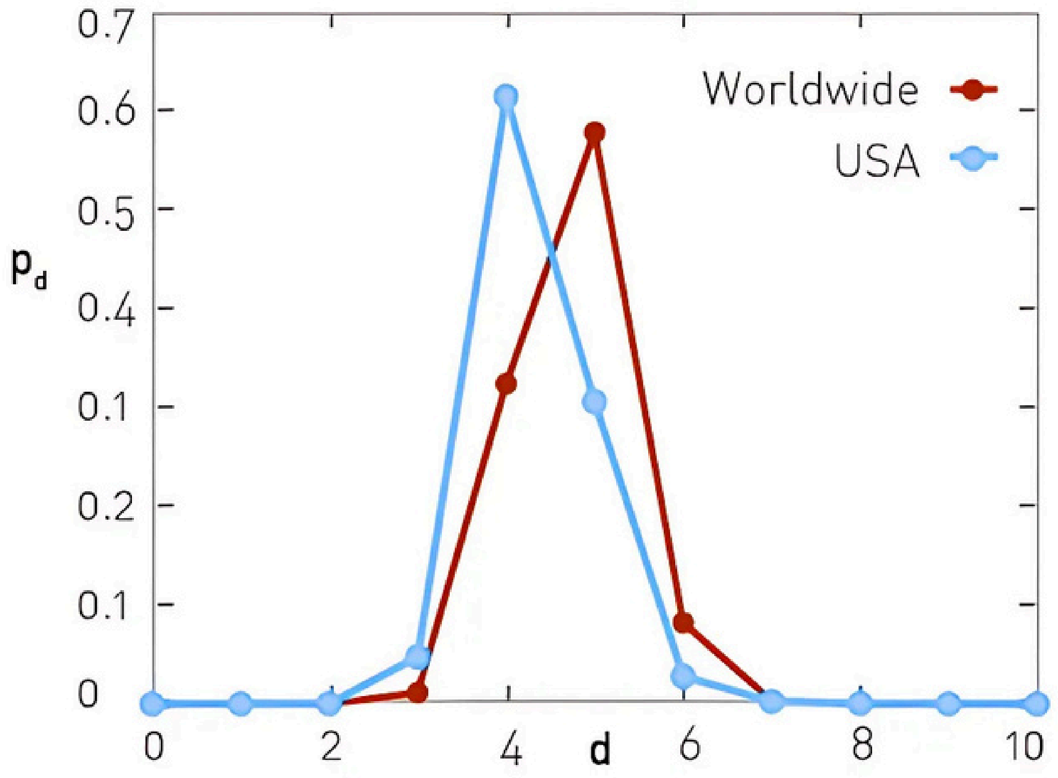
Scale-Free Networks

Network Science of Socio-Economic Systems

Giordano De Marzo

Universität Konstanz





Recap

Random Graphs

Random networks have a Poisson degree distribution a phase transition leading to a Giant Component

Small World and Clustering

Random networks present the small world property but have a small clustering

Watts-Strogatz Model

We can get both small world and high clustering using the Watts-Strogatz model

Network Robustness

The robustness of networks to failures can be computed using the Molloy-Reed criterion

Outline

1. Power Law Probability Distributions
2. Scale-Free Networks
3. Barabasi-Albert model
4. Robustness of Scale-Free Networks

Power Law Probability Distributions

The Gaussian World

We are accustomed to think in term of gaussians and average values:

- height
- weight
- speed
- performances

In the Gaussian world there are no surprises:

- a small sample is enough for knowing everything
- the future is hardly surprising

The Paretian World

However many relevant phenomena are characterized by extreme events (Pareto distribution):

- financial crises
- wars
- pandemics
- natural disasters

The Paretian world is full of surprises and strange properties:

- a large sample is not enough for knowing everything
- the future is surprising

Pareto or Power Law Distributions

Let us consider a series objects with sizes k_1, k_2 … etc.
We say that these objects follow a Pareto or Power Law distribution if the probability P(k) of observing an event with size S is of the form

$$P(k) = \frac{c}{k^\gamma}$$

In this expression

- c is a normalization constant to ensure the probability to sum to one
- γ is the power law exponent or scaling exponent

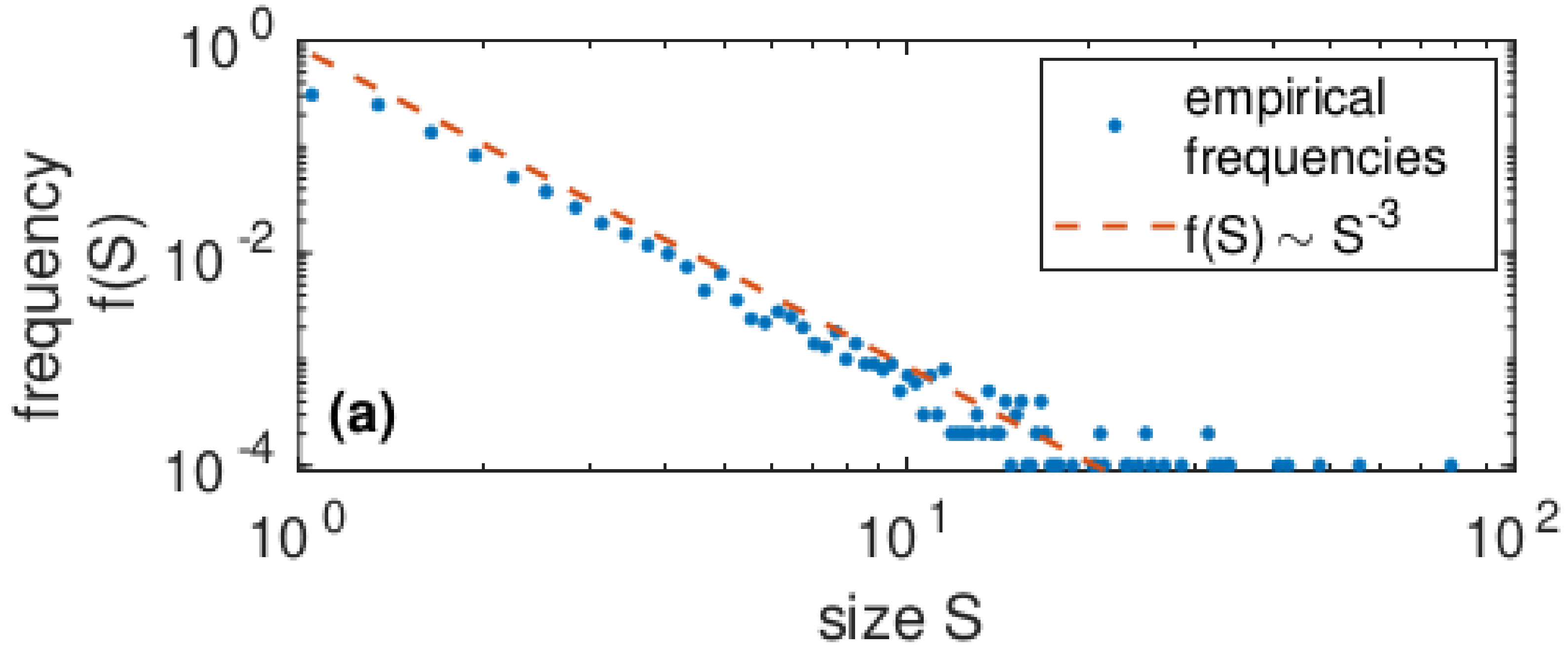
The power law shows a much slower decay with respect to a Gaussian

$$P(k) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

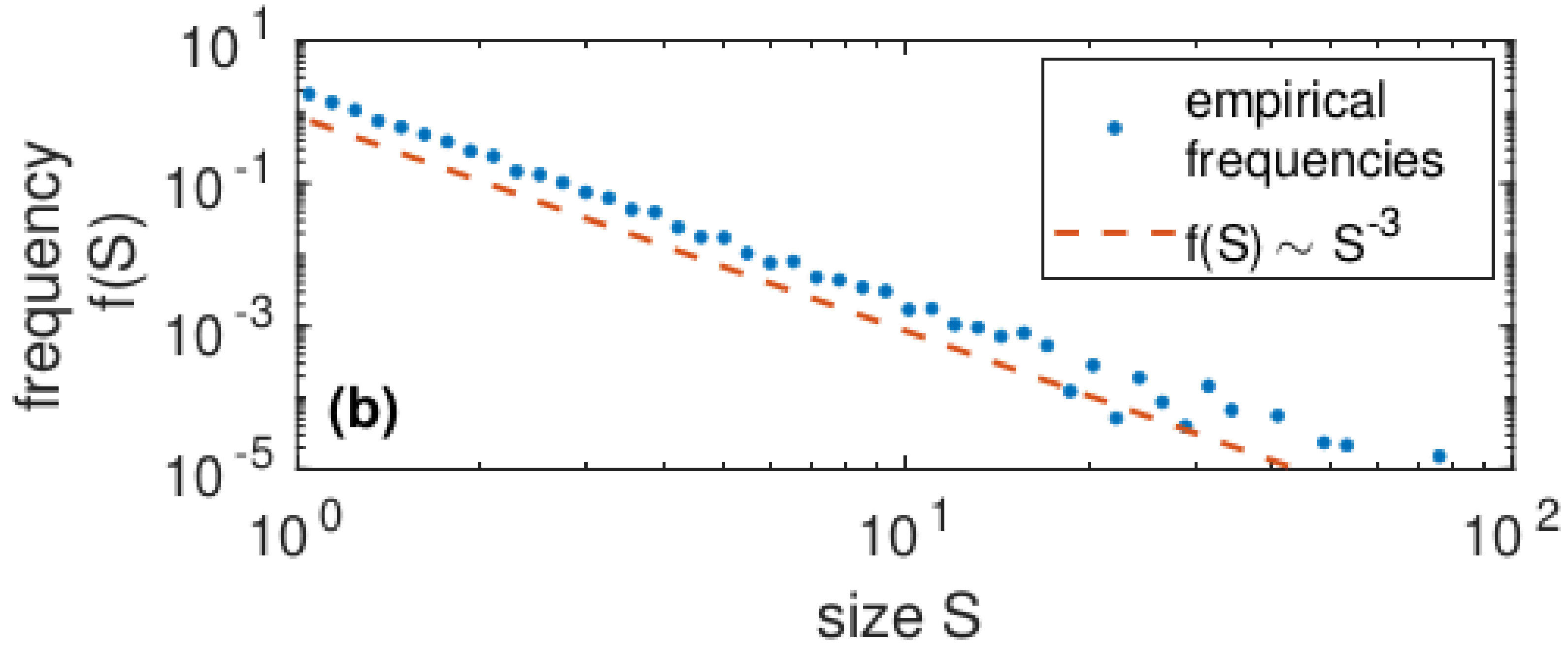
Visualizing Power Laws

Given a set of sizes we can determine their distribution performing an histogram. If the data follow a power law distribution the histogram will look like a straight line using a double logarithmic scale. In order to obtain better plots it is important to use logarithmic binning.

