INFORMATION THEORY POINT OF VIEW ON STOCHASTIC NETWORKS*

G. Wilk

The Andrzej Sołtan Institute of Nuclear Studies
Hoża 69, 00-681 Warsaw, Poland
wilk@fuw.edu.pl

and Z. Włodarczyk

Institute of Physics, Świętokrzyska Academy
Świętokrzyska 15, 25-406 Kielce, Poland
and
University of Arts and Sciences (WSU)
Wesoła 52, 25-353 Kielce, Poland
wlod@pu.kielce.pl

(Received April 11, 2005)

Stochastic networks represent very important subject of research because they have been found in almost all branches of modern science, including also sociology and economy. We provide a information theory point of view, mostly based on its nonextensive version, on their most characteristic properties illustrating it with some examples.

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

1. Introduction

Different kinds of stochastic networks show up in nature whenever one is dealing with complex systems of any kind[1]. There are two basic types of stochastic networks:

(a) Networks with constant number of nods, M, for which probability that given node has k connections with other nodes (k links) is Poissonian [3]

$$P(k) = \frac{\kappa_0^k}{k!}\, e^{-\kappa_0}\,, \qquad \kappa_0 = \langle k\rangle\,. \tag{1}$$

* Presented at the First Polish Symposium on Econo- and Sociophysics, Warsaw, Poland, November 19–20, 2004.

[1] We refer to [1] for list of relevant references to which we would like to add recent application of networks to describe some features of multiparticle production processes in high energy hadronic collisions [2].

(b) Networks in which number of nodes is not stationary and distribution of links $P(k)$ is given by dynamics of the growth of network [4] and varies between being *exponential*

$$P(k) = \frac{1}{m}\,\exp\left(-\frac{k}{m}\right)\,, \tag{2}$$

(for the case when each new node connects with the already existing ones with equal probability $\Pi(k_i) = 1/(m_0+t-1)$ independent on k_i) or being *power-like*,

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

for the case of preferential attachment (the so called "rich-get-richer" mechanism, here $m < m_0$ is the number of new nodes added in each time step) with, in this case, $\Pi(k_i) = k_i/(2mt)$ choice.

Recently we have demonstrated that using information theory approach in its nonextensive version (*i.e.*, maximizing Tsallis entropy [5])

$$H_q = \left[1 - \sum_k P_q^q(k)\right] \Big/ (q-1)\,,$$

under conditions that

$$\langle k\rangle_q = \sum_k k P_q^q(k) \Big/ \sum_k P_q^q(k) = \kappa_0$$

and $\sum_k P_q(k) = 1$ one obtains that [1]

$$P_q(k) = \frac{1}{\kappa_0}\left[1-(1-q)\frac{k}{\kappa_0}\right]^{\frac{q}{1-q}}\,, \tag{4}$$

which is very universal because for $q \to 1$ it recovers Eq. (2) whereas for $k \gg \kappa_0/(q-1)$ it leads to the power-like distribution $P_q(k) \propto k^{q/(1-q)}$. For properly chosen parameters κ_0 and q it is, therefore, able to describe data in the *whole region* of variable k. As an example, in [1] distribution of WWW network after [6] with the mean number of connections $\langle k\rangle = 5.46$ has been described by $P_q(k)$ with parameters $\kappa_0 = 1.91$ and $q = 1.65$ obtaining for large values of k power-like distribution $\propto k^{-\gamma}$ with $\gamma = q/(q-1) = 2.54$ as observed in [6]. Notice that condition of finiteness of the first moment of $P_q(k)$, equal to $\langle k\rangle = \kappa_0/(2-q)$, results in the limitation that $q < 2$, whereas similar condition imposed on the variance of $P_q(k)$, given by $\mathrm{Var}(k) = \kappa_0^2/[(3-2q)(2-q)^2]$, limits q further to $q < 3/2$ (it is worth to stress that for $q = 3/2$, at which variance $\mathrm{Var}(k)$ diverges, one gets exponent $\gamma = 3$, as in Eq. (3)).

Since then, detailed analysis of preferential attachment growth and its connection with nonextensive statistical mechanics has been performed in [7] whereas in [8] it was demonstrated that introducing notion of fluctuations to a random graph one obtains, as one of the results, the scale-free power-like networks (in very much similar way as fluctuations of parameter $1/m$ in distribution (2) given entirely by parameter q lead directly to distribution $P_q(k)$ [9]). In what follows we shall provide two additional particular examples of using methods described in [1, 9] to analysis of some stochastic network features.

