SOCIAL NETWORK ANALYTICS

Power-Law and Power-Law Degree Distributions

Prakash C O

Department of Computer Science and Engineering

SOCIAL NETWORK ANALYTICS

Power-Law Degree Distributions

Prakash C O

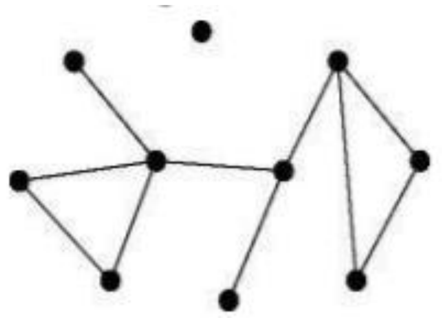
Department of Computer Science and Engineering

Power Law

- Network **degree distributions are fundamental tools** used in the study of complex networks such as online social networks.
- Not only do they reveal insights into the structure and formation of these networks, but **they also lay the foundation for modeling network dynamics** and help to guide the design of graph algorithms and applications.
- **It is widely believed that the Power-law distribution accurately captures node degrees in complex networks** such as online social networks.

Degree distributions:

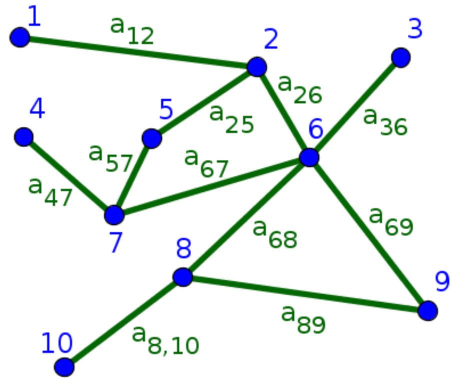
- The **degree distribution** is **the probability distribution of the degrees over the whole network.**
- Let $\mathbf{p_x}$ be **the fraction of vertices in a network that have degree *x***. That is the same as saying: $\mathbf{p_x}$ is the probability that a randomly selected node of the network will have degree ***x***.

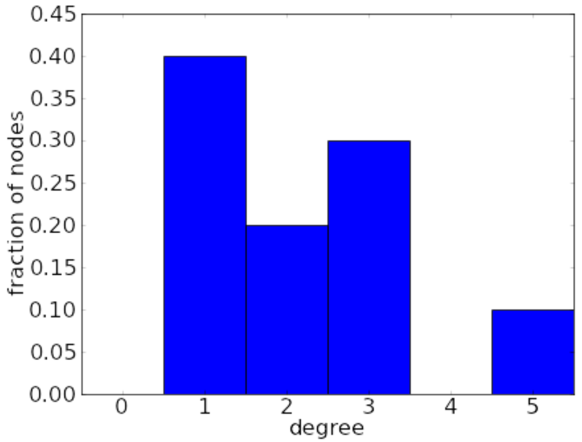


$$p_0 = \frac{1}{10}, p_1 = \frac{2}{10}, p_2 = \frac{4}{10}, p_3 = \frac{2}{10}, p_4 = \frac{1}{10}, p_k = 0 \forall k \geq 5$$

- A well-connected vertex is called a hub.

Degree distributions: Example-1



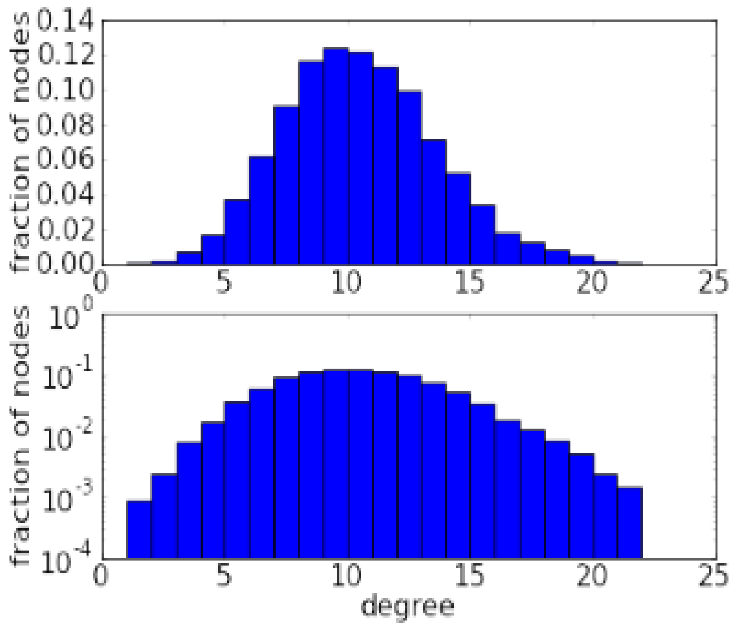


- **For this undirected network**, $(\text{node-1})_{deg}=1$, $(\text{node-2})_{deg}=3$, $(\text{node-3})_{deg}=1$, $(\text{node-4})_{deg}=1$, $(\text{node-5})_{deg}=2$, $(\text{node-6})_{deg}=5$, $(\text{node-7})_{deg}=3$, $(\text{node-8})_{deg}=3$, $(\text{node-9})_{deg}=2$, and $(\text{node-10})_{deg}=1$.

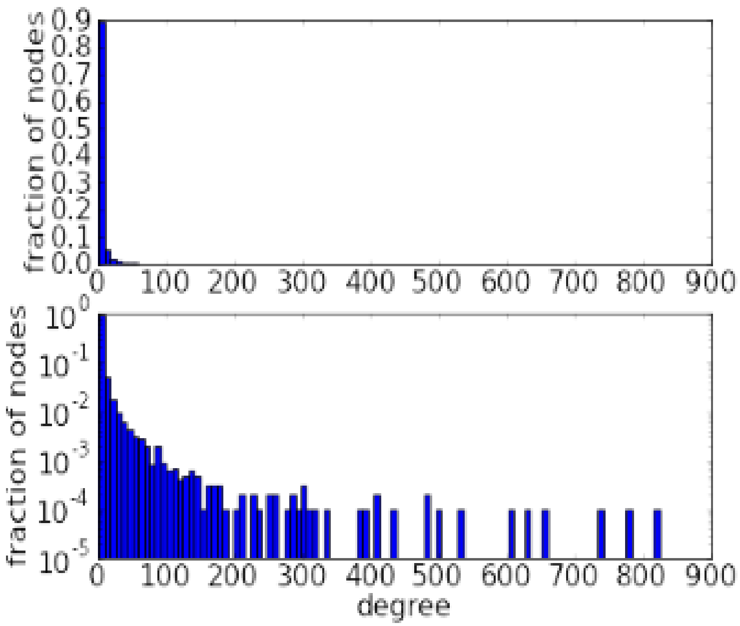
- **Its degree distribution** is $P_1=4/10$, $P_2=2/10$, $P_3=3/10$, $P_4=0$, $P_5=1/10$, and for all $k\geq6$ $P_k=0$.

Degree Distribution

Degree distributions: **Example-2**



A binomial degree distribution of a network with 10,000 nodes and average degree of 10. The top histogram is on a linear scale while the bottom shows the same data on a log scale.



A power law degree distribution of a network with 10,000 nodes and average degree of around 7. The top histogram is on a linear scale while the bottom shows the same data on a log scale.

What is Power-Law?

- The **power-law** (scaling law) is **a functional relationship between two quantities**, where a relative change in one quantity results in a proportional relative change in the other quantity, independent of the initial size of those quantities: **one quantity varies as a power of another.**

- **Example:**
  - The simplest example of the law in action is a square; if you double the length of a side (say, from 2 to 4 inches) then the area will quadruple (from 4 to 16 inches squared).

If we double the length we multiply the area by a factor of four.

$(2)^2=4*1=4$ $(8)^2=4*16=64$

$(4)^2=4*4=16$ $(16)^2=4*64=256$

What is Power-Law?

- In its empirical form, the Power Law **describes how a lot of the time, not much actually happens but more often than not, some patterns cover a wide range of magnitudes.**
- Examples:
    - Number of Comments on a Post
    - Number of Social Media Followers
    - Money Grossed at Box Offices
    - Books sold for Top Authors
    - Market Capitalization of American Companies

They all seem to fit the pattern.

Power-Law Distribution

- Most real-world networks, instead, follow a "power law" distribution for the node connectivity.
- A power-law distribution is a special kind of probability distribution.
  - The probability for a node to have x-edges connected is proportional to $\mathbf{x}^{-\alpha}$

$$p(x) = C\ x^{-\alpha}$$

where:
- p(x) and x are variables of interest,
- α is the power-law exponent,
- C is a normalization constant(probabilities over all x must sum to 1)

  - It has been found that most scale-free networks have exponents (i.e., α) between 2 and 3. Thus, they lack a characteristic degree or scale, and therefore their name. High degree nodes are called **hubs**.

PES UNIVERSITY ONLINE

Power-Law Distribution

- It is widely believed that the Power-law distribution accurately captures node degrees in complex networks such as online social networks, i.e. their degree distribution follows a Power-law function

$$p_x \sim x^{-\alpha}$$

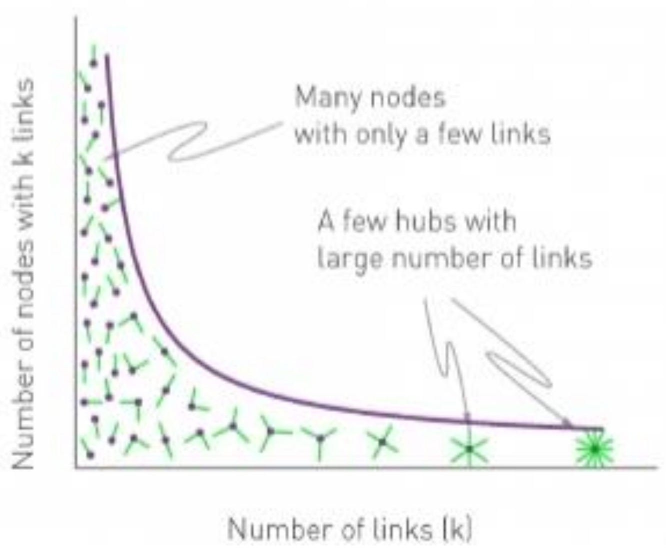
- As a result, the Power-law model has already played a significant role in guiding the design of algorithms on social network problems such as
  - influence maximization,
  - landmark selection, and
  - link privacy protection.

Ref: Revisiting Degree Distribution Models for Social Graph Analysis - Alessandra Sala et. al.

Power-Law Distribution

Examples:

- The power law can be used to describe a phenomenon where a small number of items is clustered at the top of a distribution (or at the bottom).



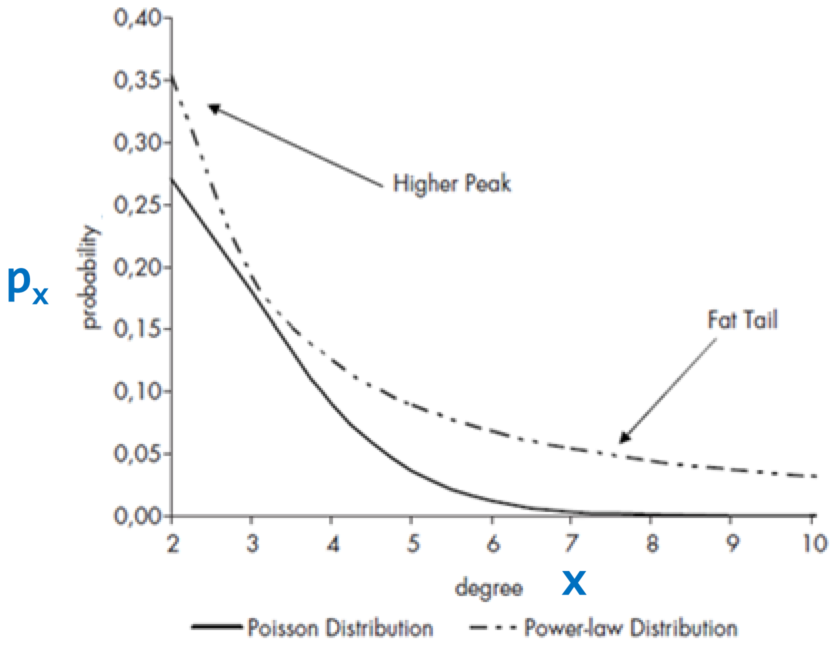
- For example, where the distribution of income is concerned, there are very few billionaires; the bulk of the population holds very modest nest eggs.