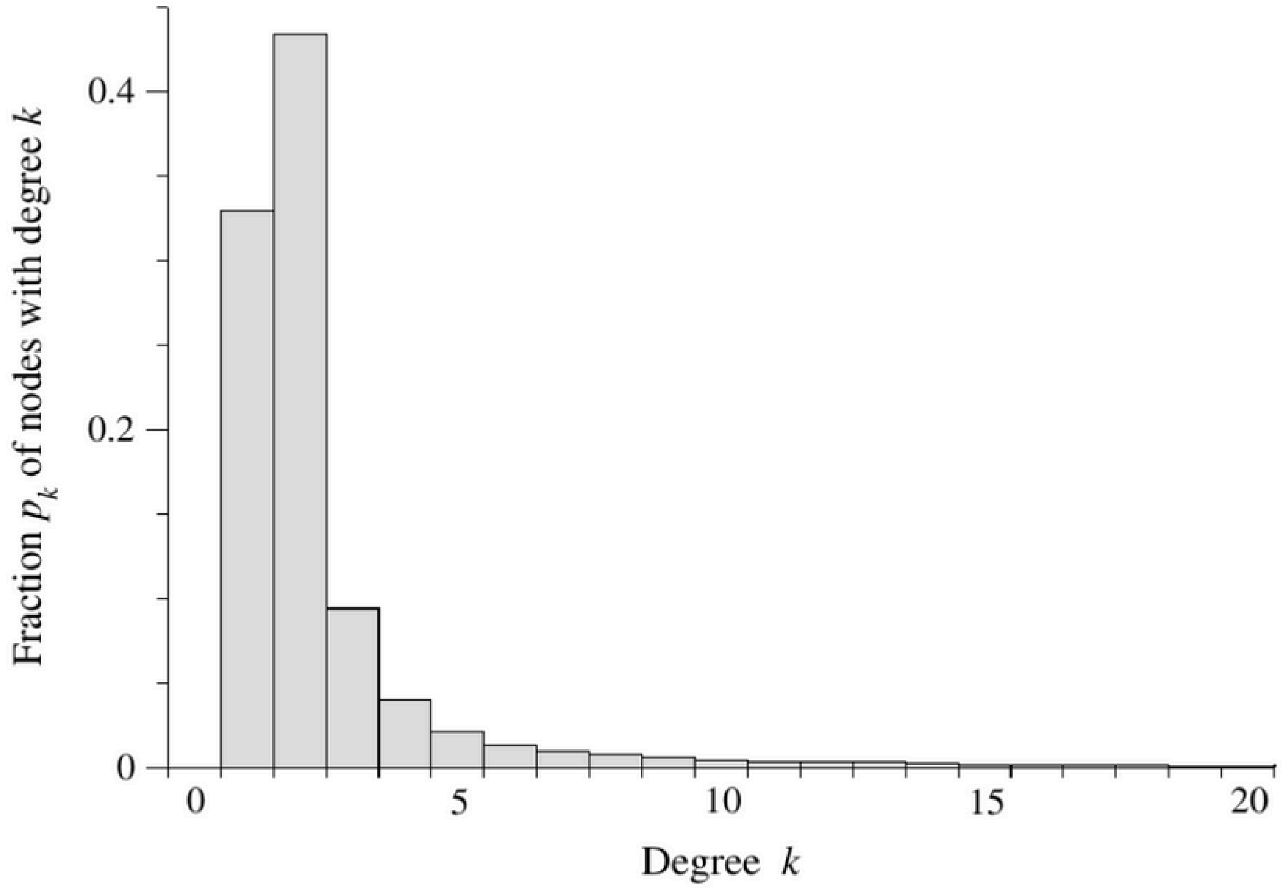
Normal Binning

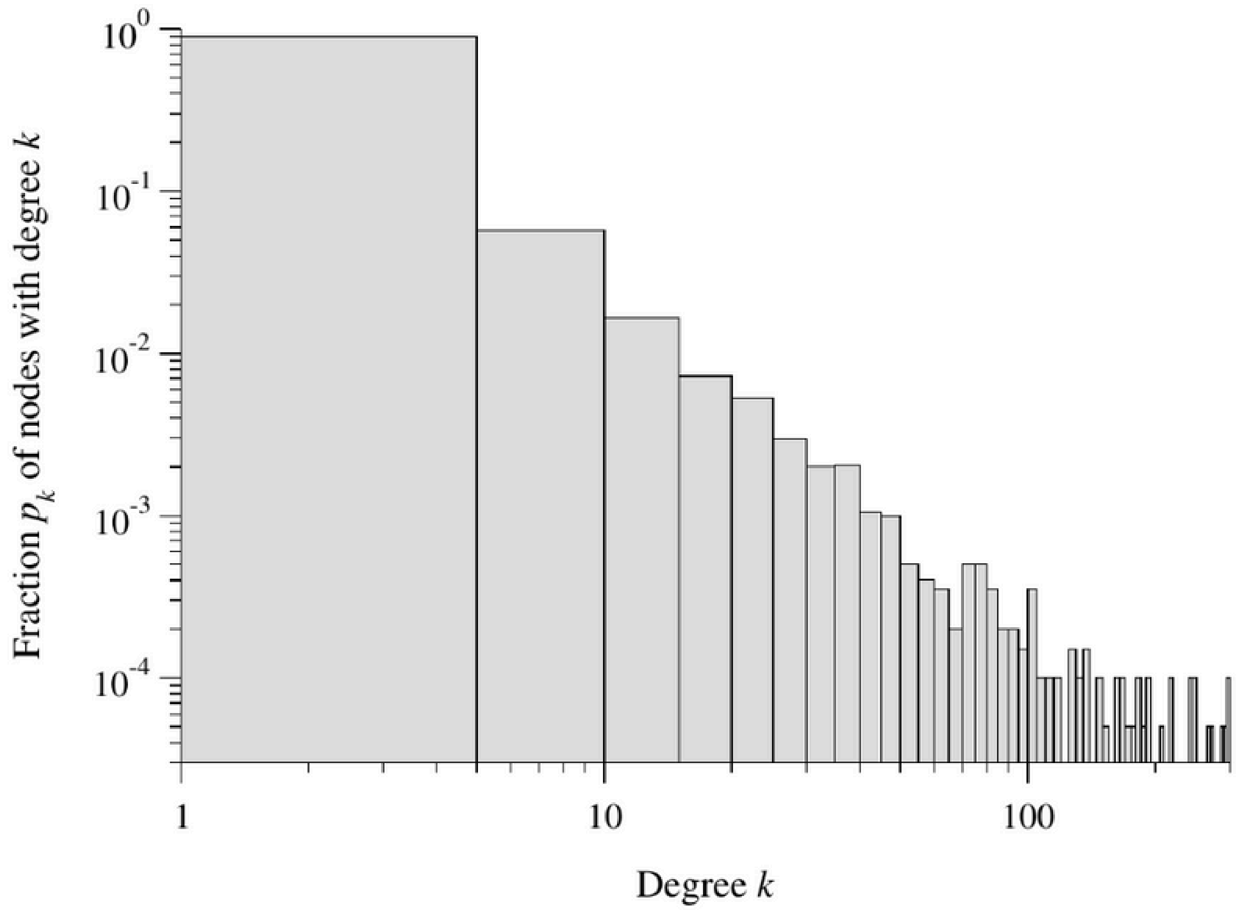


Logarithmic Binning



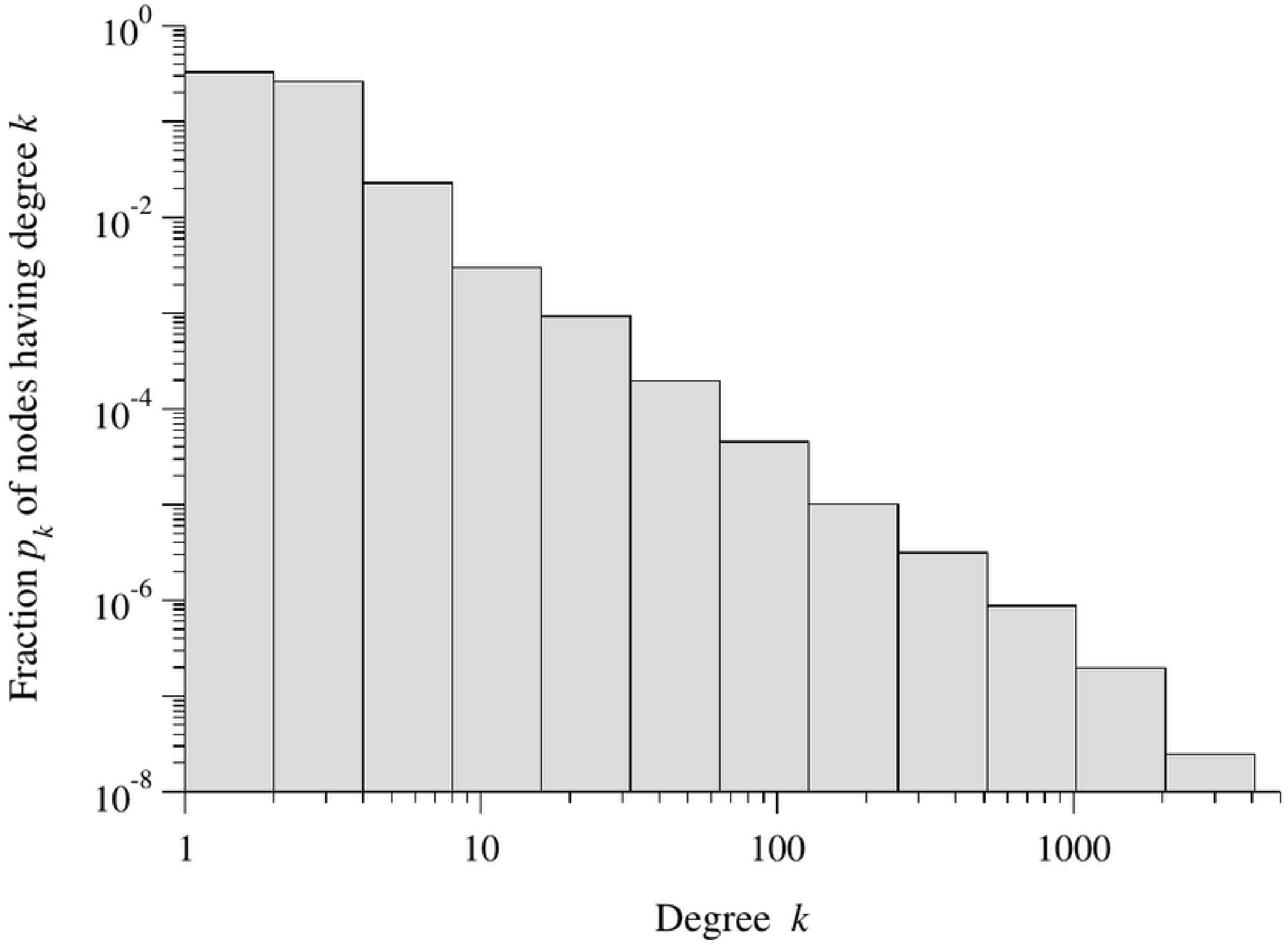
Linear **Binning**





Two histograms of the same distribution. The second one has log-transformed x and y axes and the same bins. Bins are all of the same width in linear scale, but appear different in log scale

Logarithmic Binning



Same log-log histogram but with logarithmic binning: the width of a bin is a multiple of the one on the left. Bin heights are divided by their width. Bins now look all the same in log scale.

Scale-free Property

A Power Law probability distribution is of the form

$$P(k) = \frac{c}{k^{\gamma}}$$

As a consequence if we multiply all sizes by a constant factor K, the shape of the distribution does not change

$$P(a \cdot k) = a^{-\gamma} P(k)$$

For this reason we say that power laws are scale free. They have not a typical scale like a Gaussian.

Diverging Variance

Power laws with γ<3 are particularly interesting since they have a diverging variance. Indeed we have

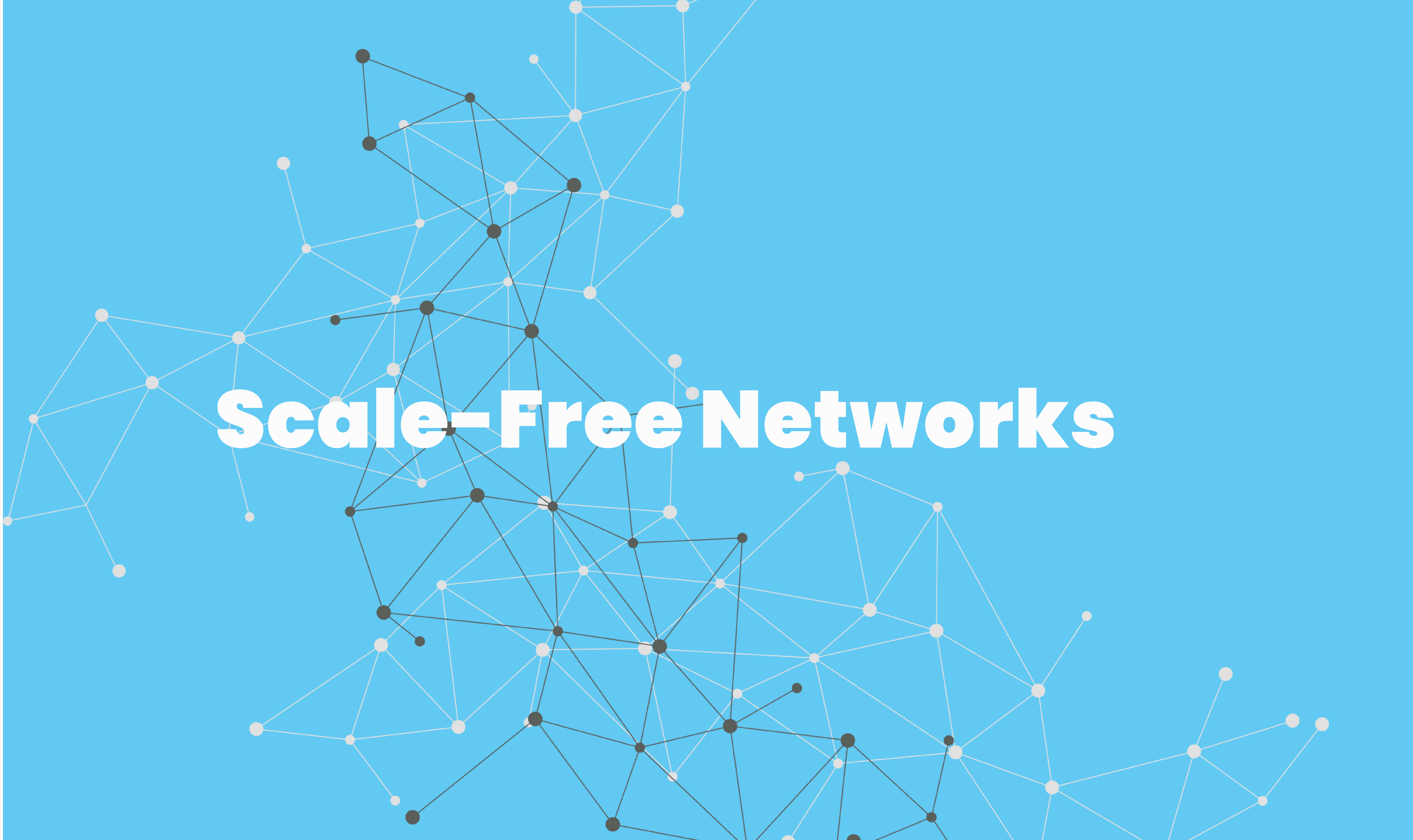
$$\mathrm{Var}[k] = \int_{k_{min}}^{\infty} k^2 P(k) dk - \langle k \rangle^2 = c \int_{k_{min}}^{\infty} k^{2-\gamma} dk - \langle k \rangle^2$$

For γ<3 the integral on the right is infinite and so is the variance

- this means that events of arbitrary large size can occur
- the average value doesn’t make much sense

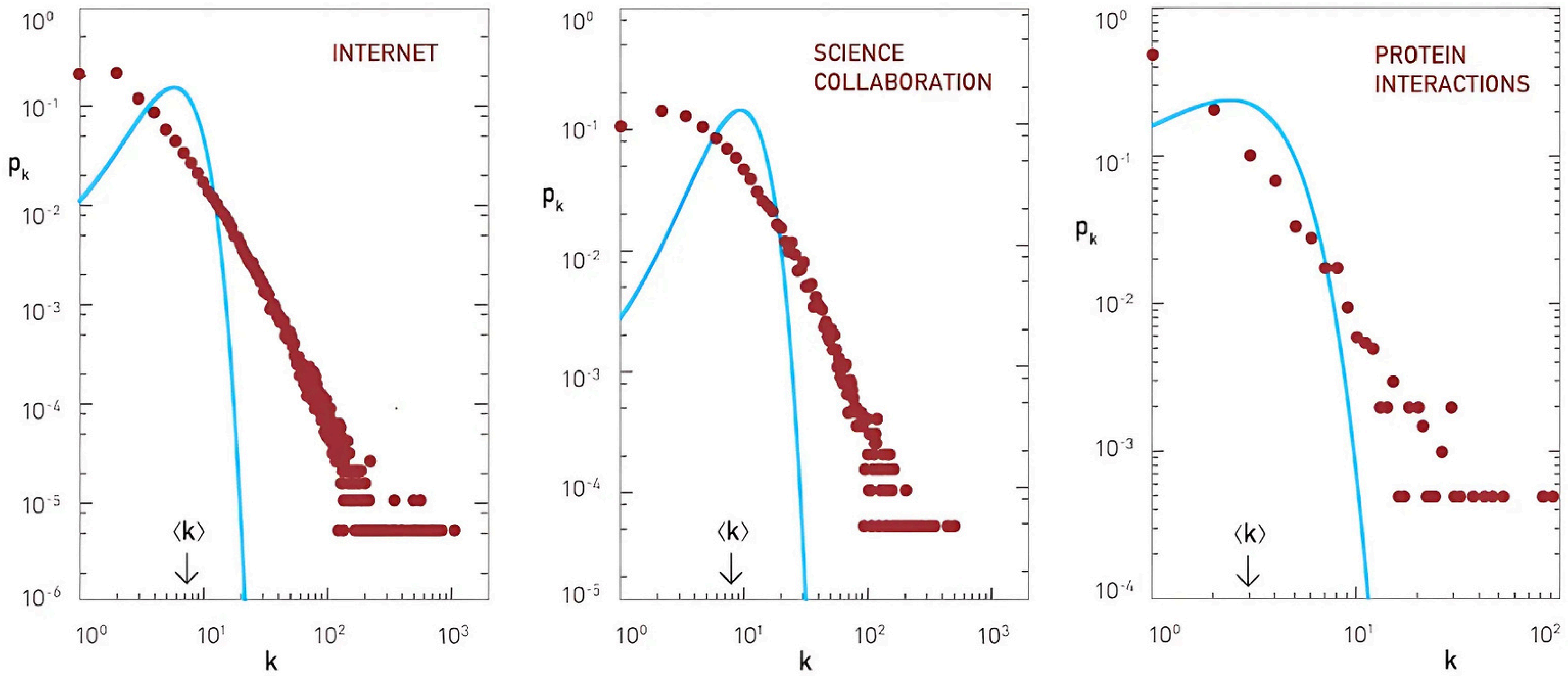
Similarly is γ<2 the mean value is diverging, but this is of less interest since it’s a more rare situation in real systems, particularly networks.

