Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 329 (2003) 499–508
PHYSICA A
www.elsevier.com/locate/physa

Power-law distributions resulting from finite resources

Thomas Wilhelm[a,*], Peter Hänggi[b]

[a] *Theoretical Systems Biology, Institute of Molecular Biotechnology, Beutenbergstrasse 11, D-07745 Jena, Germany*
[b] *Department of Physics, University of Augsburg, Universitätsstrasse 1, D-86135 Augsburg, Germany*

Received 21 March 2003; received in revised form 26 May 2003

Abstract

An elementary stochastic model, termed the normalization model, is put forward which does demonstrate that power-laws generically occur in systems with finite resources. The model is capable to exhibit power-law distributions with arbitrary power law exponents; nevertheless, for a large fraction of the parameter space power law exponents near unity are obtained.

As an application of the normalization mechanism we consider a network growth-saturation model. This model extends the scale-free network model (SF) to include the fact of finite resources. In the network growth-saturation model the scale-free property holds only for the growth period, within the stationary regime we obtain power-law distributions of the weight of the edges among the vertices. We conjecture that this pattern will be found in the Internet if it reaches the steady state.

PACS: 89.75.Da; 05.65.+b; 89.75.Hc; 89.20.Hh

Keywords: Power-law distribution; Zipf's law; Stochastic multiplicative process; Network; Internet

1. Introduction

Power-law distributions (PLDs) are characteristic for complex systems [1–16]. Early examples can be found in economics: for example, the Pareto distribution of wealth reveals that the exponent β of the complementary cumulative distribution function

* Corresponding author. Fax: +49-3641-656191.
E-mail address: wilhelm@imb-jena.de (T. Wilhelm).

0378-4371/$ - see front matter
doi:10.1016/S0378-4371(03)00609-5

$p(x_i > x) \sim x^{-\beta}$ is in the range $1 < \beta < 2$ [2]. The corresponding probability density function (PDF) reads $p(x) \sim x^{-\alpha}$, with $\alpha = \beta + 1$. More recent examples are the income distribution of individuals in Japan 1998, showing a PLD with $\alpha = 2.98$ [3], and the wealth distribution of the 400 richest people in the US 1996 with $\alpha = 2.36$ [4]. Well-known cases of social networks comprise the actor collaboration network where the number of links per actor follows a PLD with $\alpha = 2.3$ [1], and the science citation network where the citation numbers of papers follow a PLD with $\alpha = 3$ [5]. Interestingly, the size distribution of cities always follows Zipf's law (i.e. a PLD with $\alpha = 2$) quite exactly [6]. An interesting biological example is the PLD of the abundance of expressed genes (in more than 40 tissue samples of different species) where the exponent is in the range $1.6 < \alpha < 2$ [7].

Two general mechanisms are discussed to account for the abundance of PLDs in nonequilibrium systems: self-organized criticality [8] and stochastic multiplicative processes [9]. Pure stochastic multiplicative processes do not generate PLDs [10], but combined with transport processes [11], sources [12], boundary constraints [13], or reset events [10] they actually do.

Although the dynamic processes leading to the observed PLDs certainly differ a lot in detail, all the above-mentioned examples share one common feature: a finite resource is distributed among the "agents"; e.g. wealth among individuals, citation numbers among papers, inhabitants among cities, mRNA molecule numbers among genes, to name but a few. In this paper we show that finite resources generically lead to PLDs. For that purpose we study a minimal stochastic model, the normalization model. Its main feature is as follows: if some agents collect more resources, others must get fewer, because the sum remains fixed. The normalization procedure represents a new mechanism that is sufficient for the underlying stochastic multiplicative process to produce stationary PLDs.

As an application of the normalization model we expand in the second part of this paper the scale-free (SF) model [1] that has been proposed for infinitely growing networks. In particular, we include the fact that the maximum number of clicks to web-sites is finite (eventually the number of people on earth, as well as the number of clicks per person is finite). Technically, we have to generalize binary networks where a link just exists or not to nonbinary networks with weighted connections between the vertices.

In our network growth-saturation model the SF property only holds for the initial growth period; after long times a PLD of the number of clicks between different web-sites arises being in agreement with the results of the general normalization model.

