Vol 18 No 5, May 2009
1674-1056/2009/18(05)/1721-04

Chinese Physics B

© 2009 Chin. Phys. Soc.
and IOP Publishing Ltd

The principle of the Internet evolving and a conjecture on the optimal structure of the Internet*

Li Ying(李 昊)[a)b)†], Cao Hong-Duo(曹宏铎)[a)‡], Shan Xiu-Ming(山秀明)[b)], Ren Yong(任 勇)[b)], and Yuan Jian(袁 坚)[b)]

[a)] *School of Business, SUN YAT-SEN University, Guangzhou* 510275, *China*
[b)] *Department of Electronic Engineering, Tsinghua University, Beijing* 100084, *China*

(Received 14 April 2008; revised manuscript received 20 November 2008)

In this paper we will give the statistical characteristics and general principles of an optimal structure of the Internet, which is a scale-free network. Since the purpose of the Internet is to allow fast and easy communication, the average path length is used to measure the performance of the network, and the number of edges of the network is used as a metric of its cost. Based on this, the goal of this Internet optimization problem is to obtain the highest performance with the lowest cost. A multi goal optimization problem is proposed to model this problem. By using two empirical formulas of $\langle k\rangle$ and $\langle l\rangle$, we are able to find the statistical characteristics of the optimal structure. There is a critical power law exponent $\alpha_{\rm c}$ for the Internet with power law degree distribution, at which the Internet can obtain a relatively good performance with a low cost. We find that this $\alpha_{\rm c}$ is approximately 2.1.

Keywords: Internet, the evolving principle, the optimal structure, power law
PACC: 0250, 0520

1. Introduction

Almost all the information we need can be found through the Internet. The Internet has become very important in modern society. Therefore, the structure of the Internet should be well studied. However, although human-built, the Internet is not fully understood by human beings. Among the many reasons for this are that the Internet is large and complex, and was not designed or constructed by a single person or organization. Also, it constantly evolves and develops, and changes have been introduced by different groups of people from different places. As a result, many researchers have been studying the Internet from a macrostructure perspective and using statistical approaches.

It has been discovered that the Internet has three characteristics, namely scale-free, small world and high clustering. A scale-free network is a network of which the degree distribution follows the power law. This means that $P(k) \propto k^{-\alpha}$. Here $P(k)$ denotes the probability distribution function of the degrees of the nodes, with k being the degree of nodes; and α is the power law exponent, usually varying between 2 and 3. The small world characteristic means that, despite the large size of the network, there is a relatively short path between every two nodes. The path length l is defined as the number of edges along the shortest path connecting the two nodes. The average path length $\langle l\rangle$ is the mean of all l. 'Small world' is defined as $\langle l\rangle \propto \ln(N)$, the average path length $\langle l\rangle$ increases in proportion to the logarithm of the size of the network N. A high clustering network tends to form charmed circles. This feature can be quantified by the clustering coefficient C_i. All these characteristics contribute to the properties of the Internet.[1−3]

Over time, the properties of the Internet have changed. For instance, the value in 2004 is different from that in 1999.[4−6] With the development of the Internet, the size of the network continues to rise, but the characteristics of the power law degree distribution are kept constant although the exponent will change, which is shown in Table 1. Here, the routers and domains are considered to be the same. Because we studied the Internet from a physical perspective, the 'Internet' used in this study is an abstract network, and the different nodes on the Internet are considered to be the same, no matter whether they are large or

*Project supported by the National Natural Science Foundation of China (Grant Nos 70801066, 60674048, 60772053 and 60672142); the National Basic Research Program of China (Grant Nos 2007CB307100 and 2007CB307105).
†Corresponding author. E-mail: mnsliy@mail.sysu.edu.cn
‡E-mail: caohd@mail.sysu.edu.cn

small.

We have been studying the trend of the change in the Internet, and the principle of its evolution. In this paper, we show an optimal structure of the Internet, which we believe characterizes the future Internet. We found that an Internet with this structure can achieve not bad performance with a relatively small cost.

2. The principle of Internet evolution

In this section, we present the relationship between the performance and the topology of the Internet, and explain the general principles behind the evolution of the Internet.

