

Lecture 3: Small worlds, scale-free networks, generating random networks of arbitrary degrees

Prof. Sotiris Nikolettseas

University of Patras and CTI

ΥΔΑ ΜΔΕ, Patras
2020 - 2021

A. Small Worlds (I)

- The small world phenomenon: if you choose any two individuals anywhere on Earth, you will find a path of at most six acquaintances between them.
- In other words, surprisingly, even individuals on opposite sides of the globe can be connected to each other via a few acquaintances.
- This phenomenon is also known as “six degrees of separation”.

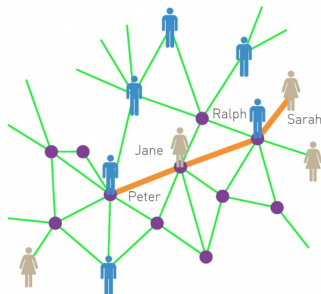


Figure: *Six degrees of separation*

Small Worlds (II)

- In formal terms, this phenomenon implies that the distance between two randomly chosen nodes in a network is short.
- But what does short mean? And how can we explain this phenomenon?
- Consider a random network with average degree $\langle K \rangle$. Then, a node has on average $\langle K \rangle$ nodes at distance $d = 1$, $\langle K \rangle^2$ nodes at distance $d = 2$, and so on, and $\langle K \rangle^d$ nodes at distance d .

Summing up, the expected number of nodes at distance d is:

$$N(d) \simeq \frac{\langle K \rangle^{d+1} - 1}{\langle K \rangle - 1}$$

- Solving for $\langle K \rangle^{d_{max}} \simeq N$ yields

$$d_{max} \simeq \frac{\ln N}{\ln \langle K \rangle}$$

for the diameter of a random network.

Small Worlds (III)

- As a matter of fact, for most networks the above formula offers a better approximation to the average distance between two randomly chosen nodes $\langle d \rangle$, rather than to d_{max} (because d_{max} is often dominated by a few extreme paths, while $\langle d \rangle$ is the average over all node pairs, a process that suppresses the fluctuations).
- Then the usual definition of the small world property is:

$$\langle d \rangle \simeq \frac{\ln N}{\ln \langle K \rangle}$$

where $\langle d \rangle$ is the average internode distance.

Small Worlds (IV)

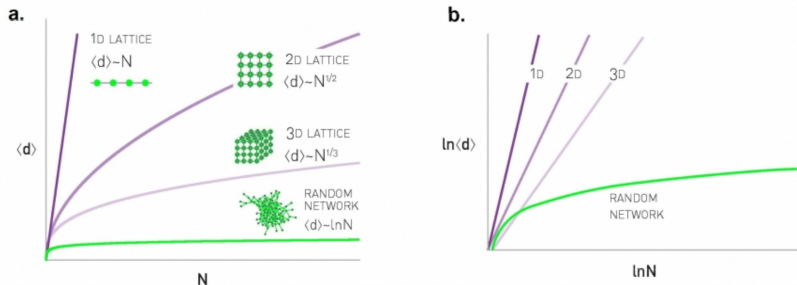
Network	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{\max}	$\ln N / \ln \langle k \rangle$
Internet	192,244	609,066	6.34	6.98	26	6.58
WWW	325,729	1,497,134	4.60	11.27	93	8.31
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile-Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	93,437	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

Figure: *Six Degrees of Separation table*

- The last column shows that the formula achieves in most cases a reasonable approximation to the measured distance $\langle d \rangle$.
- Yet the agreement is not perfect and we will see how to adjust it for many real networks.

Small Worlds (V)

- This formula basically shows that by small world we basically mean that the average path length depends logarithmically on the network size; it is proportional to $\ln N$, rather than N or some power of N .
- Also, the denser the network (large $\langle K \rangle$), the smaller the distance is.



Note: In lattices, the above branching argument does not work, so distances are bigger (a power of N).

The diameter of the WWW

- In 1999, Albert, Jeong and Barabasi suggested that the diameter of the Web is:

$$\langle d \rangle \simeq 0.35 + 0.89 \ln N,$$

where N the number of WWW nodes. At that time, that yielded $\langle d \rangle \simeq 18.69$, in other words 19 clicks sufficed to reach a randomly chosen WWW node (19 degrees of separation). In 2016, this increased to $\langle d \rangle \simeq 25$, in view of the dynamic expansion of the Web.