2. The normalization model

Our system consists of a total number of n agents, each one possessing an amount q_i of some quantity (q_i could be, for instance, the wealth of person i, or the number of mRNA molecules of gene i). We deal with normalized values $w_i = q_i / \sum_j q_j$. At each time step t, z agents (taken at random) obtain a new quantity $\tilde{w}_i(t) = x/n$, an event which we term a "mutation" in the following. The stochastic variable x is taken from an arbitrary distribution, in the following at first from the rectangular distribution ($x \in [0, 2a]$, $a \ll n$). After each time step the whole system is normalized again. Thus, we study the stochastic, effectively multiplicative process:

$$w_i(t+1) = \begin{cases} \tilde{w}_i(t)/w_{tot}(t) & \text{for } z \text{ agents }, \\ w_i(t)/w_{tot}(t) & \text{for the } n-z \text{ remaining agents} \end{cases} \quad (1)$$

with the stochastic normalization factor given by $w_{tot}(t) = \sum_j^z \tilde{w}_j(t) + \sum_{i \neq j} w_i(t)$. This normalization of all quantities accounts for a complete relaxation of the disturbance across the whole system.

After an initial time the system dynamics becomes independent of the initial conditions $w_i(0)$. Let us assume that all $q_i(0)$ are taken from a uniform distribution (random initial conditions). Then, we obtain for these *normalized* weights $\langle w_i(0) \rangle \approx 1/n$ and with large n, $\max w_i(0) \approx 2/n$. For $a = 1$ this uniform distribution remains unchanged for all t. However, for $a < 1$ the normalized weights which are not mutated grow larger on average, and for $a > 1$ they decrease instead. For all distributions with $\langle x \rangle = a$ and $z \ll n$ it then follows that

$$\langle w_{tot} \rangle = (n - z + za)/n \ . \quad (2)$$

In the following we first restrict ourselves to $z = 1$, which means that the whole system has been relaxed before a new change occurs. However, as we will show below, all our results depend only very weakly on the value of z.

For $z = 1$ it follows from (1,2) that nonmutated quantities approximately change according to

$$w_i(t) = w_i(0) \Bigg/ \left(1 - \frac{1-a}{n}\right)^t \ . \quad (3)$$

Fig. 1 depicts the evolution of all weights w_i, $i = 1, \ldots, n$ for the case with (a) $a > 1$, and (b) $a < 1$. Note that a mutation at time step t yields a discrete jump in the dynamics of that specific weight. The case $a > 1$ leads to a situation that mimics a"red queen world", whereas $a < 1$ represents a "wu wei world" (in the kingdom of the red queen one has to run as fast as possible to keep someone's position [17], whereas a wu wei world is just the counterpart: if one does nothing actively, everything is going to be alright (in Chinese Taoism wu wei means nonaction)).

After an initial evolution a stationary distribution of the normalized weights w_i arises which is *independent* of the initial values. Fig. 2 depicts the complementary cumulative distribution function $p(w_i > w)$ for different values of a.

Two different kinds of distributions are obtained for the two different "worlds", corresponding to $a > 1$ and $a < 1$, respectively. The case with $a > 1$ yields a power-law behavior for $w_i \ll 1$ and a cutoff at $w_c = 2a/n$, because for $a > 1$ and $x \in [0, 2a]$ almost no weight w_i assumes a value larger than $2a/n$. (It is only for the improbable situation that one weight w_i lies just below $2a/n$ with all remaining x-values being very small so

