Math C5.4, Networks
Week 1b

R. Lambiotte
renaud.lambiotte@maths.ox.ac.uk

UNIVERSITY OF OXFORD

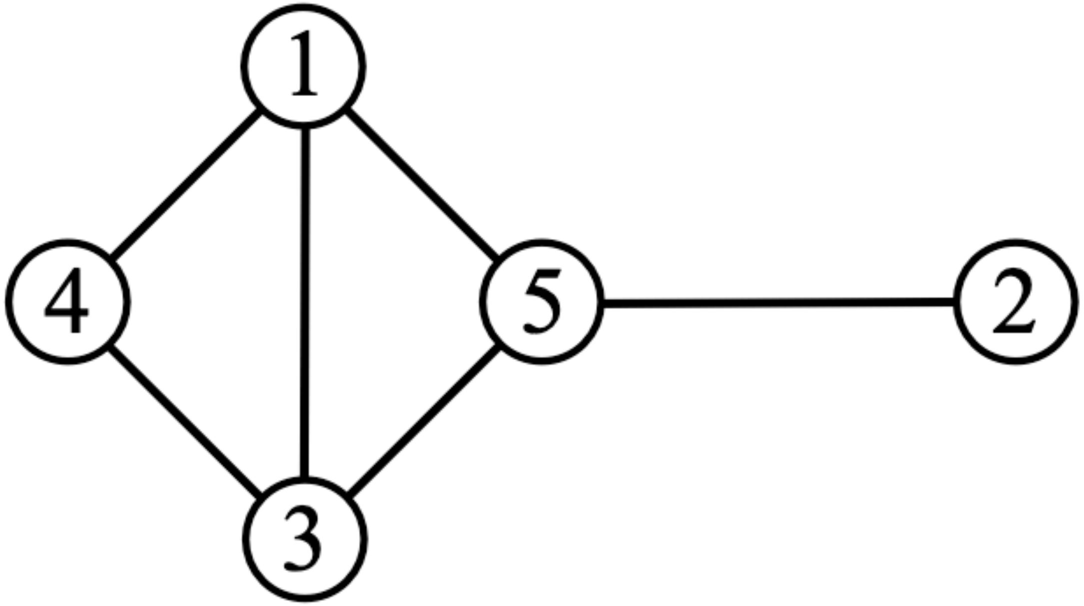
1. **What is a graph/network?**

   A network is a system made of nodes/vertices connected by links/edges. Links can be undirected or directed, and unweighted or weighted. In the mathematical literature, a network is called a graph. It is defined as

   $$\mathcal{G} = (V, E), \tag{1}$$

   where V is a set of nodes and E is a set of links. The number of nodes and that of links are denoted by N and M.

   Each link is defined by a pair of nodes, i.e., $e = (v, v') \in E$. In the case of undirected networks, the order of v and v' does not matter. In the case of directed networks, (v, v') indicates a link from v to v', and if $(v, v') \in E$ and $(v', v) \in E$, the two nodes are reciprocally connected. In the case of weighted networks, links are also assigned with a weight function, characterising the importance or weight of the link.



Example of an undirected and unweighted network with $N = 5$ nodes and $M = 6$ links.

2. **How can we store a graph? 1. Adjacency matrix**

In order to efficiently store networks and to carry out computations, it is necessary to use appropriate data structure. Each representation emphasises a certain aspect of the network and is amenable to certain types of computational or mathematical operations.

A network can be represented by the corresponding $N \times N$ adjacency matrix. Being adjacent means that two nodes are directly connected by a link. In the case of unweighted networks, the entries of the adjacency matrix are given by

$$A_{ij} = \begin{cases} 1 & \text{if node } v_i \text{ is adjacent to node } v_j, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

If the network is weighted, A_{ij} can take positive values different from one, representing the weight of the link. In general, undirected and directed networks will yield symmetric and asymmetric adjacency matrices, respectively. The adjacency matrix of the network illustrated in Fig. 1 is given by

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix}. \tag{3}$$

An adjacency matrix representation is useful for formulating and theoretically analysing structure of, and dynamical processes on networks. In particular, it is amenable to tools from linear algebra such as the analysis of eigenvectors and eigenvalues. A drawback of this representation is its memory cost because a network possessing N nodes requires $O(N^2)$ elements for storage. A majority of networks found in the real world and generated from models are sparse such that most elements of the adjacency matrix are equal to zero. Therefore, it is often preferable to use the data structure called sparse matrix. Its advantage is a significant gain in memory and faster computations because operations involving zeros are not executed.

3. How can we store a graph? 2. Link list

A matrix formulation is often unsuitable even if sparse matrix representations are employed. This is the case when, for example, all links have to be scanned repeatedly, e.g. stochastic dynamics. A representation of static networks alternative to the adjacency matrix is called the link list. In this link-centric approach, a graph is described as a list of pairs of nodes, each corresponding to a link in the network, as follows:

$$\{(u_1, v_1), (u_2, v_2), \ldots, (u_M, v_M)\}. \tag{4}$$

When the network is directed, we interpret Eq. (4) as representing directed links from u_i to v_i ($1 \leq i \leq M$). The link list of the network shown in Fig. 1 is given by

$$\{(v_1, v_3), (v_1, v_4), (v_1, v_5), (v_2, v_5), (v_3, v_4), (v_3, v_5)\}. \tag{5}$$

Link lists have the additional advantage of being efficiently used for link randomisation and numerical simulations of dynamics on sparse networks.