The objective of the Internet is to allow fast and easy communication. So one of the general principles behind its design is that any one node of the Internet can reach another one as fast as possible. Formally, the average path length $\langle l\rangle$ needs to be as small as possible. However, we also notice that a small $\langle l\rangle$ always means a large number of edges. The average path length of a completely connected network is 1, which means any node can reach any other node with only one hop. However, the cost of such a network is very high because every two nodes must be connected by an edge, and having too many edges in a network is expensive. It is always desirable to obtain the best performance while the cost is kept low. This conflict between performance and cost remains to be settled. Let $\langle k\rangle$ represent the cost, because $\langle k\rangle$ shows the average number of edges of a node. Similarly, since a short average path length is the goal of the design of the Internet, we use $\langle l\rangle$ to represent the performance. As $\langle k\rangle$ increases, the total number of edges will grow accordingly. In addition, with more edges, fewer hops will be needed for any two nodes to communicate, and therefore $\langle l\rangle$ will decrease. To achieve both the economy and performance goals of the Internet, there is a balance point, at which both $\langle k\rangle$ and $\langle l\rangle$ have satisfactory values. The search for this balance point can be modelled as a multi goal optimization as follows:

$$\begin{aligned}\min(\langle k\rangle) &= \min\left(\sum_{k=1}^{k_{\max}} P(k)\cdot k\right),\\ \min(\langle l\rangle) &= \min\left(\sum_{l=1}^{l_{\max}} P(l)\cdot l\right).\end{aligned}\tag{1}$$

Here $\langle k\rangle$ is the average degree, $k_{\max}$ is the maximum degree (which is less than the size of network N), and $P(k)$ is the probability distribution function of the degrees of nodes; $\langle l\rangle$ is the average path length, $l_{\max}$ is the maximum path length (which is also less than N), and $P(l)$ is the probability distribution function of the path length. Because $k_{\max}$ and $l_{\max}$ are both less than N, the above problem can be rewritten approximately as

$$\begin{aligned}\min(\langle k\rangle) &= \min\left(\sum_{k=1}^{N} P(k)\cdot k\right),\\ \min(\langle l\rangle) &= \min\left(\sum_{l=1}^{N} P(l)\cdot l\right).\end{aligned}\tag{2}$$

The Internet is a scale-free network, that is, the probability distribution function of the degrees satisfies $P(k)\propto k^{-\alpha}$, and the exponent α is usually in the range of 2∼3. Here $P(k)$ is a monotonically decreasing function, which means that a small k has a greater probability than a large k does. This property implies that the total number of edges can converge to a finite value when the size of the network increases to a large number. We give the mean degree empirical formula $\langle k\rangle = \sum_{k=1}^{N} P(k)\cdot k = 0.5\cdot\sum_{k=1}^{N} k^{-\alpha+1} + \frac{\alpha}{2}$, which is effective for the Internet, as shown in Table 1. This formula is a monotonically decreasing function of α, as shown in Fig.1. To minimize $\langle k\rangle$, the exponent α must be maximized.

For minimizing the average path length $\langle l\rangle$, more factors should be considered. For instance, the size of the network, i.e. the number of edges in the network, and the linking methods between all the nodes all contribute to the average path length. Among these, the linking method can have a great impact. For example, the ring network and the star network with the same numbers of nodes and edges have different average path lengths. The probability distribution function of the node degrees $P(k)$ can take account of the linking property of the network. In addition, because the node degrees follow the power law, that is $P(k)\propto k^{-\alpha}$, the exponent α can also represent the linking property of the network. Considering that the network size N is also an important factor in determining the average path length, we give the average path length an empirical formula as follows:[7–9]

$$\langle l\rangle = \left(1-\frac{1}{\alpha}\right)^{1.7}\cdot\sum_{l=1}^{N} l^{-\frac{0.6\cdot\ln(5.55-\alpha)}{\ln\alpha}}.\tag{3}$$

In Refs.[7–9], we analysed in detail the average path length of the Internet whose degree distribution