2. Transport on network

Investigations of transport in the Internet [10, 11] show that the distribution of travel times can be described by power distribution of the form $P(t) \propto t^{-(\alpha+1)}$. Such form implies specific dynamic of transport on the network in which many correlated packets travel at the same time. Detailed description of such dynamical picture is particularly complicated because of the lack of some global navigation prescription, by the fact that packets can be send parallel fashion and, finally, by dependence on the structure of network and on the algorithm of selecting the transportation path used.

To describe this complicated character of transport let us introduce transport rate distribution function $f(\tau)$ and write distribution of the transportation times as

$$P(t) = \int_0^\infty d\tau f(\tau) \exp\left(-\frac{t}{\tau}\right) . \tag{5}$$

Function $f(\tau)$ can be obtained by considering stochastic process given by the following Langevin equation

$$\frac{d\tau}{dt} + [1 + \xi(t)]\, \tau = \tau_0, \tag{6}$$

where $\xi(t)$ is white Gaussian noise with $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t + \Delta t)\rangle = 2D\delta(\Delta t)$, parameter τ_0 is the characteristic transport time on network and D denotes variation of transport times. Considering now the corresponding Fokker–Planck equation one gest (*cf.* [9]) that[2]

[2] It should be noticed that distribution just obtained is the most expected one from the point of view of the maximization of Shannon information entropy, $H = -\sum_k P(k) \ln P(k)$ under constraints that $\int f(\tau) d\tau = 1$, $\langle \tau \rangle = \tau_0$ and (because distribution we are looking for is defined only for $\tau > 0$) that $\langle \ln(\tau) \rangle = \ln(\tau_0/D)$ [12].

$$f(\tau) = \frac{\mu^\nu}{\Gamma(\nu)\tau^{\nu-1}} \ \exp\left(-\frac{\mu}{\tau}\right) , \tag{7}$$

where $\mu = \nu\tau_0$ and $\nu = 1/D$, *i.e.*, it is given by gamma distribution in the variable $1/\tau$. As was shown in [9] such function $f(\tau)$ results in the power-like distribution of $P(t)$

$$P(t) = \frac{2-q'}{\tau_0}\left[1-(1-q')\frac{t}{\tau_0}\right]^{\frac{1}{1-q'}} , \qquad \text{where} \quad q' = 1 + D\,. \tag{8}$$

In [11] it has been shown that exactly such form is observed experimentally. It means, therefore, that approach based on Tsallis statistics describes stationary states on Internet.

Power-like asymptotic form of $P(t) \propto t^{-(\alpha+1)}$ implies existence of long-range correlations in the network communication with formation of large active aggregates of transportation streams [13]. Denoting aggregates by the number of active streams existing in time t, K_t, these long-range correlations are given by auto-covariance $R(t) = \mathrm{Cov}(K_t, K_{t+s})$ (typically, for the Internet transport, the written above power-like form of $P(t)$ leads directly to similar power-like form of $R(t) \propto t^{-(\alpha-1)}$ with $\alpha \in (1,2)$)

$$R(t) = \int_t^\infty dy[1-F(y)]\,, \tag{9}$$

where $F(t)$ denotes distribuant of $P(t)$ [14]. It turns out that using formalism of random matrices (where matrix elements W_{ij} have random probability distributions) and identifying the number of decay channels with the size K of the aggregate, one can immediately obtain that [15] exponent α above is given by the mean size of the aggregate $\langle K\rangle$

$$\alpha = \frac{2-q'}{q'-1} = \frac{1}{D} - 1 = \frac{1}{2}\langle K\rangle - 1\,. \tag{10}$$

Notice that with increasing size of the aggregate, $\langle K\rangle$, fluctuations in $f(\tau)$ diminish, $\langle(\tau^{-1} - \langle\tau^{-1}\rangle)^2\rangle = 2\langle\tau^{-1}\rangle^2/\langle K\rangle$, and in the limit of large $\langle K\rangle$ the power-like behavior in (8) becomes exponential one (as now $q' \to 1$).