The first empirical study of the small world property

- In 1967 social psychologist Stanley Milgram designed an experiment to measure distances in social networks of acquaintances. A target person was chosen at random in Boston. A large enough ($N = 64$) number of randomly selected persons in Omaha, Nebraska were asked to send a letter either to the target person (if they knew him), or to a personal acquaintance more likely to know the target.
- Eventually, 64 of the 296 letters made it, with an average number of 5.2 social links (forwarding the letter) needed; thus, the "six degrees of separation" term.
- Facebook in 2011 reported an average of 4.74 links among its 721 million users (connected by 68 billion friendship links at that time).

Six degrees of separation

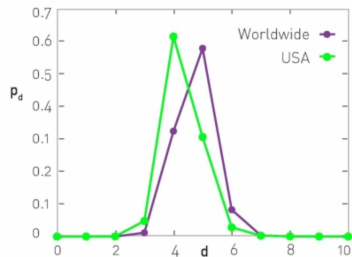
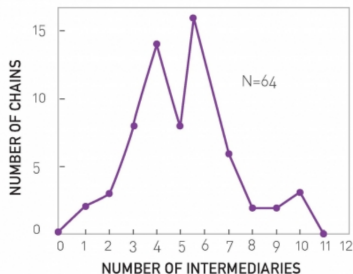


Figure: *Six Degrees? From Milgram to Facebook*

The Watts-Strogatz small world model (I)

- In 1998, Watts and Strogatz proposed an extension of the random network model motivated by two observations:
 - small world property: in both real and random networks, average node distance is logarithmic on N , rather than polynomial, as in regular lattices.
 - high clustering: in real networks the average clustering coefficient is much higher than in random networks.
- Their model (called the small-world model) interpolates between a regular lattice (which has high clustering but lacks small-world property) and a random network (which is small-world but has low clustering).

The Watts-Strogatz small world model (II)

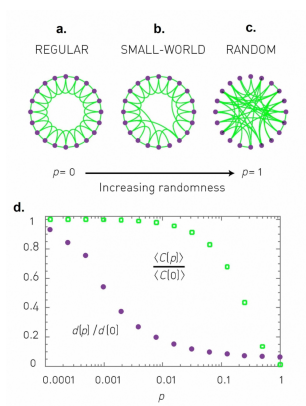


Figure: *The Watts-Strogatz Model*

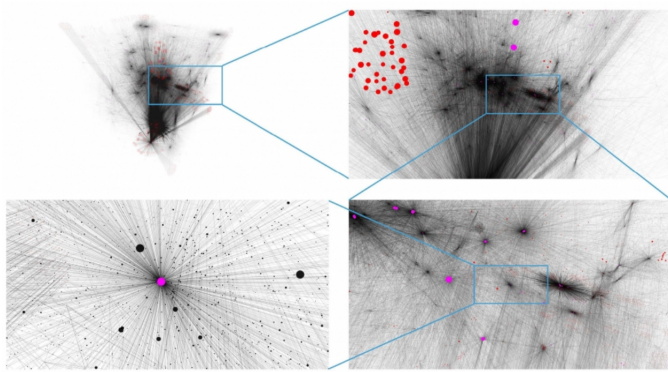
The Watts-Strogatz small world model (III)

- we start from a ring of nodes, each node connected to their immediate and next neighbors (a regular lattice), so the average clustering coefficient is $\langle c \rangle = 1/2$ (quite high).
- with probability p , each link is rewired to a randomly chosen node. For small p , the clustering remains high, but the random long-range links can drastically decrease the distance between the nodes.
- for the extreme $p = 1$, all links have been rewired, so the network turns into a random one.
- we remark a rapid drop in $d(p)$ with p , leading to the emergence of the small-world property; however, during this drop, clustering $\langle C(p) \rangle$ remains high, as desired. Overall, when $0.001 < p < 0.1$ there is both small world and high clustering! So, only little randomness suffices!

B. Deeper into the scale-free property (I)

- The WWW is a network whose nodes are documents and whose links are the URLs allowing us to move with a click from one web document to another. Its estimated size exceeds 1 trillion documents ($N \simeq 10^{12}$).
- The first "map" of the WWW was obtained in 1998 by Hawoong Jeong; he mapped the nt.edu domain (University of Notre Dame, Canada) of 300.000 documents and 1.5 million links.
- The purpose of the map was to compare the Web graph to the random network model; at that time, people believed that WWW could be well approximated by a random network (since each document reflects personal/professional interests of its creator, the links to documents might point to randomly chosen documents).