Power-Law Distribution

- The power-law distributions belong to the class of “**fat-tailed**” distributions.
- **The power-law distributions deviate from the Poisson distribution in two ways:** they have
  1. **higher peaks** and
  2. **“fat tails”** or **“heavy tails”.**

Power-Law Distribution

- **Figure below shows a power-law distribution in relation to a Poisson distribution. The peak on the left side of the chart is higher** and **the tail on the right side is thicker, therefore called “fat tail”/”heavy tail”.**

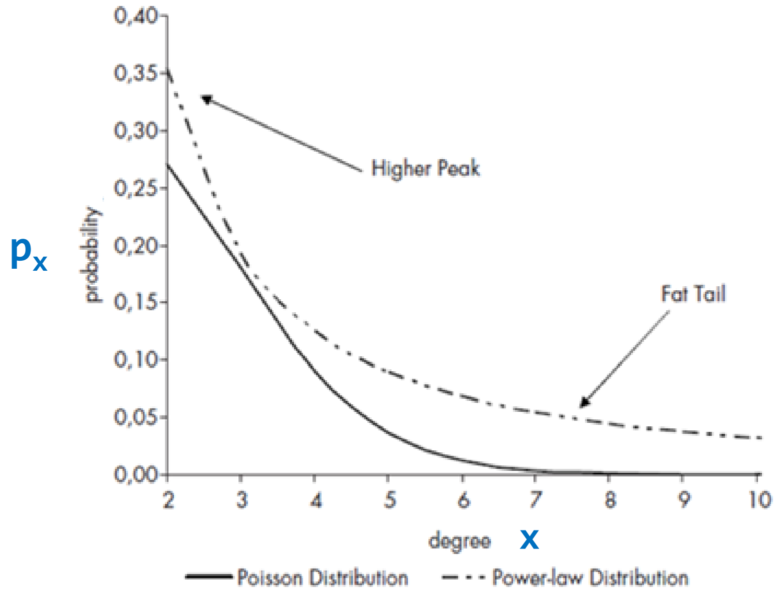


$$p_x \sim x^{-\alpha}$$

Power-Law Distribution

➢ Compared to the Poisson distribution, the power-law distribution shows

- **the greater probability of nodes with a number of links close to the average of all links** (see the peak with 2 links in the example figure), and
- **a smaller probability of nodes with many links** (4 or more links in the tail in figure).



Power-Law Distribution

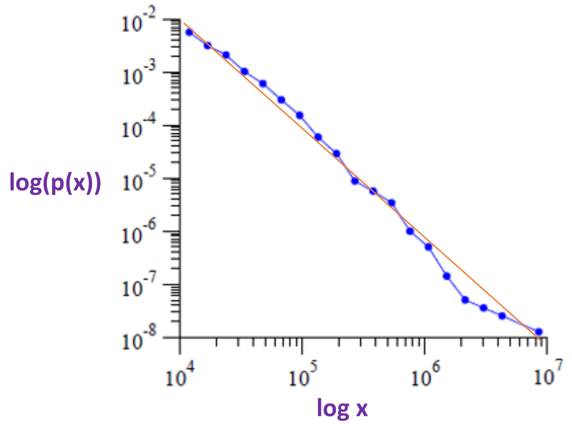
- Given a monomial equation $p(x) = C\ x^{-\alpha}$ taking the logarithm of the equation yields:

$$\log(p(x)) = \log C - \alpha \log x$$

Setting **X = log x** and **Y = log(p(x))** which corresponds to using a log–log graph, yields the equation:

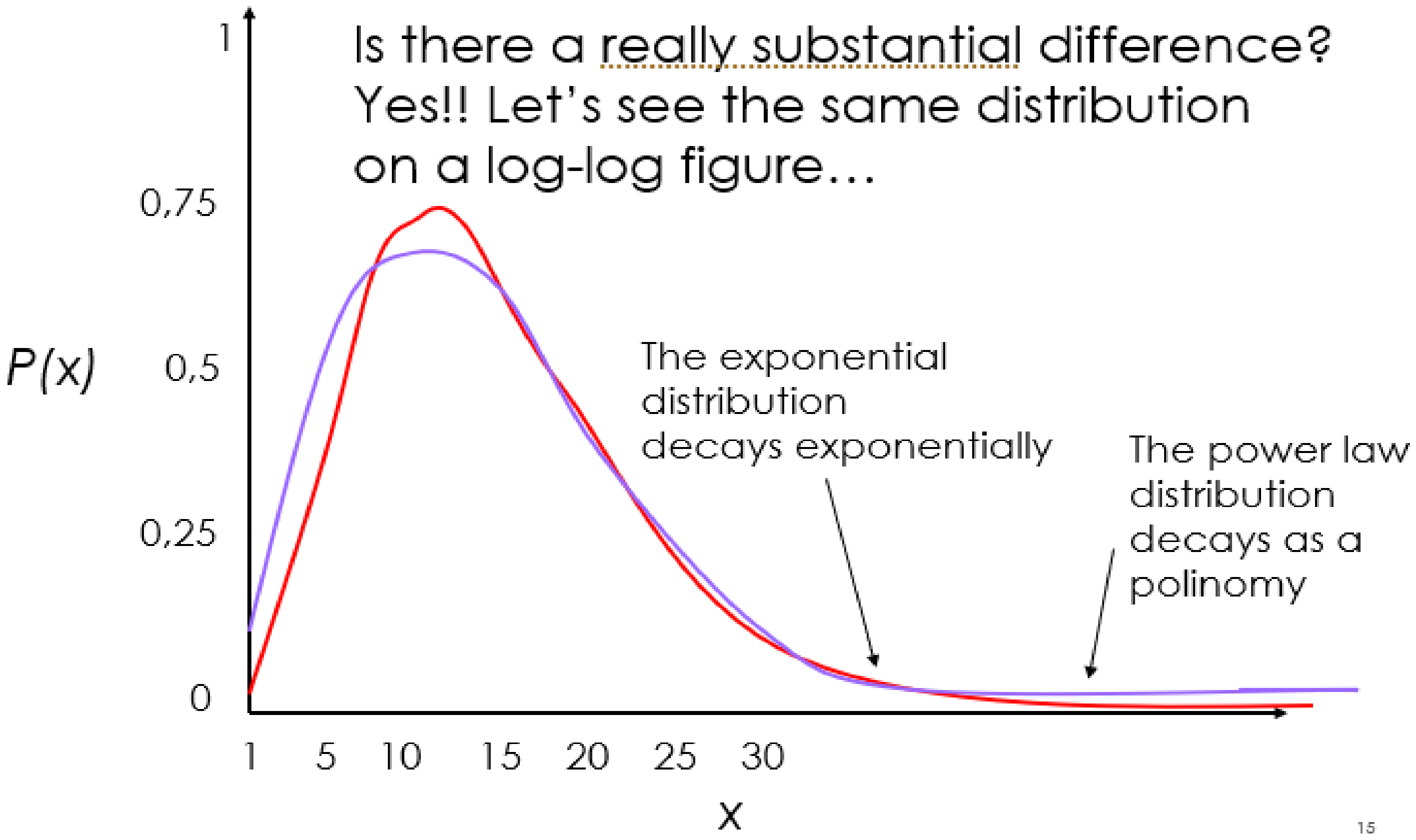
$$Y = mX + b$$

where **m = $-\alpha$** will be the slope and **b = log C** will be the y-intercept.

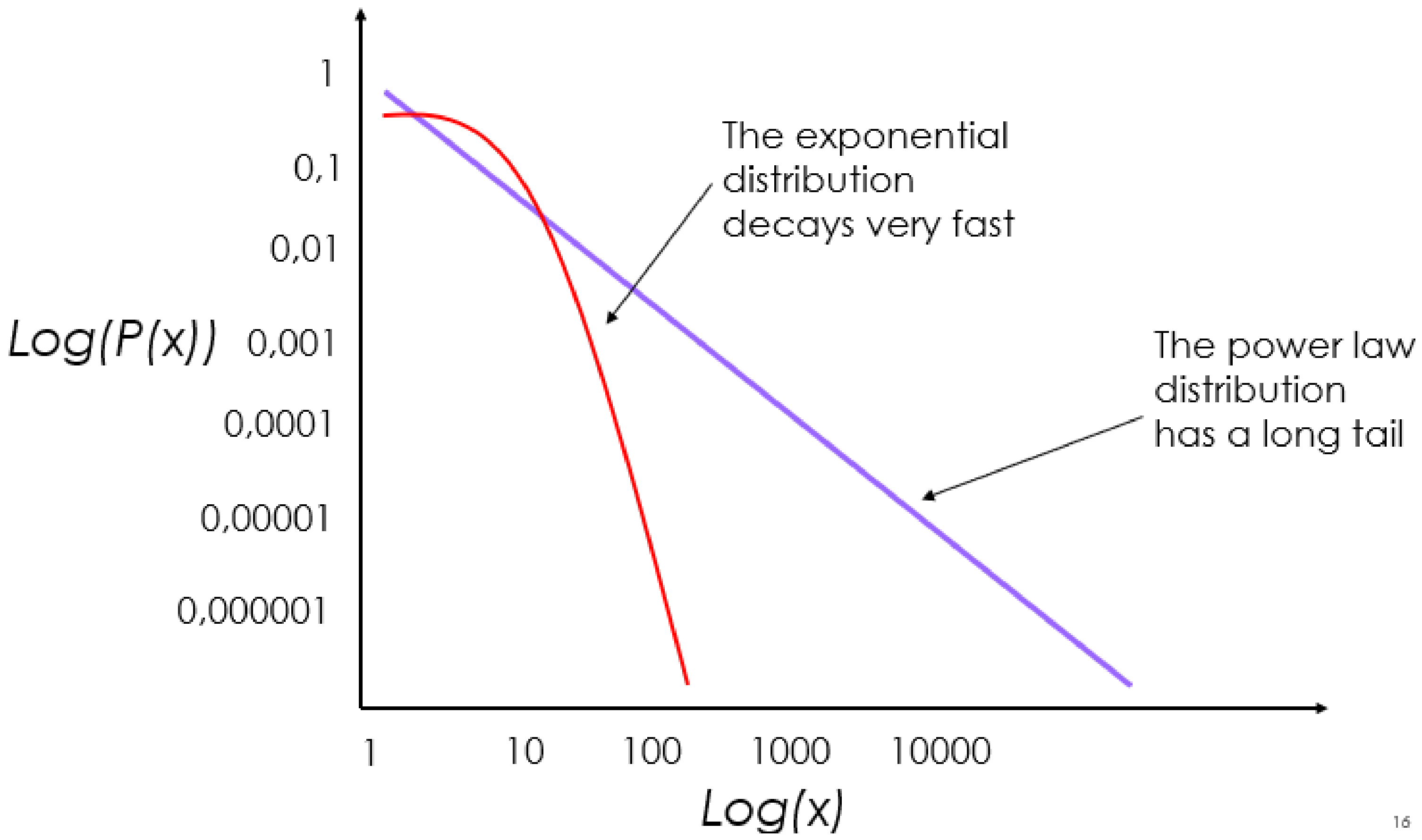


- Such a "log-log" plot thus provides a quick way to see if one's data exhibits an approximate power-law.
- **Power law distribution is a straight line on a log-log plot.**

Power-Law Vs. Exponential Distribution (Linear scale)



Power-Law Vs. Exponential Distribution (log log scale)