4. Node degree and its distribution

The degree is defined as the number of links incident to a node. We denote the degree of the ith node by k_i. For undirected networks, the degree is given by

$$k_i = \sum_{j=1}^{N} A_{ij} \left(= \sum_{j=1}^{N} A_{ji} \right). \tag{6}$$

A network is called regular if all nodes have the same degree, i.e., $k_i = k_j$ for all i and j.

For directed networks, we distinguish the in-degree, i.e., the number of links incoming to the node, and the out-degree, i.e., the number of links outgoing from the node. They are given by $k_i^{\text{in}} = \sum_{j=1}^{N} A_{ji}$ and $k_i^{\text{out}} = \sum_{j=1}^{N} A_{ij}$, respectively. These numbers are basic, local measures of the centrality, or importance, of a node in a network.

Each link has two endpoints and hence contributes to the degree of two nodes by one each. Therefore, we obtain

$$\sum_{i=1}^{N} k_i = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} = 2M \tag{7}$$

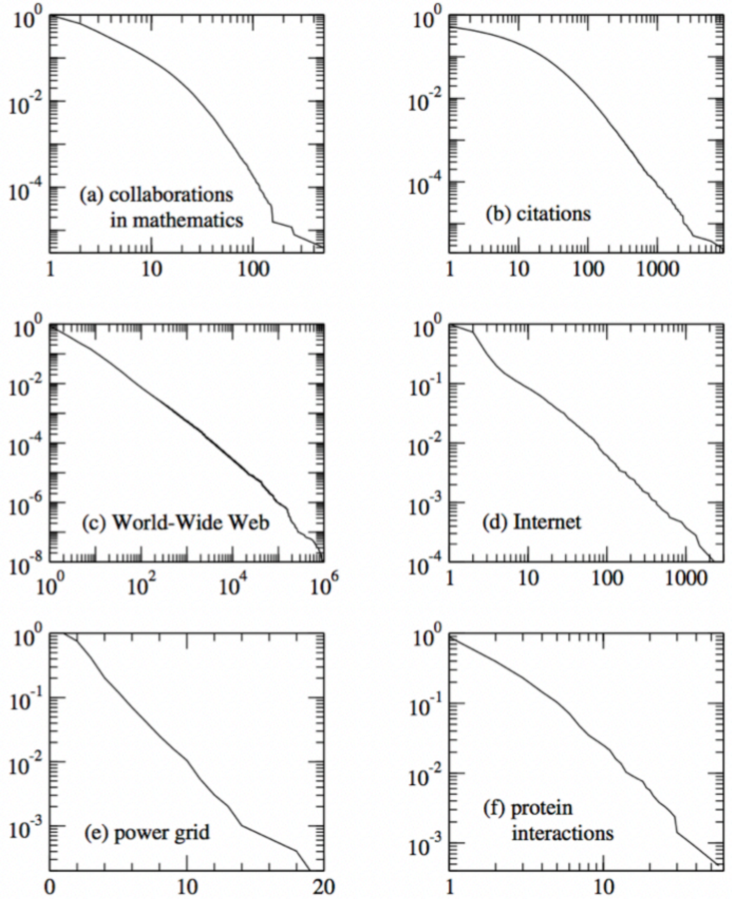
for undirected networks. Equation (7) is called the handshaking lemma. It implies that the sum of the degrees of all the nodes in any undirected network is an even number.

The degree distribution of a network is the frequency distribution of the degree and denoted by $p(k)$. A majority of networks in different domains possesses long-tailed degree distributions. In many situations, their tail is described by a power-law, i.e.,