3. Epidemic dynamics in network

Let us now concentrate on the dynamic of spreading viruses using simple model of the SIS type (susceptible-infects-susceptible). In this model each node can be either "healthy" (H) or "infected" (I) and infection spread up due to connections between nodes. On every time step node H is infected with probability ν if it is linked with at least one link with node I. At the same time node I is cured with probability δ (*i.e.*, the mean time of infection is equal $D = 1/\delta$). Denoting the effective rate of diffusive spreading up the virus by $\lambda = \nu/\delta$ it can be shown [16] that probability density of nodes I is given by

$$\begin{aligned} \rho &= 0\,, && \text{for} \quad \lambda < \lambda_c\,, \\ \rho &\propto (\lambda - \lambda_0)^{\beta}\,, && \text{for} \quad \lambda > \lambda_c\,, \end{aligned} \tag{11}$$

i.e., virus whose spreading rate λ exceeds the threshold value λ_c survives whereas it quickly vanishes when his spreading rate is below the epidemic threshold [17, 18]. This threshold depends on the variation of the number of connections in the network, $\lambda_c = \langle k\rangle/\langle k^2\rangle$. In regular networks (where $P(k) = \delta(k - k_0)$) and in stochastic networks (where $P(k)$ is Poissonian) $\langle k^2\rangle$ is always finite and, therefore, λ_c is always greater than zero. Situation changes drastically for scale-free networks for which $P(k) \propto k^{-\gamma}$, with $\gamma \le 3$. In this case for $\gamma \to 3$ one has $\lambda_c \to 0$ and there is *no threshold for the epidemic*[3].

Let us analyze this in more detail. Denoting by x_i the part of nodes of type i (*i.e.*, with i connections) which are able to be infected and by y_i the part of already infected nodes, one can write the following evolution equation (where $\nu_{ij} = ij\nu/\langle k\rangle$)

$$\frac{\partial x_i}{\partial t} = -x_i \sum_j \nu_{ij} y_j\,, \qquad \frac{\partial y_i}{\partial t} = x_i \sum_j \nu_{ij} y_j - y_i \delta\,. \tag{12}$$

Solving (12) one can show [19] that the part of nodes which are *always* infected (it is called *the final epidemic size*) is given by

$$I = \langle 1 - \exp(-k\alpha)\rangle\,, \qquad \text{where} \qquad \alpha = \rho_0 \frac{\langle k - k\exp(-k\alpha)\rangle}{\langle k^2\rangle}\,. \tag{13}$$

Here $\rho_0 = \nu D\langle k\rangle$ is the mean number of the secondary infections caused by introducing one infected node in the network (under assumption that each node has exactly $\langle k\rangle$ connections).

For regular network with $P(k)$ given by Poisson distribution (Eq. (1)) the size of epidemy is

$$I = 1 - \exp\left[-\langle k\rangle \left(1 - e^{-\alpha}\right)\right]\,, \tag{14}$$

[3] For example, for $P(k)$ given by Eq. (4) $\lambda_c = (3 - 2q)/(2\kappa_0)$, which for $q \to 3/2$ tends to zero.

where α is given by the following transcendental equation:

$$\alpha = \frac{\rho_0}{\langle k \rangle} \left\{ 1 - \exp\left[-\alpha - \langle k \rangle \left(1 - e^{-\alpha} \right) \right] \right\} . \tag{15}$$

Assuming small α and keeping in (15) only linear terms in α one obtains that

$$\alpha \simeq \frac{2\left[1 + \langle k \rangle \left(1 - \frac{1}{\rho_0} \right) \right]}{(1 + \langle k \rangle)^2 + \langle k \rangle}, \tag{16}$$

from which one can estimate the epidemic threshold (which is for $I = 0$, *i.e.*, also for $\alpha = 0$) as being given by

$$\frac{1}{\rho_0} = 1 + \frac{1}{\langle k \rangle}. \tag{17}$$

For the scale free network with $P(k)$ given by Eq. (3), for small values of ρ_0 we get the following size of epidemy:

$$I \sim 2e^{0.423} \exp\left(-\frac{2}{\rho_0} \right) . \tag{18}$$

In the general case described by $P_q(k)$ as given by Eq. (4) one has, in the approximation, that $\kappa_0 \alpha > 1$, and that