Scale-Free Networks



Scale-free Networks

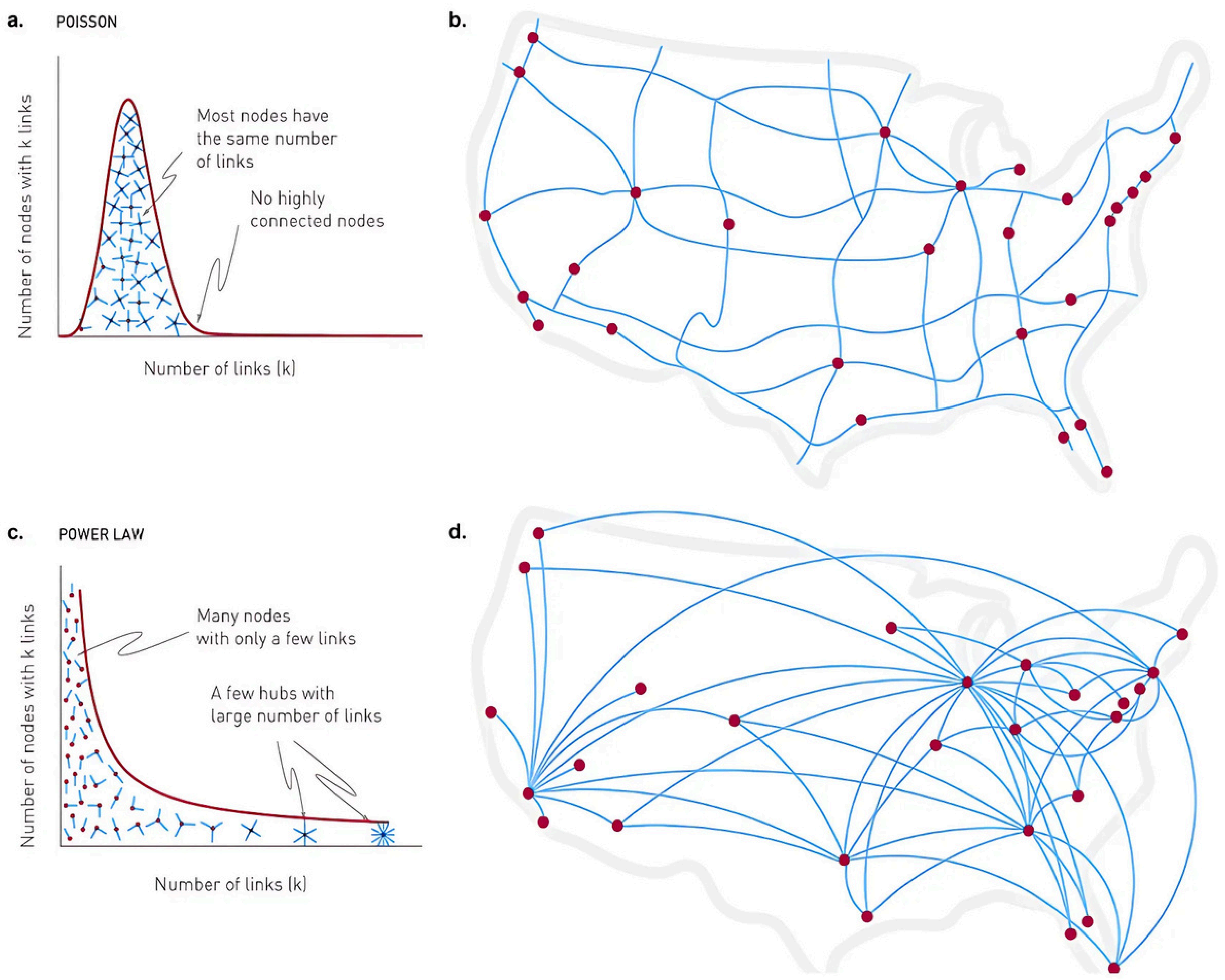
Many real world networks are characterized by a power law distribution of degrees. We call such graphs scale-free networks. In a scale-free network there are many nodes with few connections, but also few nodes with an enormous number of links.



Scale-Free vs Random

Differently from regular and random networks, in scale-free networks there are hubs with many connections

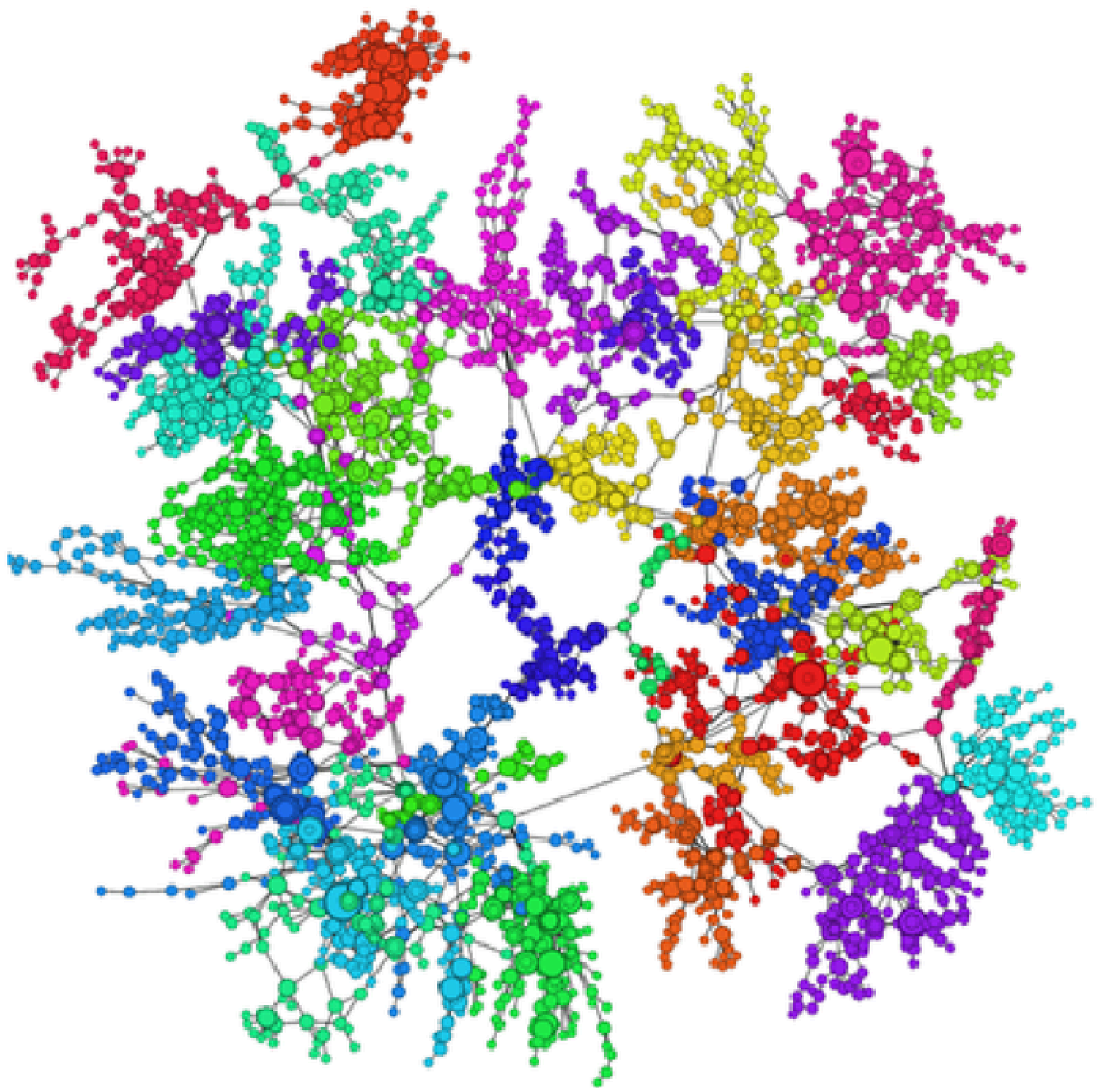
- we can visualize this having in mind an air transportation network
- there are airports connected to a large number of other cities
- this makes very easy traversing the network



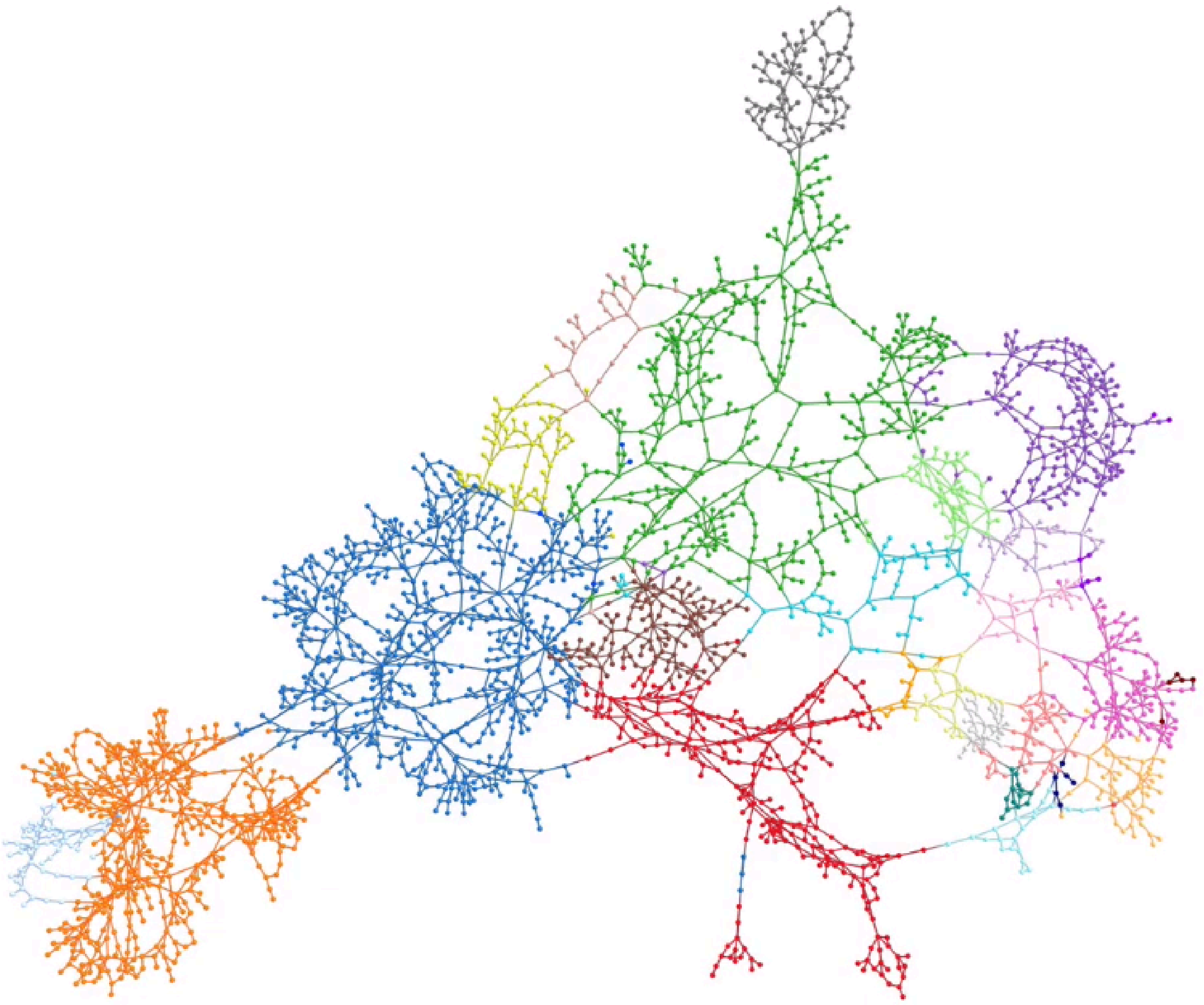
Power Grids

Power grids are an example of scale-free networks

US Power Grid



European Power Grid

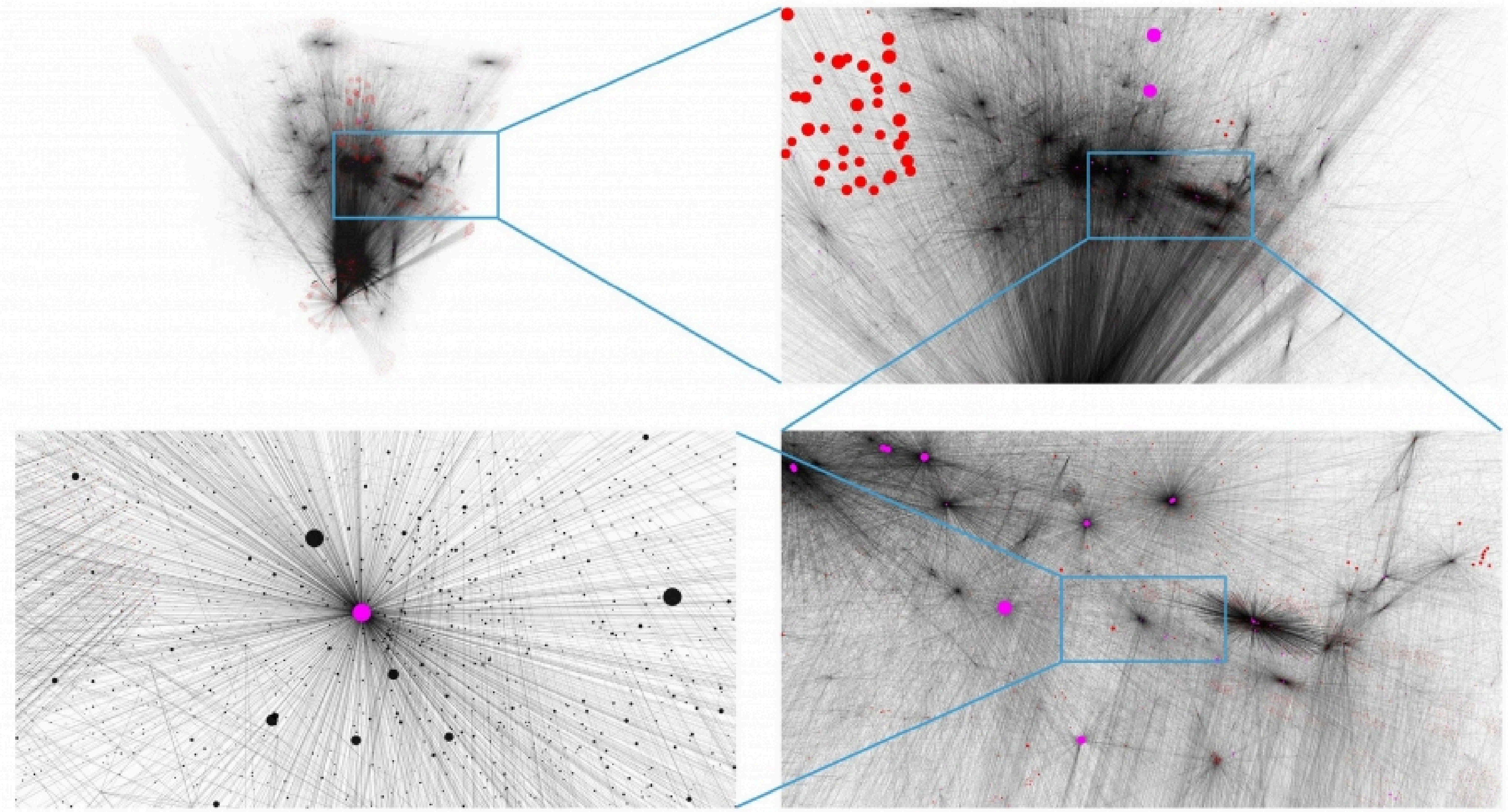


Your Bairey, Madeleine and Shanté Stowell. "US Power Grid Network Analysis." (2014). https://www.youtube.com/watch?v=_XWN53M-bxE

