Vol. 35 (2004) *ACTA PHYSICA POLONICA B* No 2

NONEXTENSIVE INFORMATION ENTROPY FOR STOCHASTIC NETWORKS

G. Wilk†

A. Sołtan Institute of Nuclear Studies
Hoża 69, 00-681 Warszawa, Poland

and Z. Włodarczyk‡

Institute of Physics, Świętokrzyska Academy
Świętokrzyska 15, 25-406 Kielce, Poland
and
The Kielce School of Science (WSU)
Wesoła 52, 25-353 Kielce, Poland

(Received September 12, 2003)

Nature is full of random networks of complex topology describing such apparently disparate systems as biological, economical or informatical ones. Their most characteristic feature is the apparent scale-free character of interconnections between nodes. Using an information theory approach, we show that maximalization of information entropy leads to a wide spectrum of possible types of distributions including, in the case of nonextensive information entropy, the power-like scale-free distributions characteristic of complex systems.

PACS numbers: 89.75.–k, 89.70.+c, 24.60.–k

Random networks have recently found applications in the description of complex systems in different, apparently very disparate branches of modern science such as, for example, molecular biology, sociology, economy and computer science [1,2]. For example, living organisms form huge genetic networks the nodes of which are proteins and links represent the corresponding chemical interactions [3]. A similarly big network is formed by the nervous system the nodes of which are connected by axons [4]. Comparable complexity show networks existing in the sociological systems in which nodes

† e-mail: wilk@fuw.edu.pl
‡ e-mail: wlod@pu.kielce.pl

are countries, organizations or single persons whereas links characterize their mutual interactions [5], in the world of finances and computer networks (with World Wide Web being the most known example where nodes are HTML documents connected via hiper-links URL [6]). For most recent reviews of random networks see [7,8].

Analysis of different random networks clearly indicate that the probability $P(k)$ of joining a given node with other nodes is described by the power law $P(k) \propto k^{-\gamma}$ [9]. For example, the most convincing analyses of computer networks with over 800 million nodes [6,10–12] lead to a power-like distribution of $P(k)$ with exponent equal $\gamma \sim 2.1 \div 2.45$. This contradicts the existing models of random networks [13,14] predicting instead exponential distributions: $P(k) \propto \exp(-k)$. The most popular model (ER) dealing with a fixed number of nodes N was proposed in [13], where the Poisson distribution was advocated to be used for probability that a given node has k links (with the mean number of connections being λ_0)[1],

$$P(k) = \frac{\lambda_0^k \, \mathrm{e}^{-\lambda_0}}{k!} . \tag{1}$$

However, in order to get the observed power-like form of $P(k)$ one has to allow for growing N and replace the democratic law of attachment a new link used in deriving (1) by a preferential one. This means that distribution $P(k)$ is determined by the dynamics of the growth of network [7,11]. Starting from a small number m_0 of nodes, adding in each time step a new node with $m \le m_0$ possible connections and assuming that this new node joins the already existing nodes with equal, k_i-independent, probability $\Pi(k_i) = \frac{1}{m_0+t-1}$, the evolution (growth) of network is described by the following equation [11]:

$$\frac{\partial k_i}{\partial t} = m\,\Pi(k_i) = \frac{m}{m_0+t-1} \tag{2}$$

leading for long times t to exponential stationary distribution:

$$P(k) = \frac{1}{m} \exp\left(-\frac{k}{m}\right) . \tag{3}$$

On the other hand, assuming that probability $\Pi(k_i)$ is selective, for example that $\Pi(k_i) = \frac{k_i}{\sum_{j=1}^{m_0+t-1} k_j} = \frac{k_i}{2mt}$, one gets instead asymptotically a simple power law for $P(k)$:

$$P(k) = \frac{2m^2 t}{m_0+t}\, k^{-3} \propto k^{-3} . \tag{4}$$

[1] Notice that in the limit of large k distribution (1) can be approximated by $P(k) = (2\pi k)^{-1/2} (\lambda_0/k)^k \exp(k-\lambda_0)$, whereas for large values of λ_0 it becomes a gaussian distribution, $P(k) = (2\pi\lambda_0)^{-1/2} \exp\left[-(k-\lambda_0)^2/2\lambda_0\right]$.