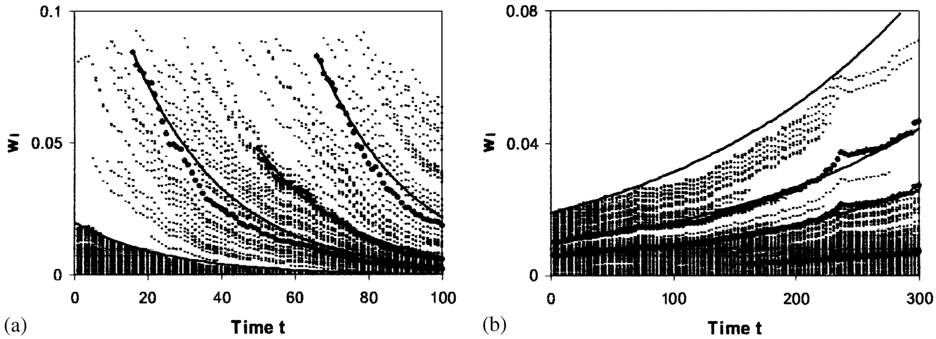

Fig. 1. Evolution of the set of normalized weights in Eq. (1) w_i with $n = 100$, $z = 1$, and random initial conditions drawn from a box distribution with parameter (a) $a = 5$ and (b) $a = 0.5$. The large dots depict the evolution of three arbitrarily selected agents, the remaining small dots present the evolution of all remaining ones. The three lines corresponding to the large dots exhibit the result of the approximation in (3) for these selected three agents w_i's. The line that starts near 0.02 corresponds to the dynamics of an agent that is prepared with an initial maximal weight $w_i(0) \approx 2/n$.

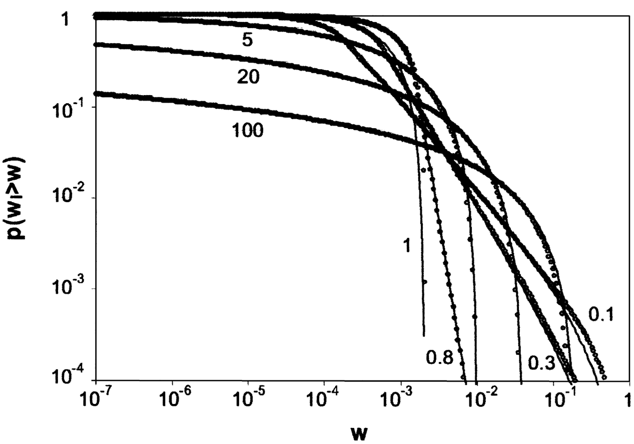

Fig. 2. Complementary cumulative distribution functions of the normalized weights w_i for different values of a, as indicated next to the corresponding lines. The small circles depict the distributions obtained with the normalization model ($n = 1000$; averaged over 500,000 mutations) after an initial time evolution of n^2 mutations. The connecting lines denote the results of the nonlinear regression put forward with Eqs. (5) and (7).

that with the next step, due to normalization, a weight above $2a/n$ may be assumed.) The simplest stationary PDF which obeys these features is constructed as follows:

$$p_{a>1}(w_i) \sim w_i^{-\alpha} \left(\frac{2a}{n} - w_i \right) . \tag{4}$$

It then follows that the complementary cumulative distribution function ($0 < \alpha < 1$) is given by

$$p_{a>1}(w_i > w) = \int_w^1 p_{a>1}(w_i)\,\mathrm{d}w_i$$
$$= 1 - \frac{w^{1-\alpha}(2a/n/(1-\alpha) - w/(2-\alpha))}{(2a/n)^{2-\alpha}(1/(1-\alpha) - 1/(2-\alpha))} \,. \tag{5}$$

Of course, the ultimate cutoff for all normalization systems is at maximal weight, i.e. $w = 1$; thus (at least for large a) it could be important to add the factor $(1 - w_i)$ in (4). Closer inspection shows that for marginal a-values ($a = 1 + \varepsilon$, $\varepsilon \ll 1$), or very large values ($a \gg 1$) these cutoff-terms should be decorated with exponents $\neq 1$. Note, however, that for all moderate a, such as the ones considered in Fig. 2, the simple expression (4) already yields indeed very accurate results already, as can be deduced from Fig. 2.

