Article

Chinese Science Bulletin

Statistical Physics and Mathematics for Complex Systems

December 2011 Vol.56 No.34: 3677–3682
doi: 10.1007/s11434-011-4697-3

Application of varentropy as a measure of probabilistic uncertainty for complex networks

JIANG Jian[1,2,3*], WANG Ru[1,4], PEZERIL Michel[3] & WANG Qiuping Alexandre[1,3]

[1] *Institut Supérieur des Matériaux et Mécaniques Avancés 44, Avenue F.A. Bartholdi, 72000 Le Mans, France;*
[2] *Complexity Science Center, Institute of Particle Physics, Central China Normal University, Wuhan 430079, China;*
[3] *LPEC, Faculté des Sciences et Techniques, Université du Maine, Ave. O. Messiaen, 72035 Le Mans, France;*
[4] *College of Information Science and Engineering, Huaqiao University, Quanzhou 362021, China*

Received April 9, 2011; accepted May 18, 2011

Varentropy is used as a general measure of probabilistic uncertainty for a complex network, inspired by the first and second laws of thermodynamics, but not limited to the equilibrium system. By exploring the relationship between the varentropy of the scale free distribution and the exponent of power laws as well as network size, we get the optimal design of a scale-free network against random failures. The behaviors of varentropy and the Shannon entropy of double Pareto law degree distribution are analyzed to compare their usefulness. Our conclusion is that varentropy is suitable and reliable.

varentropy, optimization, double Pareto law distribution, equilibrium network ensemble

Citation: Jiang J, Wang R, Pezeril M, et al. Application of varentropy as a measure of probabilistic uncertainty for complex networks. Chinese Sci Bull, 2011, 56: 3677–3682, doi: 10.1007/s11434-011-4697-3

Complex networks have recently attracted much interest from the community of physicists thanks to the possible application of various techniques inspired by statistical mechanics. Networks describe different interaction patterns in complex systems. In many cases, the interaction is changeable by time, and the topology of the network changes according to the external conditions. If a network is treated as a thermodynamic system which has an energy function depending on its topology, and entropy is introduced to describe the disorder strength, it could constitute a topological ensemble whose physical quantities are described by methods of statistical mechanics [1–11]. Equilibrium network ensembles are defined as stationary ensembles of graphs generated by restructuring processes, like edges of the graph are removed and/or inserted obeying detailed balance and ergodicity. During such processes, topological phase transitions are expected as the temperature is varied. However, this thermodynamic analogy has not yet been fully explored. In particular, the concept of energy, which is very important in the study of statistical ensemble, is a big challenge. The function form of entropy in complex networks is not given a priori. Many different expressions have been suggested in recent research. They include functions of the node degree, number of links, or some global properties of the network. For instance, Farkas et al. [4] constructed several patterns of the energy E in the network based on vertex degrees or degrees of neighboring vertices, such as $E=-\sum_{i=1}^{N}k_i^2$, $E=\sum_{(i,j)}\left(\frac{\min(k_i,k_j)}{\max(k_i,k_j)}-1\right)$ or $E=-s_{\max}$ ($s_{\max}$ is the size of the largest component in the network). Jeong et al. [12] combined two competing terms with different strengths and assigned the energy function as $E=-\frac{J_1}{M}\sum_{(i,j)}k_ik_j-\frac{J_2}{M}\sum_{(i)}k_i^2$ in which M is the total number of links and J_1, J_2 are the competitive strengths of the two parts in the function. So far,

*Corresponding author (email: jiangj2007010209@gmail.com)

there is little convincing theoretical derivation for the definition of the energy in a complex network. One of our aims in the present paper is to tentatively solve this problem.