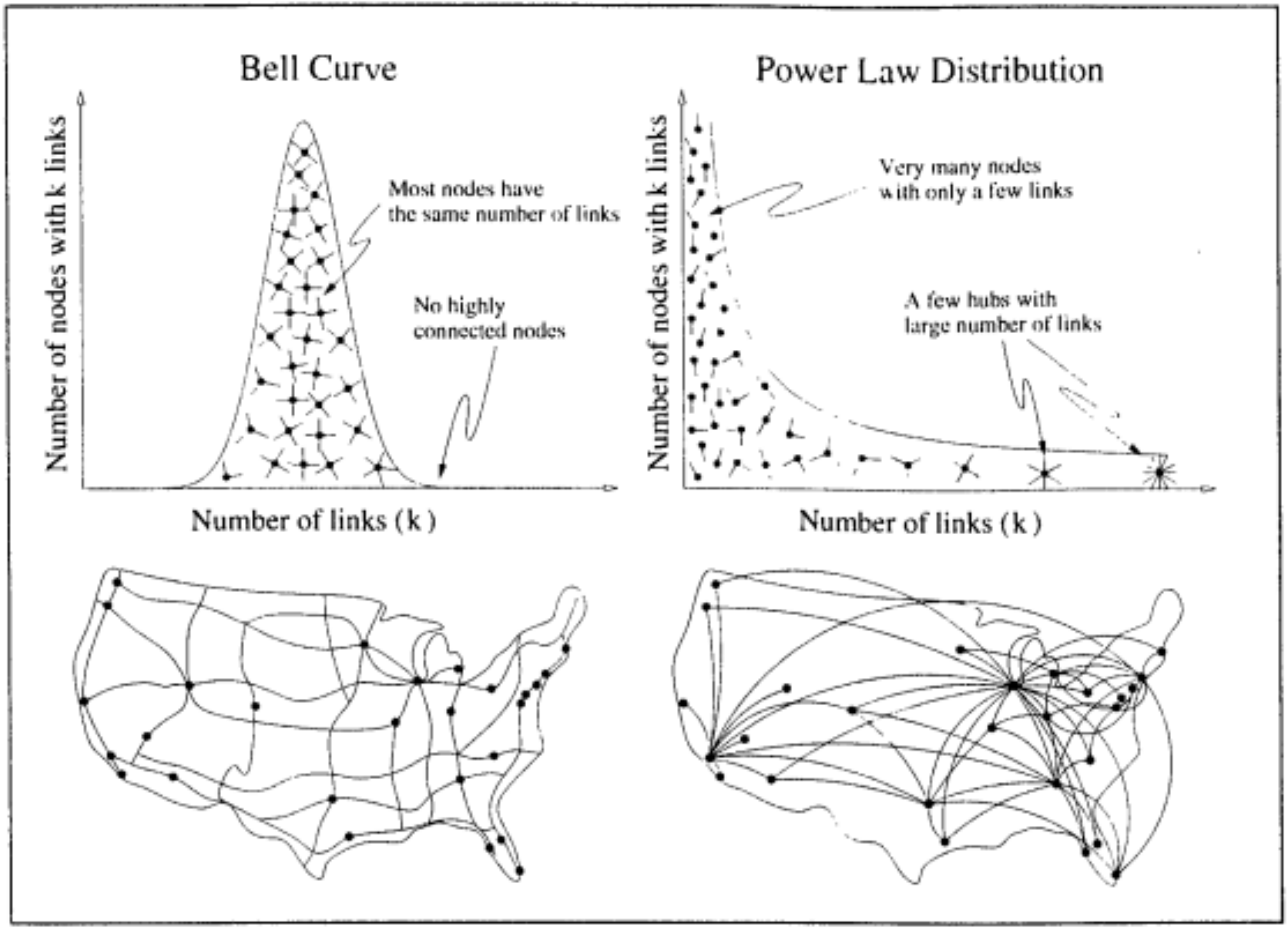
$$p(k) \propto k^{-\gamma}, \tag{8}$$

where γ is typically between two and three. Because the maximum degree is equal to $N-1$, Eq. (8) approximately holds true up to a certain cutoff degree, above which $p(k)$ rapidly decays to zero. The average degree, denoted by $\langle k \rangle$, is given by

$$\langle k \rangle = \sum_k k p(k). \tag{9}$$



Cumulative degree distributions for six different networks. The horizontal axis for each panel is vertex degree k (or in-degree for the citation and Web networks, which are directed) and the vertical axis is the cumulative probability distribution of degrees, i.e., the fraction of vertices that have degree greater than or equal to k. The networks shown are: (a) a collaboration network of mathematicians; (b) citations between 1981 and 1997 to all papers cataloged by the Institute for Scientific Information; (c) a 300 million vertex subset of the World Wide Web, *circa* 1999; (d) the Internet at the level of autonomous systems, April 1999; (e) the power grid of the western United States; (f) the interaction network of proteins in the metabolism of the yeast. Of these networks, three of them, (c), (d) and (f), appear to have power-law degree distributions, as indicated by their approximately straight-line forms on the doubly logarithmic scales. Taken from Newman, Mark EJ. "The structure and function of complex networks." SIAM review 45.2 (2003): 167-256.



The presence of hubs facilitates spreading: Pastor-Satorras, Romualdo, and Alessandro Vespignani. "Epidemic spreading in scale-free networks." *Physical review letters* 86.14 (2001): 3

Power-law distributions

Power-law distributions play a central role in network science and in the theory of complex systems in general.

Let us consider a power-law distribution for continuous variables, keeping in mind that most of the observations generalise to the case of discrete variables. Consider the Pareto distribution given by

$$p(x) = Cx^{-\alpha} \quad (x \geq x_{\min}), \tag{1}$$

where α is the power-law exponent of the distribution, $x_{\min}(> 0)$ is the minimum value taken by the random variable and $C = (\alpha - 1)x_{\min}^{\alpha-1}$ is the normalisation constant respecting

$$\int_{x_{\min}}^{\infty} Cx^{-\alpha}\mathrm{d}x = 1. \tag{2}$$

Other power-law distributions are by definition asymptotically (i.e., for large x) the same as Eq. (1) up to a normalisation constant.

Power-law distributions mainly differ from the exponential and Gaussian distributions by the significant mass of probability carried by their tail, i.e., large values of x. The exponential and Gaussian distributions have a characteristic scale such that the probability of observing instances many times larger than this scale is negligible. In contrast, under a power-law distribution, a vast majority of instances exhibits small values while few but non-negligible instances produce very large values. Power-law distributions are associated with a broad heterogeneity in the system and are said to have a fat or long tail, because the tail of the distribution is much more populated than in exponential-like distributions. Note that fat tails are also present in distributions without a power-law tail. Examples include stretched exponential distributions and log-normal distributions.