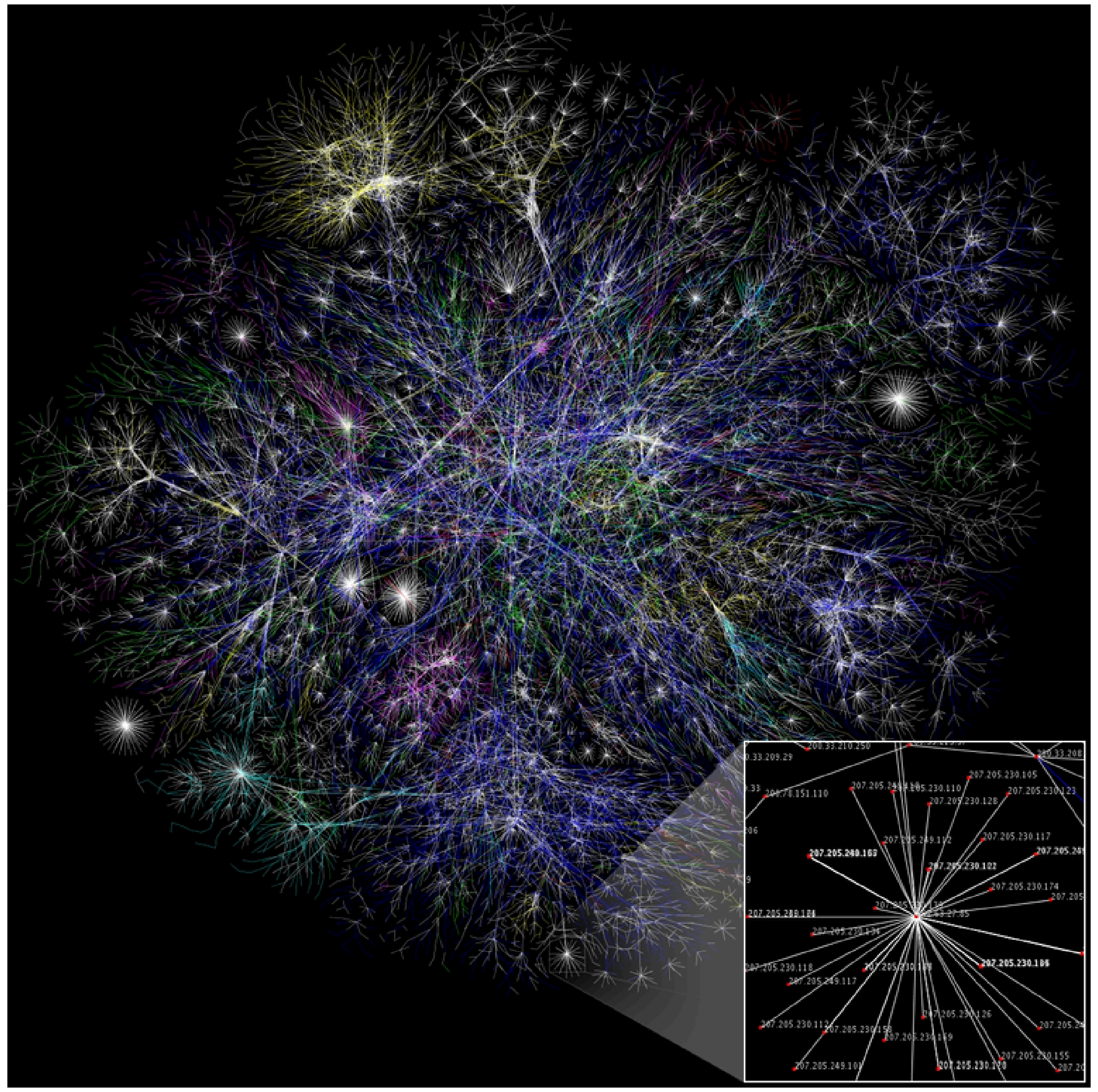
Internet and WWW

Both the WWW and the Internet present a scale-free structure

WWW



Internet



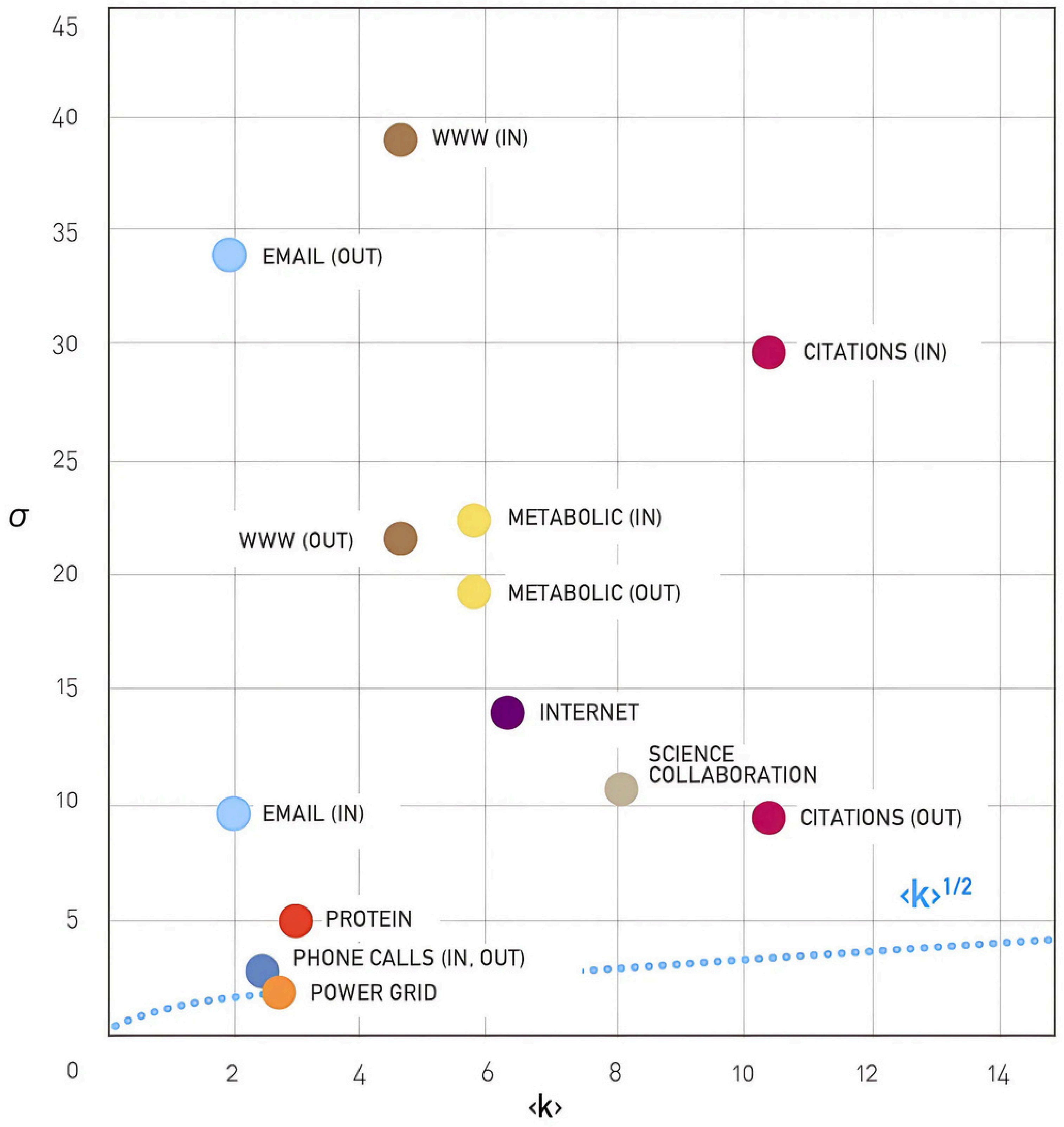
Network Science. A.L. Barabasi https://networksciencebook.com/

The Meaning of Scale-Free

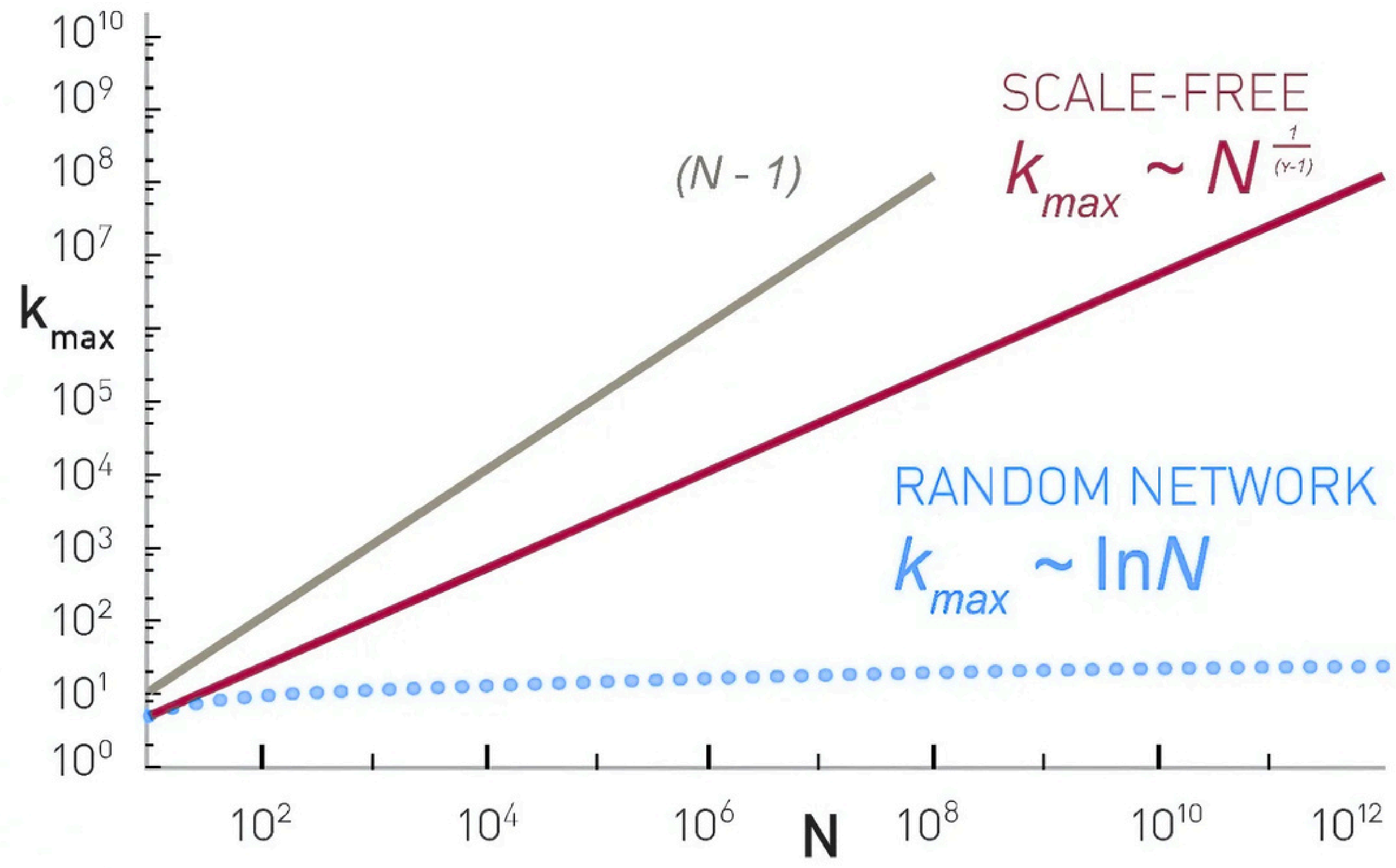
Power law distributions with exponent smaller than 3 have a diverging variance

- the variance can only diverge in an infinite system
- however also in a finite system we can observe a very large variance

Since the variance is much larger than the average degree, the network doesn't have a typical scale



Hubs in Scale-Free Networks



Given a degree distribution and N nodes, we can compute the maximal degree in the system

- for a Poisson distribution we get $k_{max} \sim \ln N$
- for a scale-free distribution instead $k_{max} \sim N^{\frac{1}{\gamma - 1}}$

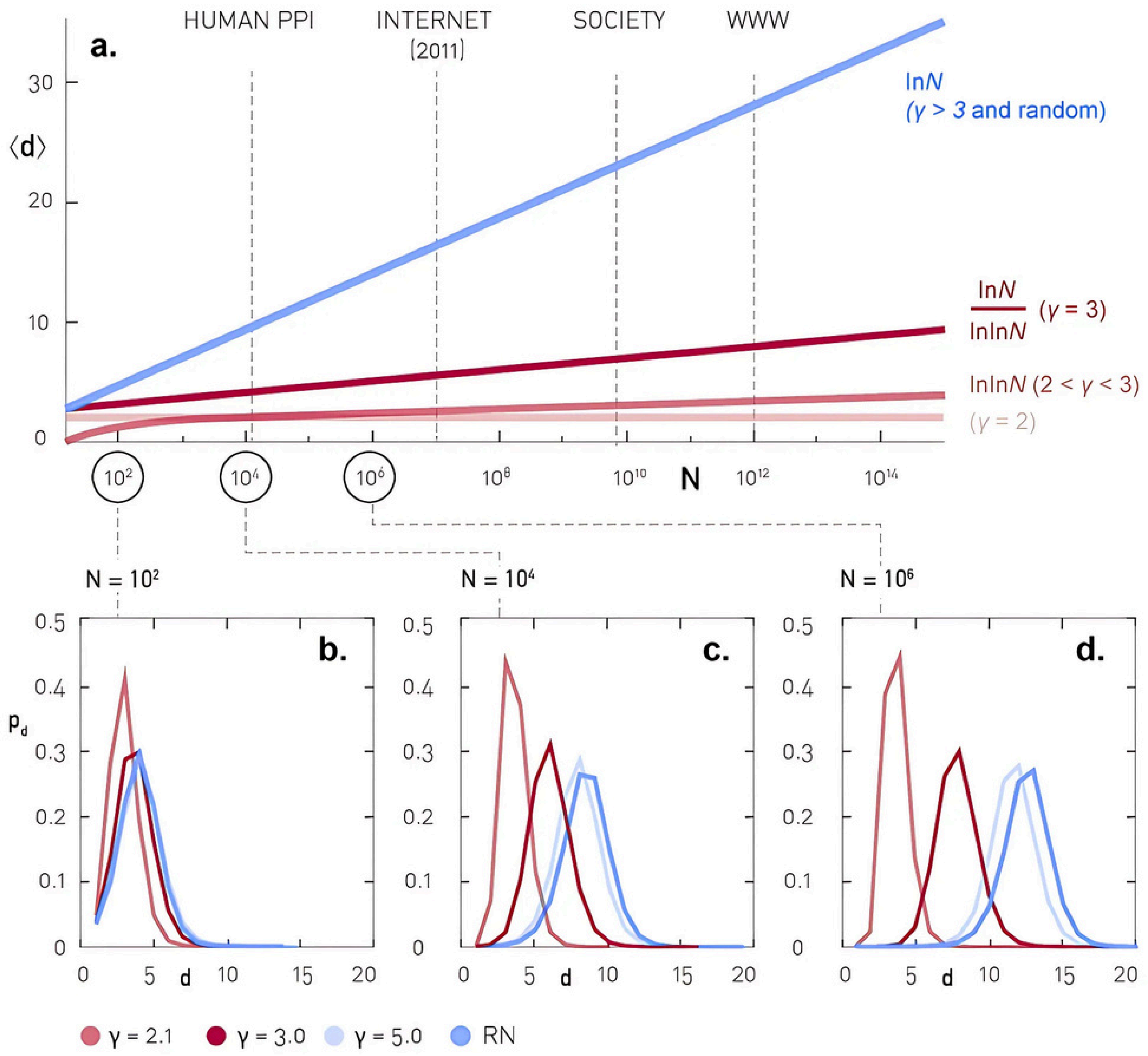
The growth of the largest degree is much faster and this leads to hubs