Power Law

Examples:

$$p(x) = Cx^{-\alpha}$$

log(p(x)) = log C – α log x

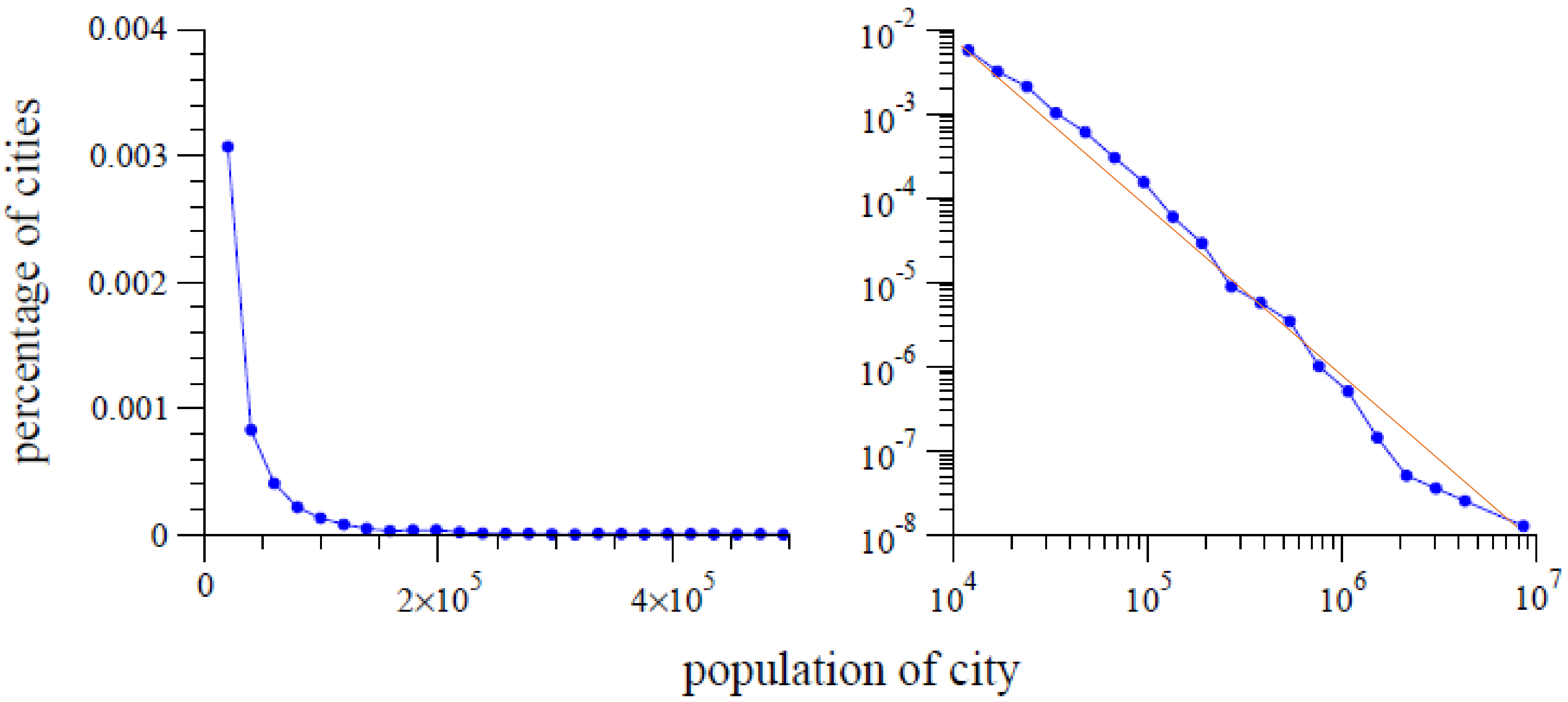


Figure 1. **Left:** histogram of the **populations of all US cities with population of 10000 or more**. **Right**: another histogram of the same data, but plotted on logarithmic scales. The approximate straight-line form of the histogram in the right panel implies that the distribution follows a power law.

Power Law

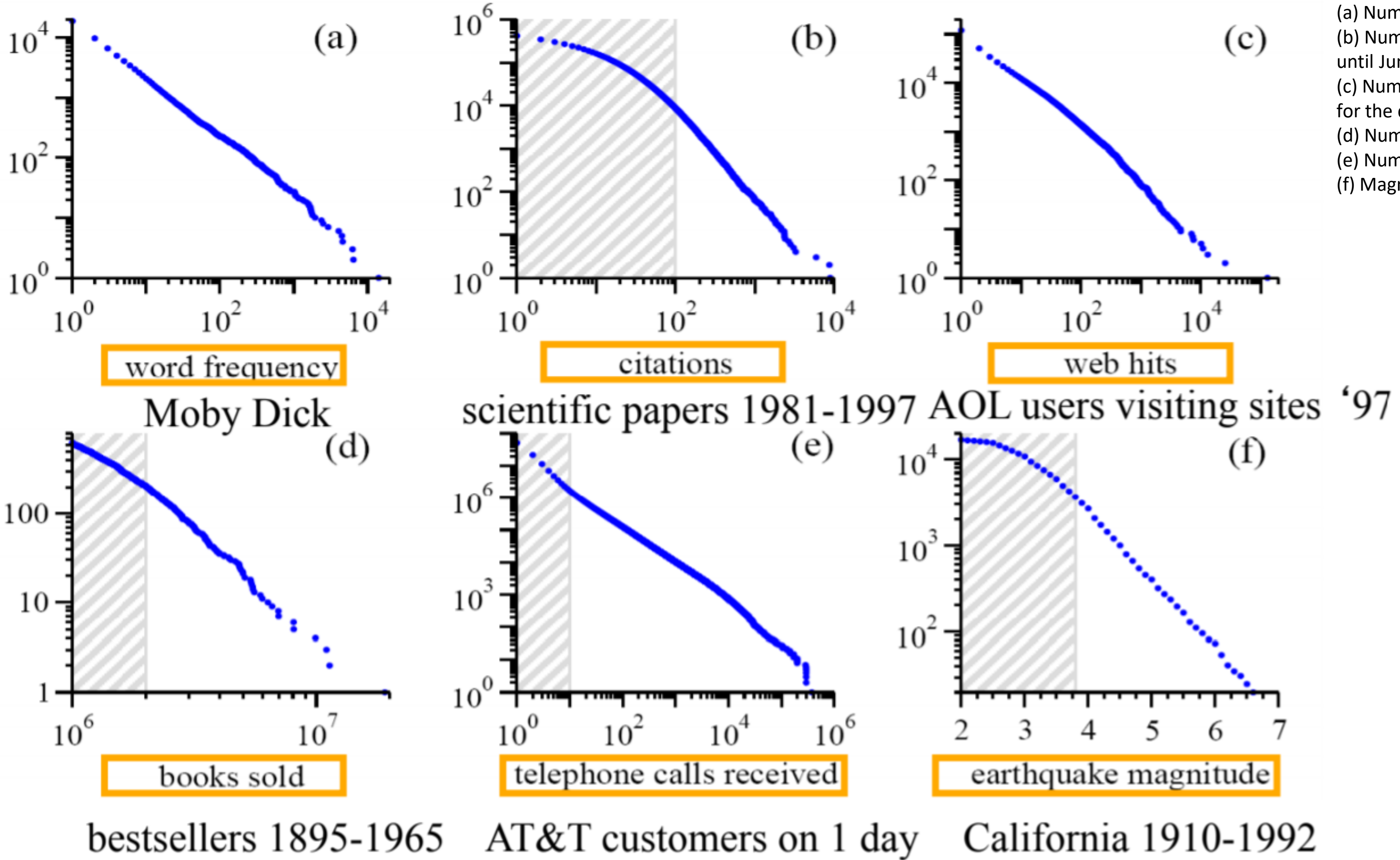
- Power-law distributions occur in a diverse range of phenomena.
  - City populations, the sizes of earthquakes, moon craters, solar flares,
  - computer files and wars, the frequency of use of words in any human language,
  - the frequency of occurrence of personal names in most cultures,
  - the numbers of papers scientists write,
  - the number of citations received by papers, the number of hits on web pages,
  - the sales of books, music recordings and almost every other branded commodity,
  - the numbers of species in biological taxa, people’s annual incomes and
  - a host of other variables all follow power-law distributions.

SOCIAL NETWORK ANALYTICS

Power Law

Examples: Source: MEJ Newman, 'Power laws, Pareto distributions and Zipf's law'

Figure 2:

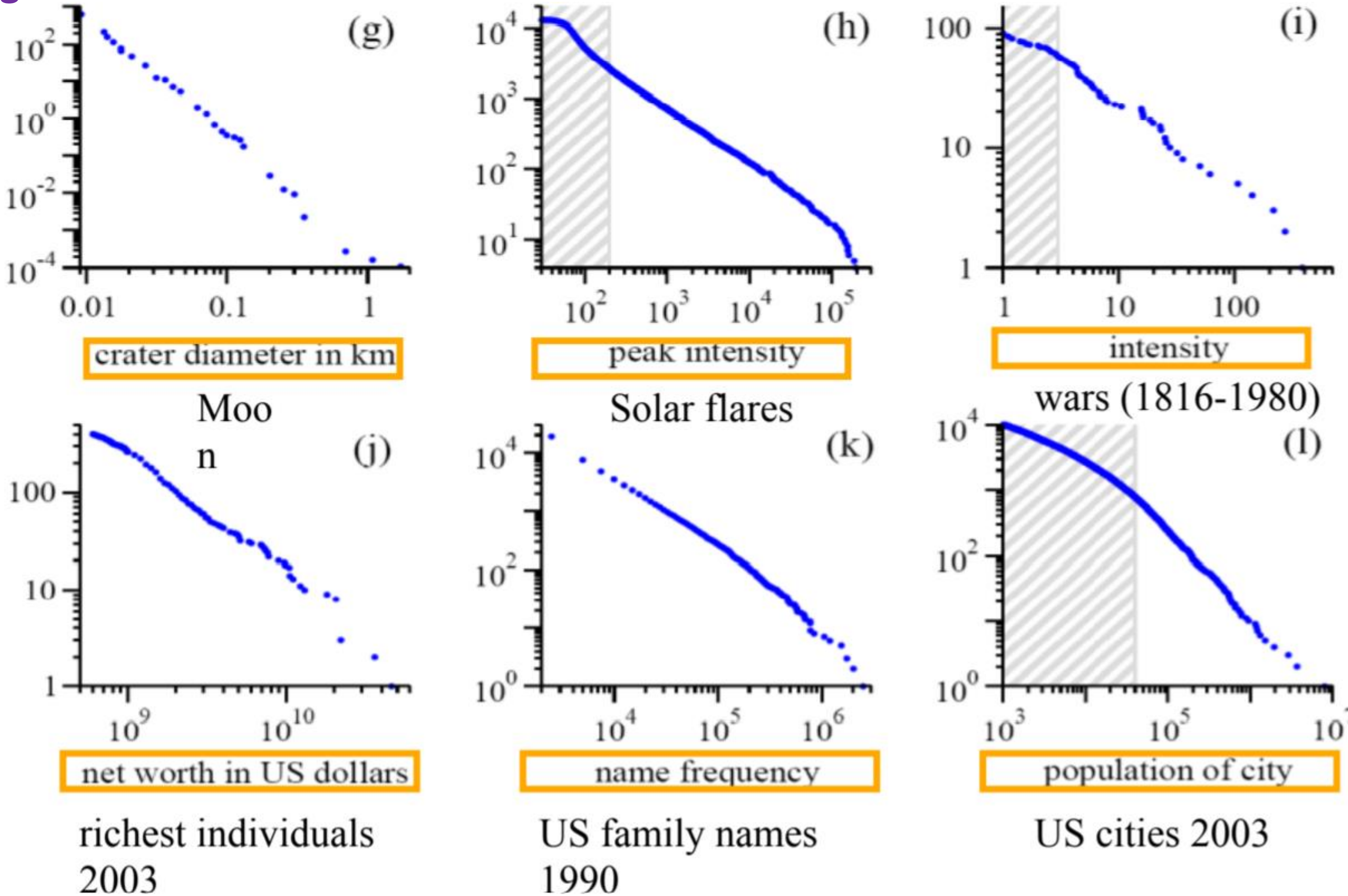


(a) Numbers of occurrences of words in the novel Moby Dick by Hermann Melville.
(b) Numbers of citations to scientific papers published in 1981, from time of publication until June 1997.
(c) Numbers of hits on web sites by 60 000 users of the America Online Internet service for the day of 1 December 1997.
(d) Numbers of copies of bestselling books sold in the US between 1895 and 1965.
(e) Number of calls received by AT&T telephone customers in the US for a single day.
(f) Magnitude of earthquakes in California between January 1910 and May 1992.

