CHIN.PHYS.LETT. Vol. 25, No. 11 (2008) 4177

Network Entropy Based on Topology Configuration and Its Computation to Random Networks *

LI Ji(李季)[1,2**], WANG Bing-Hong(汪秉宏)[2], WANG Wen-Xu(王文旭)[2], ZHOU Tao(周涛)[2]
[1]*Department of Physics, Fuyang Normal College, Fuyang 236041*
[2]*Department of Modern Physics, University of Science and Technology of China, Hefei 230026*

(Received 19 April 2008)

A definition of network entropy is presented, and as an example, the relationship between the value of network entropy of ER network model and the connect probability p as well as the total nodes N is discussed. The theoretical result and the simulation result based on the network entropy of the ER network are in agreement well with each other. The result indicated that different from the other network entropy reported before, the network entropy defined here has an obvious difference from different type of random networks or networks having different total nodes. Thus, this network entropy may portray the characters of complex networks better. It is also pointed out that, with the aid of network entropy defined, the concept of equilibrium networks and the concept of non-equilibrium networks may be introduced, and a quantitative measurement to describe the deviation to equilibrium state of a complex network is carried out.

PACS: 89.75.Hc

The complex interaction relationship between individuals in the real world can be abstractly described by a network. In the network, each node represents each individual in the system, and interactions between the individuals can be expressed in edges in the network. For a network system with a large number of nodes, Erdös and Renyi made a pioneering work in the 1960s and gave the well-known ER random network model.[1] However, the development of information technology makes it possible to perform empirical study on the actual real-world network. It was empirically found that some of the actual statistical nature of major networks is not consistent with the ER model.[2−5] In recent years, represented in the small world network model (WS model) given by Watts and Strogatz [6] and the scale-free network model (BA Model) given by Barabási and Albert,[7] many researchers have given different network models.[8−17] These models described the statistical characters of the actual network from different aspects better. In essence, these network models and the ER model are all presented as random networks. Networks generated by each test are not exactly the same. Therefore, such network systems are random systems.

A random system has some uncertainties. Entropy is usually used to describe the uncertainty of a system. In 1948, Shannon established general entropy theory.[18] For complex networks, many researchers have also given different definitions of entropy in recent years.[19−21] These network entropies have played very good roles in understanding of complex networks. The most frequently cited definition of entropy[22−28] is the Shannon entropy defined in the literature.[19] However, the specific calculation shows that the difference between these entropies of different sizes and different types of networks is not significant, and can not sensitively distinct different networks. In this Letter, we try to define an entropy based on the topology configuration of complex network, and expect that this entropy can describe the character of complex networks better. Specifically, we calculate the entropies with the ER network model, to test if the entropy is sensitive in network size and linking probability to the ER network model. In addition, it will be discussed if new physics concept can be proposed by the introduction of the entropy we defined.

The information theory tells us that, for a system that has 2^n possible states, if the probability of each state is the same, n-bit information is required to determine which state the system is in. In other words, $\log_2 \Omega$ bit information is necessary to consist of a state from Ω possible states which have the same probability. Therefore, as it reflects the uncertainty of a system, $\log_2 \Omega$ can be defined as the system's entropy. For example, a throwing up dice has 6 possible states if different faces landing on the floor represent different states. The probability for each state is 1/6. However, if the dice is not a cube shape, each probability of an upper surface appearing is no longer equal to each other. A more extreme example is that a thin sheet dice has the probability 1/2 for two faces and 0 for the remained four faces. Although the system

*Supported by the National Basic Research Programme of China under Grant No 2006CB705500, the National Natural Science Foundation of China under Grant Nos 60744003, 10635040, 10532060 and 10472116, and the Specialized Research Fund for the Doctoral Programme of Higher Education of China.
**Email: liji@fync.edu.cn

at this time has still 6 states, only one-bit information can determine the state of the dice. Obviously, $\log_2 \Omega$ can not simply be used as the system's entropy. Shannon's entropy theory has put forward a formula for the situation that the probabilities are different for different states,[18] i. e.

$$S(\Omega, P) = -K \sum_{i=1}^{\Omega} P_i \log P_i, \tag{1}$$

where Ω is the number of possible system states, P_i is the probability of the ith state, K is a constant. For the sake of brevity, $K = \log_2 10$ is suitable. Then formula (1) can be rewritten as

$$S(\Omega, P) = -\sum_{i=1}^{\Omega} P_i \log_2 P_i. \tag{2}$$

The above definition of entropy is generally called the Shannon entropy.