The Ultra-Small World Property

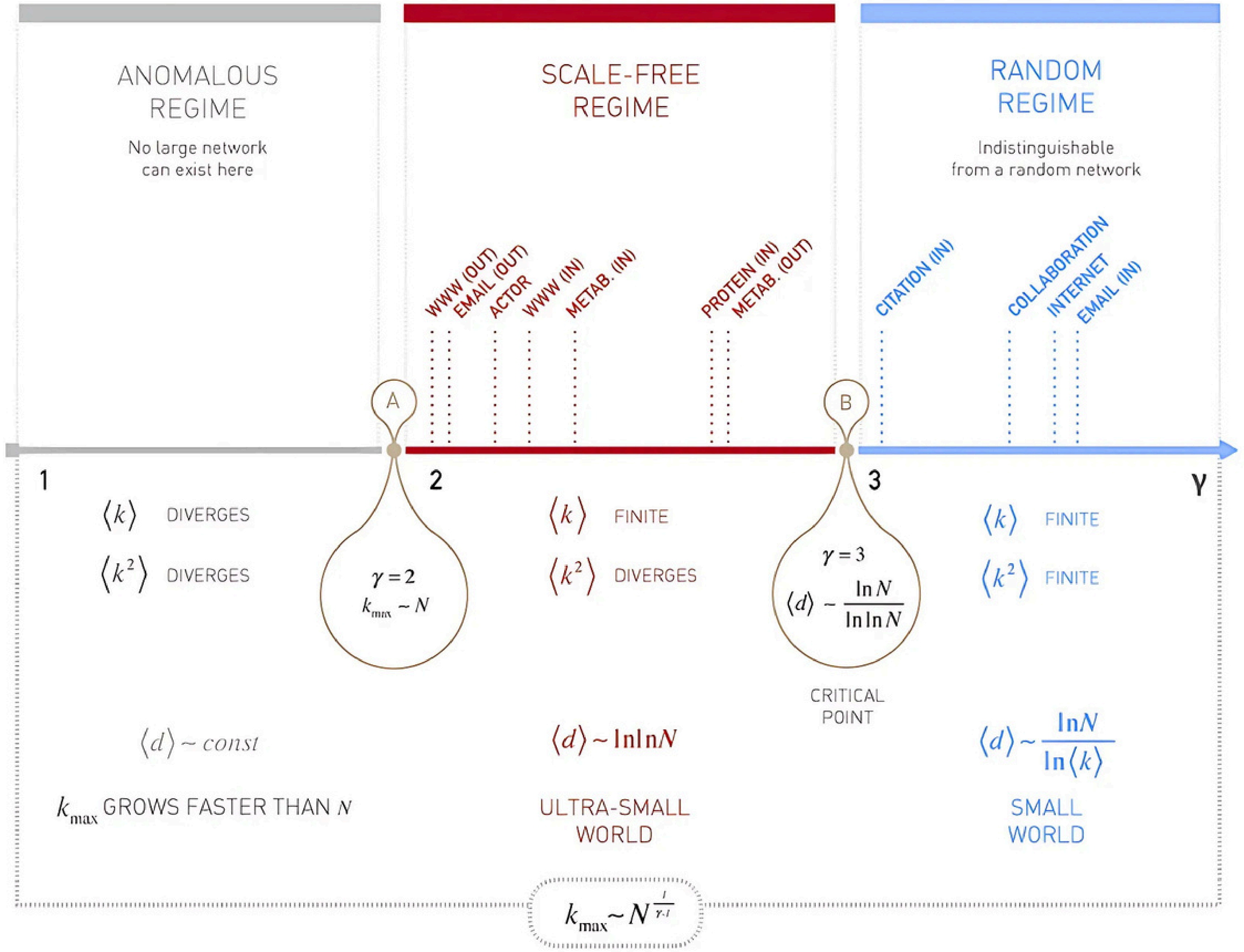
Scale-free networks present different regimes for the average path length

- for γ<3 the network is ultra-small world, the average path length is even shorter than in a small world network
- for γ>3 the behavior is the same as in random networks

$$\langle d \rangle \sim \begin{cases} const. & \gamma = 2 \\ \ln \ln N & 2 < \gamma < 3 \\ \dfrac{\ln N}{\ln \ln N} & \gamma = 3 \\ \ln N & \gamma > 3 \end{cases}$$



Regimes of Scale-Free Networks



Scale-Free networks present 3 regimes:

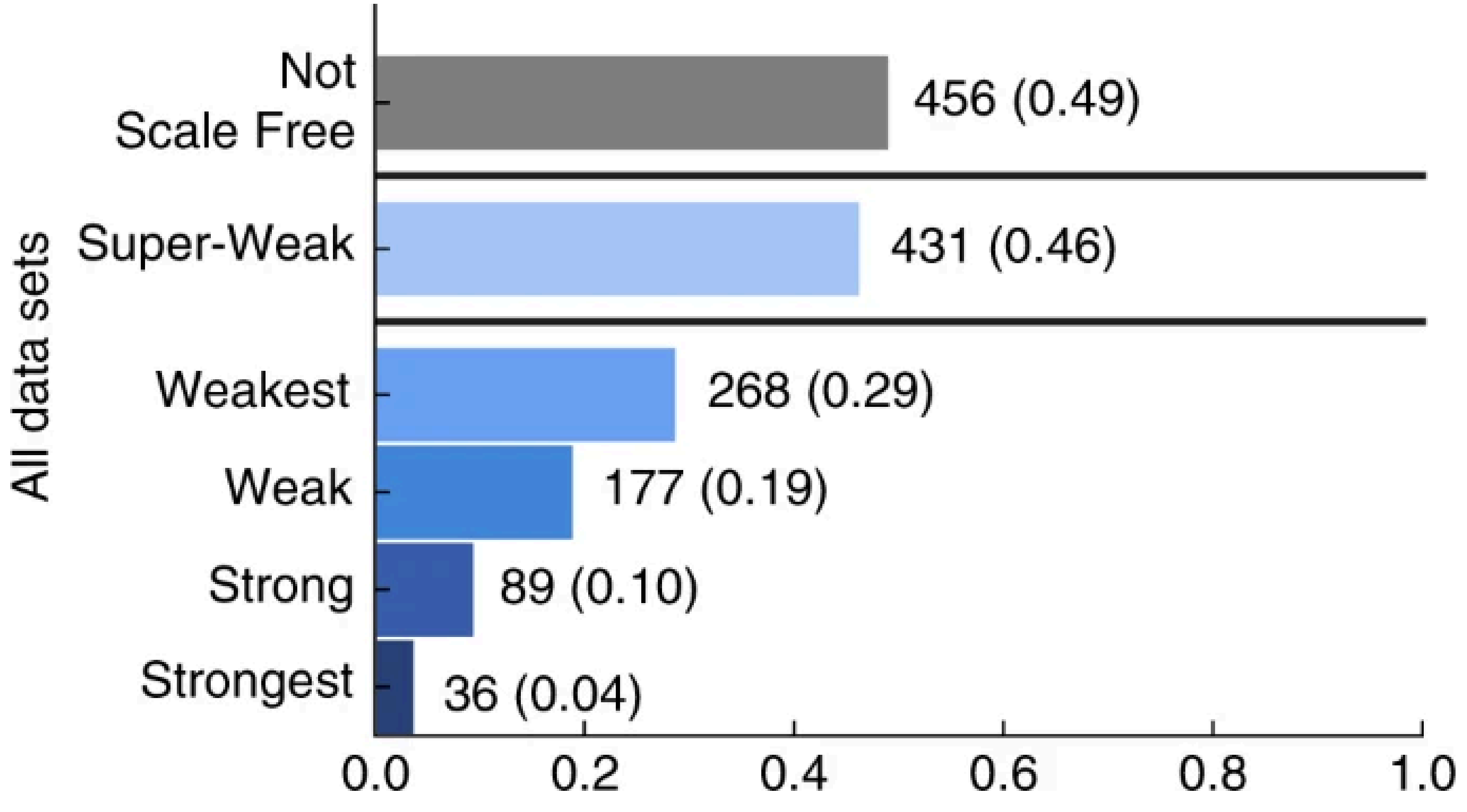
- **Anomalous Regime γ<2**
  Both the variance and the mean degree diverge. No large networks can exist.
- **Scale-Free Regime 2<γ<3**
  The variance diverges but the mean degree is finite. Networks are ultra-small world
- **Random Regime γ>3**
  Both the variance and the mean degree are finite. Networks are very similar to random networks

Are Scale-Networks Ubiquitous?

After the discovery of the first scale-free networks, the scale-free property has been attributed to hundreds and hundreds of networks

- this lead to a claim of universality of real networks
- however more recent studies found that only a limited fraction of networks is truly scale-free

Even if some network are not scale-free, they have a degree distribution much wider than in a random network



Broido, A.D., Clauset, A. Scale-free networks are rare. Nat Commun 10, 1017 (2019).

Barabasi-Albert Model

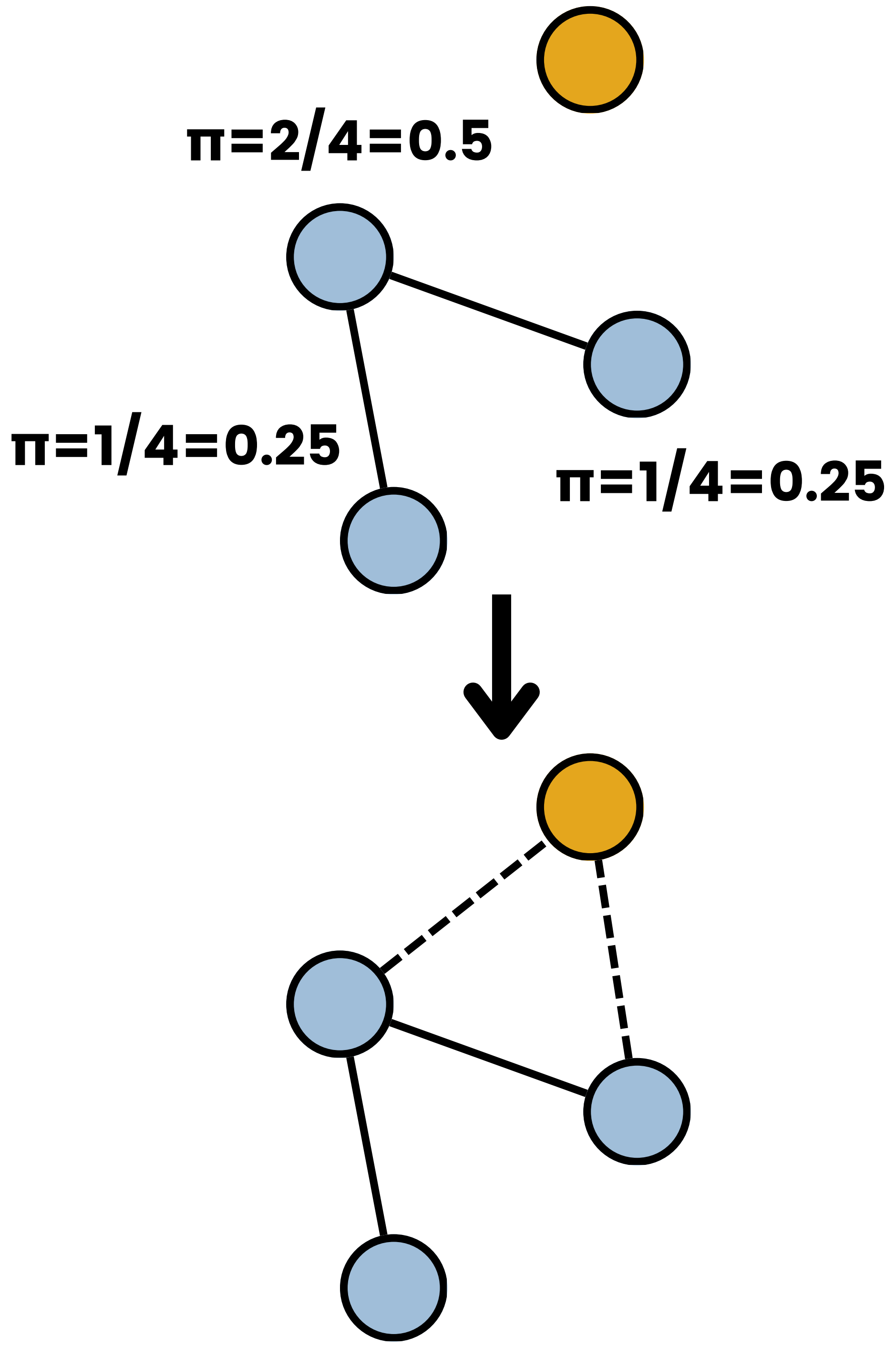
The Barabasi-Albert Model

We want to understand how scale-free networks can emerge from individual behavior.
The Barabasi-Albert model is a simple network growth process showing that scale-free networks can emerge from a simple mechanism

- we start with an initial network
- at each time step we add a new node
- this new node links to m existing nodes
- the linking probability π_i to link to node i is proportional to the node's degree

$$\pi_i = \pi(k_i) = \frac{k_i}{\sum_j k_j} = \frac{k_i}{2mN}$$

https://sarah37.github.io/barabasialbert/



Degree Dynamics

We can compute the evolution of a specific degree k_i using the linking probability

$$k_i(N+1) = k_i(N) + m\frac{k_i(N)}{\sum_j k_j(N)} \rightarrow k_i(N+1) - k_i(N) = m\frac{k_i(N)}{\sum_j k_j(N)}$$

The left side of the equation approximates the derivate and the denominator is 2mN

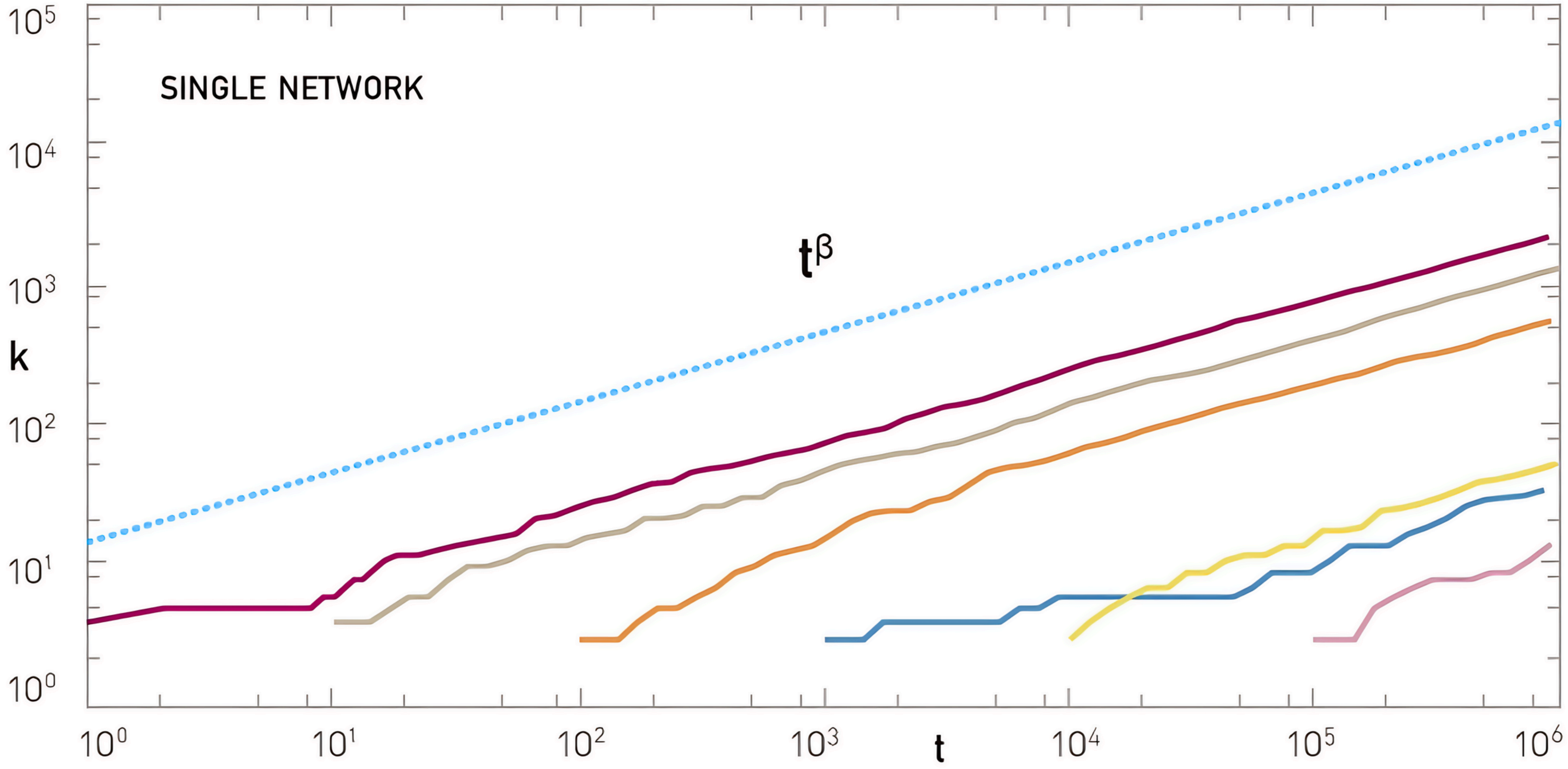
$$\frac{dk_i}{dN} = \frac{k_i}{2mN}$$

Defining as N_i the size of the network when node i entered it, the solution of this differential equation is

$$k_i(N) = m\left(\frac{N}{N_i}\right)^{\beta} \quad \text{with} \quad \beta = \frac{1}{2}$$

