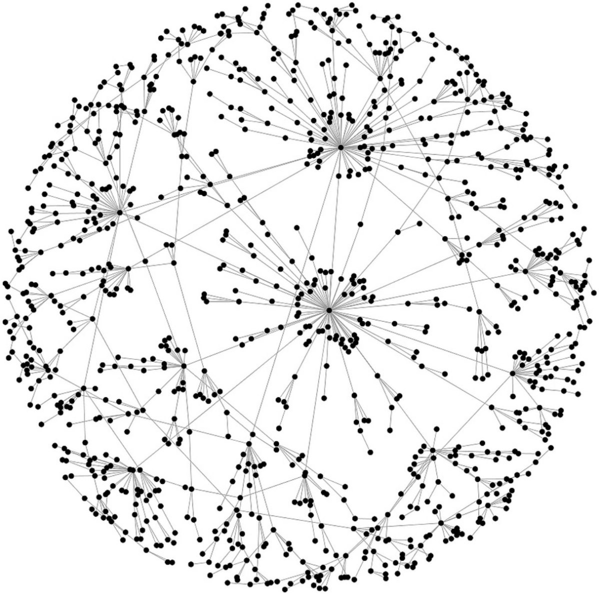
Advanced Algorithms and Computational Models (module A)

The Scale-Free Property

Giacomo Fiumara
giacomo.fiumara@unime.it

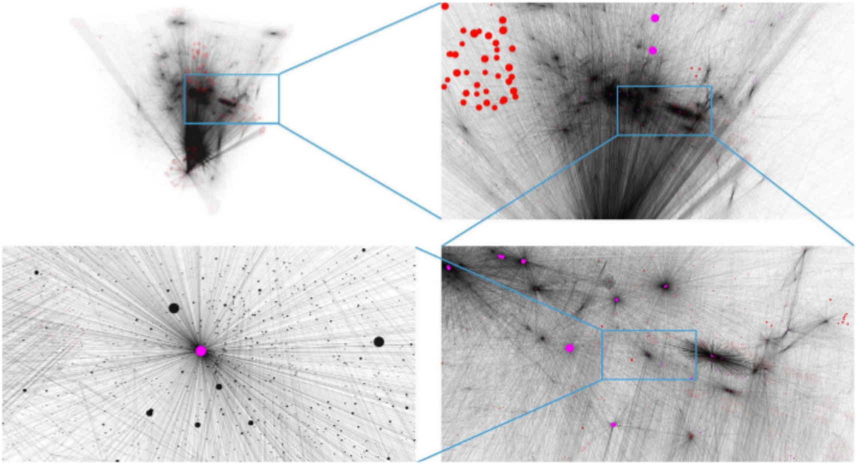
2015-2016



The Scale-Free Property

- The World Wide Web is a network whose nodes are documents and the links are the Uniform Resource Locator (URL) that allow us to “surf” with a click from one web document to another
- The Web is the largest network humanity has ever built ($N \approx 10^{12}$ documents)
- The first map of the WWW obtained with the explicit goal of understanding the structure of the network behind it was related to the `nd.edu` domain and consisted of about 300000 documents and 1.5 million links
- The purpose of the map was to compare the properties of the Web graph to the random network model
- In the WWW map are present a few hubs, nodes with an exceptionally large number of links, connected to numerous small-degree nodes

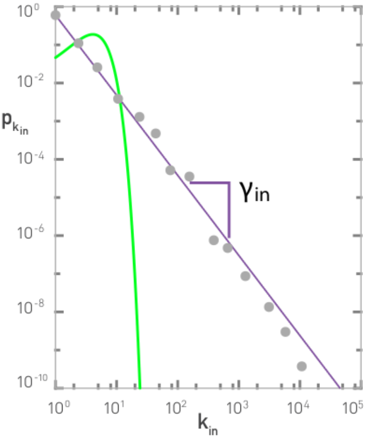
The Scale-Free Property

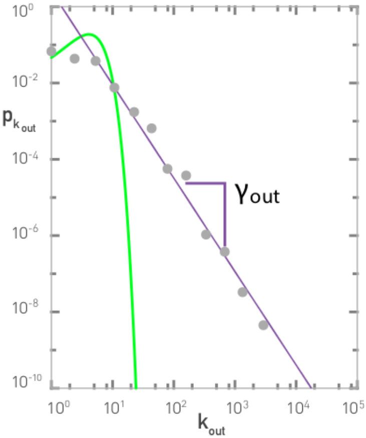


The Scale-Free Property

Power Laws and Scale-Free Networks

- If the WWW were to be a random network, the degree of the Web documents should follow a Poisson distribution