If the energy function of a complex network is given, we can obtain an optimized network ensemble which minimizes the energy through zero temperature dynamics. On the other hand, we can get various topologies of the network at finite temperature. It is well known that many real world networks are scale-free networks and they are robust to random failures but vulnerable to targeted attacks. In addition, the character of heterogeneity of a network is directly related to this feature and can be measured by entropy [13,14]. Sole et al. [14] introduced the entropy of the remaining degree and the mutual information to study networks with different heterogeneity and randomness. In the present paper, we investigate the entropic behavior of the power law degree distribution to describe a network's heterogeneity, based on a generalized information measure called varentropy [15]. We find that the optimization of a scale-free network's robustness to random failures is equivalent to maximizing the entropy of the degree distribution. Through optimization, we can get a scenario of optimal design of scale-free networks against random failures.

1 Optimization of a scale-free network against random failures

We give a simple introduction to a universal measure of probabilistic uncertainty proposed in our previous work [15]. First, we suppose that we have a random discrete variable k_i with a probability distribution $p_i = f(k_i)/Z$, where i is the state index and Z is a normalization constant. The average value of k_i is given by $\bar{k} = \sum_i k_i p_i$ and the normalization is $\sum_i p_i = 1$. The uncertainty in a probability distribution of k can be measured by many quantities. The Shannon entropy formula, or some of other known entropy forms, can also be used as a measure of the uncertainty of any p_i. However, no given entropy form, including the Shannon one, can be maximized for any distribution p_i according to MaxEnt rules [16]. Here, a general definition of the uncertainty measure underlying MaxEnt is found, in such a way that each derived entropy could be maximized to give the corresponding distribution. The general measure I is proposed by a variational definition as

$$\mathrm{d}I = \eta\left(\mathrm{d}\bar{k} - \overline{\mathrm{d}k}\right) = \eta\sum_i k_i \mathrm{d}p_i, \qquad (1)$$

where $\overline{\mathrm{d}k} = \sum_i p_i \mathrm{d}k_i$, and η is a characteristic constant.

This definition was inspired by the first and the second laws of thermodynamics in equilibrium thermodynamics. Considering the definition of internal energy, $\bar{E} = \sum_i p_i E_i$ where E_i is the energy of the microstate i with probability p_i, we obtained

$$\delta\bar{E} = \sum_i \delta p_i E_i + \sum_i p_i \delta E_i = \sum_i \delta p_i E_i + \overline{\delta E_i}, \qquad (2)$$

in which

$$\overline{\delta E_i} = \sum_i p_i \delta E_i = \sum_j \left(\sum_i p_i \frac{\partial E_i}{\partial q_j} \right) \delta q_j \qquad (3)$$

is the work done to the system by external forces $F_j = \sum_i p_i \frac{\partial E_i}{\partial q_j}$, where q_j are the extensive variables such as volume, or distance. Based on the first law of thermodynamics, the quantity $\sum_i \delta p_i E_i = \delta\bar{E} - \overline{\delta E_i}$ is the heat change in the system; that is, $\sum_i \delta p_i E_i = \delta Q = T\delta S$ for a reversible process, where S is the thermodynamic entropy and T the absolute temperature. S has the following variational relation:

$$\delta S = \frac{1}{T}\left(\delta\bar{E} - \overline{\delta E}\right). \qquad (4)$$

As is well known, S measures the dynamical disorder of the system or the uncertainty of the probability distribution of energy in the dynamics. In contrast to another measure $\sigma^2 = \overline{E^2} - \bar{E}^2$, S can be maximized in MaxEnt for an equilibrium system to derive the probability distribution. Eq. (1) is just an extension of eq. (4) to arbitrary random variables k and to arbitrary systems (even out of equilibrium), and is a generic definition of maximum uncertainty measure for any random variable.