Deeper into the scale-free property (II)



- Nodes with > 50 links are shown in red, nodes with > 500 links in purple.
- The map reveals a few highly connected nodes ("hubs"), which in a random network are effectively forbidden!
- Actually, such hubs are not unique in the WWW, but appear in most real networks. They represent a deeper organizing principle, which we call the scale-free property.

Power Laws and Scale-free Networks

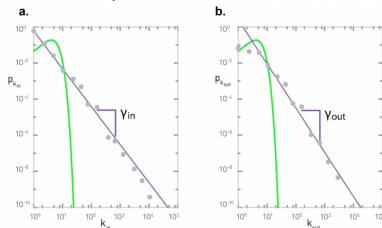
- If the WWW were to be a random net, its degrees would follow a Poisson distribution. However, it actually follows a power law distribution:

$$p_K \sim K^{-\gamma}$$

(p_K the probability that a random node has degree K , and γ is a constant degree exponent). Thus

$$\ln p_K \sim -\gamma \ln K$$

and on a log-log scale the data points form a straight line of slope γ (i.e. $\ln p_K$ depends linearly on $\ln K$).



- Since WWW is directed, we have two distributions (with corresponding exponent $\gamma_{in}, \gamma_{out}$). Also, the green line shows the Poisson distribution).

The 80/20 Rule

- A similar phenomenon was identified by the economist Vilfredo Pareto in the 19th century; he noticed that a few wealthy individuals earned most of the money, while the majority of people earned small amounts; roughly 80% of all money is earned by only 20% of the population.
- The 80/20 rule emerges in many areas:
 - 80% of profits are produced by only 20% of employees
 - 80% of citations go to only 38% of scientists
 - 80% of links in Hollywood are connected to only 30% of actors
- This 80/20 phenomenon identified by Pareto is actually the first known report of a power-law distribution.

The scale-free property

- The above empirical results for the WWW demonstrate the existence of networks whose degree distribution is quite different from the Poisson distribution characterizing random networks. We will call such networks scale-free networks.
- Definition: A scale-free network is a network whose degree distribution follows a power law.

Discrete formalism

- For $K = 0, 1, 2, \dots$ the probability p_K that a node has exactly K links is:

$$p_K = CK^{-\gamma}$$

- The constant C is determined by the normalization condition

$$\sum_{K=1}^{\infty} p_K = 1$$

which yields

$$C = \frac{1}{\sum_{K=1}^{\infty} K^{-\gamma}} = \frac{1}{\zeta(\gamma)}$$

where $\zeta(\gamma)$ is the Riemann-zeta function. Thus, the power law distribution is:

$$p_K = \frac{K^{-\gamma}}{\zeta(\gamma)}$$

(for simplicity we omitted the case $k = 0$ for which the formula diverges).

- In analytic calculations we often assume that degrees can have any positive **real** value. In this case, the power law becomes: $p(K) = CK^{-\gamma}$
Using the normalization condition:

$$\int_{K_{min}}^{\infty} p(K) dK = 1$$

we get

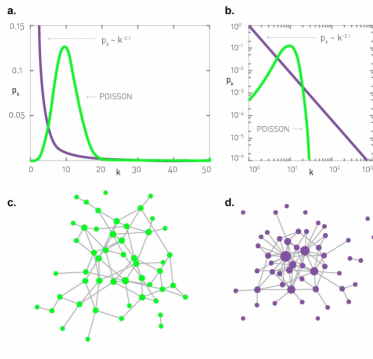
$$C = \frac{1}{\int_{K_{min}}^{\infty} K^{-\gamma} dK} = (\gamma - 1) K_{min}^{\gamma-1}$$

and finally

$$p(K) = (\gamma - 1) K_{min}^{\gamma-1} K^{-\gamma}$$

where K_{min} the smallest degree for which the power law holds. Obviously, the meaning of discrete p_K formalism (the probability that a node has exactly k links) does not make sense. Instead, only the integral of $p(K)$ has a physical meaning. $\int_{K_1}^{K_2} p(K) dK$ is the probability that a random node has degree between K_1 and K_2 .