Apparently such distribution with universal exponent $\gamma = 3$ shows up in different situations (and under different names). As Pareto distribution [15] it describes the growth of the wealth of persons living in stable economical systems, as Zipf's law [16] it is applied in linguistics and it also describes the distribution of the citations of the scientific works [17, 18].

In the limit of large t, *i.e.*, when stationary state is attained, this problem can be also studied from the information theory point of view. In it one asks the following question [19]: what is the informational content of data represented by distributions $P(k)$? In other words, what is the minimal number of parameters needed to reproduce a given shape of $P(k)$? The question asked in such approach is: suppose that we know only that a network we are interested in, which we would like to describe by $P(k)$, leads to some mean value of k, *i.e.*, we know that:

$$\sum_{k=1}^{\infty} P(k) \;=\; 1 \qquad \text{and} \qquad \langle k\rangle \;=\; \sum_{k=1}^{\infty} k\; P(k) \;=\; \lambda_0 \;=\; \text{const.}\,, \tag{5}$$

what would be then *the most probable and least biased* form of $P(k)$ in such a case (*i.e.*, describing the existing data and given entirely in terms of λ_0)? To answer this question one maximalizes the corresponding information entropy associated with probability distribution $P(k)$ under constraints (5). The usual form of such entropy is Shannon entropy [20],

$$S \;=\; -\sum_{k=1}^{\infty} P(k)\, \ln P(k)\,. \tag{6}$$

The conditions (5) representing our *a priori* knowledge of the problem lead to the exponential probability distribution

$$P(k) \;=\; \frac{1}{\lambda_0}\; \exp\left(-\frac{k}{\lambda_0}\right)\,, \tag{7}$$

closely resembling Eq. (3) [2].

There are, however, many systems with properties preventing the use of Shannon type of information entropy and calling for its generalization. For example, they posses some intrinsic fluctuations resulting in the whole spectrum of parametr λ, $P(\lambda)$, with $\lambda_0 = \langle\lambda\rangle$ being only its mean value [21] or they develop some correlations introducing memory effects, *cf.* [22] (in

[2] It is interesting to mention that additional knowledge that all entities represented by k are *indistinguishable*, which results in the necessity of introducing in the corresponding summations the additional weight factor $1/k!$, changes the above distribution to Poisson distribution of the ER model mentioned above, *cf.*, Eq. (1).

statistical physics such situation leads to the necessity of departing from the use of the usual Boltzmann–Gibbs statistics in favour of some sort of generalized one [22]). Out of many possible generalizations we shall use in this note the nonextensive Tsallis entropy defined as [22]:

$$S_q = -\frac{1}{1-q}\left\{1 - \sum_{k=1}^{\infty}[P(k)]^q\right\}. \tag{8}$$

It can be regarded as a minimal (*i.e.*, one parameter) extension of Shannon entropy (6), to which it reduces when $q \to 1$. Parameter q describes therefore summarily all effects preventing the use of Shannon entropy mentioned above.

Using S_q as a measure of information about our system, *i.e.*, maximalizing S_q with constraints (equivalent to (5) above):

$$\sum_{k=1}^{\infty} P(k) = 1 \qquad \text{and} \qquad \langle k\rangle_q = \frac{\sum_{k=1}^{\infty} k[P(k)]^q}{\sum_{k=1}^{\infty}[P(k)]^q} = \lambda_0 = \text{const.}, \tag{9}$$

one obtains as result a power-like distribution of the form:

$$P(k) = P_q(k) = C\left[1 - (1-q)\frac{k}{\lambda_0}\right]^{\frac{q}{1-q}}, \tag{10}$$

where $C = 1/\sum_{k=1}^{\infty}[1-(1-q)k/\lambda_0]^{q/(1-q)} = 1/\lambda_0$ is normalization[3]. In this case

$$\langle k\rangle = \frac{\lambda_0}{(2-q)} \quad \text{and} \quad \mathrm{Var}(k) = \frac{\lambda_0^2}{(3-2q)(2-q)^2}. \tag{11}$$

Notice that for $q \to 1$ this distribution becomes exponential, as in Eq. (7). On the other hand, for large values of k, $k >> \lambda_0/(q-1)$, it becomes a power-like distribution of the form

$$P_q(k) \propto k^{-\gamma} \qquad \text{with} \qquad \gamma = \frac{q}{q-1}, \tag{12}$$

i.e., our distribution becomes in this limit a scale-free one. It is easy to check that if we demand that $\langle k\rangle < \infty$ then $q < 2$. It is interesting to note at this point that $\gamma = 3$ in Eq. (4) corresponds precisely to $q = 3/2$ at which variation $\mathrm{Var}(k)$ diverges.