Rich-get-Richer Effect

In the Barabasi-Alber model, older nodes have an advantage over younger nodes.
This is called Rich-get-Richer effect (or cumulative advantage)



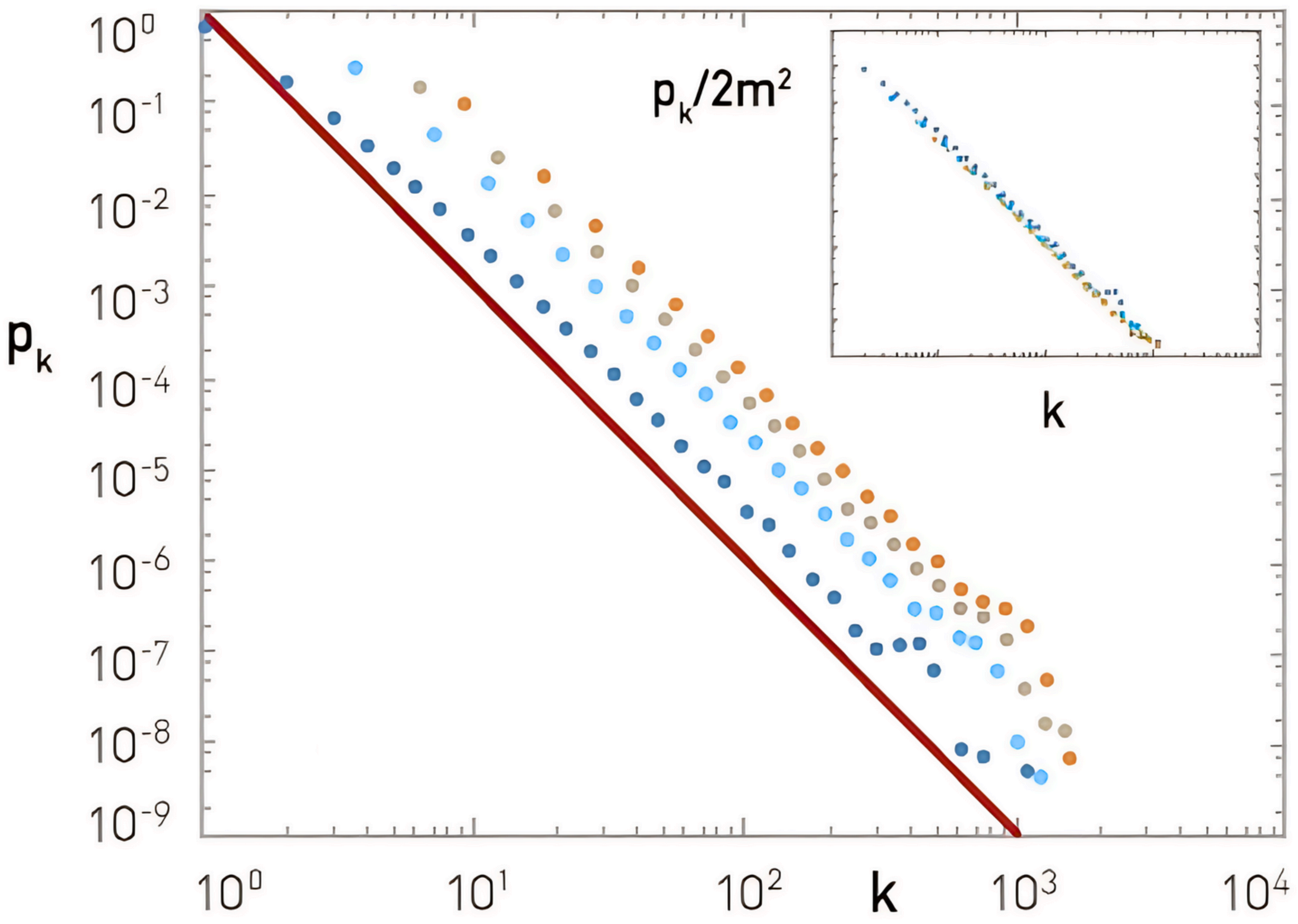
Scale-Free Degree Distribution

The Barabasi-Albert model generates scale free networks

- the power law exponent is independent of
  - the number of links m
  - the initial network
- the model asymptotically produces a degree distribution with exponent -3

$$P(k) = \frac{2m^2}{k^3}$$

- small modifications allow to get any exponent >2



Deriving the Degree Distribution

We denote by N(k, N) the number of nodes with degree k in a network with N nodes. By adding a new node this number will change following the equation below

$$N(k, N+1) = N(k, N) + m\frac{k-1}{2mN}N(k-1, N) - m\frac{k}{2mN}N(k, N)$$

N(k, N)/N is the probability P(k, N) of observing a node of degree k

$$P(k, N+1)(N+1) = P(k, N)N + \frac{k-1}{2}P(k-1, N) - \frac{k}{2}P(k, N)$$

Adding a node don't change the probability in large networks P(k, N)=P(k, N+1)=P(K)

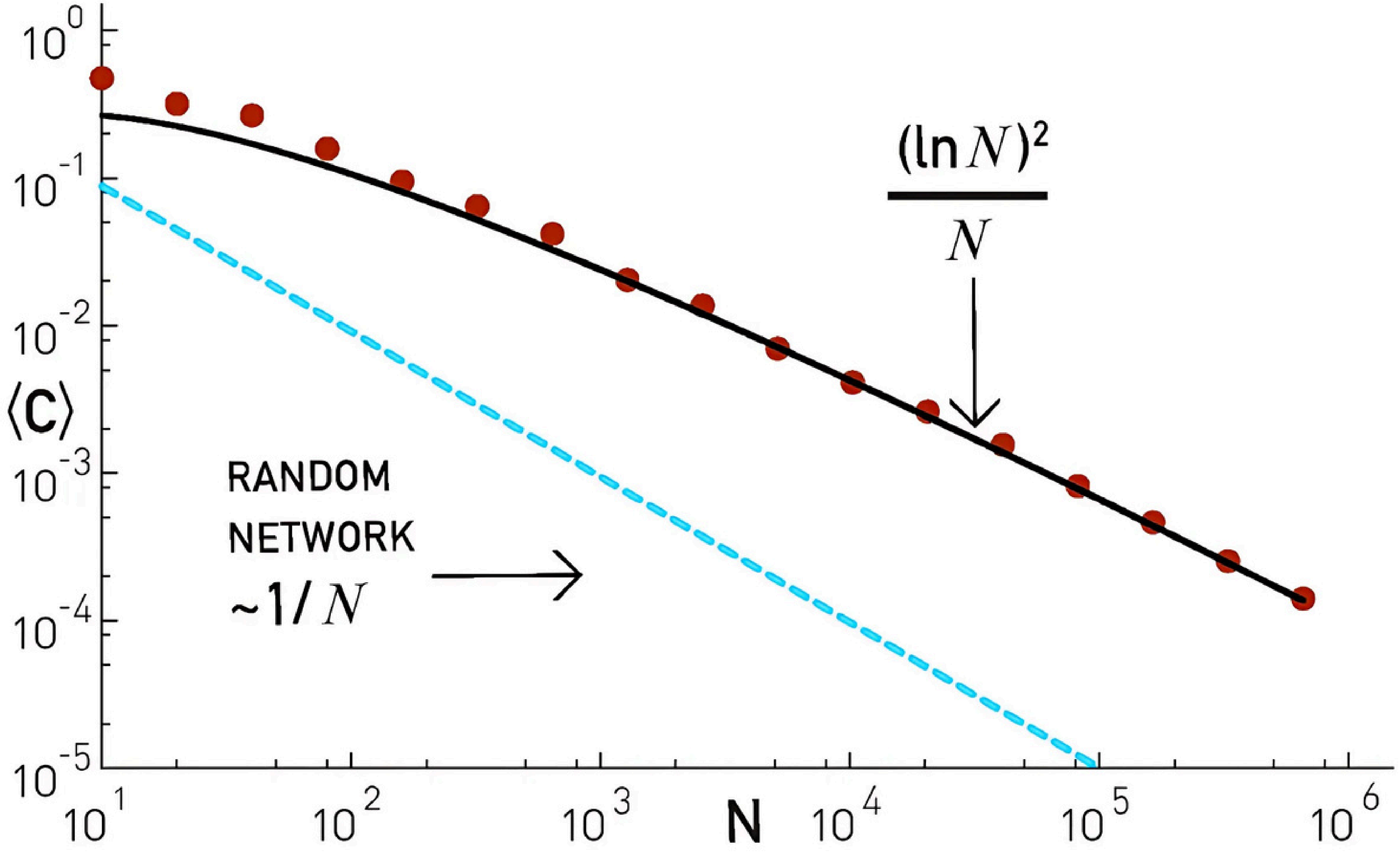
$$P(k) = -\frac{1}{2}\left[kP(k) - (k-1)P(k-1)\right] \approx -\frac{1}{2}\frac{d}{dk}\left[kP(k)\right]$$

It is easy to show that a solution to this differential equation is

$$P(k) \sim k^{-\gamma} \quad \text{with} \quad \gamma = 3$$

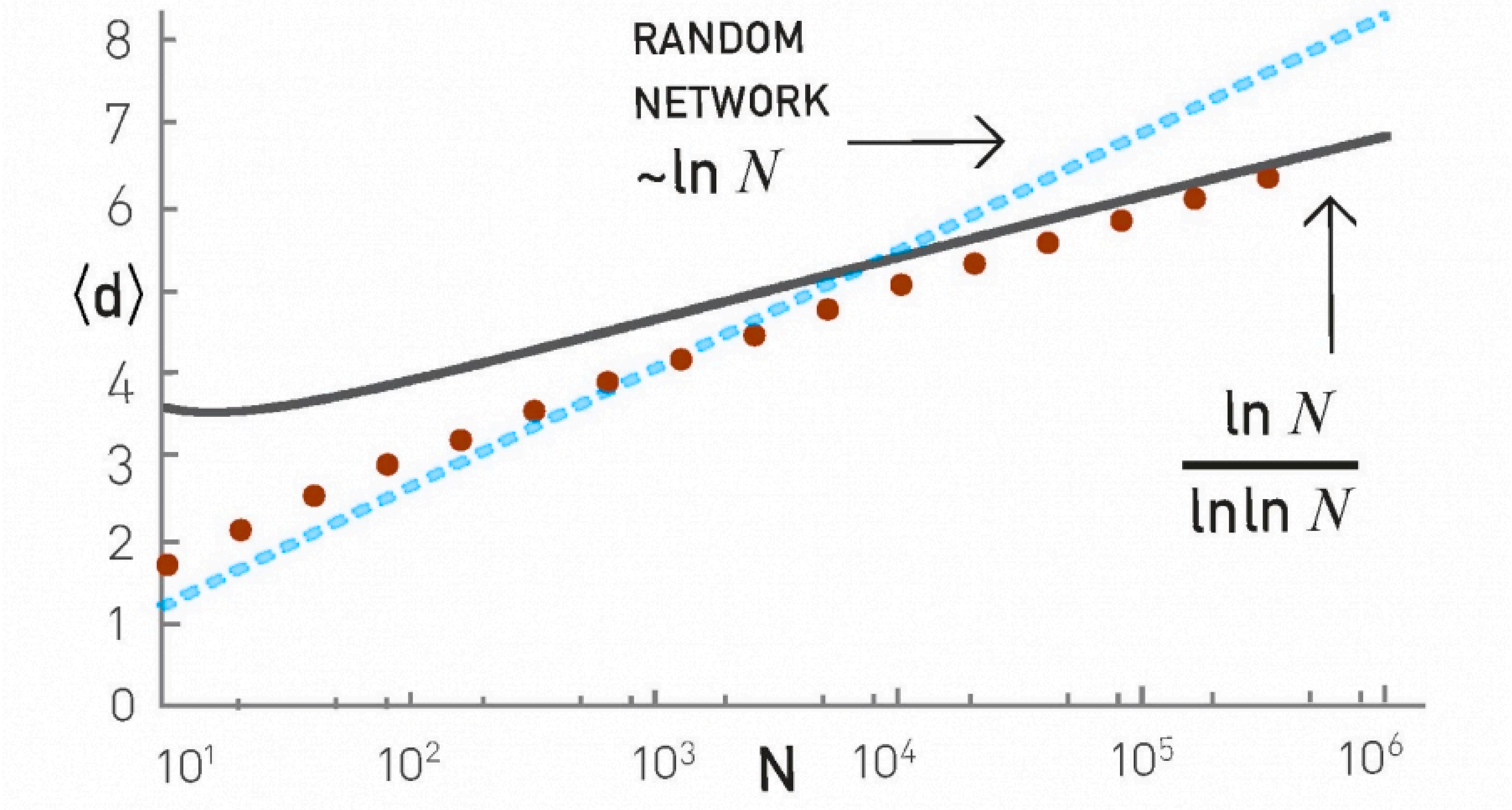
Clustering Coefficient

In the Barabasi-Albert model, the clustering coefficient is larger than in random network, but still it goes to zero for large network sizes

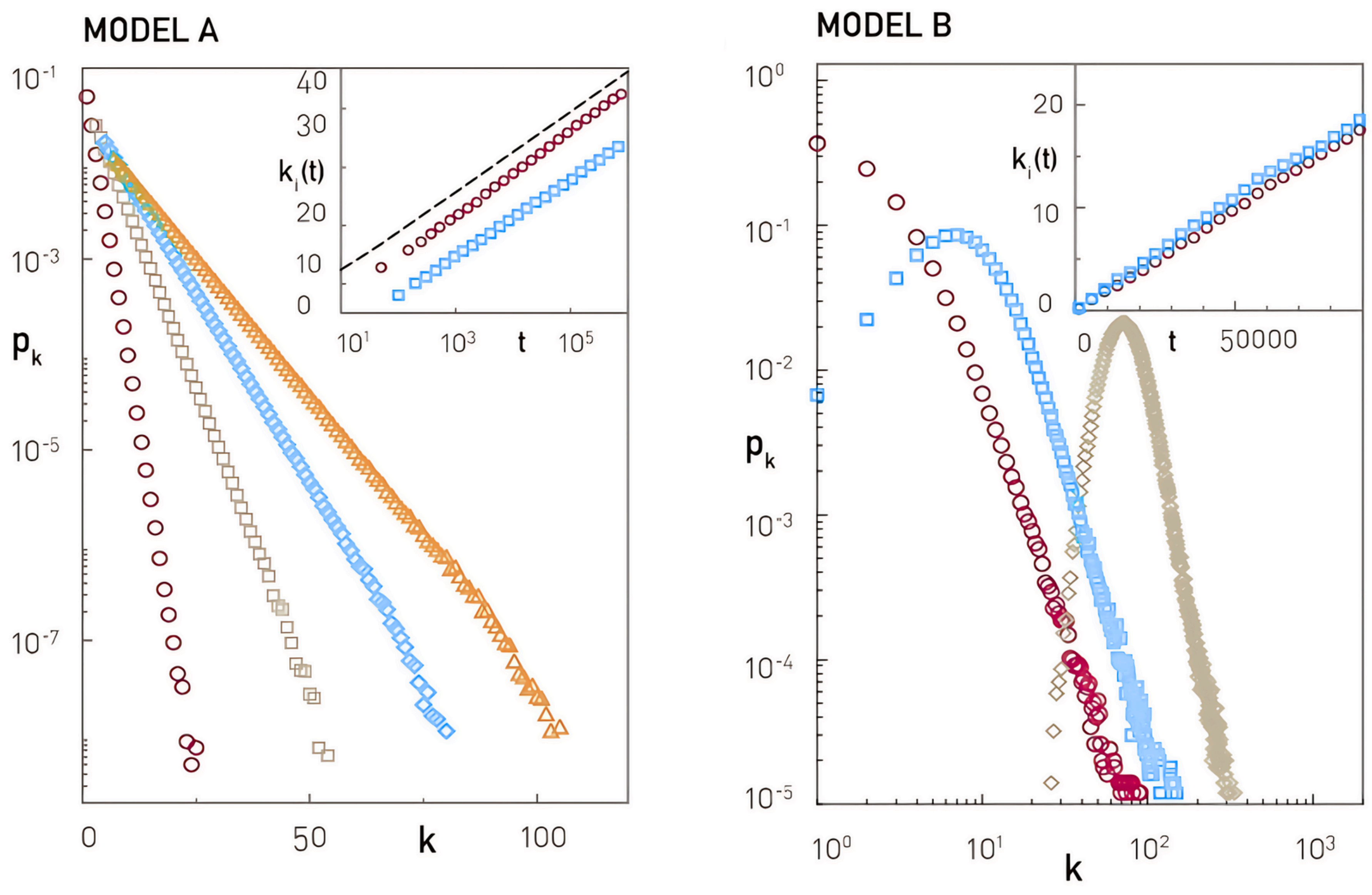


Diameter

The average path length grows slower than in a random network, but since γ=3 there is no ultra-small world