For the complex network, from different points of view, a range of definitions of entropy has been proposed. Sole and Valverde[19] also presented a definition of Shannon entropy of complex network as follows:

$$H(\boldsymbol{q}) = -\sum_{k=1}^{N} q(k) \log(q(k)), \tag{3}$$

where $\boldsymbol{q} = (q(1), \cdots, q(i), \cdots, q(N))$, and $q(k)$ is defined as

$$q(k) = \frac{(k+1)P_{k+1}}{\langle k \rangle}, \tag{4}$$

and it is named as the remaining degree, with P_k being the distribution of degrees, and $\langle k \rangle$ the average of degrees.

A number of results of Shannon entropy have been presented in the literature,[19] by computing at some real network systems and network models.

From the definition of entropy and the calculation results given in Ref. [19], we cannot believe that this definition of entropy is better from the viewpoint of characterizations of complex networks, at least it is not appropriate to name it as Shannon entropy. Firstly, with the Shannon entropy of the original definition of inconsistency, $q(k)$ in Eq. (3) is not the distribution on all possible states; secondly, for different sizes or different types of networks, this definition does not lead to significant differences when one calculates the entropy.

In order to describe the uncertainty of complex networks, we propose a new network entropy concept based on the topology configuration of network. For a complex networks generated according to certain rules, in the given parameters, each test can generate a specific network configuration. Repeated tests, a wide range of configurations will be produced. Naming all possible configuration number as Ω, the probabilities for various configurations as $P_i, (i = 1, 2, \ldots, \Omega)$, the network entropy can be defined as

$$S(\Omega, P) = -\sum_{i=1}^{\Omega} P_i \log_2 P_i. \tag{5}$$

Because this definition of entropy is suitable to the general networks, not only confined to complex networks, we call it the topology entropy of network. As an example, the ER random network model will be analysed, and the entropy will be calculated.

Table 1. Results of the network Shannon entropy defined in Ref. [19] for different networks.

Network type	N	$\langle k \rangle$	$H(q)$
Technological networks			
Software 1	168	2.81	3.04
Software 2	159	4.19	3.99
Internet AS	3200	3.56	4.77
Software 3	1993	5.00	4.82
Circuit TV	320	3.17	1.37
Circuit EC05	899	4.14	2.98
Software Linux	5285	4.29	4.47
Power grid	4941	2.67	3.01
Biological networks			
Silwood park	154	4.75	4.09
Ythan estuary	134	8.67	4.74
p53 subnetwork	139	5.09	4.00
Metabolic map	1173	4.84	3.58
Neural net (C.elegans)	297	14.5	5.12
Metabolic map	821	4.76	3.46
Romanian syntax	5916	5.65	5.45
Proteome map	1458	2.67	3.85
Theoretical systems			
Star graph	17	1.88	1.00
Barabási-Albert	3000	3.98	4.12
Erdös-Renyi	300	6.82	3.31
Modular E-R	500	10.3	3.67

The ER network model gives a typical random undirected network. We try to calculate the value of topology entropy of ER networks by means of analysis and simulation. For the ER network model, N is the total number of nodes and p is the probability of any pair of nodes connected. Generally, $p \in [0, 1]$. Especially, if $p = 1$, the total number of edges is equal to $N(N-1)/2$. Here $N(N-1)/2$ possible connected processes can be treated as $N(N-1)/2$ independent random events. We mark $M = N(N-1)/2$, the actual link edge number m is a random variable, it will comply the binomial distribution

$$P(m) = \frac{M!}{m![M-m]!} p^m \cdot (1-p)^{M-m} \quad (m = 0, 1, 2, \ldots, M). \tag{6}$$

For M possible links and m certain links, total scheme number is

$$C_M^m = \frac{M!}{m!(M-m)!}. \tag{7}$$

This result is based upon the assumption that the nodes are distinguishable.

Mark the possible configurations as $\Omega(N,p)$ when total nodes are N and probability p, there are

$$\Omega(N,p)=\sum_{m=0}^{M}C_M^m=\sum_{m=0}^{M}\frac{M!}{m!(M-m)!},\quad \text{for } M=N(N-1)/2. \tag{8}$$

For each m, there are C_M^m possible configurations correspondingly. All the configurations in these C_M^m different configurations appear in the same probability, and is equal to $1/C_M^m$ without exception. Notes that this is a conditional probability which links numbers are m, so the realization of a random network has N nodes, and link probability is p, here m edges can be regard as a random incident A_i which has occurrence probability

$$P_i=P(m)/C_M^m=p^m(1-p)^{M-m},\quad \text{for } i=1,2,\ldots,\Omega(N,p). \tag{9}$$