---

[3] It should be stressed that maximalization of entropy provides us in this case only with the shape of distribution $P(k)$, Eq. (10), and gives no information on the particular values of parameters λ_0 and q. Only knowledge of moments $\langle k\rangle$ and $\mathrm{Var}(k)$ of $P_q(k)$, as given by Eq. (11), allows for determining these two parameters.

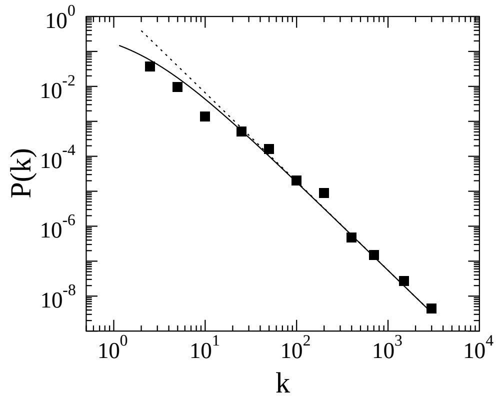

Fig. 1. The probability distribution of connections in the WWW network after [12] (full squares). The full line shows results of our fit by using Eq. (10) with $q = 1.65$ and $\lambda_0 = 1.91$. It reproduces the observed mean $\langle k\rangle = \lambda_0/(2-q) = 5.46$ and lead to the asymptotic power-like distribution $\propto k^{-\gamma}$ with $\gamma = q/(q-1) = 2.54$ (showed as dotted line).

In Fig. 1 we show (as an example) distribution of the number of connections in the WWW network [12] with 325729 nodes and the mean values of connections equal $\langle k\rangle = 5.46$ fitted by $P_q(k)$ as given by Eq. (10) with parameters $\lambda_0 = 1.91$ and $q = 1.65$ [4]. Notice that Eq. (10) describes *the whole range* of k whereas the purely power-like distribution $\propto k^{-\gamma}$ with $\gamma = q/(q-1) = 2.54$ occurs only for large values of k. In the spirit of information theory this result can therefore be interpreted in the following way: (a) the system forming network described in Fig. 1 possesses some features (mentioned above) preventing the use of Shannon information entropy and (b) data represented by $P(k)$ can be quite adequately described in terms of only two parameters: the Tsallis entropy parameter q and the mean number of links $\langle k_q\rangle = \lambda_0$, *i.e.*, their informational capacity is rather limited.

The question now is: what is the physical meaning of the q parameter in the context of stochastic networks? There is a long list of possibilities in what concerns of the origin of $q \neq 1$ to be found in the literature dealing with nonextensivity [21,22]. Out of these we shall only mention two: fluctuations and correlations. In [21] it was demonstrated that q reflects fluctuations of the parameter λ_0 in exponential distribution (1) above. In fact, it turns out

[4] Notice that although in fitting procedure both parameters were varied independently, they are connected via distribution moments $\langle k\rangle$ and $\mathrm{Var}(k)$, *cf.* (11).

that $(q-1)/q = \pm \mathrm{Var}(1/\lambda)/\langle 1/\lambda\rangle^2$. As is known from other branches of physics where Tsallis statistics is applied [22], the appearance of q can also be caused by some correlations existing in the system under consideration. Such correlations (resulting, for example, from preferential attachments and "rich-get-richer" phenomenon [11]) seem to play a decisive role in the description of stochastic networks. Therefore when choosing vertices with connectivity k, to which a new vertex is going to be connected, we shall assume that it will do so with probability that depends on the connectivity k. To illustrate this point let us introduce in the evolution equation

$$\frac{dP(k)}{dk} = -\frac{1}{\lambda(k)}P(k)\,, \tag{13}$$

parameter $\lambda = \lambda(k)$ given by a simple linear function of k:

$$\lambda(k) = \frac{[\lambda_0 + (q-1)k]}{q}\,. \tag{14}$$

It is easy to see that in this case one gets immediately $P_q(k)$ in the form of Eq. (10). Notice that for $q \to 1$ (*i.e.*, for $\lambda \to \lambda_0$) one recovers the exponential distribution (7) [5].