Regarding a scale-free degree distribution $p_i = k_i^{-\alpha}/Z$ where $Z = \sum_i k_i^{-\alpha}$ and α is an exponent, we derived the entropy formula in previous work [15]:

$$I \propto \frac{\sum_i p_i^{1-1/\alpha}}{1 - 1/\alpha}. \qquad (5)$$

I is known as the varentropy. The variation in varentropy against the exponent is shown in Figure 1. It suggests that varentropy decreases sharply with the increase of the exponent from 1.0 to 3.0. It is worth noticing that when $\alpha \to 1$, the value of varentropy increases without bound, and when $\alpha < 1$, varentropy becomes negative, which is meaningless in a physical context. Form the point of mathematics, the reason for this may be related to the divergence of the normalization constant Z at $\alpha < 1$. Thus, the behavior of varentropy for scale-free degree distribution at $\alpha < 1$ is still obscure.

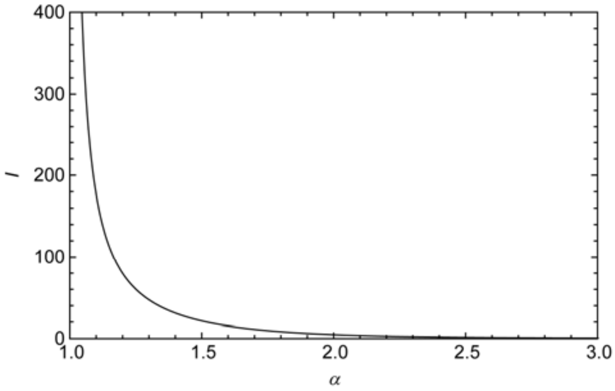

Figure 1 Variation in varentropy I versus exponent α suggests that varentropy decreases sharply with the increasing of exponent from 1.0 to 3.0. When $\alpha\rightarrow 1$, the varentropy increases without bound.

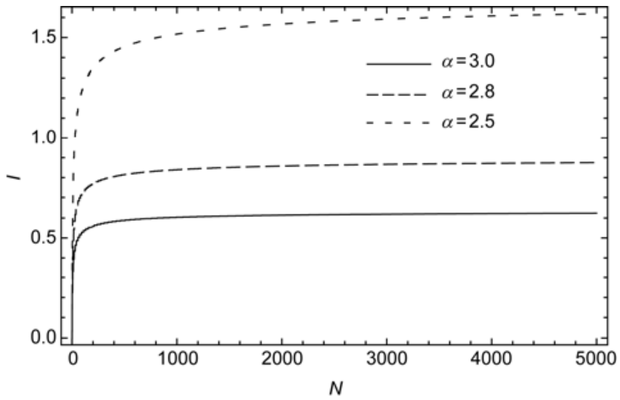

Figure 2 Variation in varentropy I versus network size N with different scaling exponent $\alpha=3.0$ (solid line), $\alpha=2.8$ (dashed line), and $\alpha=2.5$ (dotted line). The three curves are all monotonic increasing. With the increase of network size N, varentropy I increases and tends to a constant above certain N for each case. The larger the exponent, the faster the varentropy value tends to the stable value. In addition, the value of varentropy decreases when exponent α increases, which agrees with Figure 1.

In fact, the majority of scale-free networks in nature have an exponent between 2 and 3 and they are finite networks with finite degree $k=k_{\min},k_{\min}+1,...,k_{\max}$ where $k_{\min}$ is the minimum degree and $k_{\max}$ is the maximum degree cutoff present in the network, which are in the range of valid applications of varentropy.

We also can use varentropy to study the resilience of networks, which is very important to understand how to design networks that are optimally robust against both failures and attacks. For this subject, many different approaches have been proposed, such as the percolation theory [17,18], evolutionary algorithm [13], entropic principle [19], spectral and statistical measurement [20]. In the present paper, we focus on the entropy of the degree distribution to describe the network's heterogeneity, which measures the diversity of the degree distribution [14] within the context of a complex network. For a scale-free network, the effect of diversity (long tail) is an increase in the uncertainty. When the exponent increases or the cutoff degree decreases, the network becomes less heterogeneous and smaller entropy is observed [14]. Here, based on the relationship among $k_{\max}$, N and α in the network, $k_{\max}\approx N^{1/(\alpha-1)}$ ($\alpha>1$, N is the total number of vertices in network) [2]. Eq. (5) could be rewritten as