Necessary Conditions



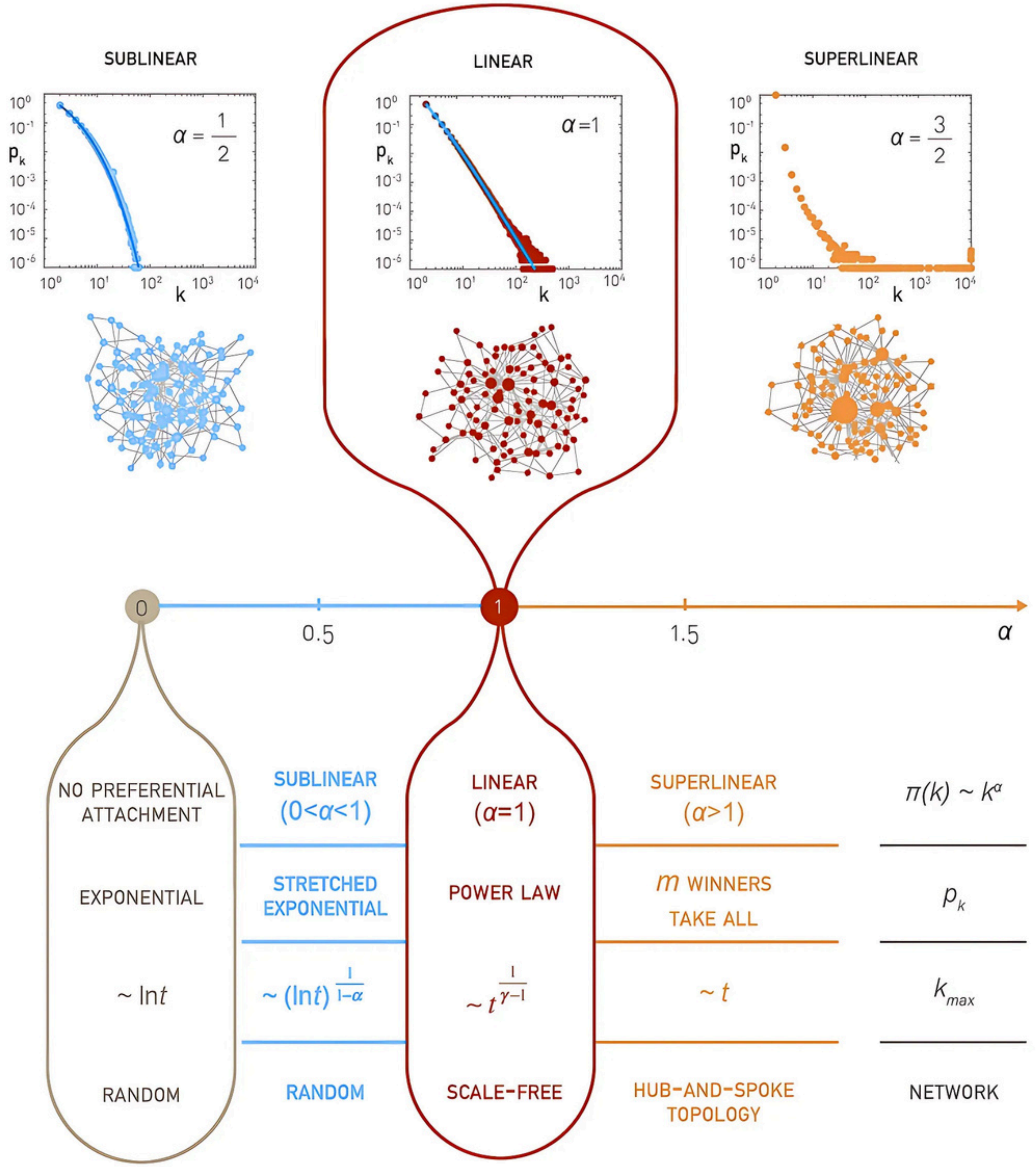
What are the necessary ingredients to get a scale-free networks? Are both growth and (linear) preferential attachment crucial?

- without (linear) preferential attachment we get random networks (exponential degree distribution)
- without growth (no new nodes) the distribution never reaches a stationary state and peaks on a specific value (depending on N)

Non-Linear Preferential Attachment

What happens if we change the exponent of the linking probability using a non-linear preferential attachment?

- **Sublinear case.** The degree distribution is a stretched exponential and the network is basically random
- **Superlinear case.** The degree distribution is peaked on the tail. The network is an hub and spoke topology



The Vertex Copy Model

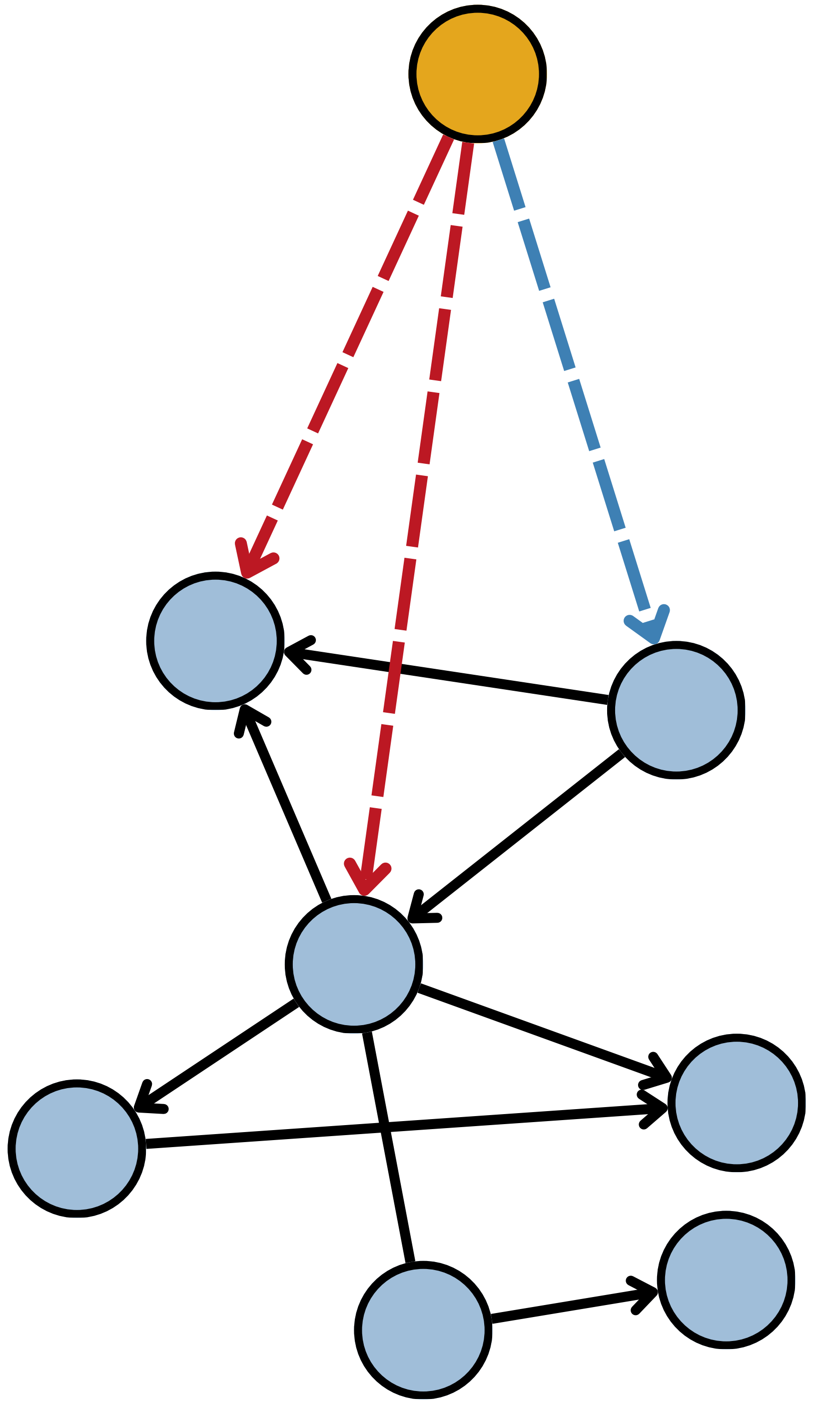
The Barabasi-Albert model is not very realistic:

- in order to compute the linking probability we have to know all degrees
- in most situation we can only observe a very limited portion of a network

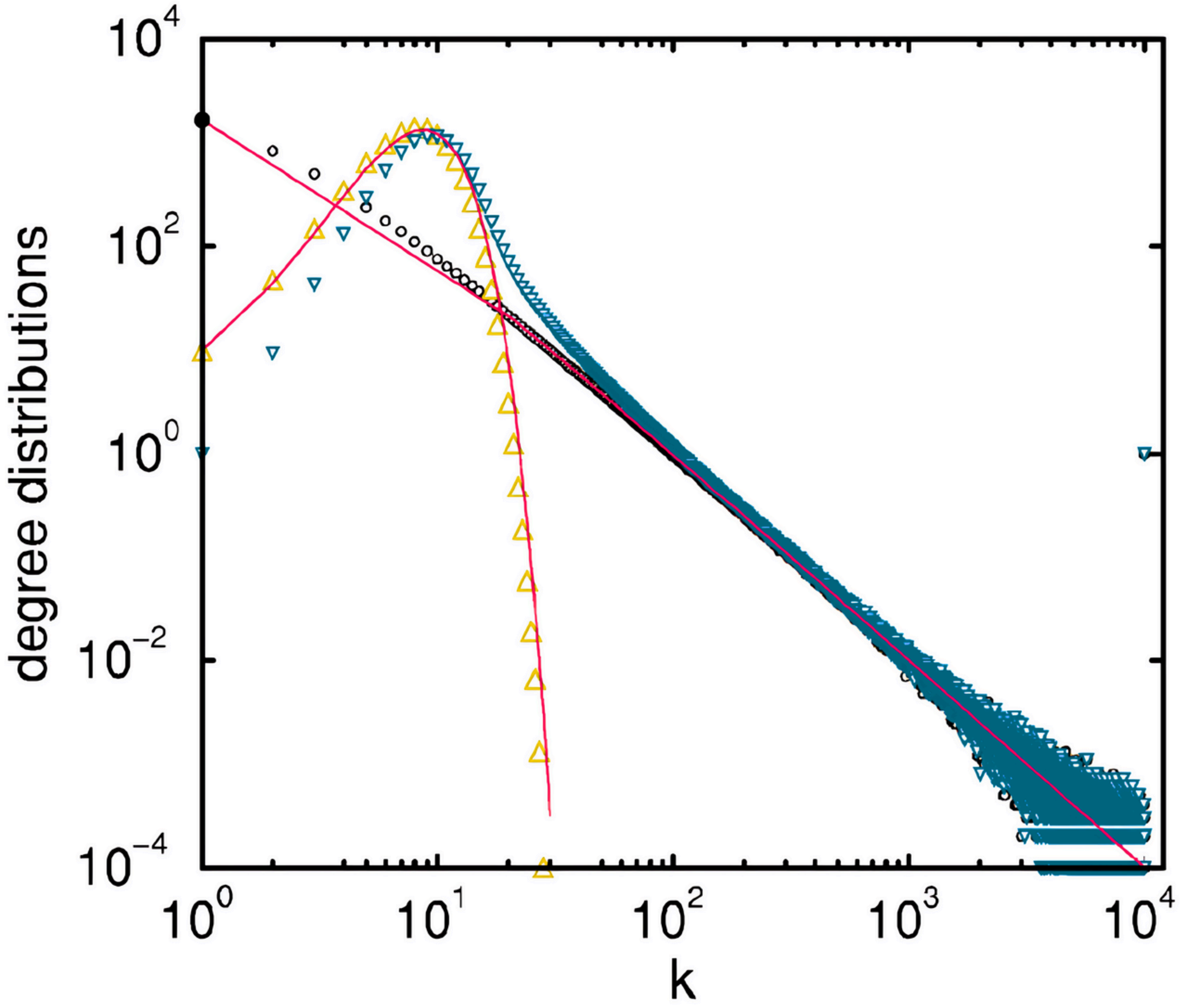
In the Vertex Copy model these limitations are overcame by exploiting a more local mechanism

- at each time step a new node is added
- this node links to a random node (blue arrow)
- it then copied all the connections of the node it has linked to (red arrows)

In this way we only need to know the local structure around a node.



Degree Distribution



The Vertex Copy model produces directed networks. The relevant property to look at is the in-degree (incoming connections)

- the out-degree distribution is peaked
- the in-degree distribution is a power law with exponent -2

This implies that it is possible to obtain scale free networks even if only local information is used. The edge copy mechanism is creating a sort of proxy of the linear preferential attachment.