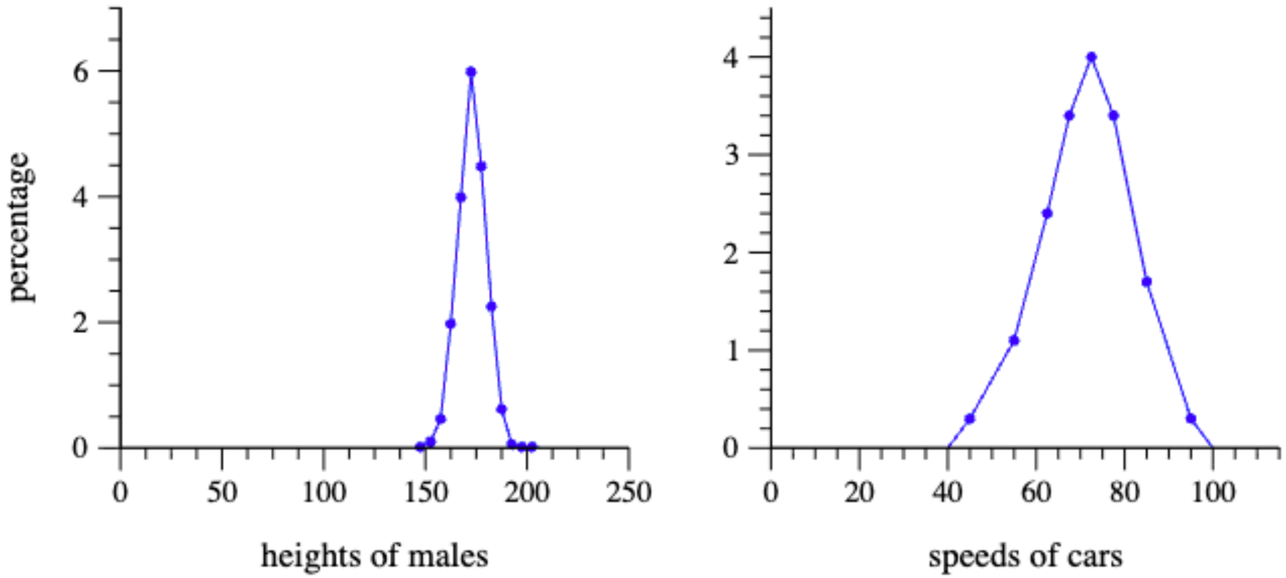


FIG. 1 Left: histogram of heights in centimetres of American males. Data from the National Health Examination Survey, 1959–1962 (US Department of Health and Human Services). Right: histogram of speeds in miles per hour of cars on UK motorways. Data from Transport Statistics 2003 (UK Department for Transport).

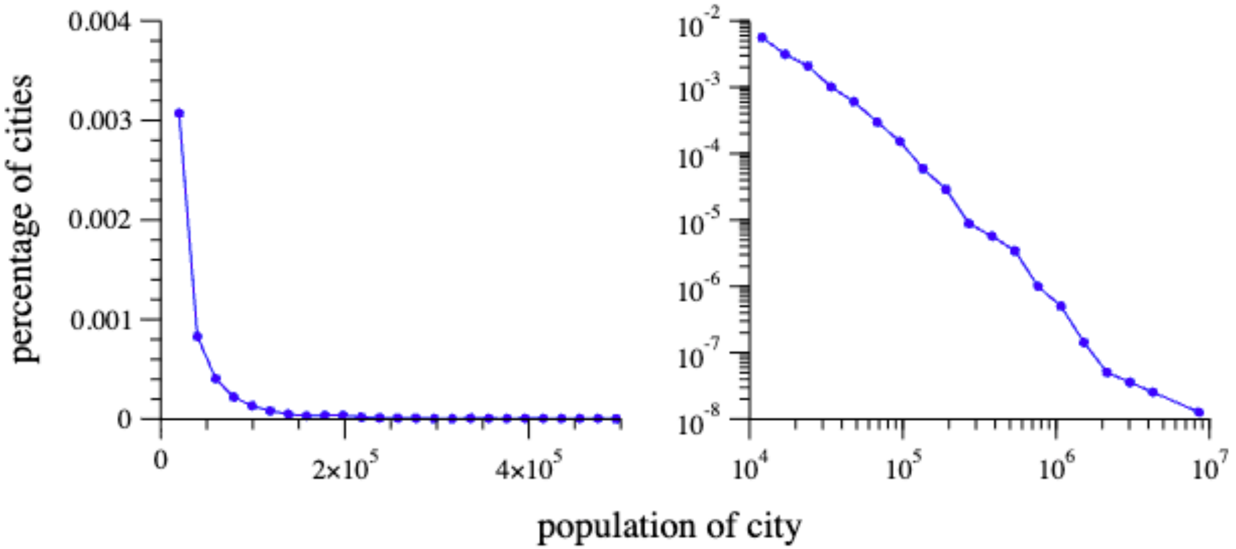
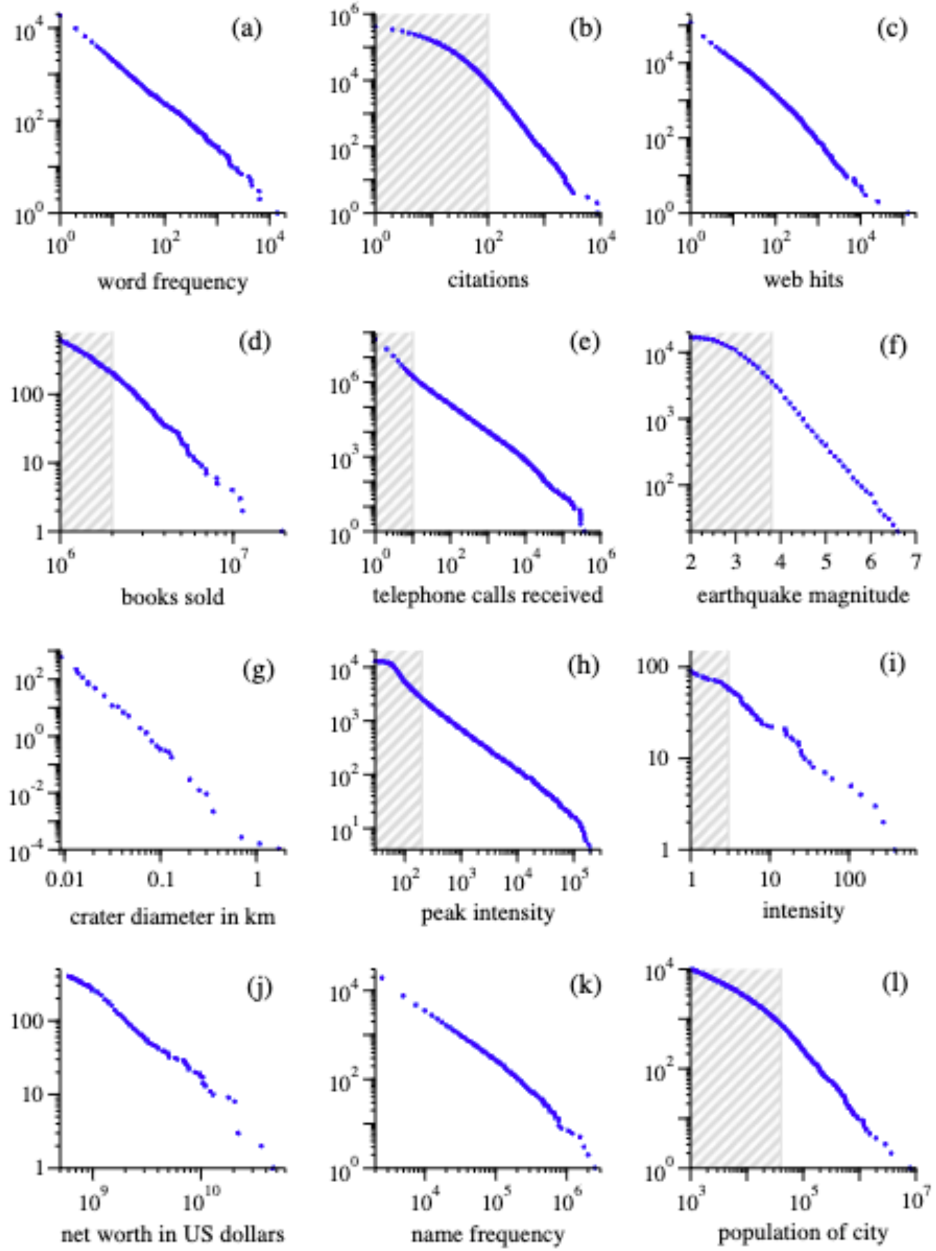