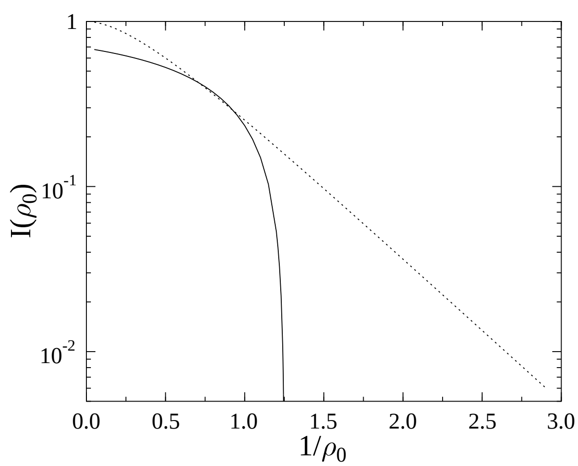

Fig. 1. Examples of behavior of the final epidemic size $I(\rho_0)$ as function of inverse of the mean number of secondary infections caused by introduction to network single infected virus, ρ_0, for network with mean number of connections equal $\langle k \rangle = 4$. Full line corresponds to Poissonian distribution of links (regular networks), dashed line is for power-like distributions of links (scale-free networks). Notice that in the latter case there are situations in which there is no threshold to epidemic.

$$I = 1 - \frac{1}{\alpha} \frac{\langle k \rangle}{q(2-q)}\,, \tag{19}$$

where α is given by the equation

$$\alpha = \frac{\rho_0}{\langle k \rangle} \left[1 - \frac{1}{\alpha^2 q(2-q)} \right]. \tag{20}$$

As is clearly demonstrated in Fig. 1 only in the former case there is epidemic threshold, in the latter one no such effect appears.

4. Conclusions

Stochastic networks represent very complex phenomena of different origin. As such they have been also investigated by using statistical mechanic approach [21]. Here we have shown that they can be also described (at least in what concerns some of their most commonly demonstrated properties) by using information theory approach based on Tsallis statistics. Such approach allows to describe by means of single formula probabilities of number of connections in a given network, $P_q(k)$, in essentially all kinds of networks, from purely exponential ones (with $q = 1$) to a scale-free power-like ones with exponent $\gamma = q/(q-1)$. In this approach it is clear that $\gamma = 3$, found in many systems with complex topologies, corresponds to $q = 3/2$, a value for which variance of $P_q(k)$ diverges.

We have presented two examples of investigations of networks by means of Tsallis power-like formula (4), transport on networks and development of epidemic on networks. Both have numerous (and still growing) practical applications to which others are added (like, for example, description of the earthquakes statistics [22]). As an example we provide here two results, one concerning description of known distributions of sexual partners, in Fig. 2, and one illustrating distribution of populations of different agglomerations, in Fig. 3. Both can be very well described by means of only two parameters, κ_0 and q. The latter parameter describes, according to [1, 7, 8], summarily the influence of dynamic of the network (resulting in intrinsic fluctuations, correlations, fractal structure and the like). Let us close with remark that different behavior of different epidemics (existence of threshold or not) is naturally explained by the different character of networks of contacts causing infection: those corresponding to regular networks (like, for example, flue epidemic mostly following earth communication network) show some threshold whereas those corresponding with scale-free networks (like, for example, AIDS epidemic) have no thresholds.

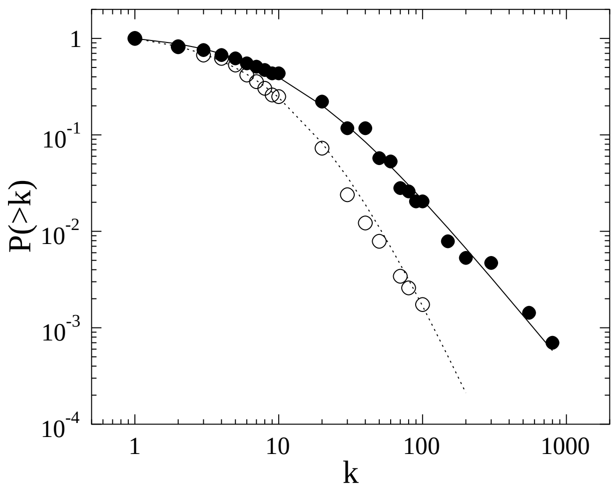

Fig. 2. Cumulative probability distribution of the number of sexual partners as given by [20] compared with Eq. (4) for men (full line and symbols; here $q = 1.55$ and $\kappa_0 = 0.68$ resulting in $\langle k\rangle = 15$) and women (dashed line, open symbols; here $q = 1.3$ and $\kappa_0 = 0.48$ resulting in $\langle k\rangle = 7$.