$$I(N,\alpha)=\frac{\left(\sum_{k=1}^{N^{1/(\alpha-1)}}k^{-\alpha}\right)^{-1/\alpha}}{1-1/\alpha}\left(\frac{\sum_{k=1}^{N^{1/(\alpha-1)}}k^{1-\alpha}}{\left(\sum_{k=1}^{N^{1/(\alpha-1)}}k^{-\alpha}\right)^{1-1/\alpha}}-1\right). \qquad (6)$$

Thus the varentropy I can be expressed as a function of α and N. Figure 2 shows the variation in varentropy I versus N with different exponents $\alpha=3.0$ (solid line), $\alpha=2.8$ (dashed line), and $\alpha=2.5$ (dotted line), and suggests that the varentropy functions are monotonically increasing with the increase of network size N. Moreover, varentropy I tends to a constant when N increases up to certain value for each case. The larger the exponent, the faster the varentropy value tends to the stable value. Varentropy decreases with the increase of exponent α, which is consistent with the behavior of Figure 1. We conclude that the network designed with larger size N is more robust with given exponent α, and the network with small exponent α will be more robust with given N.

2 Application to the double Pareto law distribution

In the real world, there are many truncated degree distribution in complex network such as the exponential truncated and double Pareto laws. Here, we pay more attention to the double Pareto law distribution and investigate its entropic behavior. First, we express the probability distribution as follows:

$$p=\begin{cases}p_1=c_1k^{-\alpha}\,(0<k<k_c),\\ p_2=c_2k^{-\beta}\,(k_c\leqslant k\leqslant k_{\max}),\end{cases} \qquad (7)$$

where k_c is the intersection of two power laws and $\alpha,\beta>0$. With the normalization condition $\sum_i p_i=1$ and $p_1(k_c)=p_2(k_c)$, we have

$$c_1=\frac{k_c^{-\beta+\alpha}}{\sum_{k<k_c}k_c^{-\beta+\alpha}k^{-\alpha}+\sum_{k\geqslant k_c}k^{-\beta}},$$
$$c_2=\frac{1}{\sum_{k<k_c}k_c^{-\beta+\alpha}k^{-\alpha}+\sum_{k\geqslant k_c}k^{-\beta}}. \qquad (8)$$

According to eq. (1), we obtain the expression of the entropy function:

$$I = \frac{c_1^{1/\alpha}}{1-1/\alpha}\sum_i \left(p_1(k_i)^{1-1/\alpha} - p_1(k_c)^{1-1/\alpha}\right) + \frac{c_2^{1/\beta}}{1-1/\beta}\sum_i \left(p_2(k_i)^{1-1/\beta} - p_{2\min}(k_{\max})^{1-1/\beta}\right) + C, \qquad (9)$$

where $p_{2\min} = c_2 k_{\max}^{-\beta}$ and C is the integration constant and is equal to $-\int_0^1 k\mathrm{d}p$ in order to keep $I = 0$ for the non-probabilistic case $p_i = 0,...1,...0$ for all possible states $i = 1, 2,\ldots$. Next, we investigate the variation in the entropy function versus different variable α, β, k_c, $k_{\max}$. In Figure 3(a), the entropy decays with the increase of α at $k_c = 100$, $\beta = 3$, $k_{\max} = 10^3$ and tends to a minimum value approximately 1.0 for larger values of α with $I_{p_1} \approx 0$. In Figure 3(b), the tendency of the entropy function is similar to that in Figure 3(a) at $k_c = 100$, $\alpha = 1.2$, $k_{\max} = 10^3$. It tends to a minimum value 15.8 for larger β with $I_{p_2} \approx 0$. In Figure 3(c), the entropy grows fast until $k_{\max}$ and increases to around 2000 at $k_c = 5$, $\alpha = 1.2$, $\beta = 2.5$, and after that, it reaches the maximum value 5.88. In general, the vertex in the network with bigger size often has larger degree, and thus this curve suggests that the size effect of the network becomes trivial when its size reaches some critical scale. In Figure 3(d), the entropy increases monotonically with increasing of k_c at $k_{\max} = 10^3$, $\alpha = 1.2$, $\beta = 1.5$. When k_c is equal to 1 or 1000, the truncated distribution p evolves into single distribution p_2 or p_1, and then the value of each entropy is equal to $I_{p_2} = 25.5$ or $I_{p_1} = 86.1$, respectively. Hence, from the above discussion, we find that for a network with a double Pareto law distribution, it will become more heterogeneous when the degree distribution has two smaller exponents and a larger cross over point with larger network size.