As the probabilities of each C_M^m configuration for certain m are the same, the entropy value aimed to $\sum_{i=0}^{M}C_M^m$ total configurations can be calculated for different m respectively, so S can be written as

$$\begin{aligned}&S(N,p)\\ =&-C_M^0[p^0(1-p)^{M-0}]\log_2[p^0(1-p)^{M-0}]\\ &-C_M^1[p^1(1-p)^{M-1}]\log_2[p^1(1-p)^{M-1}]\\ &-C_M^2[p^2(1-p)^{M-2}]\log_2[p^2(1-p)^{M-2}]-\cdots\\ &-C_M^M[p^M(1-p)^{M-M}]\log_2[p^M(1-p)^{M-M}]\\ =&-\sum_{m=0}^{M}C_M^m[p^m(1-p)^{M-m}]\\ &\cdot\log_2[p^m(1-p)^{M-m}].\end{aligned} \tag{10}$$

Given Eq. (10) a further simplification, we obtain the ER random network entropy based on topology configuration in an analytical form:

$$\begin{aligned}S(N,p)=&-\sum_{m=0}^{M}C_M^m[p^m(1-p)^{M-m}]\\ &\cdot[m\log_2 p+(M-m)\log_2(1-p)]\\ =&-\sum_{m=0}^{M}P(m)[m\log_2 p+(M-m)\log_2(1-p)]\\ =&-\langle m\rangle\log_2 p-\langle M-m\rangle\log_2(1-p),\end{aligned} \tag{11}$$

where $M\to\infty$ when $N\to\infty$, then $P(m)$ tends to a Gaussian distribution

$$P(m)=\frac{1}{\sigma_M\sqrt{2\pi}}\exp\Big(-\frac{(m-\langle m\rangle)^2}{2\sigma_M^2}\Big), \tag{12}$$

$$\langle m\rangle=Mp=\frac{1}{2}N(N-1)p,$$

$$\sigma_M=\sqrt{Mp(1-p)}=\sqrt{N(N-1)p(1-p)/2}. \tag{13}$$

Entropy $S(N,p)$ as functions of N and p can be obtained,

$$\begin{aligned}S(N,p)&=-M[p\log_2 p+(1-p)\log_2(1-p)]\\ &=-\frac{1}{2}N(N-1)[p\log_2 p+(1-p)\log_2(1-p)].\end{aligned} \tag{14}$$

In order to verify Eq. (14), we take the simulation method on ER network and obtain the experimental results of entropy (Fig. 1). The algorithms are as follows.

Given the size of the network N and the link probability p, the first test generates a random ER network. Memorize the network with an adjacency matrix storage, and as a template, set the template's matching number to be 1. The next test generates an another random ER network, and compare it with existing templates, if it fully consists of the known template, we call it matching with the template, and the corresponding template's matching number plus 1. Otherwise, memorize the ER network generated just as a new template, and set its matching number to be 1. Repeat the above steps enough times, collect various template matching numbers, the templates' probability distribution can be obtained. As each template corresponds with one topology configuration, the templates' probability distribution is just the probability distribution P_i of various configuration of the ER network for given N and p. However, P_i to Eq. (5) we can obtained the entropy in this situation.

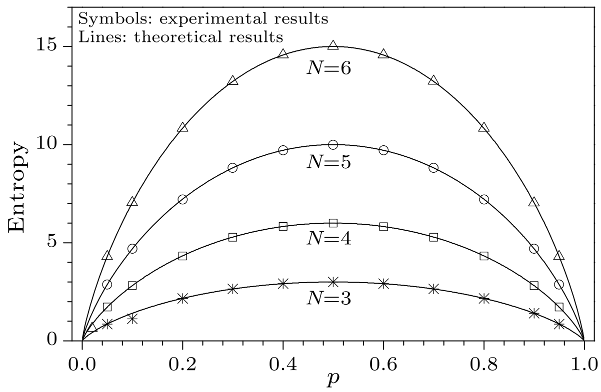

Fig. 1. Comparison of experimental results with analytical results of network entropy based on topology configuration of the ER network model.

On cases of that the network size N is respectively 3, 4, 5, 6, and link probability p is respectively 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, ER networks are generated randomly and the numbers of each configuration are recorded. The total test numbers are all $T=10^6$ when $N=3$, 4, 5, and when $N=6$ the total test number is $T=10^7$. The experimental results are marked in different symbols in

Fig. 1. The lines through the symbols in Fig. 1 are plotted according to Eq. (14). It is clear that the experimental results agree well with the theoretical results.

The network entropy based on topology configuration presented here is consistent with the original definition of Shannon entropy, in which the probability distribution is a description of possible state of network. Compared to Ref. [19], it should be more appropriate that we call the entropy defined here the Shannon entropy of network. Because of the well-known Shannon entropy defined in Ref. [19], the network entropy based on the topology configuration defined here can be called the topology entropy of network, and the entropy in Ref. [19] the degree distribution entropy.

Our result shows that even only to ER random network, in the case of different network size N and different link probability p, the topological configuration entropy values are significantly different from each other. Therefore, the network entropy based on topology configuration can be a new physical quantity describing the complex network character better.