Hubs in scale-free networks



In the above figure:

- (a) linear plot, $\langle K \rangle = 11$
- (b) same curves as in (a), but on a log-log plot
- (c) a random network with $\langle K \rangle = 3$, $N = 50 \Rightarrow$ most nodes similar degree $k \simeq \langle K \rangle$
- (d) a scale-free network with $\langle K \rangle = 3$, $N = 50 \Rightarrow$ few hubs and numerous small degree nodes

The tails of the degree distribution

The main difference between random and scale-free networks comes in the tails of the degree distribution:

- for small K , a scale-free net has a large number of small-degree nodes, most of which are absent in random networks
- for K around the mean degree $\langle K \rangle$ there is an excess of nodes with degree $K \simeq \langle K \rangle$ in random networks.
- for large K , the probability of high-degree nodes (hubs) in scale-free networks is several orders of magnitude higher than in random networks.

The Largest Hub (I)

- All real networks are finite, even when they are huge, such as in the case of the WWW or social networks ($N \simeq 7 * 10^9$ nodes). Other networks are relatively small, such as the genetic network in a human cell (around 20,000 genes).
- Natural question: how does the network size affect the size of its hubs? To answer this, we calculate the **maximum degree** K_{max} of the degree distribution p_K . It represents the expected size of the largest hub in the network.

The Largest Hub (II)

- To simplify calculations, let us start with the exponential distribution

$$p(K) = Ce^{-\lambda K}$$

For a network with minimum degree K_{min} we get:

$$\int_{K_{min}}^{\infty} p(K) dK = 1$$

which yields $C = \lambda e^{\lambda K_{min}}$

- To calculate K_{max} we assume that in a network of N nodes we expect at most one node in the (K_{max}, ∞) range. So, for the probability of having a node of degree $\geq K_{max}$ it is

$$N \int_{K_{max}}^{\infty} p(K) dK = 1$$

which yields $K_{max} = K_{min} + \frac{\ln N}{\lambda}$

- As $\ln N$ is a very slow function of the network size N , the maximum degree is not significantly different than the minimum degree.
- For a Poisson distribution, things are quite similar (actually the dependence of K_{max} on N is even slower).

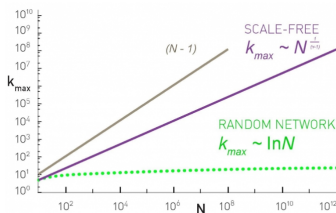
The Largest Hub (III)

- In contrast, for a scale-free network, it is:

$$K_{max} = K_{min} N^{\frac{1}{\gamma-1}}$$

i.e. the dependence of K_{max} on N is polynomial, thus the biggest hub can have size orders of magnitude larger than the smallest node K_{min} , and also, the larger the network size N , the larger the degree of its biggest hub.

- As an example, in the WWW sample, an exponential distribution would imply $K_{max} \simeq 14$, while assuming a scale-free property would imply $K_{max} \simeq 95,000$



About the term “scale-free”

- To better understand this term, we need to address the moments of the degree distribution. The n -th moment is the mean of the n -th power of the degree random variable:

$$\langle K^n \rangle = \sum_{K_{min}}^{\infty} K^n p_K \simeq \int_{K_{min}}^{\infty} K^n p(K) dK$$

- In particular, the first moments are of special importance:
 - $n = 1$: the first moment is the average degree
 - $n = 2$: the second moment is related to the variance as follows:

$$\sigma_K^2 = \langle K^2 \rangle - \langle K \rangle^2$$

where σ_K (the square root of the variance) is the standard deviation.

- $n = 3$: the third moment determines skewness, telling us how symmetric p_K is around the mean $\langle K \rangle$.