Power Law

Examples: Source: MEJ Newman, ’Power laws, Pareto distributions and Zipf’s law’

Figure 2:

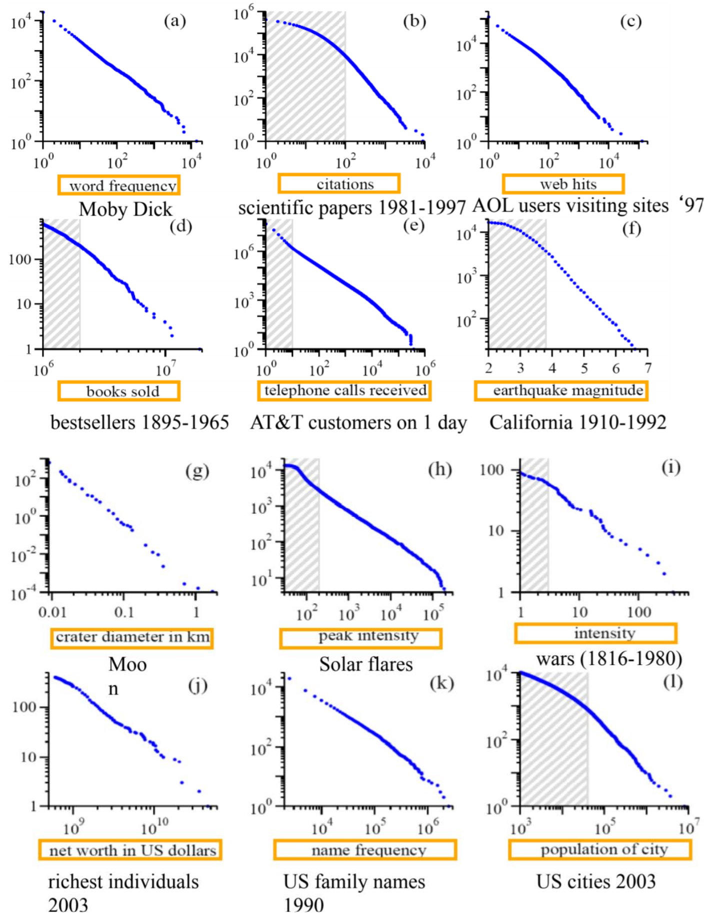


(g) Diameter of craters on the moon. Vertical axis is measured per square kilometre.
(h) Peak gamma-ray intensity of solar flares in counts per second, measured from Earth orbit between February 1980 and November 1989.
(i) Intensity of wars from 1816 to 1980, measured as battle deaths per 10 000 of the population of the participating countries.
(j) Aggregate net worth in dollars of the richest individuals in the US in October 2003.
(k) Frequency of occurrence of family names in the US in the year 1990.
(l) Populations of US cities in the year 2000.

Power Law

- **Few real-world distributions follow a power law over their entire range**, and in particular not for smaller values of the variable being measured.
- In reality, **the distribution must deviate from the power-law form below some minimum value x_{min}.**
- In our computer-generated example we simply cut off the distribution altogether below x_{min} so that $p(x) = 0$ in this region, but most real-world examples are not that abrupt.
- **Figure 2 shows distributions with a variety of behaviours for small values of the variable measured**; the straight-line power-law form asserts itself only for the higher values.
- Data in the shaded regions were excluded from the calculations of the exponents α in Table 1.

Figure 2:



Power Law

- **Extracting a value for the power-law exponent from distributions** like these can be a little tricky, since **it requires us to make a judgement**, **about the value x_{min} above which the distribution follows the power law**.
  Once this judgement is made, however, α can be calculated simply from equation below.

$$\alpha = 1 + n\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}\right]^{-1} \qquad (1)$$

- Here the quantities x_i, i = 1 . . . n are the measured values of x and x_{min} is again the minimum value of x.
- **Care must be taken to use the correct value of n in the formula**; **n is the number of samples that actually go into the calculation**, excluding those with values below x_{min}, not the overall total number of samples.
- In practical situations $\mathbf{x_{min}}$ usually corresponds not to the smallest value of x measured but to **the smallest value of x for which the power-law behaviour holds.**

Power Law

- An estimate of the expected statistical error on equation (1) is given by

$$\sigma = \sqrt{n}\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}\right]^{-1} = \frac{\alpha - 1}{\sqrt{n}}.$$

SOCIAL NETWORK ANALYTICS

Power Law

- **Table 1 lists the estimated exponents for each of the distributions of Fig. 2**, along with standard errors and also the values of x_{min} used in the calculations.
- Note that **the quoted errors correspond only to the statistical sampling error in the estimation of α**; they include no estimate of any errors introduced by the fact that a single power-law function may not be a good model for the data in some cases or for variation of the estimates with the value chosen for x_{min}.

	quantity	minimum $x_{\min}$	exponent α
(a)	frequency of use of words	1	2.20(1)
(b)	number of citations to papers	100	3.04(2)
(c)	number of hits on web sites	1	2.40(1)
(d)	copies of books sold in the US	2 000 000	3.51(16)
(e)	telephone calls received	10	2.22(1)
(f)	magnitude of earthquakes	3.8	3.04(4)
(g)	diameter of moon craters	0.01	3.14(5)
(h)	intensity of solar flares	200	1.83(2)
(i)	intensity of wars	3	1.80(9)
(j)	net worth of Americans	\$600m	2.09(4)
(k)	frequency of family names	10 000	1.94(1)
(l)	population of US cities	40 000	2.30(5)

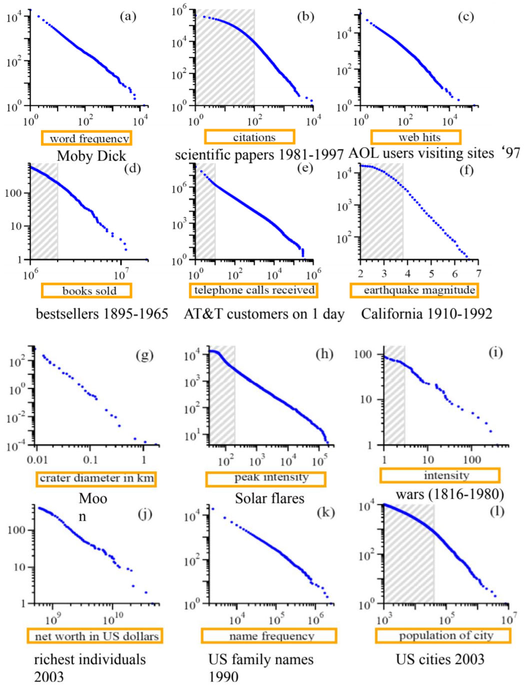
TABLE 1: Parameters for the distributions shown in Fig. 2.

Exponent values were calculated using the maximum likelihood method of Eq. (1). Numbers in parentheses give the standard error on the trailing figures.

$$\alpha = 1 + n\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}\right]^{-1} \qquad (1)$$

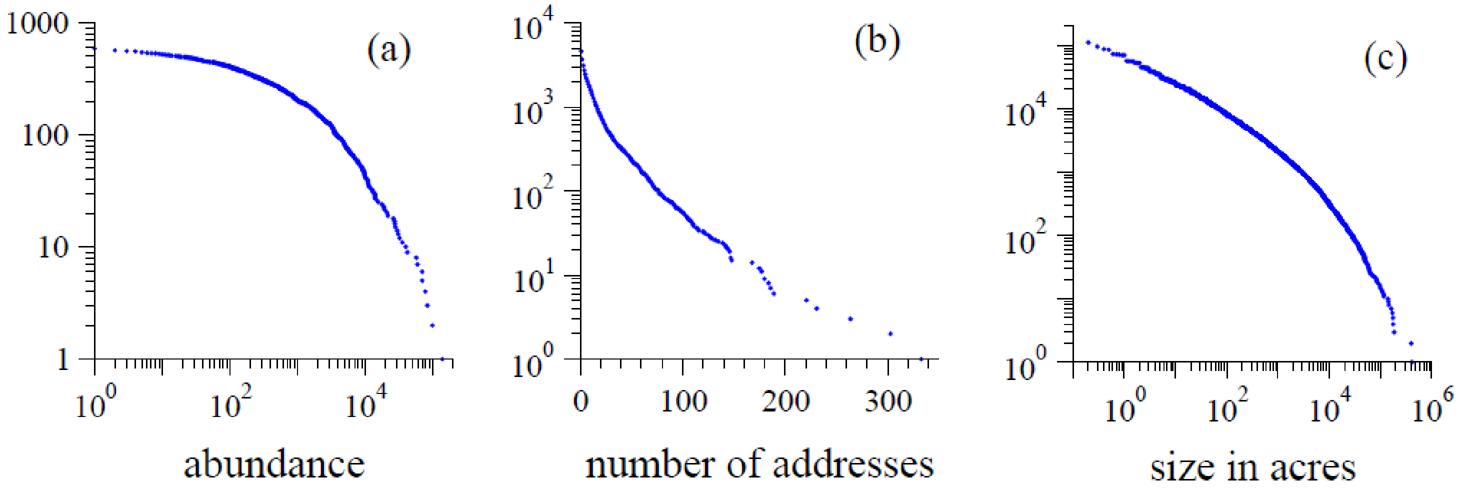
$$\sigma = \sqrt{n}\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}\right]^{-1} = \frac{\alpha - 1}{\sqrt{n}}.$$

Figure 2:



Power Law

- **Distributions that do not follow a power law**



- **FIG. 3:** Cumulative distributions of some quantities whose distributions span several orders of magnitude but that nonetheless do not follow power laws.
  (a) The number of sightings of 591 species of birds in the North American Breeding Bird Survey 2003.
  (b) The number of addresses in the email address books of 16 881 users of a large university computer system [33].
  (c) The size in acres of all wildfires occurring on US federal land between 1986 and 1996.
- Note that the horizontal axis is logarithmic in frames (a) and (c) but linear in frame (b).

Power Law in World Wide Web(WWW)

➢ **Power-Law Distribution of the World Wide Web** - Lada A. Adamic, Bernardo A. Huberman, A.-L. Barabási, R. Albert, H. Jeong, G. Bianconi

- We studied a crawl of 260,000 sites, each one representing a separate domain name. We counted how many links the sites received from other sites, and found that the distribution of links followed a power law.

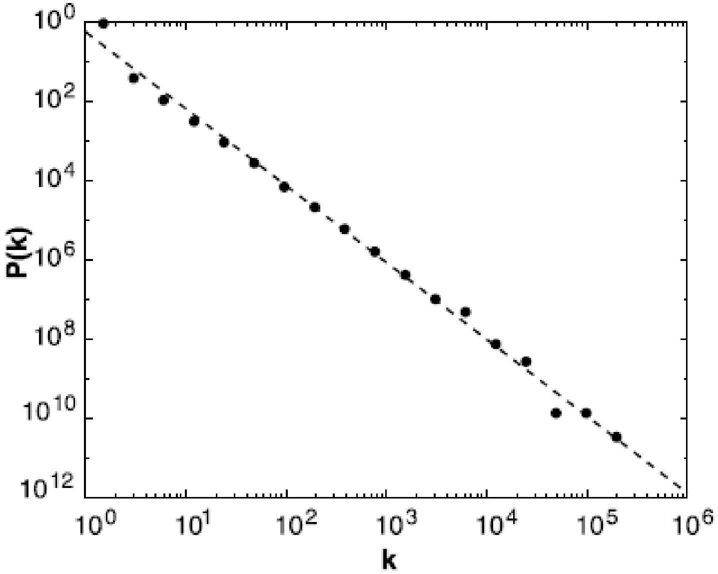


Fig. The distribution function for the number of links, k, to Web sites (from crawl in spring 1997). The dashed line has slope $\alpha = 1.94$.