Xie *et al.* studied the rewiring network model in Ref. [29], and the results have shown that a network with fixed node number N takes the random rewiring possibility $G(k) = \dfrac{k+\alpha}{2E+N\alpha}$, where k is the degree of the networks, E is the total link number, α is a non-minus constant. Taking α and $\gamma = 2E/N$ with different values, the system evolutes into the networks in Poisson's form, power-law form and exponential form after a period of time evolution processing. Inspired by the network in Ref. [29], when we introduce the rewiring mechanism, a network with fixed nodes can be treated as a thermal dynamical system, each edge in the networks rewires in certain rules. When the rewiring possibility is a constant, no extra information is needed in the processing, the system evolutes into a random network, and the degree distributes in Poisson's form. When each edge rewiring possibility is related to k, the system needs the degree distribution at present, namely needs extra information, and the system evolutes to one of other kind of networks. For instance, when $\alpha \ll 1$, the system evolutes into network with degree distribution in a power-law form. Taking the rewiring network without extra information as an equilibrium network,one needs extra information input continually as a non-equilibrium network, a statistical value is needed to describe the deviation of the system. Because of the assume that the properties of all configurations of ER network are equivalent for given network size N and edge number m, among various evolution network models rewind ceaselessly in different rules in the case of given nodes number N and edge number m, ER network entropy value will be the greatest. If we regard the ER network as a kind of evolution networks in an equilibrium state, the difference of entropy values between an evolution network and the ER network can be used as the deviation degree to describe the evolution network departing from the equilibrium state.

For example, to describe the deviation of the network, we compare the network's entropy S_x with the entropy S_0 of the ER network with the same nodes. We can define deviation $D = (S_0 - S_x)/S_0$, and D will belong to [0, 1]. When $D = 0$, the network is of equilibrium, $D \neq 0$ means that the network is not of equilibrium, the larger the D value, the more the system deviated from equivalent state.

Therefore, via the network entropy we defined and rewiring network presented in Ref. [29], we can introduce the conceptions of equilibrium network and non-equilibrium network, and estimate the degree of a network deviating away from the equilibrium state.

References

[1] Erdös P and Renyi A 1960 *Publications of the Mathematica Institute of the Hungarian Academy of Sciences* **5** 17
[2] Faloutsos M et al 1999 *Comput. Commun. Rev.* **29** 251
[3] Pastor-Satorras R et al 2001 *Phys. Rev. Lett.* **87** 258701
[4] Newman M E J 2001 *PNAS* **98** 404
[5] Li M et al 2007 *Physica* A **375** 355
[6] Watts D J and Strogatz S H 1998 *Nature* **393** 440
[7] Barabási A L et al 2000 *Science* **287** 2115a
[8] Albert R and Barabási A L 2002 *Rev. Mod. Phys.* **74** 47
[9] Newman M E J 2003 *SIAM Rev.* **45** 167
[10] Wang X F 2002 *Int. J. Bifur. Chaos* **12** 885
[11] Holme P and Kim B J 2002 *Phys. Rev.* E **65** 026107
[12] Klemm K and Eguíluz V E 2002 *Phys. Rev.* E **65** 036123
[13] Andrade Jr J S et al 2005 *Phys. Rev. Lett.* **94** 018702
[14] Wang W X et al 2005 *Phys. Rev. Lett.* **94** 188702
[15] Zhou T et al 2005 *Phys. Rev.* E **71** 046141
[16] He Y et al 2004 *Acta Phys. Sin.* **53** 1710 (in Chinese)
[17] Li J et al 2006 *Acta Phys. Sin.* **55** 4051 (in Chinese)
[18] Shannon C E A 1948 *Bell System Techn. J.* **27** 379
Shannon C E A 1948 *Bell System Techn. J.* **27** 623
[19] Sole R V and Valverde S 2004 *Information Theory of Complex Networks: on Evolution and Architectural Constraints, Lect. Notes Phys.* **650** 189 (Berlin: Springer)
[20] Allegrini P et al 2004 *Chaos Solitons Fractals* **20** 95
[21] Wang W N et al 2004 *Chin. Phys. Lett.* **21** 243
[22] Tan Y J and Wu J 2004 *The Systems Engineering Theory Practices* (6) (in Chinese)
[23] Wilk G et al 2004 *Acta Phys. Polonica* B **35** 871
[24] Marr C and Hutt M T 2005 *Physica* A **354** 641
[25] Venkatasubramanian V et al 2006 *Aiche J.* **52** 1004
[26] Fu B B and Gao Z Y 2006 *Chin. Phys. Lett.* **23** 520
[27] Wang B et al 2006 *Physica* A **363** 591
[28] Lezon T R et al 2006 *Proc. Nat. Acad. Sci.* U.S.A. **103** 19033
[29] Xie Y B, Zhou T and Wang B H 2008 *Physica* A **387** 1683