About the term “scale-free” (II)

- For a scale-free network, the n -th moment is:

$$\langle K^n \rangle = \int_{K_{min}}^{K_{max}} K^n p(K) dK = C \frac{K_{max}^{n-\gamma+1} - K_{min}^{n-\gamma+1}}{n - \gamma + 1}$$

- Since K_{min} is typically fixed, the degree of the longest hub, K_{max} , increases with the network size. For large networks, we thus need take $K_{max} \rightarrow \infty$ and the moment $\langle K^n \rangle$ depends on the interplay of n and γ .
 - if $n - \gamma + 1 > 0$ then $\langle K^n \rangle$ goes to infinity as $K_{max} \rightarrow \infty$, therefore all moments larger than $\gamma - 1$ diverge.
 - if $n - \gamma + 1 \leq 1$ then $\langle K^n \rangle$ goes to zero, therefore all moments $n \leq \gamma - 1$ are finite.
- But for many scale-free networks, the degree exponent γ is between 2 and 3, therefore the first moment $\langle K \rangle$ is finite, but the second and higher moments $\langle K^2 \rangle$ and $\langle K^3 \rangle$ go to infinity as $N \rightarrow \infty$

About the term “scale-free” (III)

- This divergence of the higher moments helps us understand the origin of the “scale-free” term. Indeed, the degrees in the normal distribution (and thus in a great variety of distributions) concentrate in a range

$$K = \langle K \rangle \pm \sigma_K$$

- In random networks with Poisson degrees $\sigma_K^2 = \langle K \rangle$ thus $\sigma_K = \langle K \rangle^{1/2}$, which is much smaller than $\langle K \rangle$, hence the degrees lie in the range $K = \langle K \rangle \pm \langle K \rangle^{1/2}$. In other words, nodes in random networks have comparable degree close to the average degree $\langle K \rangle$, thus the average degree $\langle K \rangle$ serves as the “scale” of the network.
- In contrast, scale-free networks lack a scale, since the first moment of degree K is finite, yet the second moment is infinite, thus the fluctuation of degrees around their average can be arbitrarily large. Hence networks with $\gamma < 3$ do not have a meaningful internal scale, so we can call them “scale-free”.

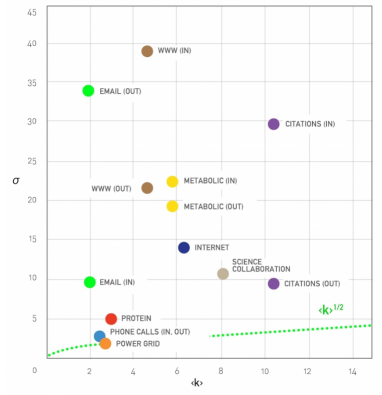
About the term “scale-free” (IV)

- Strictly speaking, $\langle K^2 \rangle$ diverges only in the $N \rightarrow \infty$ limit. Yet the divergence is relevant for finite networks as well.

Network	N	L	$\langle K \rangle$	$\langle K_{in}^2 \rangle$	$\langle K_{out}^2 \rangle$	$\langle K^2 \rangle$	Y_{in}	Y_{out}	Y
Internet	192,244	609,066	6.34	-	-	240.1	-	-	3.42*
WWW	325,729	1,497,134	4.60	1546.0	482.4	-	2.00	2.31	-
Power Grid	4,941	6,594	2.67	-	-	10.3	-	-	Exp.
Mobile-Phone Calls	36,595	91,826	2.51	12.0	11.7	-	4.69*	5.01*	-
Email	57,194	103,731	1.81	94.7	1163.9	-	3.43*	2.03*	-
Science Collaboration	23,133	93,437	8.08	-	-	178.2	-	-	3.35*
Actor Network	702,388	29,397,908	83.71	-	-	47,353.7	-	-	2.12*
Citation Network	449,673	4,689,479	10.43	971.5	198.8	-	3.03*	4.00*	-
E. Coli Metabolism	1,039	5,802	5.58	535.7	396.7	-	2.43*	2.90*	-
Protein Interactions	2,018	2,930	2.90	-	-	32.3	-	-	2.89*-

- For most of these real networks, σ is significantly larger than $\langle K \rangle$, thus allowing large variations in node degrees. The only exceptions are the power grid (which is not scale-free) and the phone-calls net (which is scale-free but has a large γ , so it can be well approximated by a random network).

About the term “scale free” (V)



- The standard deviations of these real networks are also depicted in the above figure. The green line corresponds to $\sigma_K = \langle K \rangle^{1/2}$ (the standard deviation of a random network)
- For all networks (except power grid and phone-calls) the standard deviation is much larger than what it should be in a random network.