The case with $a < 1$ yields uniform distributions for $w_i < W := 2a/n$ and power-laws for larger values with a cutoff at 1. The simplest corresponding PDF assumes the form

$$p_{a<1}(w_i) \begin{cases} = c & \forall w_i \leqslant W, \\ \sim w_i^{-\alpha}(1 - w_i) & \forall w_i \geqslant W, \end{cases} \tag{6}$$

which in turn yields the complementary cumulative distribution function ($\alpha > 1$)

$$p_{a<1}(w_i > w) = \begin{cases} cP_{\leqslant} & \forall w \leqslant W, \\ cP_{\geqslant} & \forall w \geqslant W \end{cases} \tag{7}$$

with

$$P_{\leqslant} = \left(W - w + \frac{W^{\alpha}}{1 - W} \left(\frac{1 - W^{1-\alpha}}{1 - \alpha} - \frac{1 - W^{2-\alpha}}{2 - \alpha} \right) \right),$$
$$P_{\geqslant} = \frac{W^{\alpha}}{1 - W} \left(\frac{1 - w^{1-\alpha}}{1 - \alpha} - \frac{1 - w^{2-\alpha}}{2 - \alpha} \right)$$

and

$$c = \left(W + \frac{W^{\alpha}}{1 - W} \left(\frac{1 - W^{1-\alpha}}{1 - \alpha} - \frac{1 - W^{2-\alpha}}{2 - \alpha} \right) \right)^{-1} .$$

In order to determine the exponents α as a function of a, a nonlinear regression of the complementary cumulative distribution functions based on (5) and (7) has been invoked. Fig. 3 depicts that by setting $\alpha = f(a)$ excellent agreement can be obtained in both cases, i.e. for both $a > 1$ and $a < 1$, respectively, by use of the insightful ansatz:

$$\alpha = f(a) = (1 - a^A)^B \exp g(a, C) \tag{8}$$

with $g(a, C) = C/a$ for $a > 1$, and $g(a, C) = (1 - a)^C - 1$ for $a < 1$.

As can be deduced from Fig. 3 the simple normalization model yields PLDs within a wide range of power law exponents α. However, for most values $a > 0$ the resulting exponents α are within the neighborhood of 1, which is also the case in the above

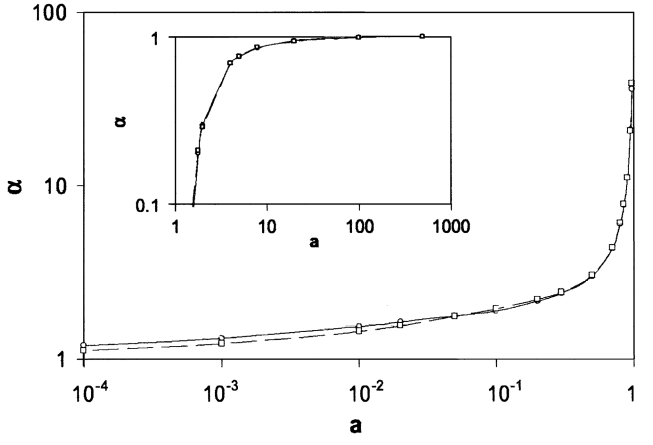

Fig. 3. Nonlinear regression fit $\alpha = f(a)$ obtained with (5) for $a > 1$, see in the inset, and for $a < 1$, cf. (7). The circles (connected by solid lines) are simulation results of the normalization model, the squares (connected by dashed lines) follow from nonlinear regression form using (8) ($a > 1 : A = -0.96,\ B = 4.7,\ C = 4.3;\ a < 1 :\ A = 0.23,\ B = -0.9,\ C = 1.4$).

mentioned examples.[1] For $a \ll 1$ and $a \gg 1$ the exponent α approaches asymptotically the value 1.

The qualitative results are nearly independent of the assumed distribution for x. Generally, if only positive quantities ($x > 0$) are considered, "$a < 1$"—characteristic distributions are obtained if $\int_0^1 p(x)\,\mathrm{d}x > \int_1^\infty p(x)\,\mathrm{d}x$. If x is taken from a Gaussian distribution with $\langle x_G \rangle = a$ and $\sigma(x_G) = a$ one obtains nearly the same function $\alpha = f(a)$ as the one shown with Fig. 3. Interestingly enough, even quantitatively, the results in Fig. 3 do depend only very weakly on the total number of agents n and mutation number z. The relative deviations of the graphs $\alpha = f(a)$ are below 1% for, e.g., $z = 200$ and 500, compared with the result for $z = 1$ that is shown in Fig. 3. For $z \gg 1$ $p(w_i)$ resembles the rectangular distribution ($\alpha = 0$).