The Scale-Free Property

Power Laws and Scale-Free Networks

- Instead, on a log-log scale the data points form an approximate straight line, suggesting that the degree distribution of the WWW is well approximated with

$$p_k \sim k^{-\gamma}$$

- This is called *power law distribution* and the exponent γ is its *degree exponent*
- Note that taking the logarithm of both sides one obtains

$$log p_k \sim -\gamma log k$$

- Therefore, the slope of this line is the degree exponent γ

The Scale-Free Property

Power Laws and Scale-Free Networks

- The WWW is a directed network, hence each document is characterized by an *out-degree* k_{out}, representing the number of links that point from the document to other documents, and an *in-degree* k_{in}, representing the number of other documents that point to the selected document
- It is necessary, therefore, to distinguish two degree distributions: the probability that a randomly chosen document points to k_{out} web documents, or $p_{k_{out}}$, and the probability that a randomly chosen node has k_{in} web documents pointing to it, or $p_{k_{in}}$

The Scale-Free Property

Power Laws and Scale-Free Networks

- In the case of WWW both $p_{k_{out}}$ and $p_{k_{in}}$ can be approximated by a power law

$$p_{k_{in}} \sim k^{-\gamma_{in}}$$

$$p_{k_{out}} \sim k^{-\gamma_{out}}$$

- where γ_{in} and γ_{out} are the degree exponents for the in- and out-degrees
- In general γ_{in} may differ from γ_{out}
- For the WWW we have $\gamma_{in} \approx 2.1$ and $\gamma_{in} \approx 2.45$

The Scale-Free Property

Power Laws and Scale-Free Networks

- These results document the existence of a network whose degree distribution is quite different from the Poisson distribution characterizing random networks
- These networks are called *scale-free networks* and follow the definition: *A scale-free network is a network whose degree distribution follows a power law*

The Scale-Free Property

Power-Law Distribution: discrete formalism

- Node degrees are positive integers, $k = 0, 1, 2, \ldots,$
- The probability p_k that a node has exactly k links is

$$p_k = Ck^{-\gamma}$$

- Where the constant C is determined by the normalization condition

$$\sum_{k=1}^{\infty} p_k = 1$$

- Namely

$$C \sum_{k=1}^{\infty} k^{-\gamma} = 1$$

The Scale-Free Property

Power-Law Distribution: discrete formalism

- Hence
$$C = \frac{1}{\sum_{k=1}^{\infty} k^{-\gamma}} = \frac{1}{\zeta(\gamma)}$$
- Where $\zeta(\gamma)$ is the Riemann-zeta function defined as
$$\zeta(\gamma) = \sum_{k=1}^{\infty} k^{-\gamma}$$
- Finally, for $k > 0$, the discrete power-law distribution has the form
$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}$$
- Note that p_k diverges at $k = 0$. Therefore, may be necessary to separate p_0, which represents the fraction of nodes that have no links to other nodes

The Scale-Free Property

Power-Law Distribution: continuum formalism

- In analytical calculations it is often convenient to assume that the degrees can heve any positive real value. In this case

$$p(k) = Ck^{-\gamma}$$

- Using the normalization condition

$$\int_{k_{min}}^{\infty} p(k)dk = 1$$

- we obtain

$$C = \frac{1}{\int_{k_{min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1)k_{min}^{\gamma-1}$$

- Where k_{min} is the smallest degree for which the degree distribution holds.