Power Law in World Wide Web(WWW)

➢ **As a function of x, what fraction of pages on the Web have x in-links?**

- In studies over many different Web snapshots, taken at different points in the Web's history, the recurring finding is that **the fraction of Web pages that have x in-links is approximately proportional to x^{-2} or $1/x^2$** .

- The crucial point is that $1/x^2$ decreases much more slowly as x increases, so pages with very large numbers of in-links are much more common.

- For example, $1/x^2$ is only one in a million for k = 1000.

  A function that decreases as x to some fixed power, such as $1/x^2$ in the present case, is called a power law; when used to measure the fraction of items having value x, it says, qualitatively, that it's possible to see very large values of x.

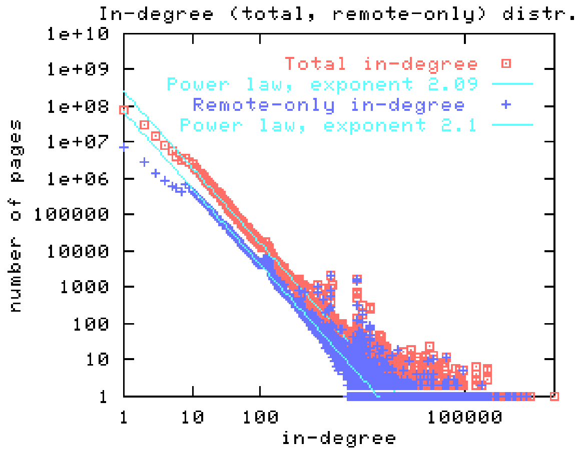


Figure: A power law distribution shows up as a straight line on a log-log plot.

➢The Power Law in Real Networks

Average k   Power law exponents

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}	ℓ_{real}	ℓ_{rand}	ℓ_{pow}	Reference
WWW	325 729	4.51	900	2.45	2.1	11.2	8.32	4.77	Albert, Jeong, and Barabási 1999
WWW	4×10^7	7		2.38	2.1				Kumar *et al.*, 1999
WWW	2×10^8	7.5	4000	2.72	2.1	16	8.85	7.61	Broder *et al.*, 2000
WWW, site	260 000				1.94				Huberman and Adamic, 2000
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2	4	6.3	5.2	Faloutsos, 1999
Internet, router*	3888	2.57	30	2.48	2.48	12.15	8.75	7.67	Faloutsos, 1999
Internet, router*	150 000	2.66	60	2.4	2.4	11	12.8	7.47	Govindan, 2000
Movie actors*	212 250	28.78	900	2.3	2.3	4.54	3.65	4.01	Barabási and Albert, 1999
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2	4	2.12	1.95	Newman, 2001b
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1	6	5.01	3.86	Barabási *et al.*, 2001
Co-authors, math.*	70 975	3.9	120	2.5	2.5	9.5	8.2	6.53	Barabási *et al.*, 2001
Sexual contacts*	2810			3.4	3.4				Liljeros *et al.*, 2001
Metabolic, *E. coli*	778	7.4	110	2.2	2.2	3.2	3.32	2.89	Jeong *et al.*, 2000
Protein, *S. cerev.**	1870	2.39		2.4	2.4				Jeong, Mason, *et al.*, 2001
Ythan estuary*	134	8.7	35	1.05	1.05	2.43	2.26	1.71	Montoya and Solé, 2000
Silwood Park*	154	4.75	27	1.13	1.13	3.4	3.23	2	Montoya and Solé, 2000
Citation	783 339	8.57			3				Redner, 1998
Phone call	53×10^6	3.16		2.1	2.1				Aiello *et al.*, 2000
Words, co-occurrence*	460 902	70.13		2.7	2.7				Ferrer i Cancho and Solé, 2001
Words, synonyms*	22 311	13.48		2.8	2.8				Yook *et al.*, 2001b

The Ubiquity of the Power Law

- **Power law distribution examples, in addition to technological and social networks**
  1. **The distribution of size of files in file systems**
  2. **The distribution of network latency in the Internet**
  3. **The networks of protein interactions (a few protein exists that interact with a large number of other proteins)**
  4. **The power of earthquakes**: statistical data tell us that the power of earthquakes follow a power-law distribution
  5. **The size of rivers**: the size of rivers in the world is is power law
  6. **The size of industries**, i.e., their overall income
  7. **The richness of people**

  In these examples, the exponent of the power law distribution is always around 2.5

PES UNIVERSITY ONLINE

Quiz:

$$p(x) = Cx^{-\alpha}$$

- As the exponent α increases, the downward slope of the line on a log-log plot

  1. stays the same
  2. becomes milder
  3. becomes steeper

Assignment: Paper Reading

1. **Power laws, Pareto distributions and Zipf's law** - M. E. J. Newman
2. **Power-Law Distribution of the World Wide Web** - Lada A. Adamic, Bernardo A. Huberman, A.-L. Barabási, R. Albert, H. Jeong, G. Bianconi

References

- Social Network Analysis: **Lada Adamic,** University of Michigan.
- Scale-Free Networks - Albert-Laszlo Barabasi and Eric Bonabeau
- Power laws, Pareto distributions and Zipf's law - M. E. J. Newman
- http://networksciencebook.com/
- Scale Free Networks - Franco Zambonelli
- Wikipedia – Current Literature

SOCIAL NETWORK ANALYTICS

Scale-Free Networks

Prakash C O
Department of Computer Science and Engineering

SOCIAL NETWORK ANALYTICS

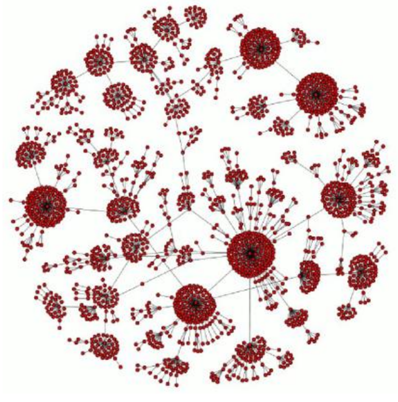
Scale-Free Networks

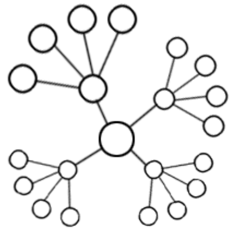
Prakash C O

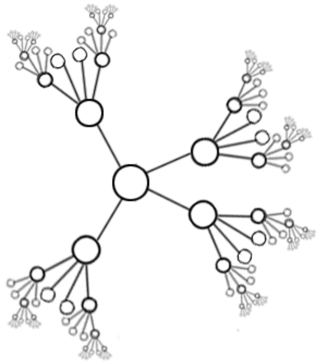
Department of Computer Science and Engineering

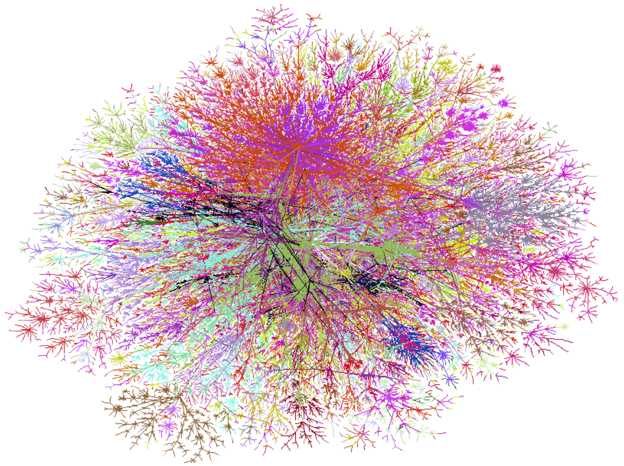
Scale-Free Network

- A variety of complex networks share an important property:
  - Some nodes have a tremendous number of connections to other nodes, whereas most nodes have just a handful.
  - The popular nodes, called hubs, can have hundreds, thousands or even millions of links. In this sense, **the network appears to have no scale.**









The Internet Routers

Scale-Free Network

- **Scale-free networks have certain important characteristics**.

  They are, for instance, robust against accidental failures but vulnerable to coordinated attacks.

- **Understanding of such characteristics could lead to new applications in many areas**.

  Example: Devising more effective strategies for preventing computer viruses in Internet.

Scale-Free Network

- **A network that has a power-law degree distribution, regardless of any other structure, is called a scale-free network.**

- A **scale-free network** is a connected graph or network with the property that **the number of links x originating from a given node exhibits a power law distribution** $p(x) = C\ x^{-\alpha}$

Scale-Free Network

- **A scale-free network can be constructed by**
  - **Growth** and
  - **Preferential Attachment**

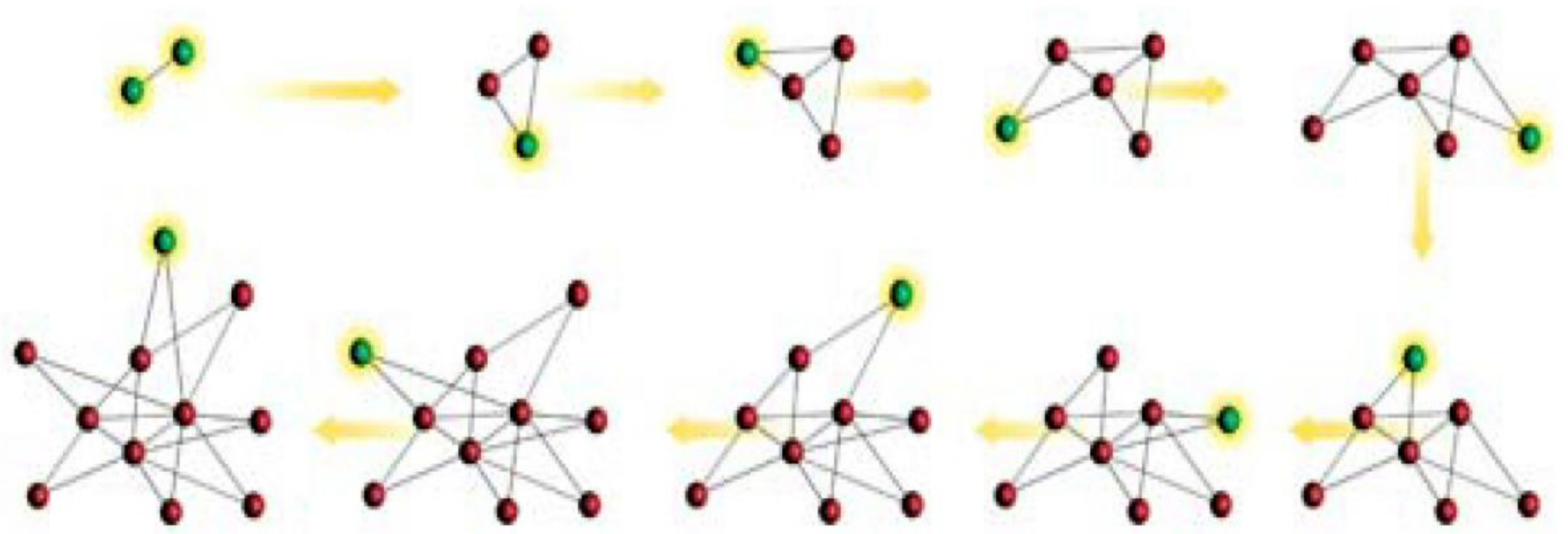
- **These two mechanisms explain the existence of hubs**: as new nodes appear, they tend to connect to the more connected sites, and these popular locations thus acquire more links over time than their less connected neighbors.

  And this "rich get richer" process will generally favor the early nodes, which are more likely to eventually become hubs.

Scale-Free Network

BIRTH OF A SCALE-FREE NETWORK

A SCALE-FREE NETWORK grows incrementally from two to 11 nodes in this example. When deciding where to establish a link, a new node (*green*) prefers to attach to an existing node (*red*) that already has many other connections. These two basic mechanisms—growth and preferential attachment—will eventually lead to the system's being dominated by hubs, nodes having an enormous number of links.



