Random models of Networks

Dr. Franck KM

University of Lubumbashi
AIMS-Senegal

franckm@aims.ac.za

July 6, 2018

Model of random network

There are two definitions of a random network:

1. $G(N, L)$ Model: N labeled nodes are connected with L randomly placed links. This model was introduced by Erdos and Rényi and their paper.
2. $G(N, p)$ Model: Each pair of N labeled nodes is connected with probability p, a model introduced by Gilbert.

.

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.

.

While in the $G(N, L)$ model the average degree of a node is simply $\langle k \rangle = 2L/N$, other network characteristics are easier to calculate in the $G(N, p)$ model.

remark

we will be exploring the $G(N, p)$ model, not only for the ease that it allows us to calculate key network characteristics, but also because in real networks the number of links rarely stays fixed.

Model of random network

(a) Paul Erdos

(b) Alfred Renyi

Paul Erdos: 1913-1996

Hungarian mathematician known for both his exceptional scientific output and eccentricity. Indeed, Erds published more papers than any other mathematician in the history of mathematics.

He co-authored papers with over five hundred mathematicians, inspiring the concept of Erds number.

(c) Paul Erdos

(d) Alfred Renyi

Alfred Renyi

Hungarian mathematician with fundamental contributions to combinatorics, graph theory, and number theory. His impact goes beyond mathematics: The Renyi entropy is widely used in chaos theory and the random network theory he co-developed is at the heart of network science.

construction of random network model

To construct a random network we follow these steps:

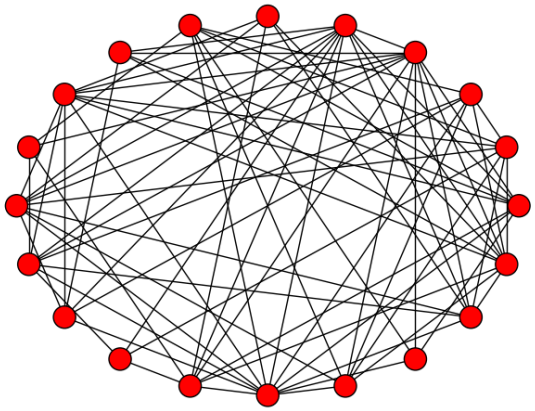
1. Start with N isolated nodes.
2. Select a node pair and generate a random number between 0 and 1. If the number exceeds p,
3. connect the selected node pair with a link, otherwise leave them disconnected.
4. Repeat step (2) for each of the $\frac{N(N-1)}{2}$ node pairs.

Erdos-Renyi network

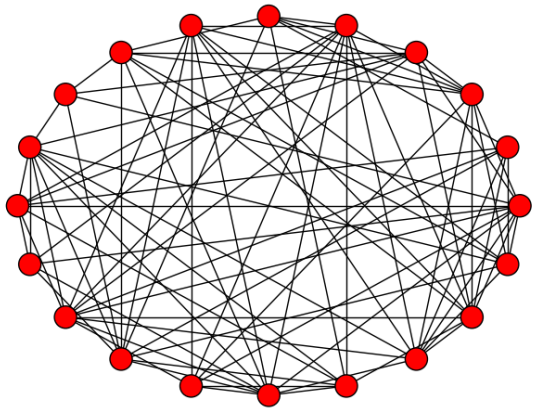
The network obtained after this procedure is called a random graph or a random network. Two mathematicians, Paul Erdos and Alfred Renyi, have played an important role in understanding the properties of these networks. In their honor a random network is called the Erdos-Renyi network.

Random network model: number of links

N=20, p=0.4, L=77



N=20, p=0.4, L=88



Realization

Each random network generated with the same parameters $N = 20$, $p = 0.4$ looks slightly different as show above. Not only the detailed wiring diagram changes between realizations, but so does the number of links L. It is useful, therefore, to determine how many links we expect for a particular realization of a random network with fixed N and p.

Random network: expected number of links

The probability that a particular realization of a random network has exactly L links is given by:

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

1. p^L is the probability that L of the attempts to connect the N(N-1)/2 pairs of nodes are successful.
2. $(1-p)^{\frac{N(N-1)}{2}-L}$ is the probability that the remaining N(N-1)/2 - L attempts have not resulted in a link.
3. $\binom{\frac{N(N-1)}{2}}{L}$ counts the number of different ways we can place L links among $N(N-1)/2$ node pairs.

Random network: expected number of links

Following the binomial distribution the expected number of links is given by

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

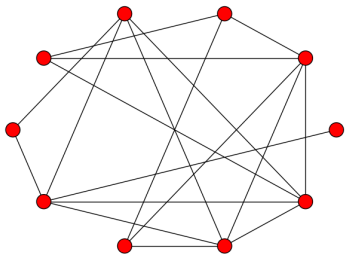
and the mean degree is given by

$$\langle k \rangle = \frac{2L}{\langle N \rangle} = p(N-1) \tag{3}$$

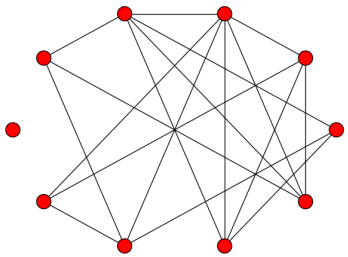
The number of links in a random network varies between realizations. Its expected value is determined by N and p. If we increase p a random network becomes denser: The average number of links increase linearly from $\langle L \rangle = 0$ to $Lmax$ and the average degree of a node increases from $\langle k \rangle = 0$ to $\langle k \rangle = N - 1$.