The Scale-Free Property

Power-Law Distribution: continuum formalism

- Therefore the degree distribution has the form

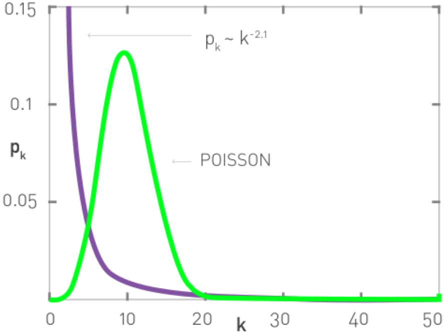
$$p(k) = (\gamma - 1)k_{min}^{\gamma-1}k^{-\gamma}$$

- Note that p_k encountered in the discrete formalism is the probability that a randomly selected node has degree k
- In contrast, only the integral of $p(k)$ encountered in the continuum formalism has a physical interpretation

$$\int_{k_1}^{k_2} p(k)dk$$

- is the probability that a randomly chosen node has degree between k_1 and k_2

The Scale-Free Property
Hubs



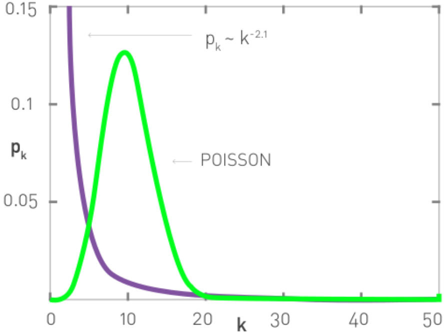
The Scale-Free Property

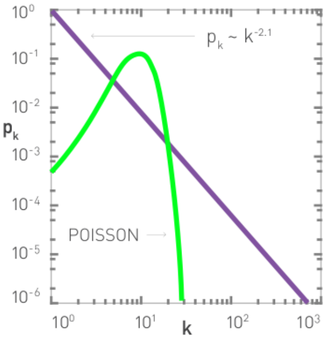
Hubs

- The main difference between a random and a scale-free network comes in the *tail* of the degree distribution, representing the high-k region of p_k
- For small k the power law is above the Poisson function, indicating that a scale-free network has a large number of small degree nodes, most of which are absent in a random network
- For k in the vicinity of $\langle k \rangle$ the Poisson distribution is above the power law, indicating that in a random network there is an excess of nodes with degree $k = \langle k \rangle$
- For large k the power law is again above the Poisson curve. The difference is even more clear if p_k is shown on a log-log plot

The Scale-Free Property

Hubs





The Scale-Free Property

Hubs

- Let us use the WWW to illustrate the magnitude of these differences
- The probability to have a node with $k = 100$ is about $p_{100} \approx 10^{-94}$ in a Poisson distribution while it is about $p_{100} \approx 4 \times 10^{-4}$ is p_k follows a power law
- As a consequence, if the WWW were to be a random network with $\langle k \rangle = 4.6$ and size $N \approx 10^{12}$, we would expect

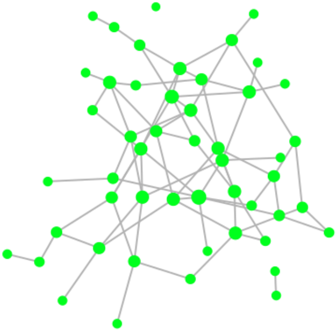
$$N_{k \geq 100} = 10^{12} \sum_{k=100}^{\infty} \frac{(4.6)^k}{k!} e^{-4.6} \approx 10^{-82}$$

- nodes with at least 100 links (namely none)
- In contrast, given the WWW's power law degree distribution, with $\gamma_{in} = 2.1$, we have

$$N_{k \geq 100} = 4 \times 10^9$$

The Scale-Free Property

Hubs



The Scale-Free Property

Hubs

- An interesting question concerns the relation existing between the size of a network vs the size of the largest hub
- In other words: how does the network size affect the size of the hubs?
- To answer this we calculate the maximum degree k_{max}, called the *natural cutoff* of the degree distribution p_k
- It represents the expected size of the largest hub in a network

The Scale-Free Property

Hubs

- Let us consider the exponential distribution

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

- For a network with minimum degree K_{min} the normalization condition

$$\int_{k_{min}}^{\infty} p(k)dk = 1$$

- provides $C = \lambda e^{\lambda k_{min}}$.