It must be stressed, however, that the information theory approach leads in a natural way (via maximalization of the respective information entropy) only to equilibrium (or stationary) distributions $P_q(k)$, whereas in models describing evolving complex networks [7, 11] the functional form of $P(k)$ is determined by the growth equation $\partial k/\partial t$. Introducing more complicated network evolution than the one presented above when deriving Eq. (4), for example allowing for the occurence of local events in the form of internal edges and rewirings, one gets (see Eq. (111) of [7])

$$P(k) \sim [k + \kappa(p,r,m)]^{-\gamma(p,r,m)}\,, \tag{15}$$

where p is the probability that one is adding m new edges to the system, r is probability that one is rewiring m edges and $1-p-r$ is probability that

[5] Eq. (13) can be derived by dividing master equation, $\partial P(k)/\partial t = -c\,P(k)$, by the "growth of network" $\partial k/\partial t$. One gets then evolution equation, $\partial P(k)/\partial k = -c\,P(k)\partial t/\partial k$, which for the linear dependence of the growth of network assumed here, $\partial k/\partial t = a + bk$, leads to Eq. (13) with $\lambda(k) = (a+bk)/c$. Our example considered here corresponds to the choice: $c = 1/\lambda_0$, $a = 1/q$ and $b = (1-1/q)/\lambda_0$. Notice that $1/q$ plays role of weight with which we select the constant and linear terms in the equation describing the growth of network. Notice also that this kind of the growth of network, *i.e.*, its dependence on k (*cf.* [11]) corresponds to selective probability $\Pi(k)$, which leads to power-like distribution (4).

one is adding a new node to the system. The κ and γ in Eq. (15) are given by [7]:

$$\kappa(p,r,m) = A(p,r,m) + 1 \qquad \text{and} \qquad \gamma(p,r,m) = B(p,r,m) + 1\,, \tag{16}$$

where, in turn,

$$\begin{aligned} A(p,r,m) &= (p-r)\left[\frac{2m(1-r)}{1-p-r} + 1\right], \\ B(p,r,m) &= \frac{2m(1-r) + 1 - p - r}{m}. \end{aligned} \tag{17}$$

One can now write formally Eq. (15) in the form resembling Eq. (10), *i.e.*, as

$$P(k) \sim [\kappa(p,r,m)]^{-\gamma(p,r,m)} \left[1 - (1-q)\,\frac{k}{(1-q)\kappa(p,r,m)}\right]^{\frac{q}{1-q}} \tag{18}$$

and identify $\gamma(p,r,m) = q/(q-1)$ and $\lambda_0 = (1-q)\kappa(p,r,m)$, or

$$q = 1 + \frac{1}{B} \qquad \text{and} \qquad \lambda_0 = \frac{A+1}{B}\,. \tag{19}$$

However, it must be noticed that, at least in the example considered here, the limit $q = 1$ cannot be achieved because the quantity B above is finite (for all reasonable values of parameters [7]). It means then that formula (10) is more general and captures (by means of parameters q and λ_0) some additional feature of complex networks, not present in its simple formulation (as, for example, given by Eq. (15)).

To summarize: We have demonstrated that, in order to apply the information theory approach to analysis of stochastic networks one has to use the nonextensive Tsallis information entropy S_q [22] leading to distribution $P_q(k)$ as given by Eq. (10). As shown in Fig. 1, such distribution provides satisfactory description of data on number of links in random networks *in the whole range of variable* k by means of only two parameters: the mean value of k and the parameter characterizing the type of information entropy to be chosen, q. In this way one describes such disparate situations as the exponential model ER [13] (for $q = 1$) and the scale-free, power-like models [7, 11] (with $q = \gamma/(\gamma-1)$). For the value of nonextensivity parameter $q = 3/2$, for which variance of our system is divergent, one obtains the exponent $\gamma = 3$, which seems to be limited value observed in analyses of diverse systems displaying complex topology. Although only one example has been shown here in Fig. 1, it is obvious that one can just as easily also fit other, similar results discussed in the literature (*cf.*, for example, [7, 8]) [6]. The other point is the possible systematics of the q parameter emerging from such a search, but this problem is outside of the scope of our presentation.