Random network: expected number of links

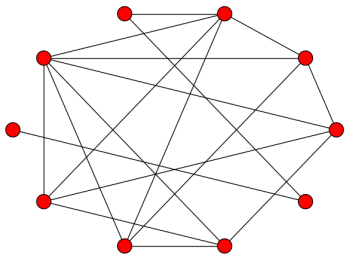
the degree distribution of a random network follows the binomial distribution



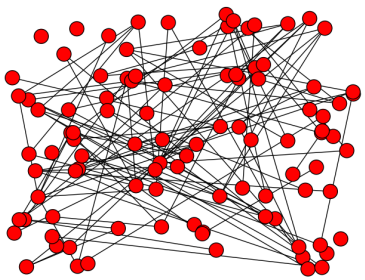
(e)



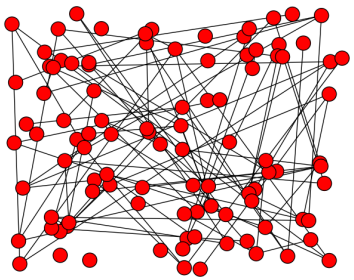
(f)



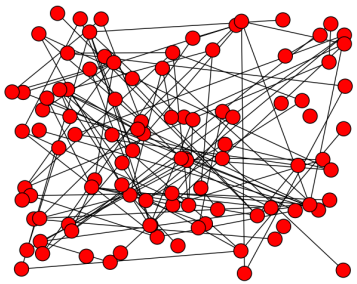
(g)



(h)



(i)



(j)

First row: 3 realizations of a random network generated with the same

Binomial distribution

If we toss a fair coin N times, tails and heads occur with the same probability p $= 1/2$. The binomial distribution provides the probability p_x 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:

$$p_x = \binom{N}{x} p^x (1-p)^{N-x} \tag{4}$$

The mean of the distribution (first moment) is

$$\langle x \rangle = \sum_{x=}^{N} x p_x = Np \tag{5}$$

Its second moment is

$$\langle x^2 \rangle = \sum_{x=0}^{N} x^2 p_x = p(1-p)N + p^2 N^2 \tag{6}$$

providing its standard deviation as

$$\sigma_x = \left\{ \langle x^2 \rangle - \langle x \rangle^2 \right\}^{1/2} = (p(1-p)N)^{1/2} \tag{7}$$

Degree distribution

.

The exact form of the degree distribution of a random network is the binomial distribution. For $N \gg \langle k \rangle$ the binomial is well approximated by a Poisson distribution. As both formulas describe the same distribution, they have the identical properties, but they are expressed in terms of different parameters:

.

The binomial distribution depends on p and N, while the Poisson distribution has only one parameter, $\langle k \rangle$. It is this simplicity that makes the Poisson form preferred in calculations.

Binomial distribution

The degree distribution of a random network follows the binomial distribution given by:

$$p_x = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{8}$$

where,

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

$\diamondsuit$

The shape of this distribution depends on the system size N and the probability p. The binomial distribution allows us to calculate the network's average degree $\langle k \rangle$, as well as its second moment $\langle k^2 \rangle$ and variance σ_k

Poisson distribution

Most real networks are sparse, meaning that for them $\langle k \rangle \ll N$. In this limit the degree distribution is well approximated by the Poisson distribution.

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \tag{9}$$

which is often called, the degree distribution of a random network. The binomial and the Poisson distribution describe the same quantity, hence they have similar properties:

1. Both distributions have a peak around $\langle k \rangle$. If we increase p the network becomes denser, increasing $\langle k \rangle$ and moving the peak to the right.
2. The width of the distribution (dispersion) is also controlled by p or $\langle k \rangle$. The denser the network, the wider is the distribution, hence the larger are the differences in the degrees.

Poisson distribution

When we use the Poisson form (3.8), we need to keep in mind that

1. the exact result for the degree distribution is the binomial form

$$p_x = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{10}$$

2. the expression below

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \tag{11}$$

   represents only an approximation valid in the $\langle k \rangle \ll N$ limit.

3. As most networks of practical importance are sparse, the condition $\langle k \rangle \ll N$ is typically satisfied.

4. The advantage of the Poisson form is that key network characteristics, like $\langle k \rangle$, $\langle k^2 \rangle$ and σ_k, have a much simpler form. depending on a single parameter, $\langle k \rangle$.