As one application of our general normalization model let us draw the new value of the mutated agents, x, as well as the number of mutated agents at a given time, z, from two Gaussian distributions with $\langle x_G \rangle = \sigma(x_G) = 0.15$, and $\langle z \rangle = \sigma(z) = 1000$. For this case one finds the mean Pareto exponent to read $\beta = \alpha - 1 = 1.5$ [2]. Note here the robustness of the result: with a rectangular distribution of x and $z = 1$ we find for $a = 0.15$ $\alpha = 2.3$; i.e. a corresponding value for $\beta = 1.3$. Thus, our normalization model is indeed capable of explaining the main features of the underlying dynamical processes that lead to the Pareto distribution of wealth: in short; if some people become richer, others necessarily have to become poorer, and vice versa.

[1] Note that the power-law exponents α corresponding to $a < 1$ refer to the right branch of the complementary cumulative distribution function shown in Fig. 2. The exponents α corresponding to $a > 1$ refer to the left branch. Therefore, in the limit case $a = 1$ we find two exponents: $\alpha = 0$ (referring to the left, the constant, branch, and $\alpha = \infty$, referring to the right, here the vertical branch.

3. The network growth-saturation model

In this subsection yet another application of the normalization model procedure is studied in greater detail. Many recent studies of network topology deal with non-weighted links between the vertices (a link is either present or absent) as found, for instance, in studies involving the internet [14], metabolic networks [15], or food webs [16] and alike, [1]. Here we generalize the network consideration to a case with weighted links (weight of a link $\in [0,1]$). We take into account the fact that also the resources which are distributed among the links of the network are finite. The celebrated SF model [1] which inherently considers infinitely growing undirected binary networks is generalized accordingly. Our extension naturally comprises a saturation effect of the network growth at long times and even changes the network topology during the network evolution dynamics. The original SF model works as follows: (i) Starting with a small number (m_0) of vertices, at every time step a new vertex with $m(\leqslant m_0)$ edges is added, and, (ii) the probability p that a new vertex will be connected to the vertex i depends on the connectivity k_i (i.e. the number of links of vertex i) of that vertex and has the form: $p(k_i) = k_i / \sum_j k_j$. The main result yields a linear growth of the number of vertices $N = m_0 + t$ accompanied by a corresponding power-law connectivity distribution [1].

We next add the following saturation mechanism: In real networks usually some connections are stronger than others. In the WWW-example the vertices denote the different web-sites and a connection between web-sites exists if there is a link between them. The edges (links) can be weighted by the different number of the corresponding clicks on the links (averaged over some time span). The quantity T_{ij} counts the number of clicks from site i to site j in a fixed time interval (in contrast to the original SF model we consider here directed links). With the total number of clicks $T_t = \sum_{ij} T_{ij}$ one obtains normalized values $t_{ij} = T_{ij}/T_t$. The nonbinary SF model now works as the original binary one with the *additional* feature that the connections of the new vertex to and from the old vertices assume a value $t_{ij} = x/N^2$ (x uniformly distributed, $x \in [0, 2a]$). Note that we work with N^2 different weighted directed links between the N vertices, including also the (return)-links to the individual vertex. After each addition of a new vertex with its new connections we normalize all t_{ij} again. Thus, if N grows, $\langle t_{ij} \rangle$ becomes smaller. Of course, during the growth evolution also T_t grows in reality and thus during this time span the t_{ij}'s do not necessarily decrease. However, after some initial transient time the normalized values t_{ij} necessarily do decrease. For the sake of simplicity we assume the normalization procedure working from the beginning.

A binary network is then obtained from this non-binary one in the following way: if $t_{ij} > t^*$ the link is counted, otherwise it is not counted (the case $t^* = 0$ corresponds to the original SF mechanism). However, note that very small values for t_{ij} can be neglected (e.g. one click in 100 years). We assume that each vertex with no inputs dies out. Thus, after long times a steady state is reached for this corresponding binary web.

Fig. 4 depicts the number of vertices (nodes) $N = f(t)$ for different values of the parameter a and t^*. For $a = 10$, $t^* = 10^{-4}$, for instance, different large growth-intervals again and again alternate with decline-intervals.