The Scale-Free Property

Hubs

- To calculate k_{max} we assume that in a network of N nodes we expect at most one node in the (k_{max}, ∞) regime
- In other words, the probability to observe a node whose degree exceeds k_{max} is $1/N$:

$$\int_{k_{max}}^{\infty} p(k) dk = \frac{1}{N}$$

- Namely

$$\int_{k_{max}}^{\infty} \lambda e^{\lambda k_{min}} e^{-\lambda k} dk = \frac{1}{N}$$

- From which we obtain

$$k_{max} = k_{min} + \frac{lnN}{\lambda}$$

The Scale-Free Property

Hubs

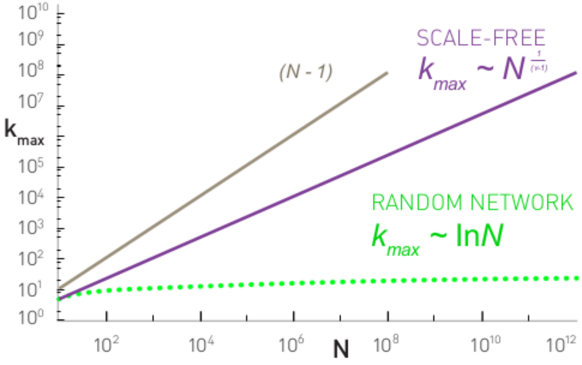
- For a Poisson degree distribution the dependence of k_{max} on N is even slower than the logarithmic dependence of the exponential distribution
- For a scale-free network, the natural cutoff follows

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

- Hence, the large a network, the larger is the degree of its biggest hub
- The polynomial dependence of k_{max} on N implies that in a large scale-free network there can be orders of magnitude differences in size between the smallest node, k_{min}, and the biggest hub, k_{max}

The Scale-Free Property

Hubs



The Scale-Free Property

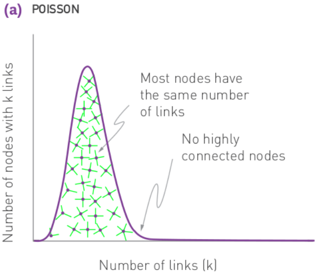
Hubs

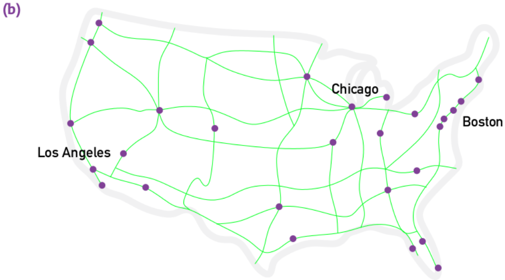
- In the case of the WWW sample with $N \approx 3 \times 10^5$ nodes, $k_{min} = 1$
- If the degree distribution were to follow an exponential distribution, the maximum degree should be $k_{max} \approx 14$ for $\lambda = 1$
- In a scale-free distribution of similar size and $\gamma = 2.1$, we have $k_{max} \approx 95000$

- In a random network hubs are effectively forbidden, while in scale-free networks they are naturally present

The Scale-Free Property

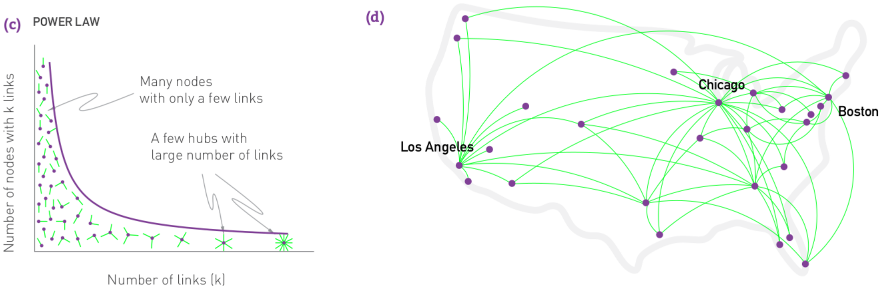
Hubs





The Scale-Free Property

Hubs



The Scale-Free Property

The Meaning of Scale-Free

- The term "scale-free" derives from a branch of statistical physics called the *theory of phase transitions* that explored power laws in the 1960s and 1970s
- The meaning of the scale-free term may be more clear after a brief recall of the moments of the degree distribution