Poisson distribution

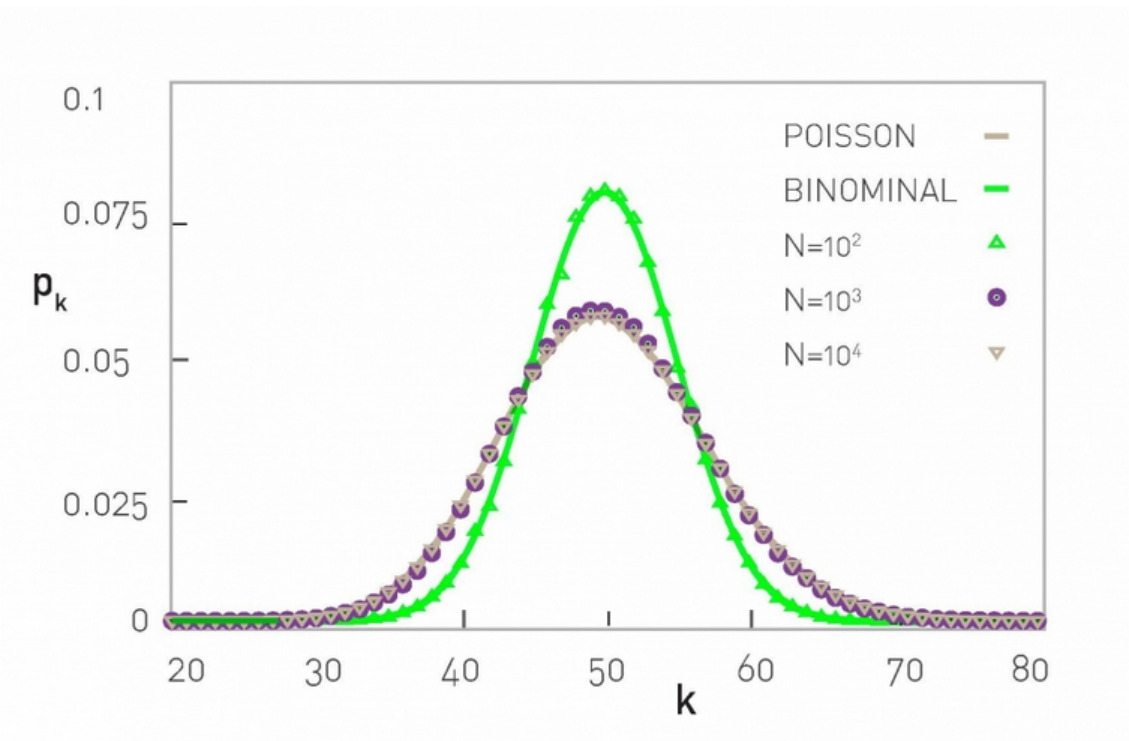
$\partial\partial$

The Poisson distribution does not explicitly depend on the number of nodes N. Therefore, $p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$ predicts that the degree distribution of networks of different sizes but the same average degree $\langle k \rangle$ are indistinguishable from each other.

Degree Distribution is Independent of the Network Size

Small Networks: Binomial

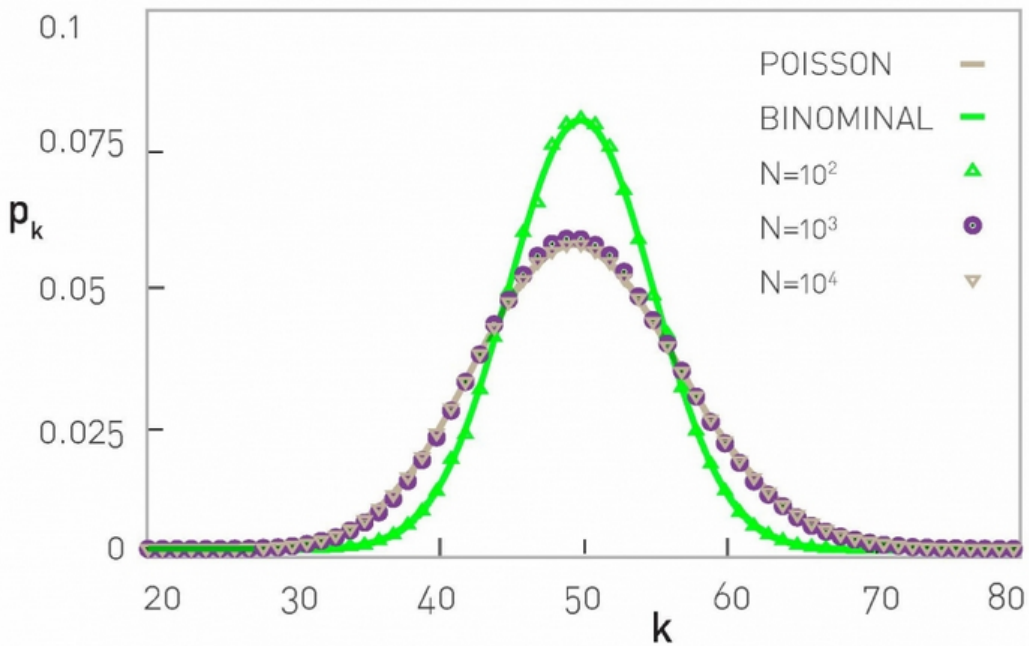
For a small network ($N = 10^2$) the degree distribution deviates significantly from the Poisson form as the condition for the Poisson approximation, $N \gg \langle k \rangle$, is not satisfied. Hence for small networks one needs to use the exact binomial.