In addition, we make a comparison with the Shannon entropy formula using the conditions of Figure 3(a). According to the expression of Shannon entropy $S = -\sum_i P_i \ln P_i$. Substituting eq. (7) into it, we get

$$S = -c_1 \ln(c_1)\sum_{k<k_c} k^{-\alpha} - c_1 \sum_{k<k_c}\left(k^{-\alpha}\ln(k^{-\alpha})\right) - c_2 \ln(c_2)\sum_{k\geqslant k_c} k^{-\beta} - c_2 \sum_{k\geqslant k_c}\left(k^{-\beta}\ln(k^{-\beta})\right), \qquad (10)$$

where c_1 and c_2 are as in eq. (8). Here, we investigate the variation in Shannon entropy versus α. In Figure 4, we can find that the tendency of Shannon entropy at $k_c = 100$, $\beta = 3$, $k_{\max} = 10^3$ is similar to that in Figure 3(a). Both are monotonically decreasing functions. However, the difference between the two pictures is that when α increases without bound, the Shannon entropy trends to 0, but the varentropy tends to 1. We believe that the latter result is more reasonable because the value of the entropy of S_{p_2} is not 0, but 1 when $\alpha \to \infty$. Thus the Shannon entropy may be invalid for this kind of double Pareto law distribution in this condition.

3 Equilibrium network ensembles

Energy is often an important concept in optimization problems. It is not possible to derive an energy expression for graphs from first principles. We can find analogies with well-established equilibrium graph ensembles by assigning a statistical weight to each allowed graph. Phenomenological and heuristic arguments could lead to such various energy functions such as those described in the introduction. As is well known, the entropy is the bridge between the microscopic dynamics and macroscopic behaviors. According to the first law in equilibrium thermodynamics,

$$\mathrm{d}\bar{E} = T\mathrm{d}S + \mathrm{d}W, \qquad (11)$$

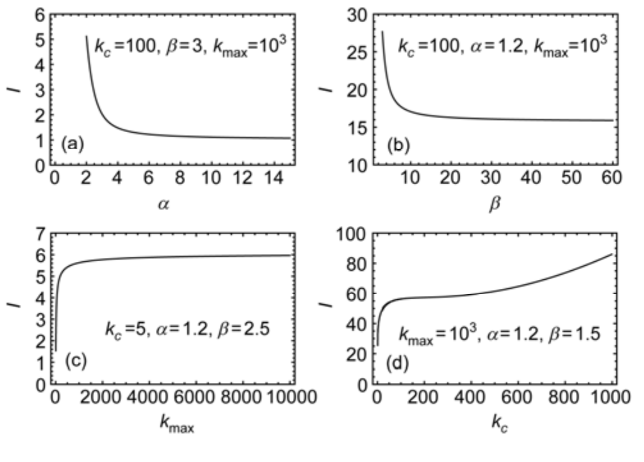

Figure 3 Variation in entropy function versus various parameters of a double Pareto law degree distribution.