- The n^{th} moment of the degree distribution is defined as

$$\langle k^n \rangle = \sum_{k_{min}}^{\infty} k^n p_k = \int_{k_{min}}^{\infty} k^n p(k) dk$$

The Scale-Free Property

The Meaning of Scale-Free

The lower moments have important interpretation:

- The first moment is the average degree, $\langle k \rangle$
- The second moment $\langle k^2 \rangle$ is useful for the calculation of the *variance* $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$
- The third moment, $\langle k^3 \rangle$, determines the *skewness* of a distribution (how symmetric is p_k around the average $\langle k \rangle$)

- For a scale-free network, the n^{th} moment of the degree distribution is

$$\langle k^n \rangle = \sum_{k_{min}}^{k_{max}} k^n p_k = C \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n-\gamma+1}$$

The Scale-Free Property

The Meaning of Scale-Free

$$\langle k^n \rangle = \sum_{k_{min}}^{k_{max}} k^n p_k = C \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n-\gamma+1}$$

- Typically k_{min} is fixed, while the degree of the largest hub k_{max} increases with the system size
- To understand the behavior of $\langle k^n \rangle$ we need to take the asymptotic limit $k_{max} \to \infty$, probing the properties of very large networks

The Scale-Free Property

The Meaning of Scale-Free

$$\langle k^n \rangle = \sum_{k_{min}}^{k_{max}} k^n p_k = C \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n-\gamma+1}$$

- If $n - \gamma + 1 \leq 0$, the the first term of the rhs goes to zero as k_{max} increases. Therefore all the moments that satisty $n \leq \gamma - 1$ are finite
- If $n - \gamma + 1 > 0$ then $\langle k^n \rangle$ goes to infinity as $k_{max} \to \infty$. Therefore all the moments larger than $\gamma - 1$ diverge

The Scale-Free Property

The Meaning of Scale-Free

NETWORK	N	L	$\langle k \rangle$	$\langle k_{in}^2 \rangle$	$\langle k_{out}^2 \rangle$	$\langle k^2 \rangle$	γ_{in}	γ_{out}	γ
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	93,439	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*	2.9 0*	-
Protein Interactions	2,018	2,930	2.9 0	-	-	32.3	-	-	2.89*

The Scale-Free Property

The Meaning of Scale-Free

- For many scale-free networks the degree exponent γ is between 2 and 3
- Hence for these in the $N \to \infty$ limit the first moment $\langle k \rangle$ is finite, but the second and higher moments go to infinity
- This divergence helps us understand the origin of the "scale-free" term
- Indeed, if the degrees follow a normal distribution, then the degree of a randomly chosen node is typically in the range

$$k = \langle k \rangle \pm \sigma_k$$

The Scale-Free Property

The Meaning of Scale-Free

- For a random network with a Poisson degree distribution, $\sigma_k = \langle k \rangle^{1/2}$, which is always smaller than $\langle k \rangle$
- Hence the network's nodes have degrees in the range

$$k = \langle k \rangle \pm \langle k \rangle^{1/2}$$

- In other words, nodes in a random network have comparable degrees and the average degree $\langle k \rangle$ is the "scale" of the network