Degree Distribution is Independent of the Network Size

Large Networks: Poisson

For larger networks ($N = 10^3, 10^4$) the degree distribution becomes indistinguishable from the Poisson prediction, shown as a continuous grey line. Therefore for large N the degree distribution is independent of the network size. In the figure an average over $1,000$ independently generated random networks to decrease the noise.



Why hubs are missing?

The factor $1/k!$ significantly decreases the chances of observing large degree nodes. Indeed, the Stirling approximation

$$k! \sim (2\pi k)\left(\frac{e}{k}\right)^k \tag{12}$$

we can then rewrite the poisson as:

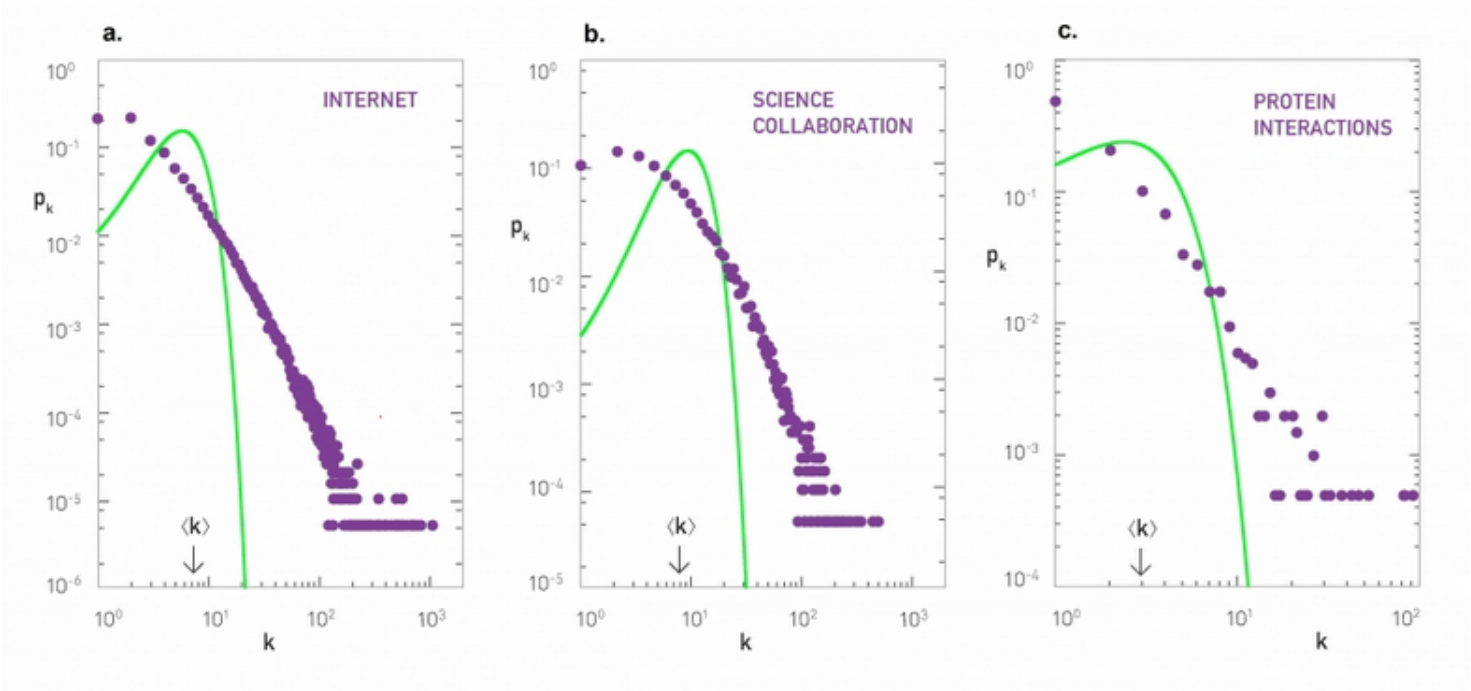
$$p_k = e^{-\langle k\rangle}\frac{\langle k\rangle^k}{k!} = \frac{e^{\langle k\rangle}}{(2\pi k)}\left(\frac{e\,\langle k\rangle}{k}\right)^k \tag{13}$$

For degrees $k > e\,\langle k\rangle$ the term in the parenthesis is smaller than one, hence for large k both k-dependent terms in (13), i.e. $1/\sqrt{k}$ and $(e\,\langle k\rangle\,/k)^k$ decrease rapidly with increasing k. (13) predicts that in a random network the chance of observing a hub decreases faster than exponentially.