FIG. 2 Left: histogram of the populations of all US cities with population of 10 000 or more. Right: another histogram of the same data, but plotted on logarithmic scales. The approximate straight-line form of the histogram in the right panel implies that the distribution follows a power law. Data from the 2000 US Census.



Power-law distributions are typically found in the wealth of individuals, populations of cities, the frequency of words in text, sales of books and music, citations that a scientific paper receives and so forth. Since the advent of the Pareto distribution and the associated Zipf's law, power-law distributions have been studied over a century.

The moments of power-law distributions are given by

$$\langle x^\beta \rangle = \int_{x_{\min}}^{\infty} x^\beta p(x) \mathrm{d}x = \frac{\alpha - 1}{\alpha - 1 - \beta} x_{\min}^\beta \quad (\beta < \alpha - 1). \tag{3}$$

The moments for $\beta \geq \alpha - 1$ are divergent. In particular, the mean $\langle x \rangle$ does not exist for $1 < \alpha \leq 2$, and the variance does not exist for $2 < \alpha \leq 3$. These features impact various structural and dynamical properties of complex systems including networks, as we will see throughout these notes. When $\alpha \leq 1$, the distribution is ill-defined because $\int_{x_{\min}}^{\infty} p(x) \mathrm{d}x$ is divergent such that $p(x)$ cannot be normalised. When a moment, $\langle x^\beta \rangle$, diverges, its empirical measurement diverges as the number of samples increases and $\langle x^\beta \rangle$ with β only slightly smaller than $\alpha - 1$ converges very slowly. Both the divergence and slow convergence of moments are due to the appearance of extreme values. For example, the sample mean for the power-law distribution with $\alpha = 2$ diverges as we accumulate samples.

In a majority of empirical data, the distribution can be close to Eq. (1) only in a certain range of the variable. However, key observations such as the divergent moments hold true as long as a distribution behaves the same as Eq. (1) when $x \to \infty$ up to a normalisation constant. For example, the Cauchy distribution given by $p(x) = 1/\left[\pi(1 + x^2)\right]$ is qualitatively the same as Eq. (1) with $\alpha = 2$ as $x \to \infty$. It should also be noted that the tail of an empirical distribution ceases to be a power-law beyond a certain scale because of the finiteness of the system. The finite size effect typically leads to exponential cut-offs. Therefore, the power-law regime, if present, usually dominates for values that are neither too small nor too large.