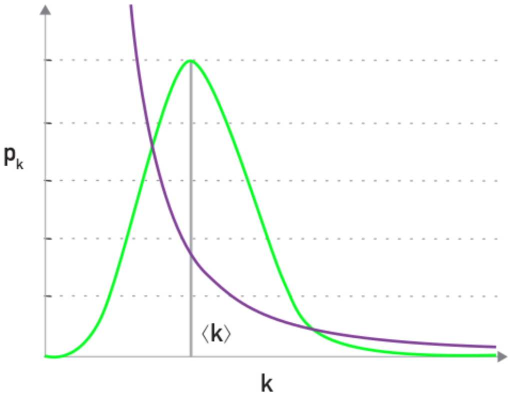
The Scale-Free Property

The Meaning of Scale-Free

- For a network with a power-law degree distribution with $\gamma < 3$ the first moment is finite, but the second moment is infinite
- The divergence of $\langle k^2 \rangle$ for large N indicates that the fluctuations around the average can be arbitrary large
- This means that when we randomly choose a node, we do not know what to expect: the selected node's degree could be tiny or arbitrarily large
- Hence networks with $\gamma < 3$ do not have a meaningful internal scale, but are "scale-free"

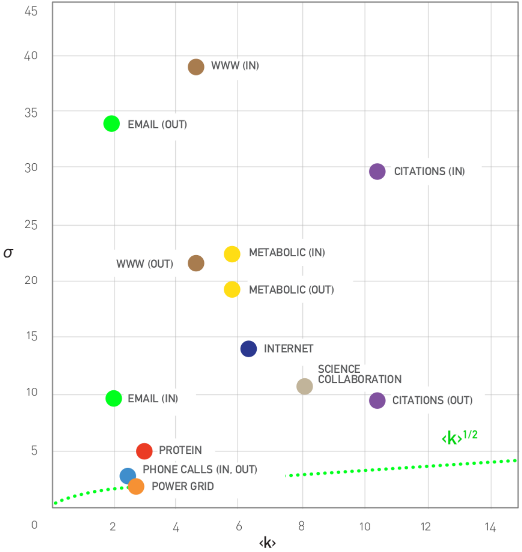
The Scale-Free Property

The Meaning of Scale-Free



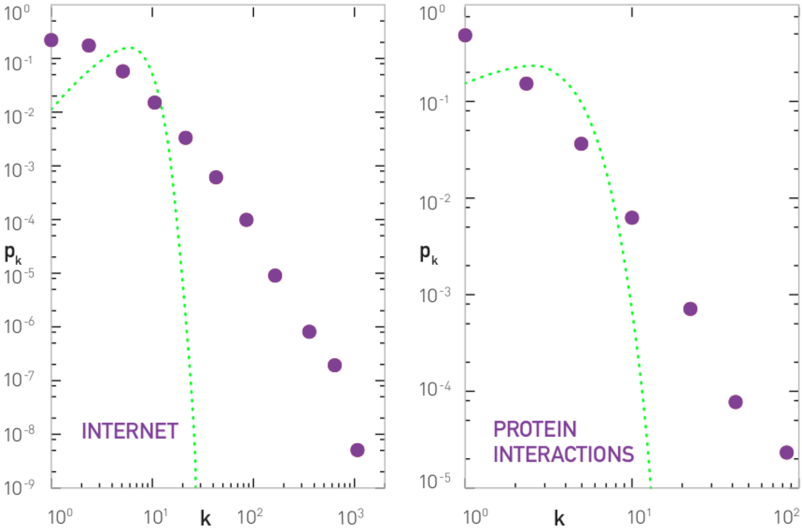
The Scale-Free Property

The Meaning of Scale-Free



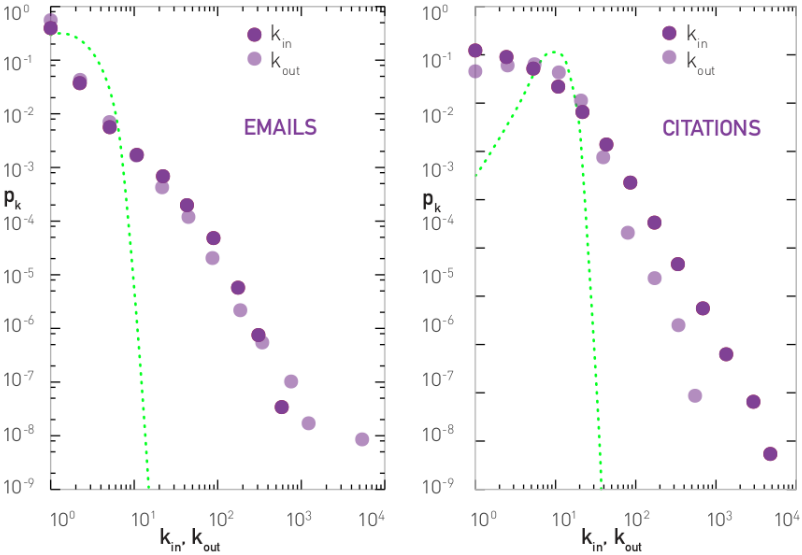
The Scale-Free Property

Universality



The Scale-Free Property

Universality



The Scale-Free Property

Universality

- In the past decade many real networks of major scientific, technological and societal importance were found to display the scale-free property
- The diversity of the systems that share the scale-free property is remarkable
- Think, for example, to the WWW which is a man-made network with a history of about 25 years and to the protein interaction network which is the product of 2 billion years of evolution
- This diversity lead us to call the scale-free property a *universal* network characteristic

The Scale-Free Property

Ultra-Small World Property

- The presence of hubs in scale-free networks raises an interesting question: do hubs affect small world property?
- The dependence of the average distance $\langle d \rangle$ on the system size N and the degree exponent γ is

$$\langle d \rangle \sim \begin{cases} const & \gamma = 2 \\ lnlnN & 2 < \gamma < 3 \\ \frac{lnN}{lnlnN} & \gamma = 3 \\ lnN & \gamma > 3 \end{cases}$$

The Scale-Free Property

Ultra-Small World Property

Anomalous Regime $\gamma = 2$

- According to

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

  for $\gamma = 2$ the degree of the biggest hub grows linearly with the system size, i.e. $k_{max} \sim N$