Features of Scale-free Networks

1. A greater number of nodes with high degree known as hubs, they appear as a result of preferential attachment.
   Hubs are Large in Scale-free Networks.

2. **The degree distribution follows a power law.**

3. Hubs usually have links from all around the network, serving as links between different parts of the network, therefore showing **a small world property**.

PES UNIVERSITY ONLINE

Properties of Scale-free Networks

- Scale-free networks have qualitatively different properties from strictly random, Erdos and Renyi, networks. These are:

1. Scale-free networks **display an amazing robustness against accidental failures, a property that is rooted in their inhomogeneous topology**.
   This means that the network is more likely to stay connected than a random network after the removal of randomly chosen nodes.

2. Scale-free networks are **more vulnerable against non-random attacks**. This means that the network quickly disintegrates when nodes are removed according to their degree.

3. Scale-free networks have **short average path lengths**. In fact the average path length goes as $L \sim \log N / \log\log k$

PES UNIVERSITY ONLINE

Examples:

- **Scale-free networks occur in many areas of science and engineering**, including

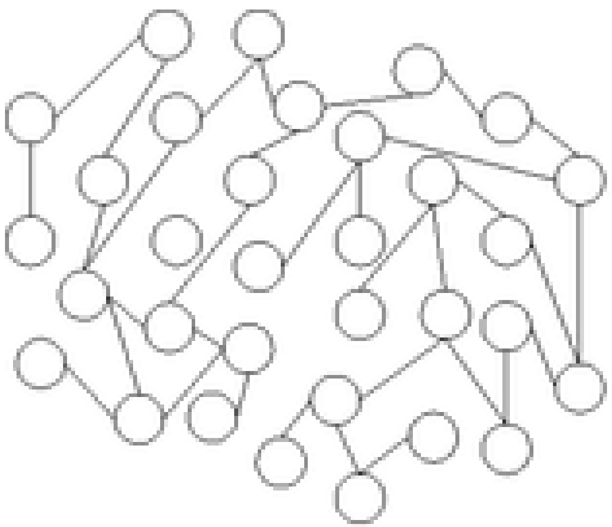
    1. the **topology of web pages** (where the nodes are individual web pages and the links are hyper-links),
    2. the **peer-reviewed scientific literature** (where the nodes are publications and the links are citations).
    3. the **collaborative network of Hollywood actors** (where the nodes actors and the links are co-stars in the same movie),

Scale-Free Network

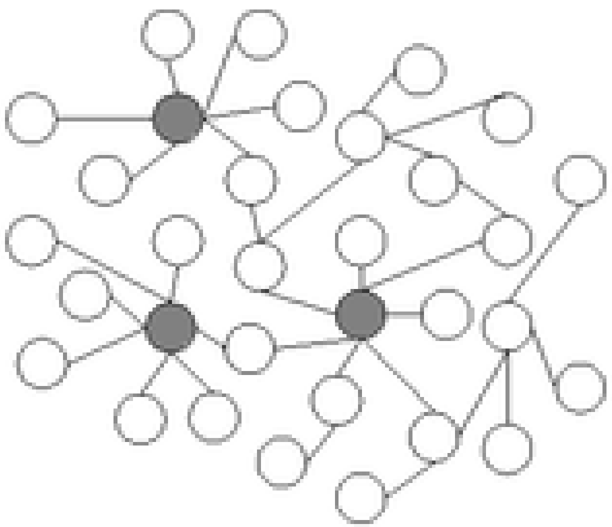
Examples of Scale-Free Networks

NETWORK	NODES	LINKS
Cellular metabolism	Molecules involved in burning food for energy	Participation in the same biochemical reaction
Hollywood	Actors	Appearance in the same movie
Internet	Routers	Optical and other physical connections
Protein regulatory network	Proteins that help to regulate a cell's activities	Interactions among proteins
Research collaborations	Scientists	Co-authorship of papers
Sexual relationships	People	Sexual contact
World Wide Web	Web pages	URLs