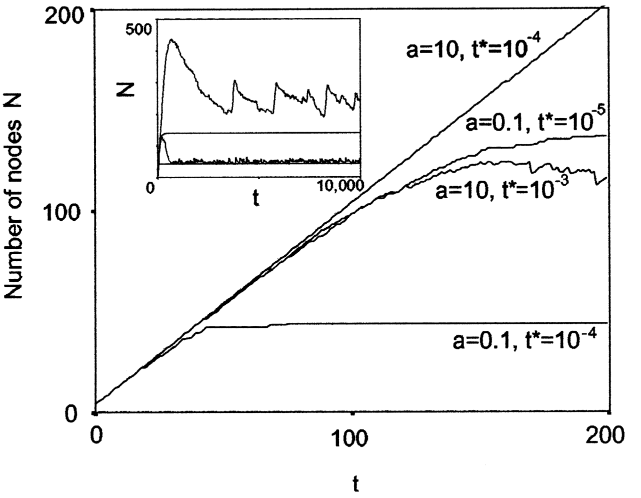

Fig. 4. The number of vertices $N(t)$ of the networks for different parameters a and t^*. The inset depicts the results for the same set a and t^* up to asymptotic times $t = 10,000$.

During the initial time evolution the growth mechanism dominates and the dynamics works essentially equivalent to the original SF model. It comes at no surprise that within this growth regime a power-law connectivity distribution emerges as well. However, later on the average number of vertices begins to saturate, and the network does not grow any further. A corresponding analysis shows that within the saturation regime the maximum of the connectivity distribution is shifted towards the right. This is in accordance with Barabási's "Model B" [1], where it was shown that without network growth the corresponding connectivity distributions exhibit a distinct maximum. In our network growth-saturation model, after saturation is reached, the normalized weights t_{ij} follow a power-law, cf. Fig. 2. Just as in the normalization model we obtain $\alpha < 1$ for $a = 10$ and $\alpha > 1$ for $a = 0.1$. This result corroborates the recent finding that today's Internet seemingly also reaches a saturation stage. [2]

4. Discussion

With this work we could demonstrate that power-law distributions naturally arise in systems with finite resources. We thereby conjecture that the normalization mechanism is central and at work in many processes with PLDs in different real life, complex systems. A main finding is that the power law exponents α assume values in a large regime within the neighborhood of one. Other exponents can be obtained only for a small fraction of the parameter space. To our knowledge this is the first model that

[2] CEO of Nortel J. Roth (July 2001): "Data transfer in the web does not grow anymore."; Press officer of Denic (organization that performs registration of German Internet addresses) K. Herzig (April 2002): "There is a saturation effect—The big euphoria is over."

principally can yield any value for the power law exponents $\alpha > 0$. The general result is robust and quite independent of all the details of the mechanism.

As we have shown (Fig. 2, Eqs. (4)–(7)), in steady state our normalization model yields two different classes of complementary cumulative probability distributions. For $a > 1$ we obtain a distribution function with a power-law part with $\alpha < 1$, and for $a < 1$ another distribution function with a power-law part with $\alpha > 1$. It is interesting to note that the latter, without the cutoff term, can also well be described by a function which is the complementary cumulative of a generalized power-law function, the Zipf–Mandelbrot distribution $p(x) = b/(c + x)^{\alpha}$ $(b, c, \alpha > 0)$, which can be normalized for $\alpha > 1$. This function can be rewritten as the so-called q-exponential function $p(x) = p_0 \exp_{q'}(-Ax) = p_0[1 - (1 - q')Ax]^{1/(1-q')}$, where $p_0 = bc^{-\alpha}$, $A = \alpha/c$, and $q' = 1 + 1/\alpha$. For $q' \to 1$ $\exp_{q'}(-x)$ reduces to the usual exponential function $\exp(-x)$. The q-exponential function arises naturally in the context of generalized, nonextensive statistical mechanics [18]. This theory has successfully been applied to many different complex systems under nonequilibrium conditions. Recent examples deal with urban agglomeration [19], and the Internet [20], which additionally contains further references of successful applications of Tsallis statistics.