Real Networks are Not Poisson

Degree Distribution of Real Networks

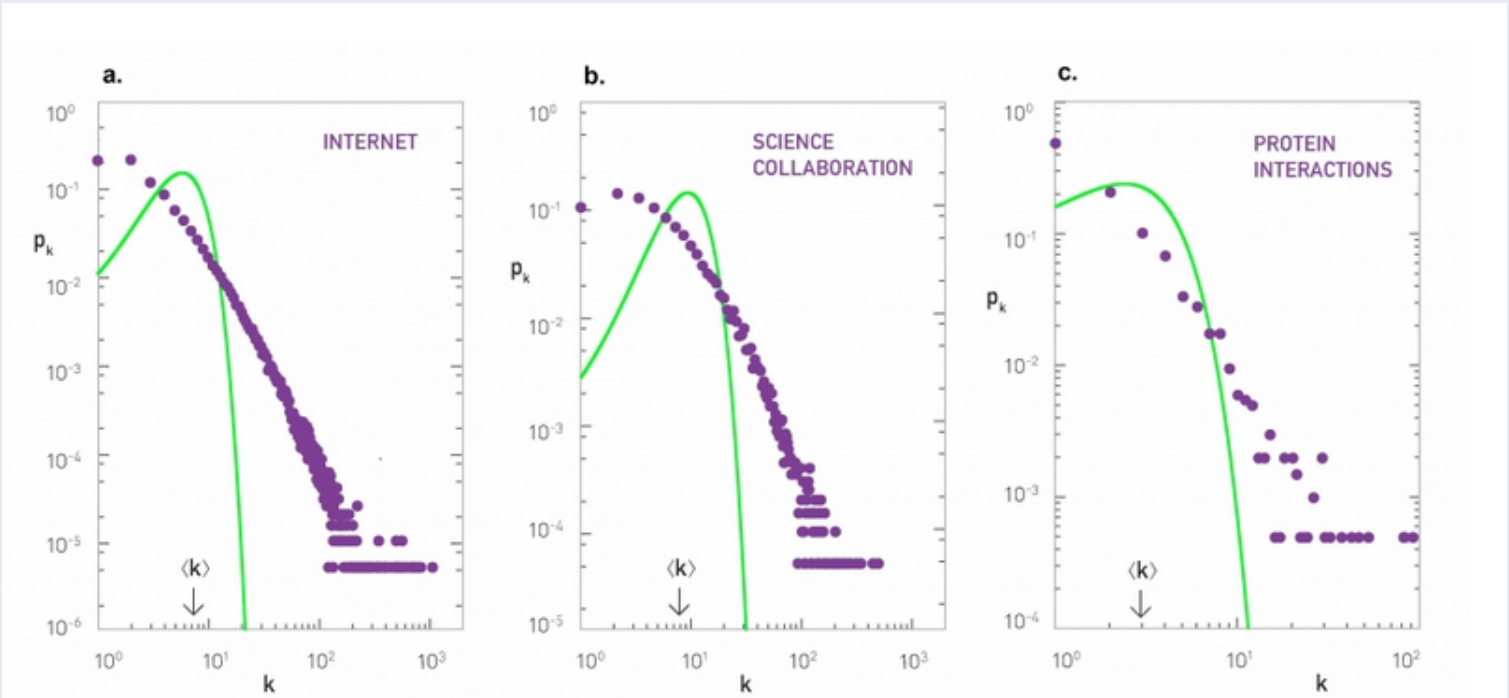
The degree distribution of the (a) Internet, (b) science collaboration network, and (c) protein interaction network. The green line corresponds to the Poisson prediction, obtained by measuring $\langle k \rangle$ for the real network and then plotting $p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$.



Real Networks are Not Poisson

Degree Distribution of Real Networks

The significant deviation between the data and the Poisson fit indicates that the random network model underestimates the size and the frequency of the high degree nodes, as well as the number of low degree nodes. Instead the random network model predicts a larger number of nodes in the vicinity of $\langle k \rangle$ than seen in real networks.



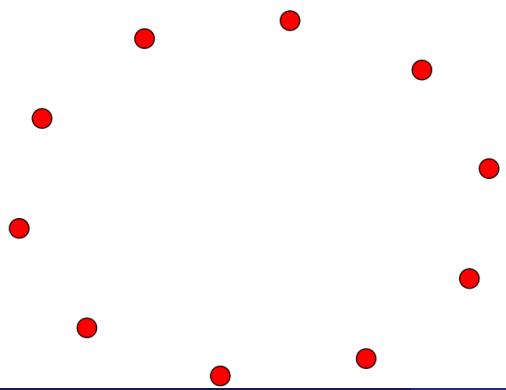
The Evolution of a Random Network

We inspect how the size of the largest connected cluster within the network, N_G, varies with $\langle k \rangle$.

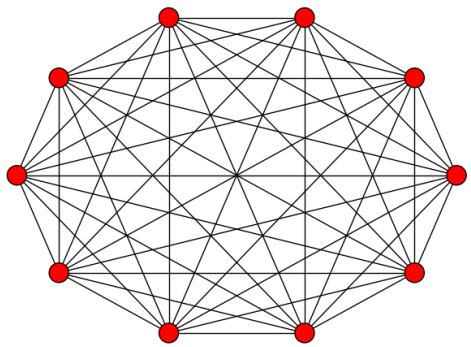
two extreme cases

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 $\langle k \rangle = N - 1$, hence the network is a complete graph and all nodes belong to a single component. Therefore $N_G = N$ and $N_G/N = 1$.