C. Ultra-Small-World property

- Important question: do hubs affect the small-world property? The answer is yes; distances in scale-free networks are smaller than the distances observed in an equivalent random network.
- The dependence of average distance $\langle d \rangle$ on the system size N and the degree exponent γ is:

$$\langle d \rangle \sim \left\{ \begin{array}{l} \bullet \text{ constant, } \gamma = 2 \\ \bullet \ln \ln N, \quad 2 < \gamma < 3 \\ \bullet \frac{\ln N}{\ln \ln N}, \quad \gamma = 3 \\ \bullet \ln N, \quad \gamma > 3 \end{array} \right.$$

i.e. there are four scaling regimes which correspond to a characteristic impact to the average path length.

Four regimes

We now discuss each one out of the four regimes.

- Anomalous regime ($\gamma = 2$). According to the formula

$$K_{max} = K_{min} \cdot N^{\frac{1}{\gamma-1}}$$

for $\gamma = 2$ the size of the biggest hub is linear in N , almost all nodes are connected to the same central hub, thus they are very close to each other and the path lengths do not depend on N .

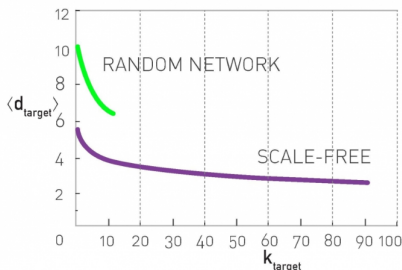
- Ultra-small-world ($2 < \gamma < 3$): The average path length increases as $\ln \ln N$, which is significantly smaller than $\ln N$ in random networks. This is due to the hubs that radically reduce path lengths by connecting with each other a large number of small degree nodes. As an example, for the world's social network ($N \simeq 7 * 10^9$), the random network model would give $\ln N = 22.66$ while the fact that it is scale-free gives the actual average path length of $\ln \ln N = 3.12$ only.

Four regimes (II)

- Critical point ($\gamma = 3$). The second moment of degrees does not diverge any longer and the $\ln N$ dependence of random networks returns; however, a double logarithmic correction $\ln \ln N$ occurs, shrinking the distances compared to random nets.
- Small world ($\gamma > 3$). In this regime, $\langle K^2 \rangle$ is finite and the average path length exhibits similar small-world properties as for random networks. This is because, although hubs continue to be present, for $\gamma > 3$ their size and number do not suffice to drastically reduce path lengths.

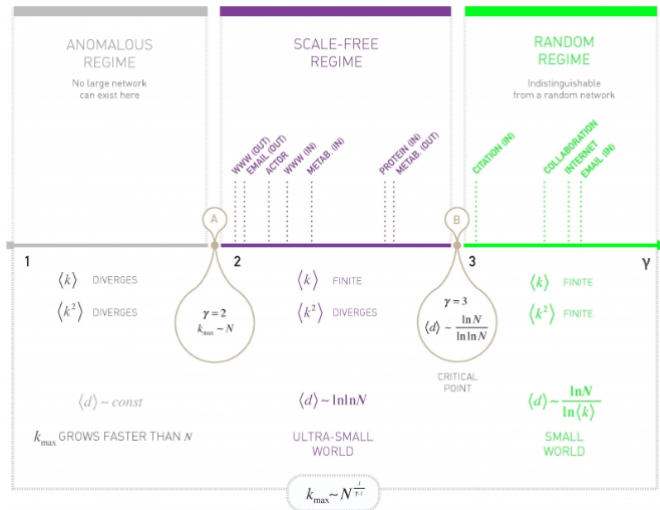
We are always close to the hubs

- F. Kavanthy (1929) claimed, counterintuitively, that "it is always easier to find someone who knows a famous person than some insignificant person". This is particularly the case in scale-free networks:



- The figure shows the distance ($\langle d_{target} \rangle$) of a node with degree $K \simeq \langle K \rangle$ from a target node with degree K_{target} . We remark that:
 - in scale-free nets we are closer to the hubs of high degree
 - path lengths are visibly longer in random networks

The role of the degree exponent



The role of the degree exponent (II)