Robustness of Scale-Free Networks

Disrupting Scale-Free Networks

During last lecture we studied the robustness of networks focusing on random graphs. However many real networks are scale-free

- in scale-free networks there are hub providing connections to easily traverse the network
- these hubs will make the network much more tolerant to random failures
- however target attacks to the hubs can seriously compromise the functioning of the whole networks

Recap: Molloy-Reed Criterion

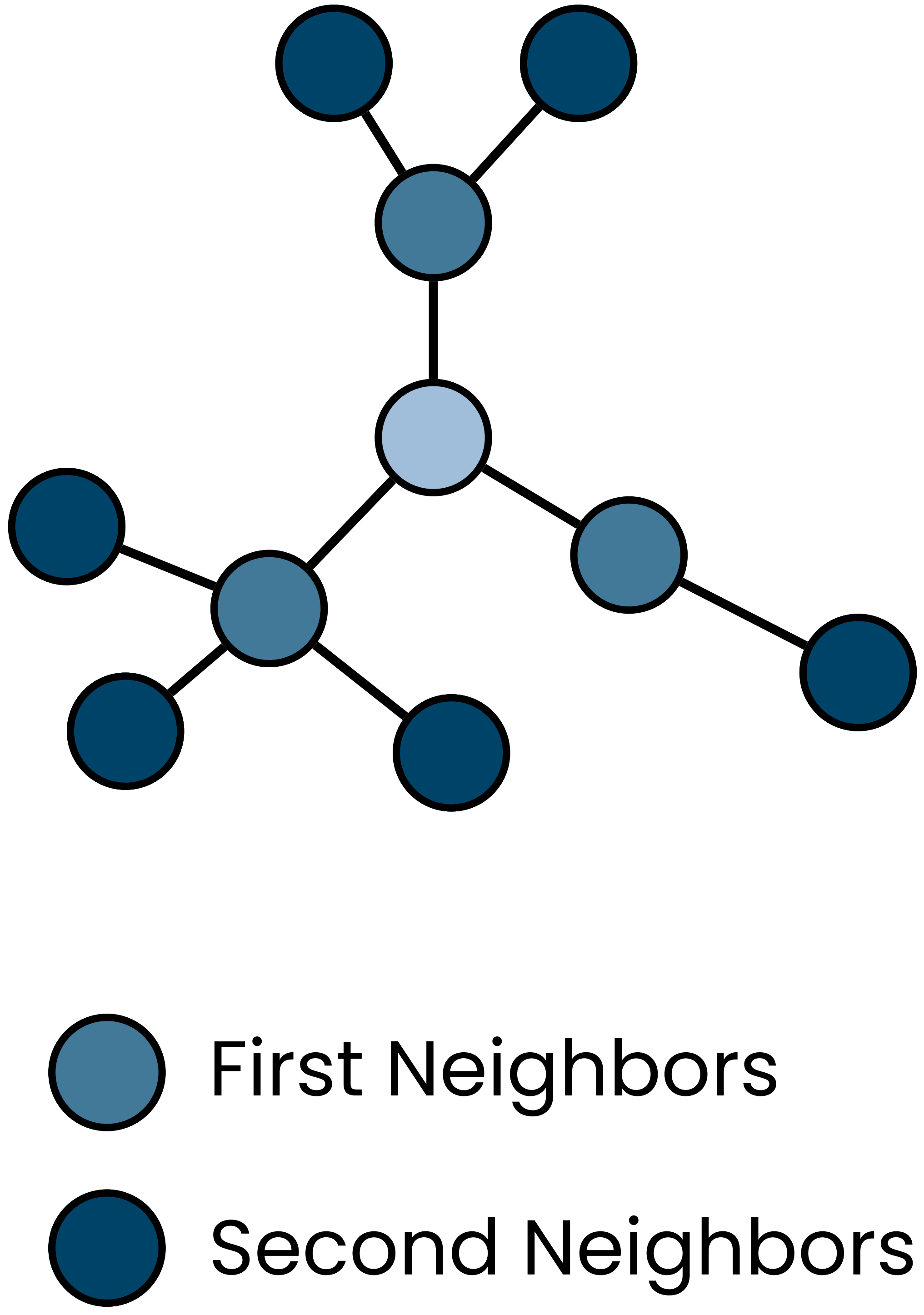
Molloy-Reed criterion allows to determine if a network contains a giant component by comparing the number of first and second neighbors

$$\langle k^2 \rangle - \langle k \rangle > \langle k \rangle \rightarrow \frac{\langle k^2 \rangle}{\langle k \rangle} > 2$$

When performing a random node removal this lead to the following expression for the critical threshold above which the giant component gets destroyed

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

From this expression we can already understand that scale-free networks are very resistant!



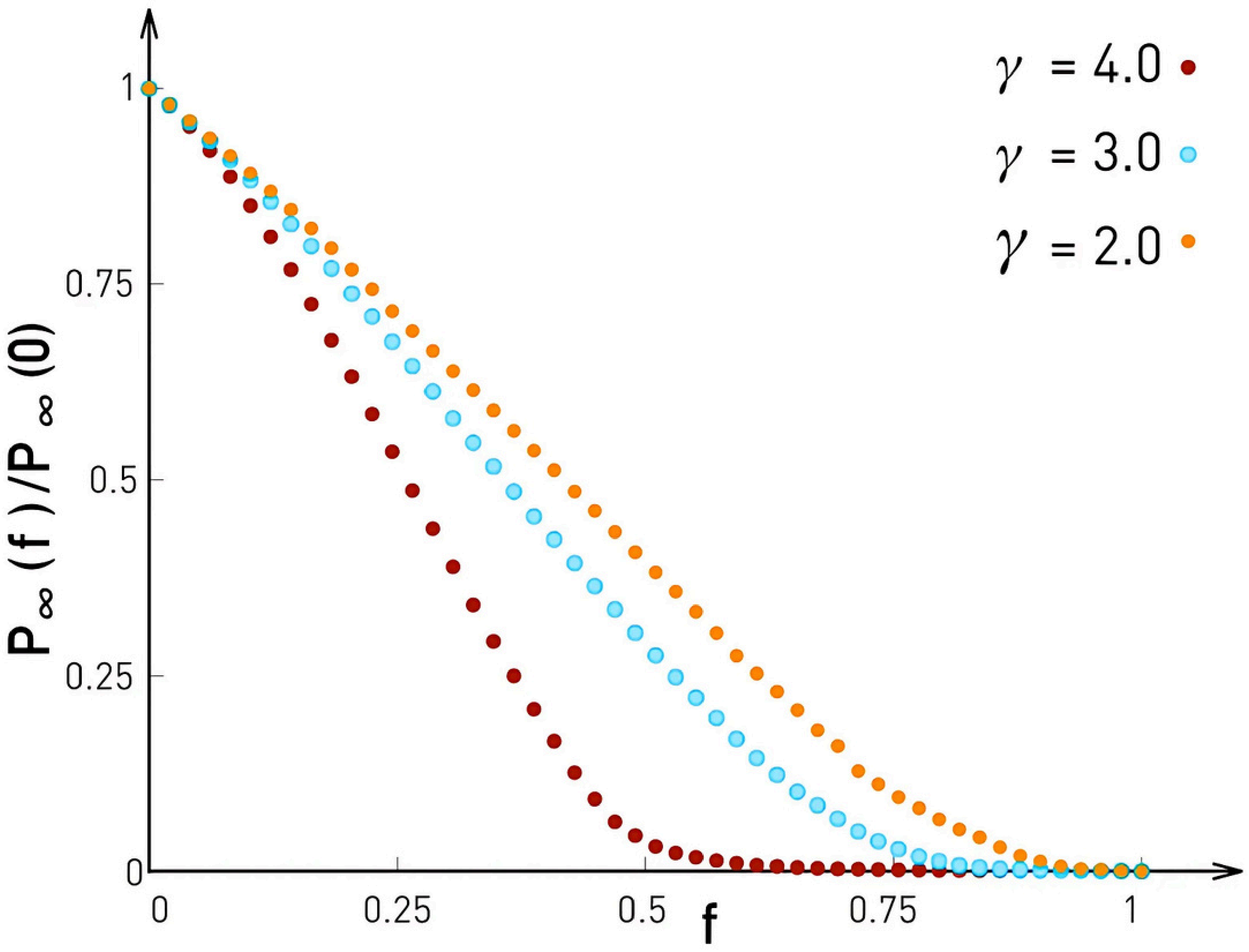
Tolerance to **Failures**

By using Molloy-Reed criterion we can compute explicitly the critical threshold for scale-free networks

$$f_c = \begin{cases} 1 - \dfrac{1}{\frac{\gamma-2}{3-\gamma} k_{\min}^{\gamma-2} k_{\max}^{3-\gamma} - 1} & 2 < \gamma < 3 \\ 1 - \dfrac{1}{\frac{\gamma-2}{\gamma-3} k_{\min} - 1} & \gamma > 3 \end{cases}$$

For γ<3 the largest degree diverges and so the critical threshold is 1 for infinite networks. In finite networks instead

$$f_c \approx 1 - \frac{C}{N^{\frac{3-\gamma}{\gamma-1}}}$$

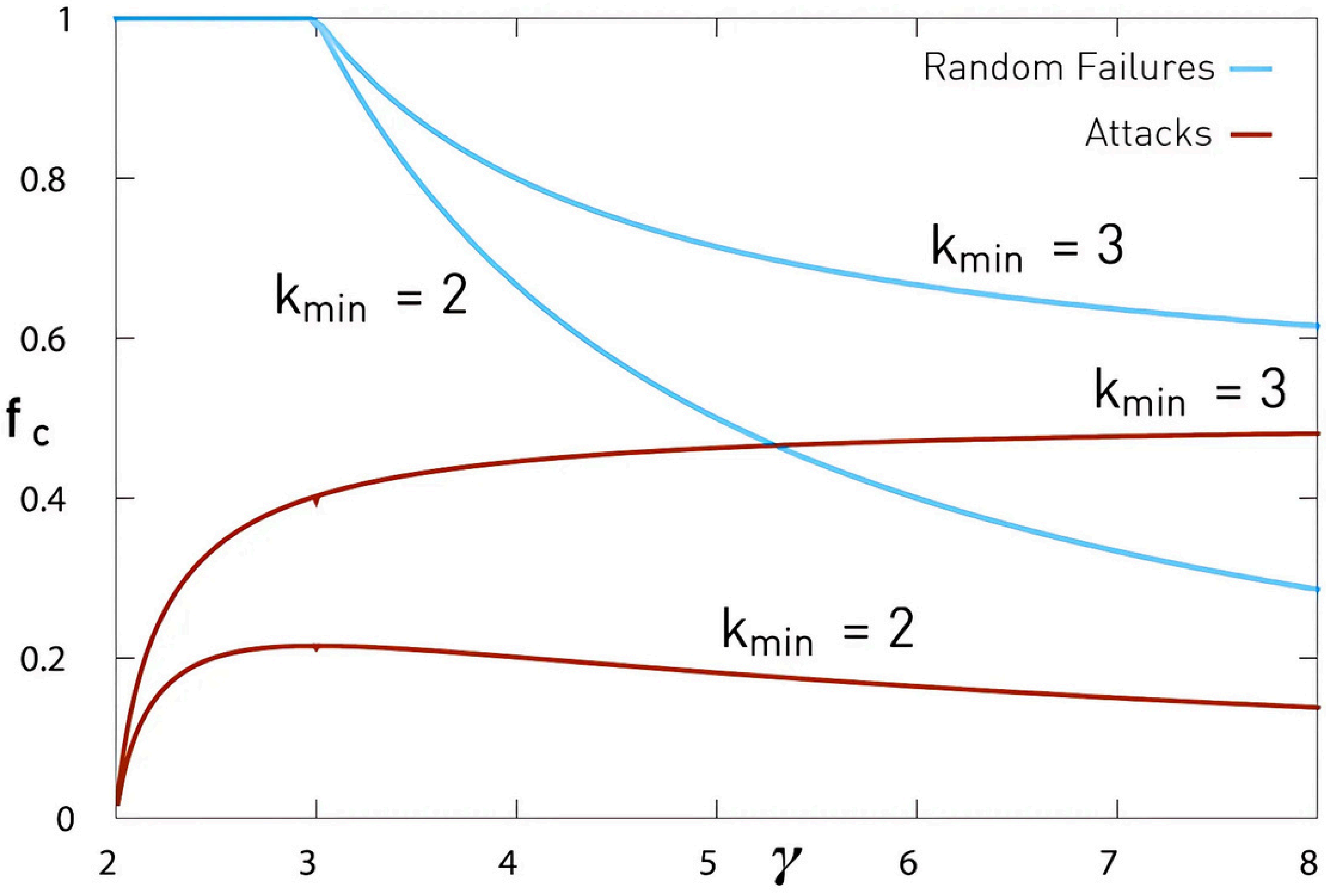


Tolerance to Attacks

We can use Molloy-Reed criterion also for studying attacks, but in this case computations are more complex. The equation that one obtains is

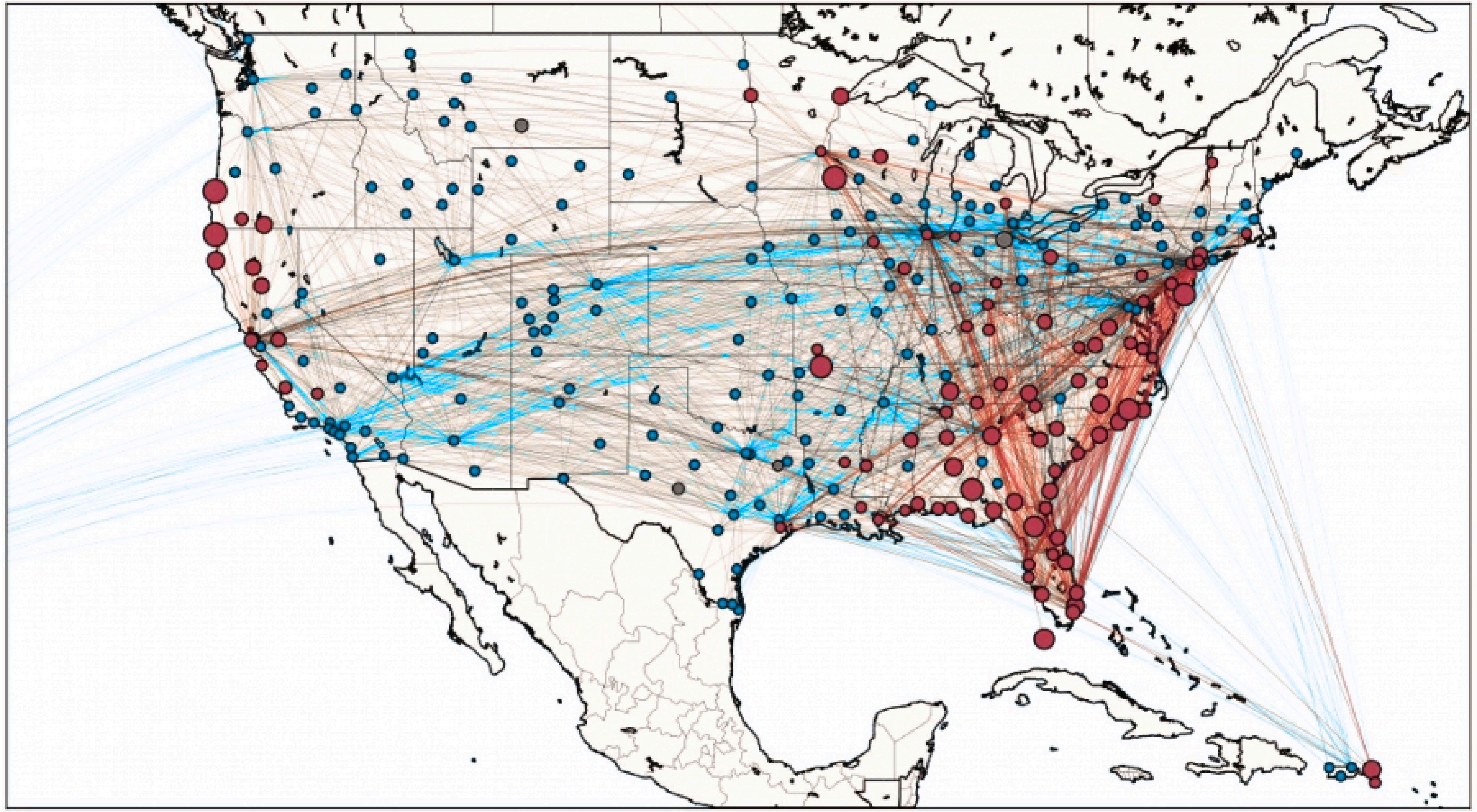
$$f_c^{\frac{2-\gamma}{1-\gamma}} = 2 + \frac{2-\gamma}{3-\gamma} k_{\min} \left(f_c^{\frac{3-\gamma}{1-\gamma}} - 1 \right)$$

As shown in the plot the tolerance to attacks is much smaller than the tolerance to failures, that for γ<3 is maximal. Note that the minimal degree plays a role in this problem.



Cascading Failures

This is only a first approximation to the problem, since in many real case scenario a single failure may cause a cascade of failures in the network



Conclusions

Power Law Probability Distributions

Many real life phenomena are characterized by extreme events described by power law probability distributions

Scale-Free Networks

Real world networks tend to be scale-free, i.e.their degree distribution is a power law. These networks are (ultra) small world, but the clustering goes to zero

Barabasi-Albert Model

Scale-free networks can be generated using a simple microscopic mechanism based on linear preferential attachment

Robustness of Scale-Free Networks

Scale-free networks are more tolerant to failures than random networks, but they tend to be more susceptible to attacks

Quiz

- Can you list some extreme events that are not explained by a Gaussian distribution?
- Do you know the concept of Black Swan?
- What do you think is the largest daily fluctuation ever in the US stock market?
- Do you know any scale free network?
- What are the implausible assumptions of the Barabasi-Albert model?
- Do you think the Barabasi-Albert model captures the dynamics of online social networks?
- What are some examples of cascading processes?