As an application of our normalization model we have elucidated that the normalization mechanism can be utilized to extend the SF network model to cover the case with a saturation stage, too. The resulting network growth-saturation model exhibits a power-law connectivity distribution only during the growth period, at later times it crosses over towards a PLD for the weights of the corresponding links. In short, our model predicts a PLD of the number of clicks between web-sites (after saturation of the Internet is reached), a fact that can be put to a crucial test of our prediction in the future.

Acknowledgements

We thank S. C. Manrubia for critical reading of the manuscript and valuable comments, and an anonymous referee for indicating a possible connection to Tsallis statistics.

References

[1] A.-L. Barabási, R. Albert, Science 286 (1999) 509;
A.-L. Barabási, R. Albert, H. Jeong, Physica A 272 (1999) 173;
R. Albert, A.-L. Barabási, Rev. Mod. Phys. 74 (2002) 47;
S. Bornholdt, H.-G. Schuster (Eds.), Handbook of Graphs and Network: From the Genome to the Internet, Wiley-VCH, Weinheim, 2003.

[2] V. Pareto, Cours d'économie politique (reprinted as volume of Oeuvres Complètes), Droz, Geneva, 1896–1965.

[3] H. Aoyama, Y. Nagahara, M.P. Okazaki, W. Souma, H. Takayasu, M. Takayasu, Fractals 8 (2000) 293.

[4] M. Levy, S. Solomon, Physica A 242 (1997) 90.

[5] S. Redner, Eur. Phys. J. B 4 (1998) 131.

[6] G.K. Zipf, Human Behavior and the Principle of Least Effort, Addison-Wesley, Cambridge MA, 1949; D.H. Zanette, S.C. Manrubia, Phys. Rev. Lett. 79 (1997) 523.
[7] C. Furusawa, K. Kaneko, Zipf's law in gene expression, Phys. Rev. Lett., submitted for publication; preprint physics/0209103, 2003.
[8] P. Bak, How Nature Works, Oxford University Press, Oxford, 1997.
[9] P. Hänggi, H. Thomas, Phys. Rep. 88 (1982) 207;
N.G. van Kampen, Stochastic Processes in Physics and Chemistry, North-Holland, Amsterdam, 1992;
P. Hänggi, F. Marchesoni, P. Grigolini, Z. Phys. B 56 (1984) 333.
[10] S.C. Manrubia, D.H. Zanette, Phys. Rev. E 59 (1999) 4945.
[11] S.C. Manrubia, D.H. Zanette, Phys. Rev. E 58 (1998) 295.
[12] G. Grinstein, M.A. Muñoz, Y. Tu, Phys. Rev. Lett. 76 (1996) 4376;
H. Takayasu, A.-H. Sato, M. Takayasu, Phys. Rev. Lett. 79 (1997) 966;
D. Sornette, Phys. Rev. E 57 (1998) 4811.
[13] M. Levy, S. Solomon, Int. J. Mod. Phys. C 7 (1996) 745;
D. Sornette, R. Cont, J. Phys. I 7 (1997) 431.
[14] R. Albert, H. Jeong, A.-L. Barabási, Nature 401 (1999) 130;
B.A. Huberman, L.A. Adamic, Nature 401 (1999) 131;
L.A. Adamic, B.A. Huberman, Quart. J. Electron. Comm. 1 (2000) 5.
[15] A. Wagner, D. Fell, Proc. Roy. Soc. London B 268 (2001) 1803.
[16] J.M. Montoya, R.V. Solé, J. Theor. Biol. 214 (2002) 405.
[17] L. Carrol, Alice's Adventures in Wonderland, Penguin Books, Aarmondsworth, 1994.
[18] C. Tsallis, J. Stat. Phys. 52 (1988) 479;
S. Abe, Y. Okamoto (Eds.), Nonextensive Statistical Mechanics and Its Applications, Springer-Verlag, Heidelberg, 2001;
T. Yamano, Physica A 305 (2002) 486.
[19] L.C. Malacarne, R.S. Mendes, E.K. Lenzi, Phys. Rev. E 65 (2002) 017106.
[20] S. Abe, N. Suzuki, Phys. Rev. E 67 (2003) 016106.