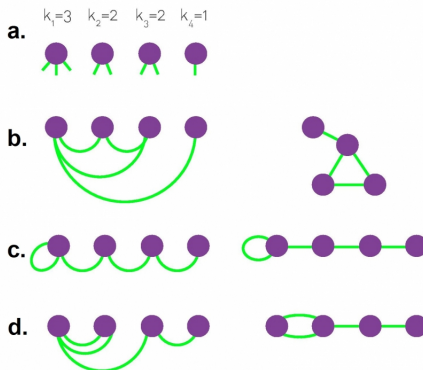
- Note that when $\gamma < 2$ then $K_{max} = K_{min} N^{\frac{1}{\gamma-1}}$ would lead to K_{max} bigger than N ! This is not possible without self-loops / multiple links, so such degree distributions do not correspond to real networks! In other words, there exist no scale-free networks for $\gamma < 2$ (anomalous regime).
- for $\gamma > 3$, the probability $p_K \sim K^{-\gamma}$ for nodes of degree K is small for big K , so hubs are small and not so many. Thus, the scale-free network is hard to distinguish from a random network (as an example, path lengths are logarithmic in the network size).
- Summarizing, the most interesting regime is $2 < \gamma < 3$, when scale-free nets become ultra-small. Interestingly, many important real networks, such as the WWW and protein interaction networks, are in this regime!

D. Generating networks with arbitrary degree distribution

- Random networks generated by the Erdős–Rényi model have a Poisson degree distribution.
- However, the degree distributions of real networks significantly deviate from a Poisson form.
- Important question: Can we improve random networks so that their degree distribution becomes closer to the one of real networks?
- We will present 2 frequently used methods.

a. Configuration Model

- It creates random networks with a pre-defined degree sequence. This is done via making sure that each node has a pre-defined degree K_i but otherwise the network is wired randomly.



Configuration Model (II)

The configuration model algorithm includes the following steps:

- Degree Sequence: Assign a degree to each node, represented as stubs or half-links. We obviously must start with an even number of stubs. Note that the degree sequence is either generated from a pre-selected p_K distribution or by the fixed degrees of a real network. Also note that, if L is the number of network links, then the sum of degrees (stubs) is $2L$.
- Network Assembly: Randomly select a stub pair and connect the two stubs. Then, randomly choose another pair from the remaining $2L - 2$ stubs and connect the two stubs, and so on, until all stubs are paired up.

Configuration Model (III)

- Note 1: The probability of having a link between nodes of degree K_i and K_j is:

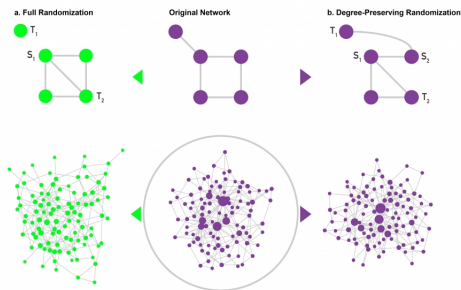
$$P_{ij} = \frac{K_i K_j}{2L - 1}$$

Indeed, a stub of node i can connect to $2L - 1$ stubs, among which K_j are attached to node j , so that the probability that a particular i -stub is connected to node j is $\frac{K_j}{2L-1}$, and node i has K_i stubs (attempts to connect to node j).

- Note 2: The obtained network may contain self-loops and multi-links. Yet, their number remains negligible, as the number of connection choices increases with N .
- Note 3: The network obtained is inherently random and this simplifies analytic calculations.

b. Degree-Preserving Randomization

- This is another method of obtaining a random rewiring of an original scale-free network, which however remains scale-free and preserves degrees.
- We randomly select two sources (S_1, S_2) and two targets (T_1, T_2) such that initially there is a $S_1 - T_1$ link and a $S_2 - T_2$ link. We then swap the two links, creating an $S_1 - T_2$ link and an $S_2 - T_1$ link. The swap leaves the degrees unchanged, yet random rewiring is introduced. This procedure is repeated until we rewire each link at least once.



The exact power-law form is rarely seen in real systems. Instead, the scale-free property tells us that we must distinguish two rather different classes of networks.

- Exponentially bounded networks: their degree distributions decreases exponentially or faster for high K , so we lack significant degree variations (since $\langle K^2 \rangle$ is smaller than $\langle K \rangle$). Examples of such p_K include the Poisson, Gaussian, or exponential distributions. Erdős–Rényi, Watts-Strogatz models are the best known models in this class. Real networks include highway networks and the power grid.
- Fat-tailed networks: their degrees have a power law in the high- K region, and $\langle K^2 \rangle$ is much larger than $\langle K \rangle$, resulting in considerable degree variations and big hubs. Scale-free nets with a power-law degree distribution offer the best-known example. Real networks include the WWW, the Internet, protein interaction networks, social networks.