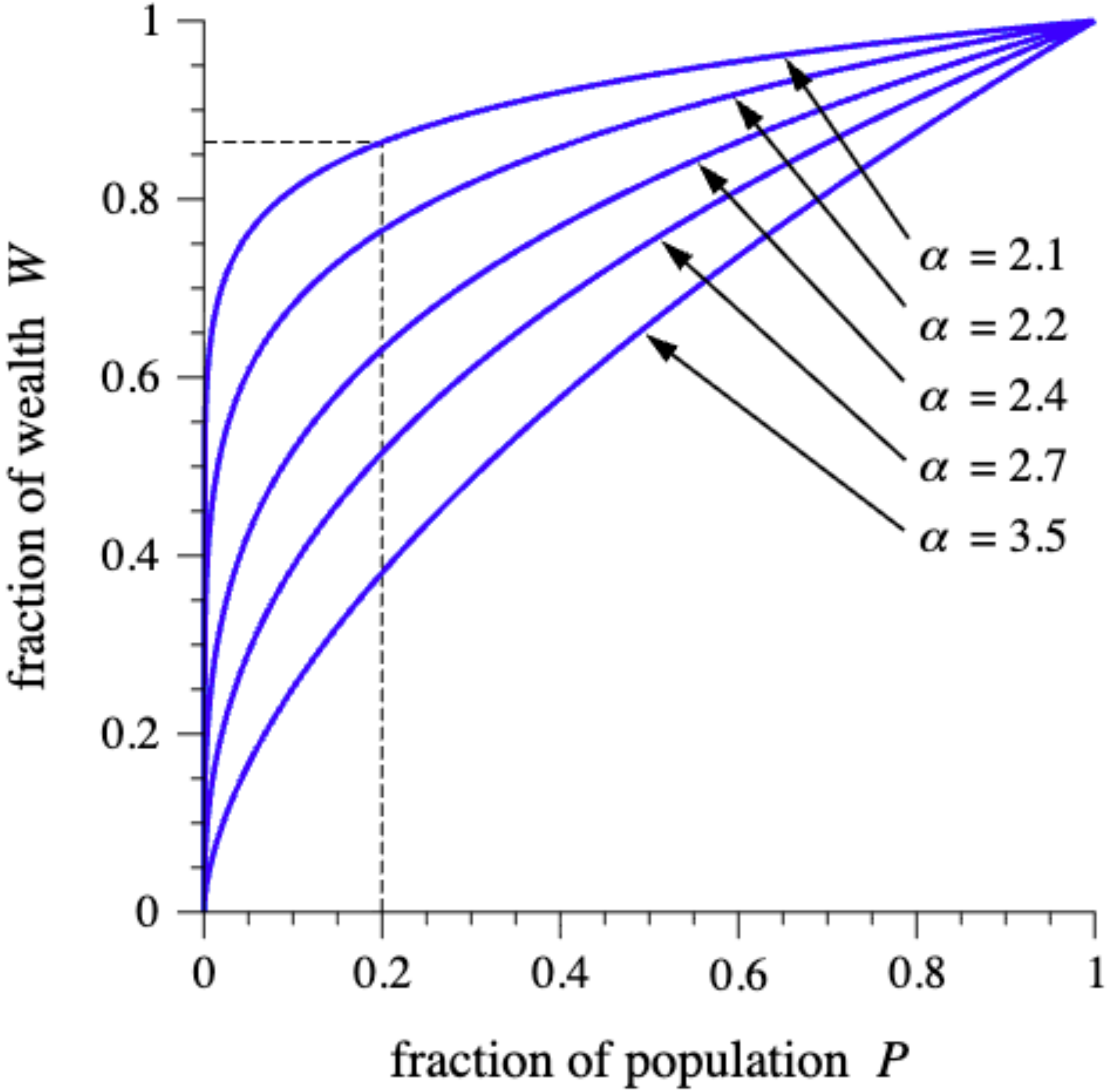
The heterogeneity of power-law distributions is often associated with the presence of inequalities in the system. What fraction w of the total wealth is held by a certain fraction of the richest people when the wealth distribution is given by Eq. (1)? To answer this question, let us first calculate the fraction of the people whose wealth is at least x_0:

$$p(x \geq x_0) = \int_{x_0}^{\infty} Cx^{-\alpha}\mathrm{d}x = \left(\frac{x_0}{x_{\mathrm{min}}}\right)^{-\alpha+1}. \tag{4}$$

The fraction of wealth held by these richest people is given by

$$\begin{aligned} w(x_0) &= \frac{\int_{x_0}^{\infty} x \cdot Cx^{-\alpha}\mathrm{d}x}{\int_{x_{\mathrm{min}}}^{\infty} x \cdot Cx^{-\alpha}\mathrm{d}x} = \left(\frac{x_0}{x_{\mathrm{min}}}\right)^{-\alpha+2} \\ &= \left[p(x \geq x_0)\right]^{\frac{\alpha-2}{\alpha-1}}, \end{aligned} \tag{5}$$

where we have assumed that $\alpha > 2$ so that the average wealth is finite. Equation (5) neither depends on x_0 nor x_{min} explicitly, and it provides a direct relation between $w(x_0)$ and $p(x \geq x_0)$. This relation is often called the "80-20 rule", anecdotally meaning that 80% of the wealth is in the hands of the richest 20%. More precisely, setting $p(x \geq x_0) = 0.2$, $w(x_0) = 0.2^{(\alpha-2)/(\alpha-1)}$ can take any value between 0.2 and 1 depending on the value of α. In the limit $\alpha \to \infty$, the system does not exhibit a power-law tail, and we obtain $w(x_0) = 0.2$. In this case, the system is egalitarian. As α decreases, the tail of the distribution becomes fat and inequality grows. In the extreme situation with $\alpha \to 2$, the total wealth belongs to an infinitesimally small fraction of the richest people. In the econometrics literature, the measurement of this effect in empirical data can be done with the Gini coefficient.



1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
fraction of wealth W
fraction of population P
α = 2.1
α = 2.2
α = 2.4
α = 2.7
α = 3.5

Richest 1% bag nearly twice as much wealth as the rest of the world put together over the past two years

Published: 16th January 2023

- *Super-rich outstrip their extraordinary grab of half of all new wealth in past decade.*
- *Billionaire fortunes are increasing by $2.7 billion a day even as at least 1.7 billion workers now live in countries where inflation is outpacing wages.*

- *A tax of up to 5 percent on the world's multi-millionaires and billionaires could raise $1.7 trillion a year, enough to lift 2 billion people out of poverty.*

The richest 1 percent grabbed nearly two-thirds of all new wealth worth $42 trillion created since 2020, almost twice as much money as the bottom 99 percent of the world's population, reveals a new Oxfam report today. During the past decade, the richest 1 percent had captured around half of all new wealth.