➢Random Vs. Scale-free networks

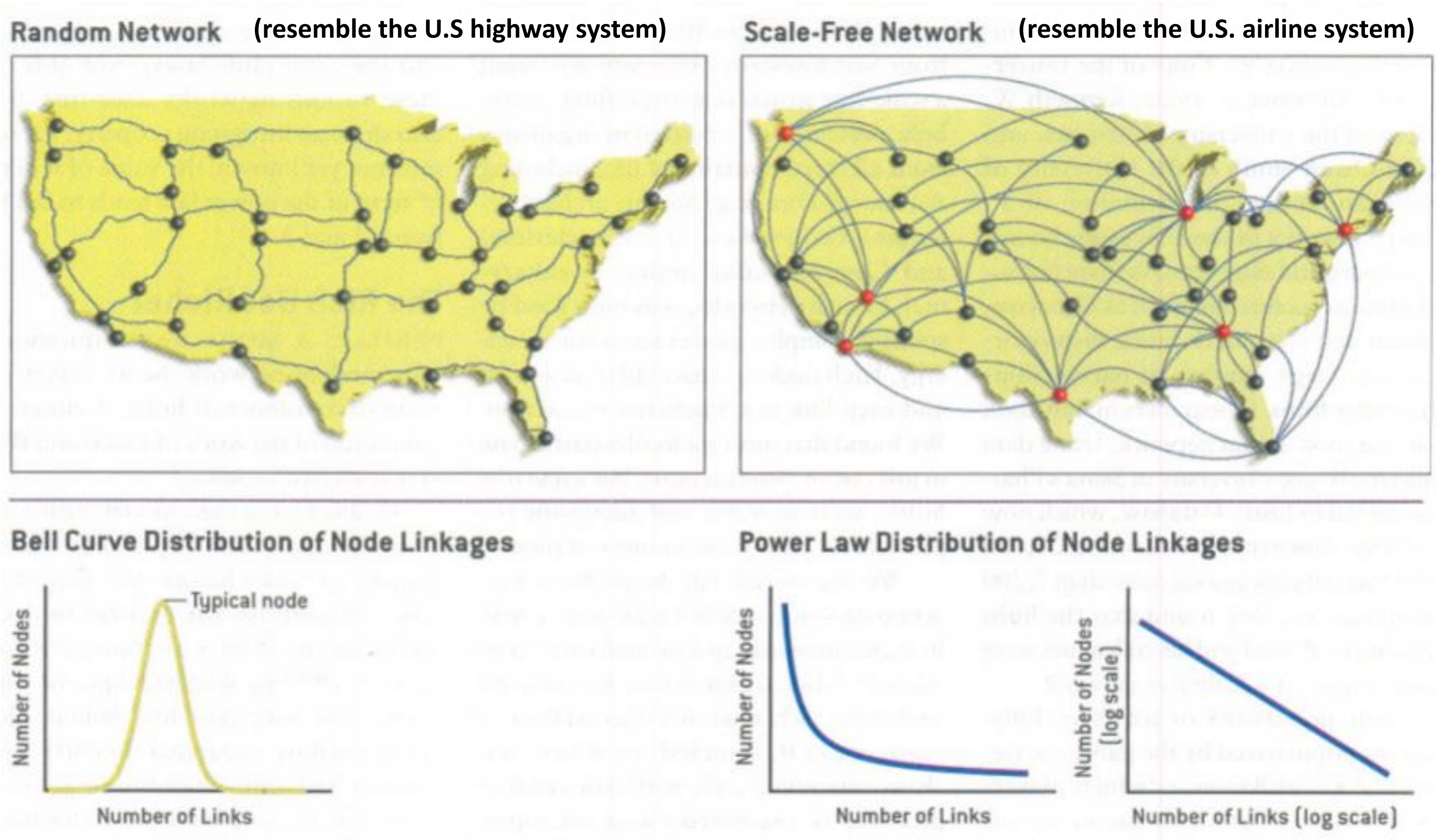


(a) Random network



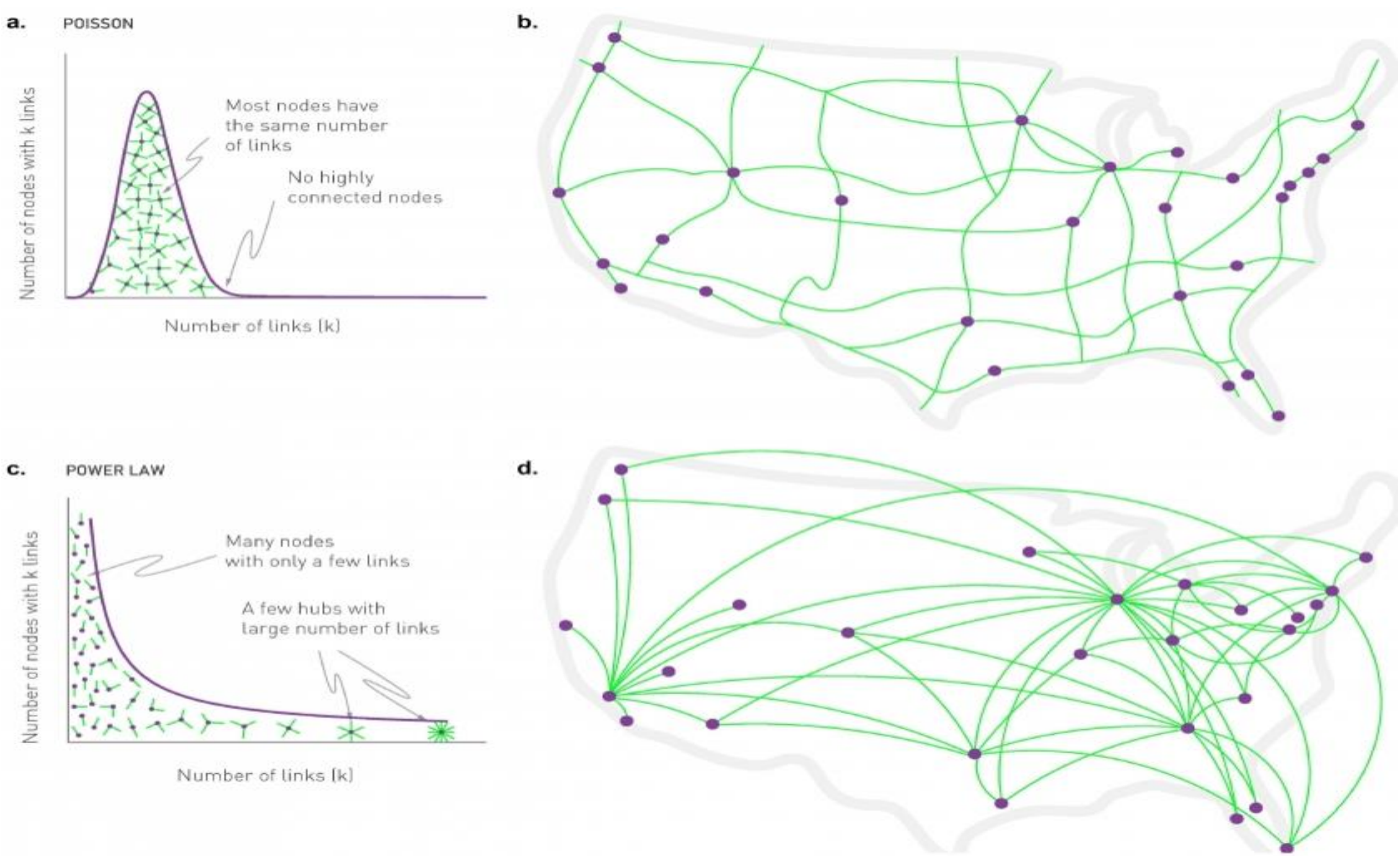
(b) Scale-free network

➢Random Vs. Scale-free networks



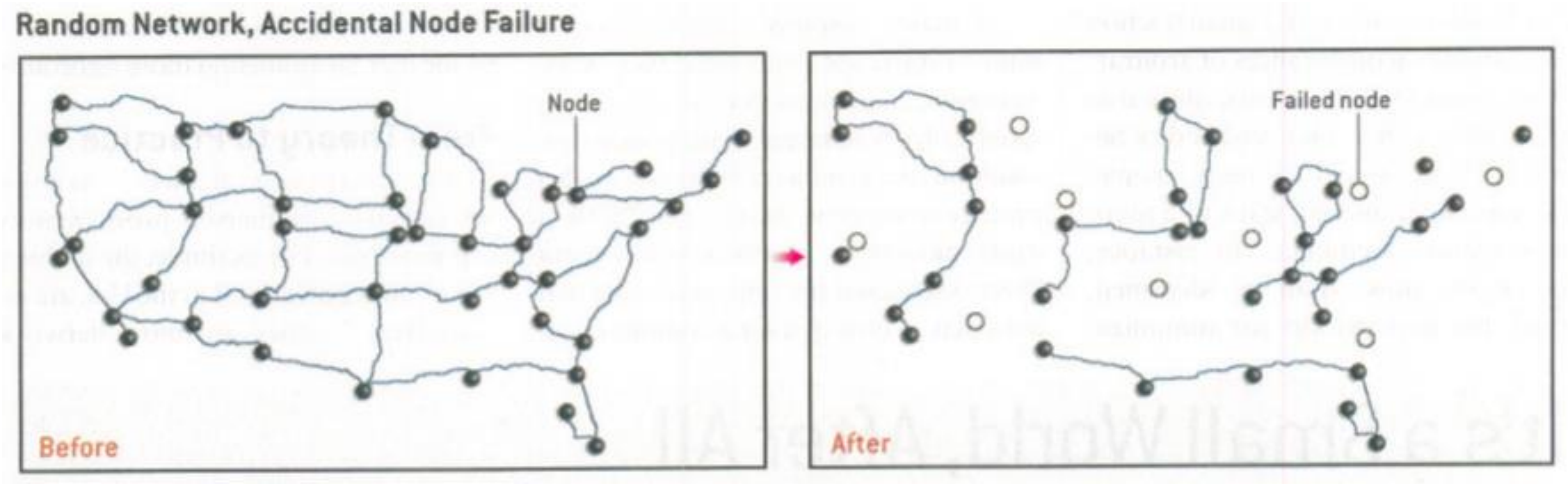
Scale-Free Network

➢Random Vs. Scale-free networks



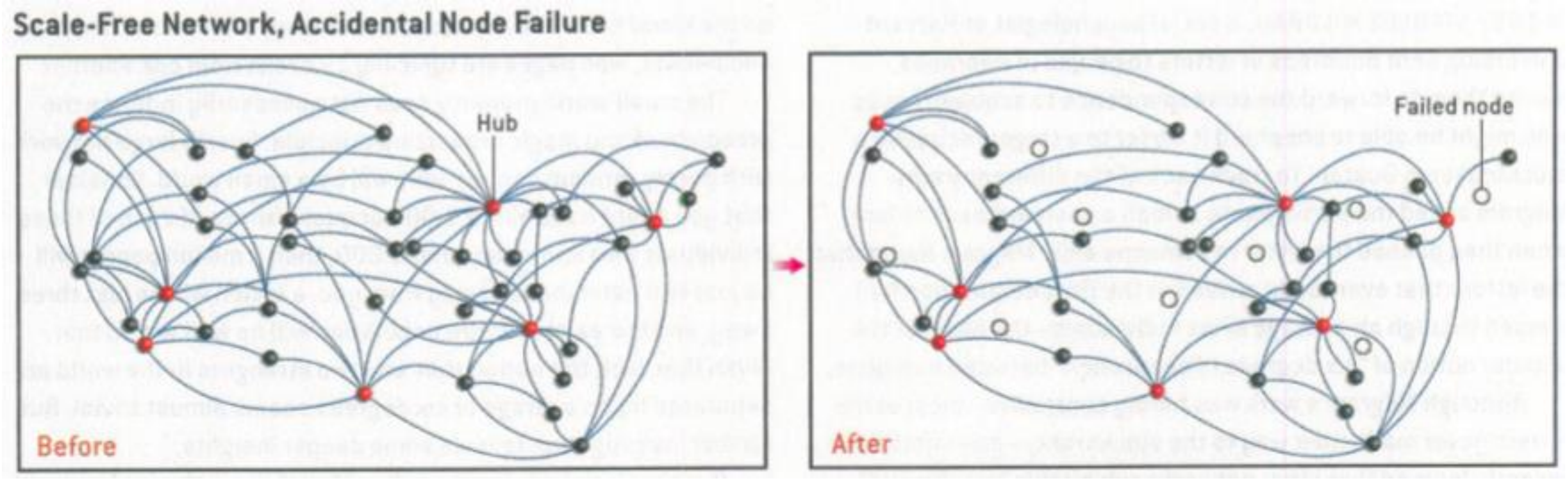
Scale-Free Network

➢How robust are Random and Scale-free networks?



For random networks: if a critical fraction of nodes is removed, these systems break into tiny, noncommunicating islands. **Yet simulations of scale-free networks tell a different story:** as many as 80 percent of randomly selected Internet routers can fail and the remaining ones will still form a compact cluster in which there will still be a path between any two nodes.

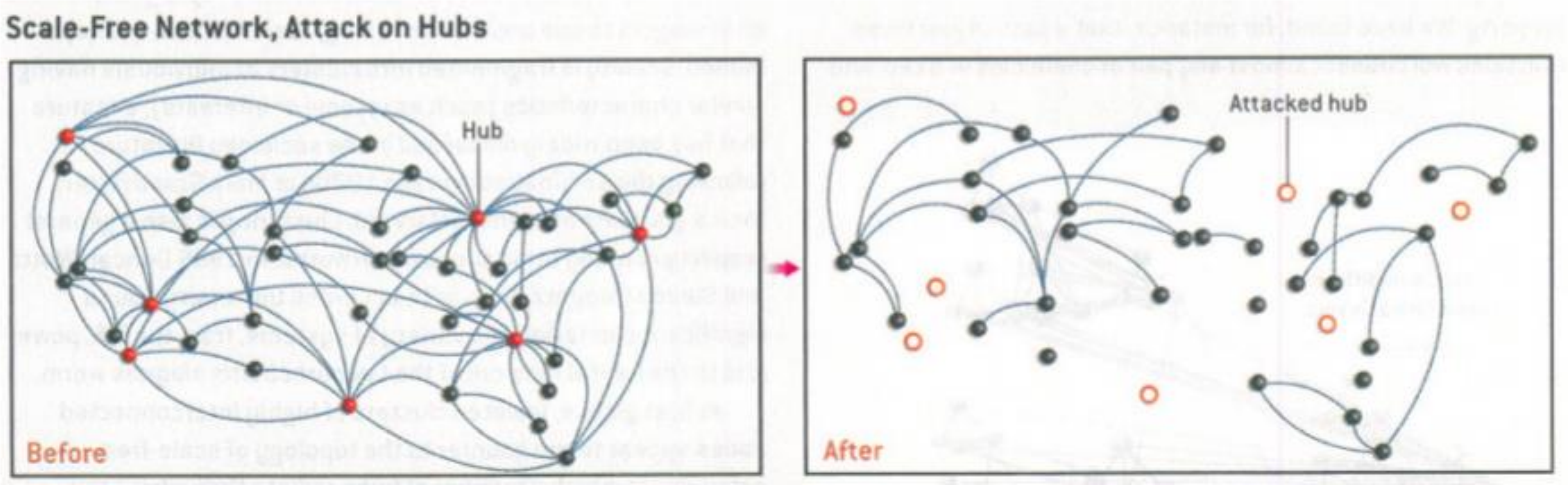
➢How robust are Random and Scale-free networks?



In general, scale-free networks display an amazing robustness against accidental failures.
The random removal of nodes will take out mainly the small ones because they are much more plentiful than hubs.
And the elimination of small nodes will not disrupt the network topology significantly, because they contain few links compared with the hubs, which connect to nearly everything.

Scale-Free Network

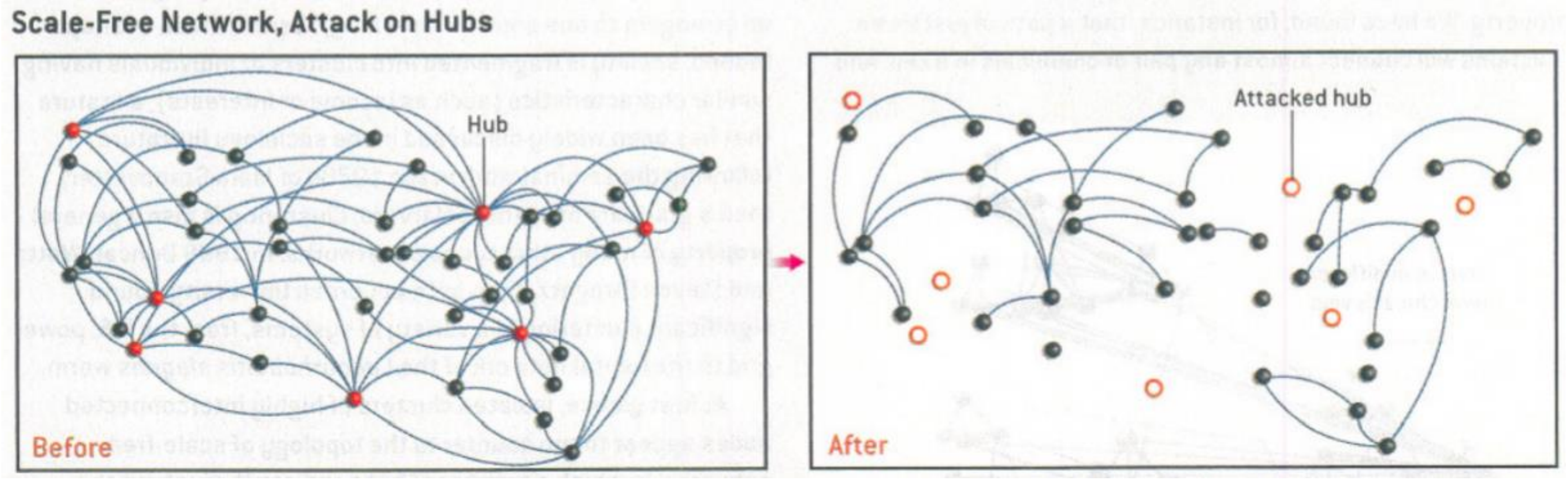
➢How robust are Random and Scale-free networks?



A reliance on hubs in Scale-free network has a serious drawback: vulnerability to attacks. In a series of simulations, we found that the removal of just a few key hubs from the Internet splintered the system into tiny groups of hopelessly isolated routers.

Scale-Free Network

- **How robust are Random and Scale-free networks?**



For the Internet, our experiments imply that a highly coordinated attack-first removing the largest hub, then the next largest, and so on-could cause significant disruptions after the elimination of just several hubs. Therefore, protecting the hubs is perhaps the most effective way to avoid large-scale disruptions caused by malicious cyber-attacks.

Scale-Free Network

The Potential Implications of Scale-Free Networks for...

Computing

- Computer networks with scale-free architectures, such as the World Wide Web, are highly resistant to accidental failures. But they are very vulnerable to deliberate attacks and sabotage.
- Eradicating viruses, even known ones, from the Internet will be effectively impossible.

Medicine

- Vaccination campaigns against serious viruses, such as smallpox, might be most effective if they concentrate on treating hubs—people who have many connections to others. But identifying such individuals can be difficult.
- Mapping out the networks within the human cell could aid researchers in uncovering and controlling the side effects of drugs. Furthermore, identifying the hub molecules involved in certain diseases could lead to new drugs that would target those hubs.

Business

- Understanding how companies, industries and economies are interlinked could help researchers monitor and avoid cascading financial failures.
- Studying the spread of a contagion on a scale-free network could offer new ways for marketers to propagate consumer buzz about their products.