$N = 10, p = 0, \langle k \rangle = 0$
isolated nodes.



$N = 10, p = 1, \langle k \rangle = 9$
complete network.



The Evolution of a Random Network

We can show the following facts:

1. N_G/N remains zero for small $\langle k \rangle$, indicating the lack of a large cluster.
2. Once $\langle k \rangle$ exceeds a critical value, N_G/N increases, signaling the rapid emergence of a large cluster that we call the giant component.

emergence of the giant component

Erdos and Renyi predicted in 1959 that the condition for the emergence of the giant component is

$$\langle k \rangle = 1 \tag{14}$$

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

for a giant component to exist, each of its nodes must be linked to at least one other node. It is somewhat counter intuitive, however, that one link is sufficient for its emergence.

The Evolution of a Random Network

The fact that we need at least one link per node to observe a giant component is not unexpected. Indeed,

$$\langle k \rangle = 1 \tag{15}$$

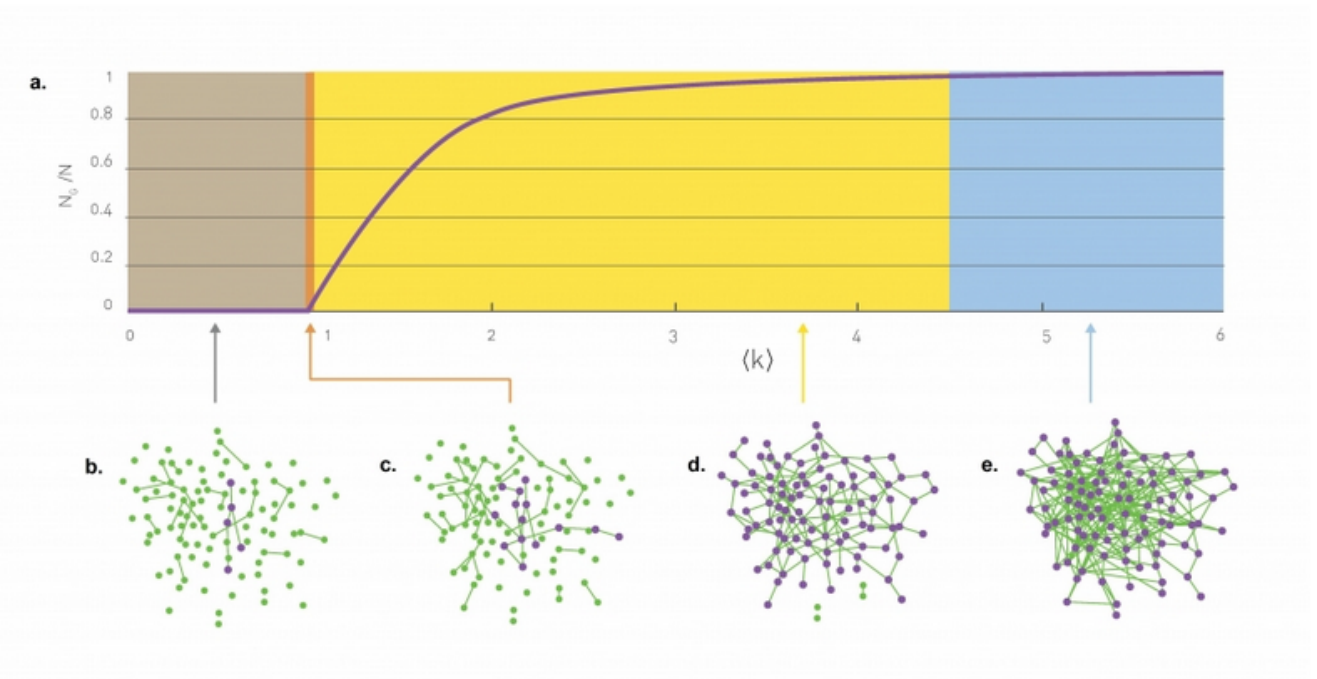
can be expressed in term of p as:

$$p_c = \frac{1}{N-1} \approx \frac{1}{N} \tag{16}$$

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

The Evolution of a Random Network

The emergence of the giant component is only one of the transitions characterizing a random network as we change $\langle k \rangle$. We can distinguish four topologically distinct regimes, each with its unique characteristics:



Subcritical Regime: $0 < \langle k \rangle < 1(p1/N$

For $\langle k \rangle = 0$ the network consists of N isolated nodes. Increasing $\langle k \rangle$ means that we are adding

$$N \langle k \rangle = pN(N-1)/2 \tag{17}$$

links to the network. since, $\langle k \rangle < 1$, we have only a small number of links in this regime, hence we mainly observe tiny clusters.

1. We can designate at any moment the largest cluster to be the giant component.
2. In this regime the relative size of the largest cluster, NG/N, remains zero.
3. The reason is that for $\langle k \rangle < 1$ the largest cluster is a tree with size $N_G \sim \ln N$, hence its size increases much slower than the size of the network. Therefore $N_G/N \simeq \ln N/N \to 0$ in the $N \to \infty$ limit.