- Power-law distributions are scale-invariant because they satisfy

$$p(c_1 x) = c_2 p(x) \tag{6}$$

for large x, where c_1 and c_2 are constants. Equation (6) implies that multiplying the variable, or equivalently, changing the unit in which it is measured, does not affect properties of the system.

- Power-law distributions conveniently take the form of a straight line in a log-log plot because Eq. (1) is equivalent to

$$\log p(x) = \log C - \alpha x. \tag{7}$$

When testing if empirical data are power-law distributed, it is instructive (but not conclusive) to plot their distribution on the log-log scale.

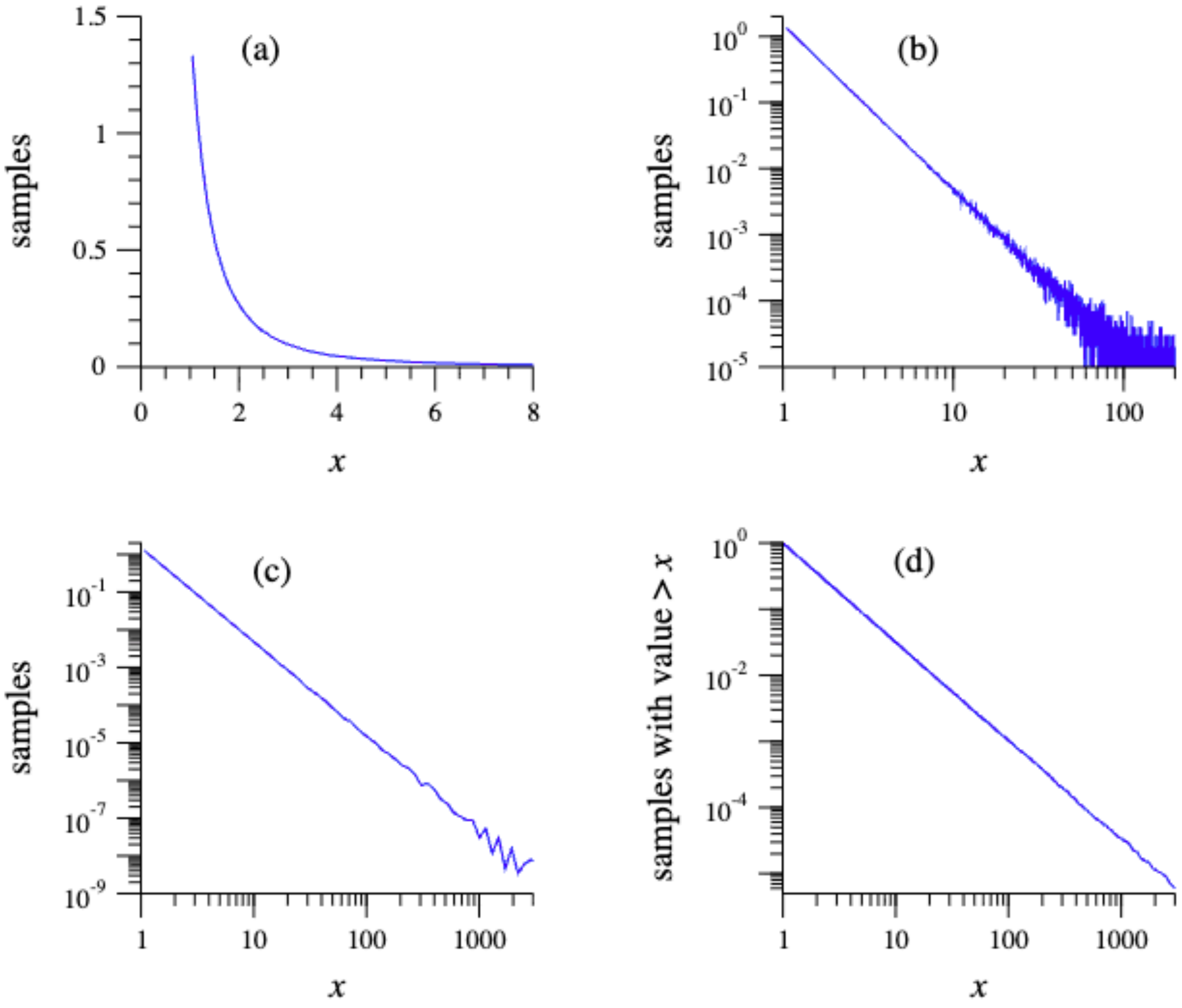


FIG. 3 (a) Histogram of the set of 1 million random numbers described in the text, which have a power-law distribution with exponent $\alpha = 2.5$. (b) The same histogram on logarithmic scales. Notice how noisy the results get in the tail towards the right-hand side of the panel. This happens because the number of samples in the bins becomes small and statistical fluctuations are therefore large as a fraction of sample number. (c) A histogram constructed using "logarithmic binning". (d) A cumulative histogram or rank/frequency plot of the same data. The cumulative distribution also follows a power law, but with an exponent of $\alpha - 1 = 1.5$.

Contemporary Physics, Vol. 46, No. 5, September–October 2005, 323 – 351

Taylor & Francis
Taylor & Francis Group

Power laws, Pareto distributions and Zipf's law

M.E.J. NEWMAN*

Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109, USA

(*Received 28 October 2004; in final form 23 November 2004*)

When the probability of measuring a particular value of some quantity varies inversely as a power of that value, the quantity is said to follow a power law, also known variously as Zipf's law or the Pareto distribution. Power laws appear widely in physics, biology, earth and planetary sciences, economics and finance, computer science, demography and the social sciences. For instance, the distributions of the sizes of cities, earthquakes, forest fires, solar flares, moon craters and people's personal fortunes all appear to follow power laws. The origin of power-law behaviour has been a topic of debate in the scientific community for more than a century. Here we review some of the empirical evidence for the existence of power-law forms and the theories proposed to explain them.