SOCIAL NETWORK ANALYTICS

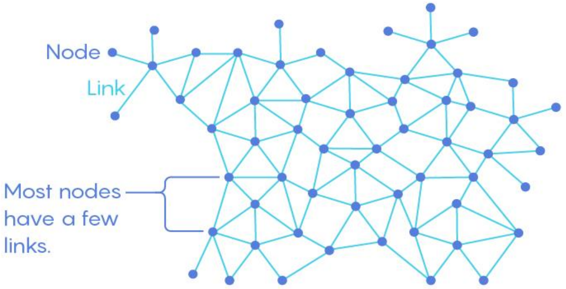
Scale-Free Network

To Be or Not to Be Scale-Free

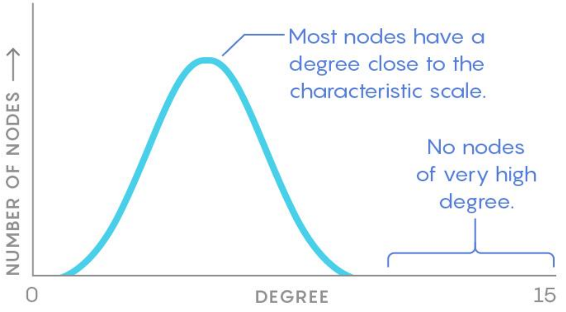
Scientists study complex networks by looking at the distribution of the number of links (or "degree") of each node.
Some experts see so-called scale-free networks everywhere. But a new study suggests greater diversity in real-world networks.

Random Network

Randomly connected networks have nodes with similar degrees. There are no (or virtually no) "hubs" — nodes with many times the average number of links.

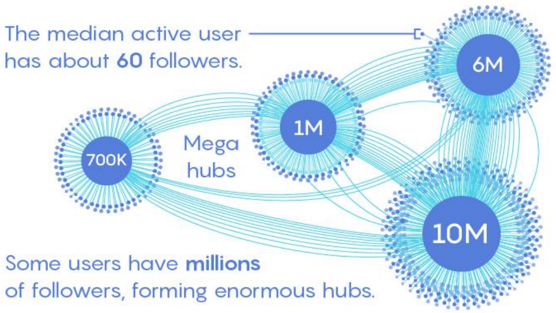


The distribution of degrees is shaped roughly like a bell curve that peaks at the network's "characteristic scale."

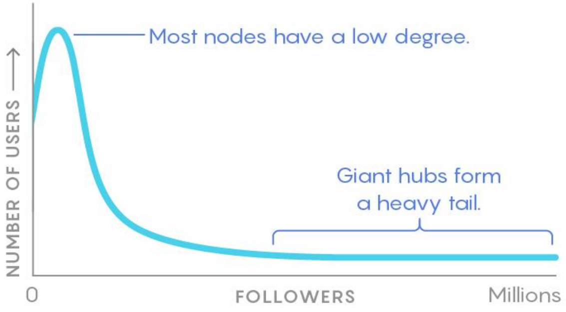


Twitter's Scale-Free Network

Most real-world networks of interest are not random. Some nonrandom networks have massive hubs with vastly higher degrees than other nodes.

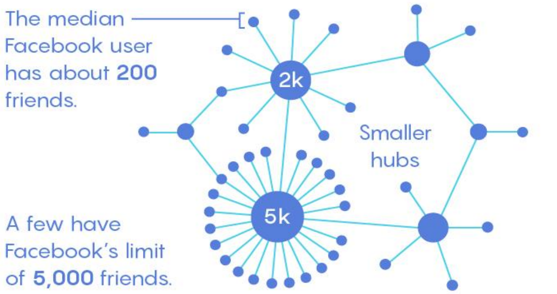


The degrees roughly follow a power law distribution that has a "heavy tail." The distribution has no characteristic scale, making it scale-free.

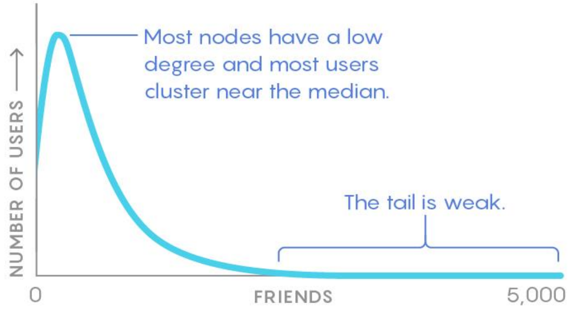


Facebook's In-Between Network

Researchers have found that most nonrandom networks are not strictly scale-free. Many have a weak heavy tail and a rough characteristic scale.



This network has fewer and smaller hubs than in a scale-free network. The distribution of nodes has a scale and does not follow a pure power law.



Scale-Free Network

$$p(x) = Cx^{-\alpha}$$

Table 1 A Choice of Some Real World Networks within the IS Domain that Share the Scale-Free Property [WaCh03]

Network	Size n	Clustering coefficient C	Degree exponent γ
Internet, domain level [VaPV02]	32711	0.24	2.1
Internet, router level [VaPV02]	228298	0.03	2.1
WWW [AlJB99]	153127	0.11	$\gamma_{in} = 2.1$ $\gamma_{out} = 2.45$
E-mail [EbMB02]	56969	0.03	1.81
Software [VaFS02]	1376	0.06	2.5

Scale-Free Network

Assignment: Paper Reading

- Scale-Free Networks - Albert-László Barabási and Eric Bonabeau

References

- Social Network Analysis: **Lada Adamic,** University of Michigan.
- Scale-Free Networks - Albert-Laszlo Barabasi and Eric Bonabeau
- Power laws, Pareto distributions and Zipf's law - M. E. J. Newman
- http://networksciencebook.com/
- Scale Free Networks - Franco Zambonelli
- Wikipedia – Current Literature

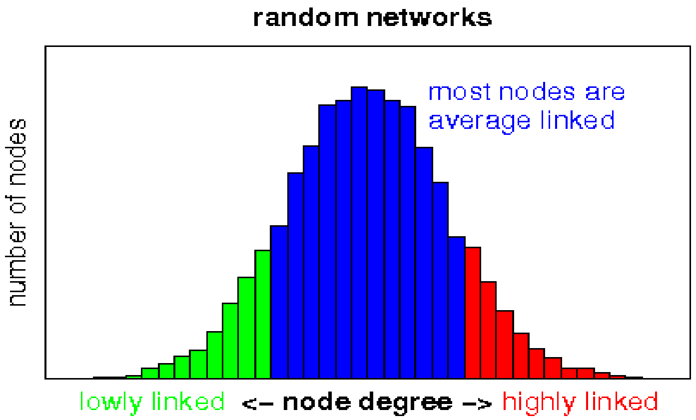
THANK YOU

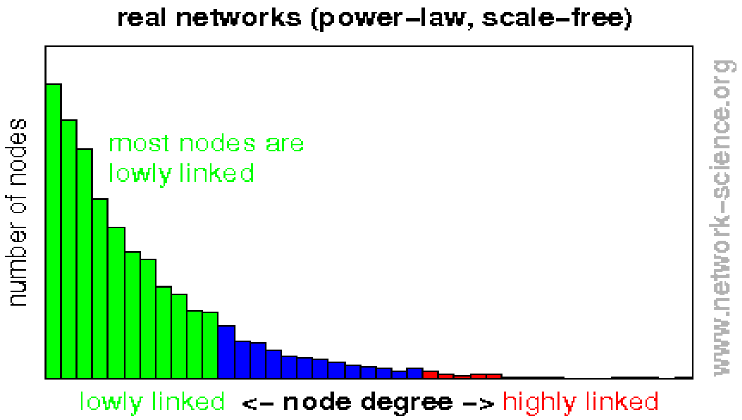
Prakash C O
Department of Computer Science and Engineering

coprakasha@pes.edu

+91 98 8059 1946

Degree distributions: Random Vs. Real-world networks.





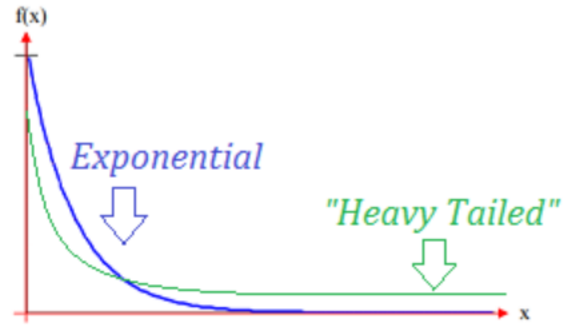
In **traditional random networks** most nodes have a medium node degree. The **degrees of all nodes are distributed around the average**.

Real networks often show a **skewed node-degree distribution** in which **most nodes have only few links** but, by contrast, **there exist some nodes which are extremely linked**. A highly connected node, a node with remarkably high degree, is called **hub.**

[Source: http://www.network-science.org/powerlaw_scalefree_node_degree_distribution.html](http://www.network-science.org/powerlaw_scalefree_node_degree_distribution.html)

Fat-tailed distribution

- A **fat-tailed distribution** is a probability distribution that exhibits a large skewness or kurtosis, relative to that of either a normal distribution or an exponential distribution.
- In common usage, the term fat-tailed and heavy-tailed are synonymous.
- A heavy tailed distribution has tails that are heavier than an exponential distribution (Bryson, 1974). In other words, the tails simply look fatter. As the tails have more bulk, the probability of extreme events is higher compared to the normal.
- The class of fat-tailed distributions includes those whose tails decay like a power law, which is a common point of reference in their use in the scientific literature.



Power Law - Statistics for real networks

	Network	Type	n	m	c	S	ℓ	α	C	C_{WS}	
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980	3.48	2.3	0.20	0.78	0.:
	Company directors	Undirected	7 673	55 392	14.44	0.876	4.60	–	0.59	0.88	0.:
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822	7.57	–	0.15	0.34	0.
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838	6.19	–	0.45	0.56	0.:
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918	4.92	–	0.088	0.60	0.
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16			2.1			
	Email messages	Directed	59 812	86 300	1.44	0.952	4.95	1.5/2.0		0.16	
	Email address books	Directed	16 881	57 029	3.38	0.590	5.22	–	0.17	0.13	0.
	Student dating	Undirected	573	477	1.66	0.503	16.01	–	0.005	0.001	−0.
	Sexual contacts	Undirected	2 810					3.2			
Information	WWW nd.edu	Directed	269 504	1 497 135	5.55	1.000	11.27	2.1/2.4	0.11	0.29	−0.
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	0.914	16.18	2.1/2.7			
	Citation network	Directed	783 339	6 716 198	8.57			3.0/–			
	Roget's Thesaurus	Directed	1 022	5 103	4.99	0.977	4.87	–	0.13	0.15	0.
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000		2.7		0.44	
ogical	Internet	Undirected	10 697	31 992	5.98	1.000	3.31	2.5	0.035	0.39	−0.
	Power grid	Undirected	4 941	6 594	2.67	1.000	18.99	–	0.10	0.080	−0.
	Train routes	Undirected	587	19 603	66.79	1.000	2.16	–		0.69	−0.

Power Law in World Wide Web(WWW)

- This study provides a quantitative form where popularity seems to exhibit extreme imbalances, with very large values likely to arise.
  And **it accords with our intuition about the Web, where there are certainly a reasonably large number of extremely popular pages**

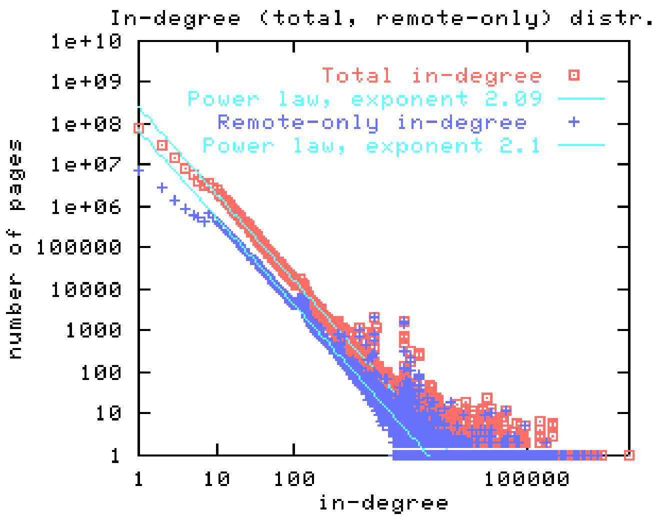


Figure: A power law distribution (the number of Web page in-links) shows up as a straight line on a log-log plot.