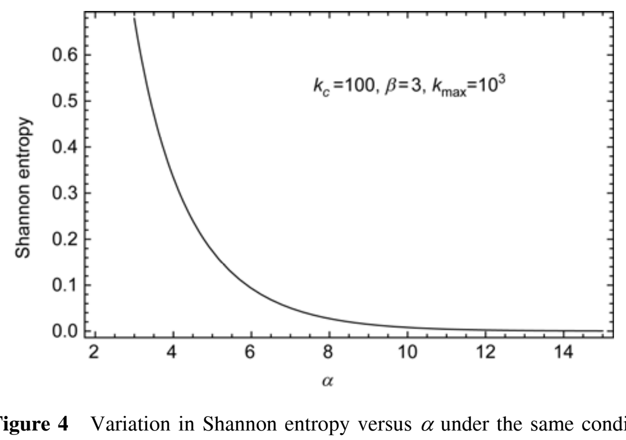

Figure 4 Variation in Shannon entropy versus α under the same conditions as in Figure 3(a) with $k_c = 100$, $\beta = 3$, $k_{\max} = 10^3$. Shannon entropy decays with the increasing of α, which is similar to that of varentropy. However it tends to zero at larger α, which is different from that in Figure 3(a) with limiting value 1.0. In fact, when α increases without bound, S_{p_1} is close to zero, the Shannon entropy of function p should be equal to S_{p_2} which is not zero.

where E is the internal or mean energy of the system, T is the temperature of the system, S is the entropy of the system, and $\mathrm{d}W$ is the work done to the system by external forces. It could be rewritten as $\mathrm{d}S=(\mathrm{d}\bar{E}-\mathrm{d}W)/T$ and in another form of probability $p_i=f(E_i)$, we have

$$\mathrm{d}S=\frac{1}{T}\sum_i E_i \mathrm{d}p_i. \quad (12)$$

If $p_i=\exp(-\beta E_i)/Z$ (Z is the partition function), after integrating of the above equation, we will get the Shannon-Gibbs entropy $S=-\sum_i P_i \ln P_i$. To investigate what will happen if p_i is not an exponential distribution or E_i is not the energy of the system of the state i, we get the extensive expression of the entropy in the formation of eq. (1), substitute the scale-free distribution $p_i=k_i^{-\alpha}/Z$ (k is the degree of the node in the network) into it, and obtain

$$I=\frac{1}{Z^{1/\alpha}}\frac{\sum_i p_i^{1-1/\alpha}-1}{1-1/\alpha}. \quad (13)$$

Comparing with $\mathrm{d}S=(\mathrm{d}\bar{E}-\mathrm{d}W)/T$, the above equation is rewritten as

$$I=\frac{1}{1-1/\alpha}\left(\bar{k}-Z^{-1/\alpha}\right). \quad (14)$$

Then, the thermodynamic equation of the scale-free network could be

$$\vartheta=\bar{k},\quad \varsigma=1-1/\alpha,\quad \zeta=Z^{-1/\alpha}, \quad (15)$$

where parameters ϑ, ς, ζ are analogous to the system energy, temperature, and free energy in thermodynamics respectively. $\bar{k}$ is the average degree of a scale-free network. Thus, based on the foundation of the first and second laws of thermodynamics, thermodynamical quantities of scale-free networks such as "energy", "temperature", and "free energy" are given. Although the assumption of an "equilibrium network ensemble" is used here, we believe that it is still a reasonable approach to apply statistical mechanisms and thermodynamics to complex networks.

4 Discussion and conclusions

In the present paper, we introduced an alternative method for the optimization of scale-free networks in terms of network entropy. This network entropy is not the traditional Shannon entropy. It is another definition of uncertainty measure of the system called varentropy, resulting from the first and second laws of thermodynamics, proposed in our previous work. Our findings indicate that small exponent and large network size can facilitate the optimal design of the scale-free network against random failures.

In addition, we studied the entropic behavior of the double Pareto law distribution with varentropy and Shannon entropy formula. The theoretical expression of each entropy function is obtained and their variance versus different parameters of distribution is analyzed. Through the comparison, we find that Shannon entropy may be less reasonable for calculation with this kind of distribution in some conditions.