- This forces the network into a *hub and spoke* configuration in which all nodes are close to each other because they all connect to the same central hun
- In this regime the average path length does not depend on N

Ultra-Small World $2 < \gamma < 3$

- In this regime

$$\langle d \rangle \sim lnlnN$$

  the average distance increases as $lnlnN$, a significantly slower growth that the lnN derived for random networks
- We call networks in this regime *ultra-small*, as the hubs radically reduce the path length
- They do so by linking to a large number of small-degree nodes, creating short distances between them
- Let us consider the world's social network with $N \approx 7 \times 10^9$. If the society is described by a random network, the N-dependent term is $lnN = 22.66$, while for a scale-free network the N-dependent term is $lnlnN = 3.12$

The Scale-Free Property

Ultra-Small World Property

Critical Point $\gamma = 3$

- This value is particularly important, as the second moment of the degree distribution does not diverge
- In fact
$$\langle k^n \rangle = C \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n-\gamma+1}$$
does not diverge when $n - \gamma + 1 \leq 0$ (as for $\gamma = 3$)
- Therefore $\gamma = 3$ is called the *critical point*
- At the critical point the *lnN* dependence encountered for random networks returns
- Yet, the calculations indicate the presence of a double logarithm correction *lnlnN* which shrinks the distances compared to a random network of similar size

Small World $\gamma > 3$

- In this regime $\langle k^2 \rangle$ is finite and the average distance follows the small-world result derived for random networks
- While hubs continue to be present, for $\gamma > 3$ they are not sufficiently large and numerous to have a significant impact on the distance between the nodes

The Scale-Free Property

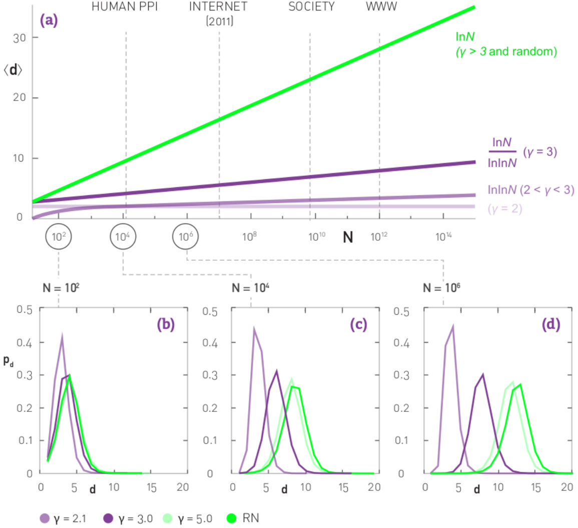
Ultra-Small World Property

$$\langle d \rangle \sim \begin{cases} const & \gamma = 2 \\ lnlnN & 2 < \gamma < 3 \\ \frac{lnN}{lnlnN} & \gamma = 3 \\ lnN & \gamma > 3 \end{cases}$$