Subcritical Regime: $0 < \langle k \rangle < 1(p1/N)$

Subcritical Regime

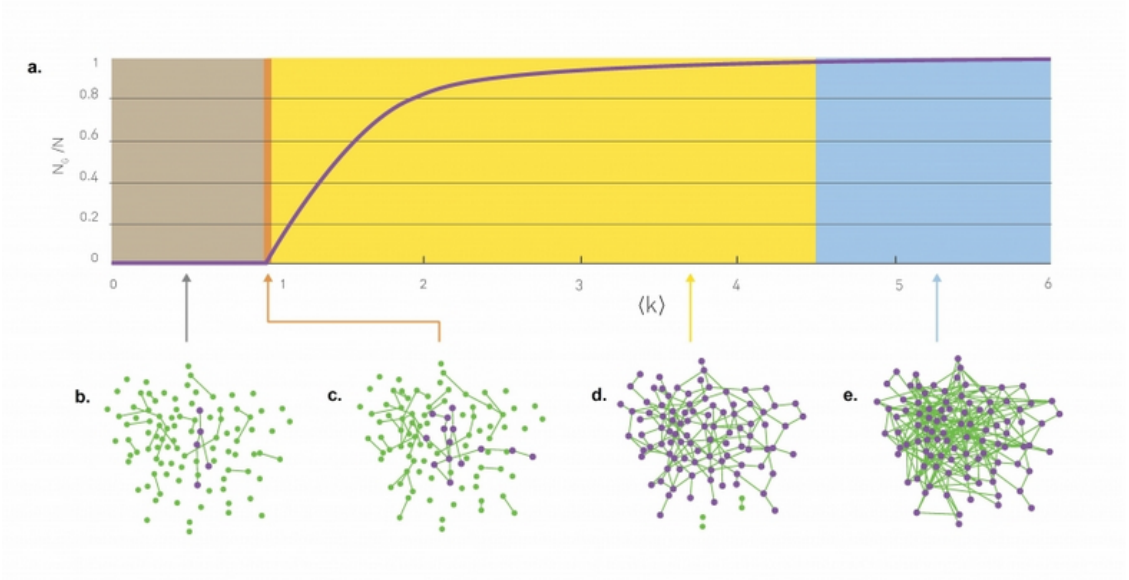
In summary, in the subcritical regime the network consists of numerous tiny components, whose size follows an exponential distribution. Hence these components have comparable sizes, lacking a clear winner that we could designate as a giant component.

Critical Point: $\langle k \rangle = 1$ ($p = 1/N$)

Critical Point

1. The critical point separates the regime where there is not yet a giant component ($\langle k \rangle < 1$) from the regime where there is one ($\langle k \rangle > 1$).
2. At this point the relative size of the largest component is still zero.
3. Indeed, the size of the largest component is $N_G \sim N^{2/3}$.
4. Consequently N_G grows much slower than the network's size, so its relative size decreases as $N_G/N \sim N^{-1/3}$ in the $N \to \infty$ limit.

Evolution of a Random Network



Evolution of a Random Network

1. The relative size of the giant component in function of the average degree $\langle k \rangle$ in the Erdos-Renyi model.
2. The figure illustrates the phase tranisition at $\langle k \rangle = 1$, responsible for the emergence of a giant component with nonzero N_G.
3. A sample network and its properties in the four regimes that characterize a random network.

Critical Point: $\langle k \rangle = 1$ ($p = 1/N$)

Critical point

1. There is a significant jump in the size of the largest component at $\langle k \rangle = 1$.
2. For example, for a random network with $N = 7 \times 10^9$ nodes, comparable to the globe's social network,
3. for for $\langle k \rangle < 1$ the largest cluster is of the order of $N_G \simeq \ln N = \ln(7 \times 10^9) \simeq 22.7$.
4. In contrast at $\langle k \rangle = 1$ we expect $N_G \sim N^{2/3} = (7 \times 10^9)^{2/3} \simeq 3 \times 10^6$, a jump of about five orders of magnitude.
5. Yet, both in the subcritical regime and at the critical point the largest component contains only a vanishing fraction of the total number of nodes in the network.

Critical Point: $\langle k \rangle = 1$ ($p = 1/N$)

In summary, at the critical point most nodes are located in numerous small components, whose size distribution follows a power law form. The power law form indicates that components of rather different sizes coexist. These numerous small components are mainly trees, while the giant component may contain loops. Note that many properties of the network at the critical point resemble the properties of a physical system undergoing a phase transition.

Supercritical Regime: $\langle k \rangle > 1 (p > 1/N)$

$\langle k \rangle > 1 (p > 1/N$

This regime has the most relevance to real systems, as for the first time we have a giant component that looks like a network. In the vicinity of the critical point the size of the giant component varies as

$$\frac{N_G}{N} \sim \langle k \rangle - 1 \tag{18}$$

$$N_G \sim (p - p_c)N \tag{19}$$

where $p_c = \frac{1}{N-1} \approx \frac{1}{N}$. 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. Note that (18) is valid only in the vicinity of $\langle k \rangle = 1$. For large $\langle k \rangle$ the dependence between N_G and $\langle k \rangle$ is nonlinear.