Lastly, we have shown that scale-free network ensembles can be described naturally by statistical mechanics and the thermodynamics methods, and provided the theoretical estimation of the quantities like "energy", "temperature" and "free energy" of the network ensemble, which is different from many previous works with various arbitrary definitions without any derivation. The results presented in this paper are only a tiny fraction of what can be done with the thermodynamics equation of scale-free network ensembles. They can facilitate the application of statistical physics and thermodynamics tools in the field of complex networks, and can also help to expand the analysis towards problems, such as optimization, extracting information and topological phase transitions.

This work was supported by the Chinese Ministry of Education (*2008677010*), *the National Natural Science Foundation of China* (*10647125, 10635020, 10975057, 10975062*) *and the Programme of Introducing Talents of Discipline to Universities* (*B08033*). *M.P. thanks the Region des Pays de la Loire of France for their support* (*2009-9397*).

1 Albert R, Barabasi A L. Statistical mechanics of complex networks. Rev Mod Phys, 2002, 74: 47–97
2 Newman M E J. The structure and function of complex networks. SIAM Rev, 2003, 45: 167–256
3 Palla G, Derenyi I, Farkas I, et al. Statistical mechanics of topological phase transitions in networks. Phys Rev E, 2004, 69: 046117
4 Farkas I, Derenyi I, Palla G, et al. Equilibrium statistical mechanics of network structures. Lect Notes Phys, 2004, 650: 163–187
5 Park J, Newman M E J. The statistical mechanics of networks. Phys Rev E, 2004, 70: 066117
6 Zhao J, Yu H, Luo J H, et al. Complex networks theory for analyzing metabolic networks. Chinese Sci Bull, 2006, 51: 1529–1537
7 Wang W, Zeng G S. Bayesian cognitive trust model based self-clustering algorithm for MANETs. Sci China Inf Sci, 2010, 53: 494–505
8 Liang X, Wang J H, He Y. Human connectome: Structural and functional brain networks. Chinese Sci Bull, 2010, 55: 1565–1583
9 Fu L D, Gao L, Ma X K. A centrality measure based on spectral optimization of modularity density. Sci China Inf Sci, 2010, 53: 1727–1737
10 Liu H T, Li W W. Language clusters based on linguistic complex networks. Chinese Sci Bull, 2010, 55: 3458–3465
11 Ma J, Wu Y, Ying H P, et al. Channel noise-induced phase transition of spiral wave in networks of Hodgkin-Huxley neurons. Chinese Sci Bull, 2011, 56: 151–157
12 Jeong D, Choi M Y, Park H. Construction of equilibrium networks with an energy function. arXiv: 07070531
13 Cancho R F I, Sole R V. Optimization in complex networks. Lect Notes Phys, 2003, 625: 114–126
14 Sole R V, Valverde S. Information theory of complex networks: On evolution and architectural constraints. Lect Notes Phys, 2004, 650: 189–207
15 Wang Q A. Probability distribution and entropy as a measure of un-

certainty. J Phys A: Math Theor, 2005, 41: 065004

16 Jaynes E T. The evolution of Carnot's principle. In: The Opening Talk at the EMBO Workshop on Maximum Entropy Methods in X-ray Crystallographic and Biological Macromolecule Structure Determination. Orsay, France, April 24–28, 1984

17 Cohen R, Erez K, Ben-Avraham D, et al. Resilience of the internet to random breakdowns. Phys Rev Lett, 2000, 85: 4626–4628

18 Zhou T, Yan G, Wang B H. Maximal planar networks with large clustering coefficient and power-law degree distribution. Phys Rev E, 2005, 71: 046141

19 Demetrius L, Manke T. Robustness and network evolution. Physica A, 2005, 346: 682–696

20 Yazdani A, Jeffrey P. A note on measurement of network vulnerability under random and intentional attacks. arXiv: 1006.2791