Indicates that the more pronounced the hubs are, the more effectively they shrink the distances between nodes

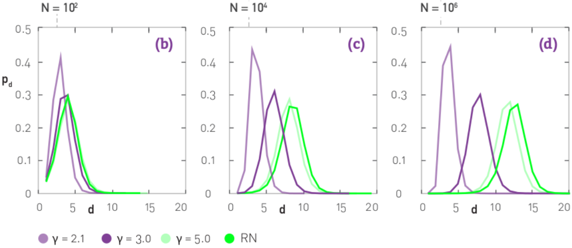
The Scale-Free Property

Ultra-Small World Property



The Scale-Free Property

Ultra-Small World Property



- For $N = 10^2$ the path length distributions overlap, indicating that at this size differences in γ result in undetectable differences in the path length
- For $N = 10^6$, p_d observed for different γ are well separated
- The larger the degree exponent, the larger are the differences between the nodes

The Scale-Free Property

Ultra-Small World Property

The scale-free property has several effects on network distances:

- Shrinks the average path lengths. Therefore most scale-free networks of practical interest are not only "small", but are "ultra-small". This is the consequence of the hubs, that act as bridges between many small degree nodes
- Changes the dependence of $\langle d \rangle$ on the system size. The smaller is γ, the shorter are the distances between the nodes
- Only for $\gamma > 3$ we recover the *lnN* dependence, the signature of the small-world property characterizing random networks

The Scale-Free Property

The Role of the Degree Exponent

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

Many properties of a scale-free network depend on the value of the degree exponent γ:

- γ varies from system to system, prompting us to explore how the properties of a network change with γ
- For most real systems the degree exponent is above 2, making us wonder: Why don't we see networks with $\gamma < 2$?

To address these questions we discuss how the properties of a scale-free network change with γ

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

Anomalous Regime ($\gamma \leq 2$)

- For $\gamma < 2$ the exponent $1/(\gamma - 1)$ is larger than one, hence the number of links connected to the largest hub grows faster than the size of the network
- This means that for sufficiently large N the degree of the largest hub must exceed the total number of nodes in the network, hence it will run out of nodes to connect to
- Similarly, for $\gamma < 2$ the average degree $\langle k \rangle$ diverges in the $N \to \infty$ limit
- These predictions are only two of the many anomalous features of scale-free networks in this regime: large scale-free networks with $\gamma < 2$ cannot exist

The Scale-Free Property

The Role of the Degree Exponent

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

Scale-free Regime $(2 < \gamma \leq 3)$

- In this regime the first moment of the degree distribution is finite but the second and higher moments diverge as $N \to \infty$
- Consequently scale-free networks in this regime are ultra-small
- k_{max} grows with the size of the network with exponent $1/(\gamma - 1)$ smaller than one
- Hence the fraction of nodes that connect to the largest hub decreases as

$$\frac{k_{max}}{N} \sim N^{-\frac{\gamma - 2}{\gamma - 1}}$$

The Scale-Free Property

The Role of the Degree Exponent

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

Random Network Regime ($\gamma > 3$)

- According to
$$\langle k^n \rangle = C \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n-\gamma+1}$$
both the first and the second moment of the degree distribution are finite
- For all practical purposes the properties of a scale-free network in this regime are difficult to distinguish from the properties of a random network of similar size
- For example, the average distance $\langle d \rangle$ goes as lnN as in the small-world regime
- The reason is that for large γ the degree distribution p_k decays sufficiently fast to make the hubs small and less numerous