SIAM REVIEW
Vol. 51, No. 4, pp. 661–703

Power-Law Distributions in Empirical Data*

Aaron Clauset†
Cosma Rohilla Shalizi‡
M. E. J. Newman§

nature communications

Article | Open Access | Published: 04 March 2019

Scale-free networks are rare

Anna D. Broido & Aaron Clauset

Nature Communications **10**, Article number: 1017 (2019) | Cite this article

Growing complex networks

Preferential attachment and the emergence of scale-free networks

EXPLORABLES

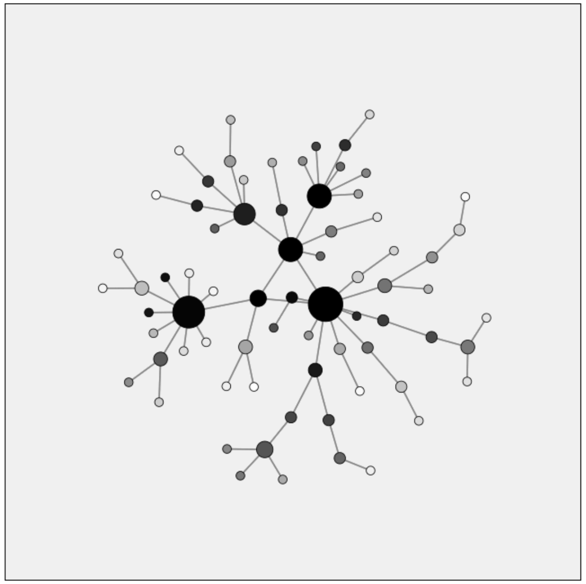
by Dirk Brockmann

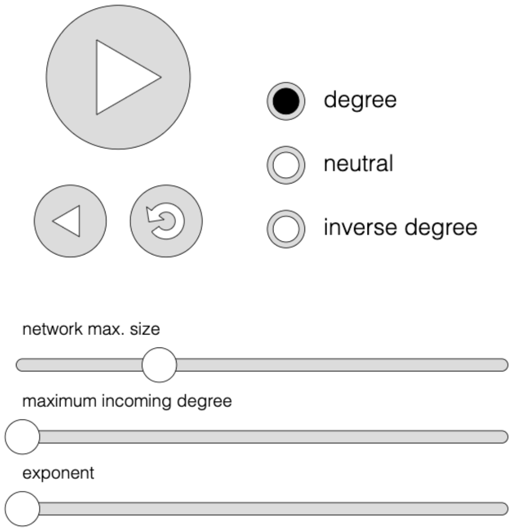
26 June, 2018

This explorable illustrates network growth based on ***preferential attachment***, a variant of the **Barabasi-Albert model** that was introduced to capture strong heterogeneities observed in many natural and technological networks. It has become a popular model for scale-free networks in nature.

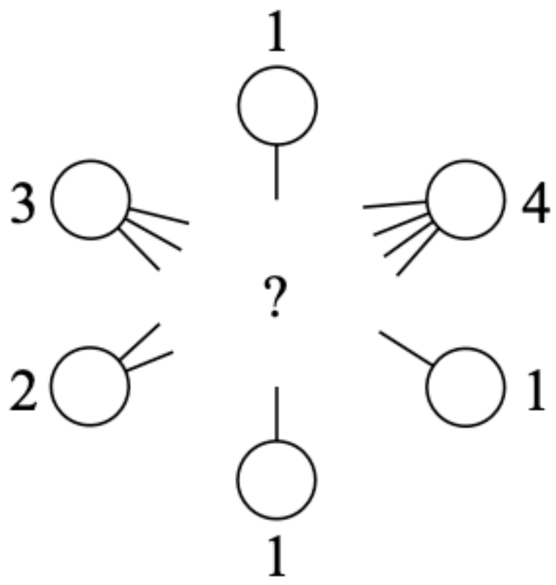
Preferential attachment means that nodes that enter the network during a growth process preferentially connect to nodes with specific properties. In the original system, they preferentially connect to existing nodes that are already well connected, increasing their connectivity even further. This *rich get richer* effect generates networks in which a few nodes are very strongly connected and very many nodes poorly.

Press play and keep on reading…





https://www.complexity-explorables.org/explorables/knitworks/



The friendship paradox is a phenomenon, in which, anecdotally, the average number of friends of a friend is greater than the average number of friends of an individual. The paradox originates from the fact that nodes with a large degree contribute disproportionately to the average degree of a friend, as they have a higher probability of being friends than low degree nodes do. The network has $N = 6$ nodes and the average degree of a randomly selected individual is equal to

$$\frac{1}{6}(1 + 2 + 3 + 1 + 4 + 1) \equiv \langle k \rangle = 2. \tag{10}$$

To calculate the average number of friends of a friend, we have instead to perform a weighted average, accounting for the fact that a node with degree k will appear k times in the calculation of the average. The weighted average degree is equal to

$$\frac{(1 \times 1 + 2 \times 2 + 3 \times 3 + 1 \times 1 + 4 \times 4 + 1 \times 1)}{(1 + 2 + 3 + 1 + 4 + 1)} \approx 2.67, \tag{11}$$

where we have allowed self-loops and multiple edges for simplicity. In general, for sufficiently large and random networks, the mean degree of a neighbour is given by

$$\sum_k k \times \frac{kp(k)}{\sum_{k'} k'p(k')} = \frac{\langle k^2 \rangle}{\langle k \rangle}. \tag{12}$$