Connected Regime: $\langle k \rangle > \ln N (p > \ln N/N)$

For sufficiently large p the giant component absorbs all nodes and components, hence $N_G \cong N$. In the absence of isolated nodes the network becomes connected. The average degree at which this happens depends on N as $\langle k \rangle = \ln N$

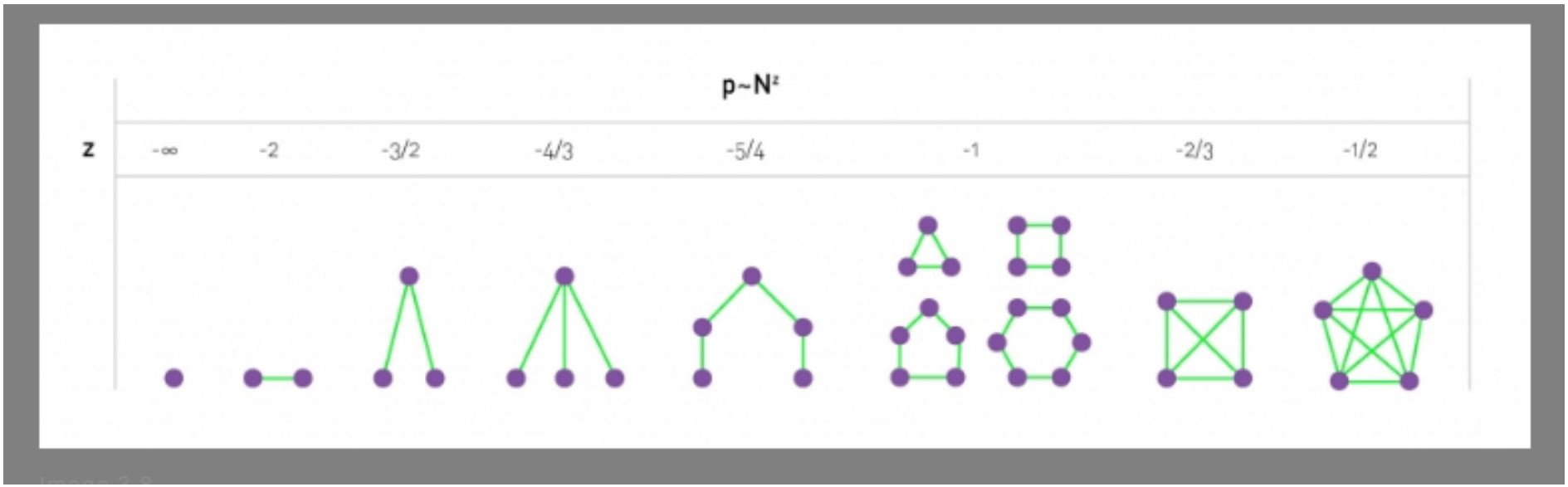
Note that when we enter the connected regime the network is still relatively sparse, as $\ln N/N \to 0$ for large N. The network turns into a complete graph only at $\langle k \rangle = N - 1$.

summary

The random network model predicts that the emergence of a network is not a smooth, gradual process: The isolated nodes and tiny components observed for small $\langle k \rangle$ collapse into a giant component through a phase transition. As we vary $\langle k \rangle$ we encounter four topologically distinct regime.

Network Evolution in Graph Theory

It is often assumed that the connection probability $p(N)$ scales as N^z, where z is a tunable parameter between $-\infty$ and 0.



Network Evolution in Graph Theory

it is often assumed that the connection probability $p(N)$ scales as N^z, where z is a tunable parameter between $-\infty$ and 0.

$\wedge$

The threshold probabilities at which different subgraphs appear in a random graph, as defined by the exponent z in the $p(N) \sim N^z$ relationship. For$z < -3/2$ the graph consists of isolated nodes and edges. When z passes $-3/2$ trees of order 3 appear, while at $z = -4/3$ trees of order 4 appear. At $z = 1$ trees of all orders are present, together with cycles of all orders. Complete subgraphs of order 4 appear at $z = -2/3$, and as z increases further, complete subgraphs of larger and larger order emerge.

References

Watts D. J. 2000 (2000)

Small world: The Dynamic of Networks between Order and Randomness

The American Mathematical Monthly, 107, 664–668.

Watts, D. J. and Strogatz, S. H.(1998)

Collective dynamic of small-world networks

Nature,393,440–442.

Barabási, A.-L. and Albert, R. (1999)

Statistical mechanics of complex networks

Science, 286, 509–512.

GPrettejohn, B. J. and Berryman, M. J. and McDonnell, M. D. (2011)

Methods for generating complex networks with selected structural properties for simulations: A review and tutorial for neuroscientists

Front. Comput. Neurosci.,5

Fronczak, A. and Fronczak, P. and Holyst, J. A. (2004)

Average path length in random networks.

Phys. Rev. E.,70, 056110.

The End