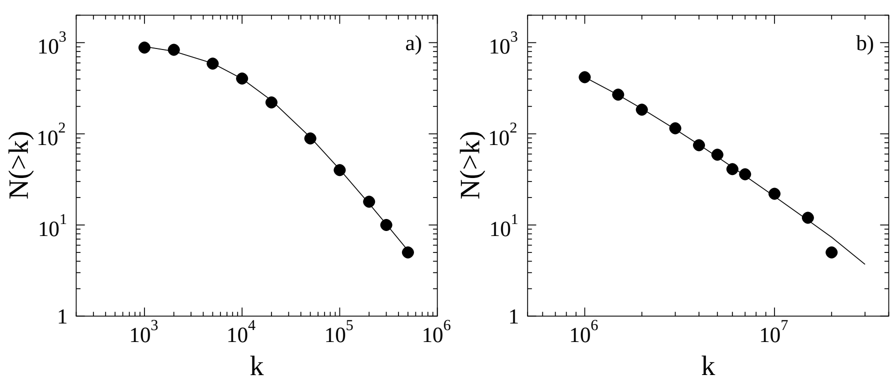

Fig. 3. Examples of unnormalized cumulative probability distributions of cities in Poland with population greater than some given value k on this value, (a), and the same for agglomerations in the whole world, (b). Data (points) were taken directly from [23], whereas curves represent fits to Tsallis formula: $N(>k) = C\,P_q(q)$ (with the following sets of parameters: $C = 1032$, $\kappa_0 = 7280$ and $q = 1.75$ for (a) and $C = 2362$, $\kappa_0 = 310000$ and $q = 1.65$ for (b)). For other similar results see [24].

REFERENCES

[1] G. Wilk, Z. Włodarczyk, *Acta Phys. Pol. B* **35**, 871 (2004).

[2] G. Wilk, Z. Włodarczyk, *Acta Phys. Pol. B* **35**, 2143 (2004).

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

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

[5] C. Tsallis, *J. Stat. Phys.* **52**, 479 (1988); *cf.* also C. Tsallis, in *Nonextensive Statistical Mechanics and Its Applications*, S. Abe and Y.Okamoto (Eds.), Lecture Notes in Physics LPN560, Springer 2000. For updated bibliography on this subject see `http://tsallis.cat.cbpf.br/biblio.htm`.

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

[7] D.J.B. Soares, C. Tsallis, A.M. Mariz, L.R. da Silva, *Europhys. Lett.* **70**, 70 (2005); *cf.* also C. Tsallis, M. Gell-Mann, Y. Sato, *Special Scale-invariant Occupancy of Phase Space Makes the Entropy S_q Additive*, `cond-mat/0502274`.

[8] S. Abe, S. Thurner, *Analytic Formula for Hidden Variable Distribution: Complex Networks Arising from Fluctuating Random Graphs*, `cond-mat/0501429`.

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

[10] B. Tadic, G.J. Rodgers, *Complex Syst.* **5**, 445 (2002).

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

[12] E.W. Monroll, M.F. Schlesinger, *J. Stat. Phys.* **32**, 209 (1983).

[13] D. Heath, S. Resnik, G. Samorodnitsky, *Math. Oper. Tres.* **23**, 145 (1998).

[14] S. Resnick, G. Smorodnitsky, *Queueing Syst.* **33**, 43 (1999).

[15] G. Wilk, Z. Włodarczyk, *Phys. Lett.* **A290**, 55 (2001).

[16] Z. Dezso, A.L. Barabasi, *Phys. Rev.* **E65**, 055103(R) (2002).

[17] R. Pastor-Satorras, A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).

[18] R. Albert, H. Jeong, A.L. Barabasi, *Nature* **406**, 378 (2000).

[19] R.M. Anderson, R.M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, Oxford 1991.

[20] F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley, Y. Aberg, *Nature* **411**, 907 (2001).

[21] *Cf.*, for example, A. Majka, W. Wiślicki, *Physica A* **337**, 645 (2004) and references therein.

[22] S. Abe, N. Suzuki, *Physica A* **337**, 357 (2004).

[23] See data basis in `http://www.citypopulation.de`.

[24] L.C. Malacarne, R.S. Mendes, E.K. Lenzi, *Phys. Rev.* **E65**, 017107 (2001).