We conclude by saying that from the information theory point of view Eq. (10) could be used to fit different data providing a pair of numbers (q, λ_0) for each example. All competing models could be then checked for their ability to correctly reproduce these (q, λ_0) and all models reproducing them correctly should be regarded as *equally good* from the point of view of distribution $P(k)$ because, according to the philosophy of infromation theory approach, they apparently contain the same amount of information existing in data which have been used. To distinguish between them further one would have to use some additional information contained in other network measures like, for example, clustering coefficient, distance between nodes, cycles or graph spectra.

REFERENCES

[1] R. Gallagher, T. Appenzeller, *Science* **284**, 79 (1999).

[2] R.F. Service, *Science* **284**, 80 (1999).

[3] G. Weng, U.S. Bhalla, R. Iyengar, *Science* **284**, 92 (1999).

[4] C. Koch, G. Laurent, *Science* **284**, 96 (1999).

[5] S. Wasserman, K. Faust, *Social Network Analysis*, Cambridge University Press, Cambridge 1994.

[6] R. Albert, H. Jeong, A.L. Barabasi, *Nature* **400**, 130 (1999).

[7] R. Albert, A.L. Barabasi, *Rev. Mod. Phys.* **74**, 47 (2002).

[8] S.N. Dorogovtsev, J.F.F. Mendes, *Adv. Phys.* **51**, 1079 (2002).

[9] A.L. Barabasi, R. Albert, *Science* **286**, 509 (1999).

[10] S. Lawrance, C.L. Giles, *Nature* **400**, 107 (1999).

[11] A.L. Barabasi, R. Albert, H. Jeong, *Physica A* **272**, 173 (1999).

[12] H. Jeong, B. Kahn, Complex Scale-Free Networks, in Asia Pacific Center for Theoretical Physics Bulletin 07-08, 3 (2001).

[13] P. Erdös, A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960).

[14] B. Bollobasi, *Random Graphs*, Academic Press, London 1985.

[15] V. Pareto, *Cours d'Economic Politique*, Lausanne, Paris 1897.

[16] G.K. Zipf, *Human Behaviour, the Principle of Least Effort*, Addison-Wesley, Cambridge (MA) 1949.

[17] S. Render, *Eur. Phys. J.* **B4**, 131 (1998).

[6] For the most recent application of Tsallis statistics to investigation of Internet traffic problems see [23], for the previous one see [18].

[18] C. Tsallis, M.P. de Albuquerque, *Eur. Phys. J.* **B13**, 777 (2000).

[19] *Cf.*, for example, A. Hobson, *Concepts in Statistical Mechanics*, Gordon and Breach, New York 1971; H. Haken, *Information and Self-Organization*, Springer-Verlag, Berlin 1988; J. Skiling, *Nature* **309**, 748 (1984).

[20] C.E. Shannon, *Bell System Technical J.*, **27** (1948) 379; *cf.* also S. Kulback, *Information Theory and Statistics*, Wiley, New York 1959; R.J. McEliece, *The Theory of Information and Coding*, Addison-Wesley, Reading (MA) 1977.

[21] G. Wilk, Z. Włodarczyk, *Phys. Rev. Lett.* **84**, 2770 (2000); *Chaos, Solitons and Fractals* **13/3**, 581 (2001); *Physica A* **305**, 227 (2002).

[22] C. Tsallis, *J. Stat. Phys.* **52**, 479 (1988); *cf.* also C. Tsallis, *Chaos, Solitons and Fractals* **13**, 371 (2002); *Physica A* **305**, 1 (2002) and in *Nonextensive Statistical Mechanics and its Applications*, S. Abe, Y. Okamoto (Eds.), *Lecture Notes in Physics* LPN560, Springer, 2000. For updated bibliography on this subject see http://tsallis.cat.cbpf.br/biblio.htm. Its recent summary is provided in the special issue of *Braz. J. Phys.* **29**, No 1 (1999) (available also at http://sbf.if.usp.br/ WWW_pages/ Journals/BJP/ Vol29/Num1/index.htm).

[23] S. Abe, N. Suzuki, *Phys. Rev.* **E67**, 016106 (2003).