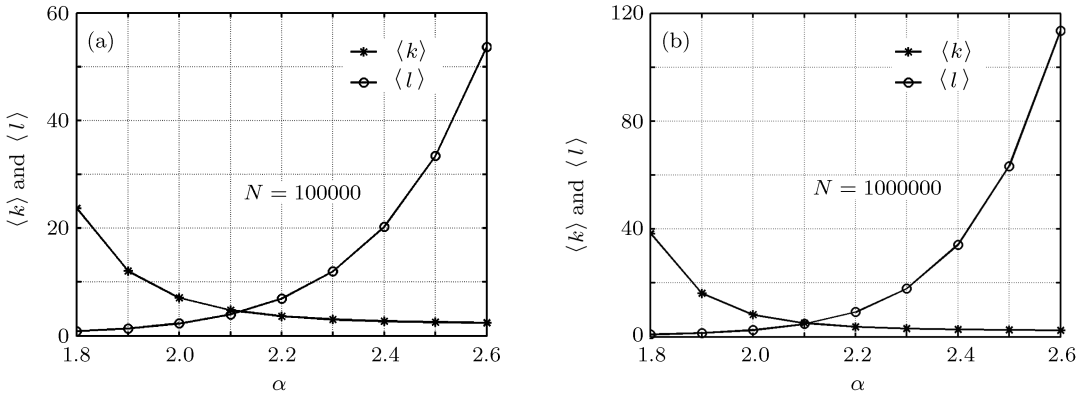


Fig.1. The variation of $\langle k\rangle$ & $\langle l\rangle$ with α. (a) The size of the Internet N is 10^4. (b) The size of the Internet N is 10^5. As α increases, $\langle k\rangle$ decreases but $\langle l\rangle$ increasess, i.e. $\langle k\rangle$ is a decreasing function of α, and $\langle l\rangle$ is an increasing function of α. The $\langle k\rangle$ curve intersects the $\langle l\rangle$ curve at one point. But at different sizes of N, the point of intersection of $\langle k\rangle$ and $\langle l\rangle$ is different.

follows the power law. One conclusion was that the path length distribution function $P(l)$ must be a decreasing function of l, and also a function of α. In Refs.[7–9], we also gave the formula as above, which is applicable to the Internet. The formula indicates the 'small world' characteristic, which says that the proportion of short paths is large and that of long paths is small. It has been tested to be valid for the Internet, as shown in Table 1. Figure 1 shows the relation between α and $\langle l\rangle$. From Fig.1 we can see that α increases as $\langle l\rangle$ increases. Therefore, to minimize $\langle l\rangle$, the exponent α must also be minimized.

Table 1. Statistics for a number of published networks. For each network we have indicated the number of nodes N, degree exponent α, the average degree $\langle k\rangle$, and the average path length $\langle l\rangle$. For comparison we have also included predictions of the average degree $k_{\rm power}$ and average path length $l_{\rm power}$, $k_{\rm power} = 0.5\cdot\sum_{k=1}^{N} k^{-\alpha+1}+\frac{\alpha}{2}, l_{\rm power} = \left(1-\frac{1}{\alpha}\right)^{1.7}\cdot\sum_{k=1}^{N} l^{-\frac{0.6\cdot\ln(5.55-\alpha)}{\ln\alpha}}$. The two predictions are valid for the Internet.

network	size N	α	$\langle k\rangle$	$\langle l\rangle$	$k_{\rm power}$	$l_{\rm power}$	Ref.	Nr.
Internet, domain	3,015-4,389	2.1-2.2	3.42-3.76	4	3.7550	3.6496	[4]	1
Internet, router	3888	2.48	2.57	12.15	2.5675	12.2838	[1]	2
Internet, domain	3530	2.16	3.65	3.77	3.6537	3.6883	[4]	3
Internet, domain	11174	2.1	4.18	3.62	4.3737	3.2763	[5]	4
Internet, domain	11461	2.1	5.71	3.56	4.3787	3.2846	[5]	5

3. The optimal structure of the Internet

From Fig.1 we can see that with increasing α, $\langle k\rangle$ decreases and $\langle l\rangle$ increases. Therefore, to make both goals optimized, a tradeoff between $\langle k\rangle$ and $\langle l\rangle$ must be made.

A balance point can be found in Fig.1. From this figure, both $\langle k\rangle$ and $\langle l\rangle$ are monotonic functions, and we choose the single intersection point to be the balance point. Away from this point, when α is greater than this point, $\langle l\rangle$ will increase, which means that the performance of the Internet will decrease; on the other hand, when α is less than this point, $\langle k\rangle$ will increase, which means that the cost of the Internet will increase. Only at this intersection point is a tradeoff made and a satisfactory value for both $\langle k\rangle$ and $\langle l\rangle$ found. We call this point the critical α, denoted by $\alpha_{\rm c}$.

Figure 1 also shows that different N's have different $\alpha_{\rm c}$. This implies that optimal structures for networks of different sizes are different. To calculate the optimal α_c for networks of different sizes, the optimization problem is rewritten as

$$\min(\langle k\rangle) = \min\left(0.5\cdot\sum_{k=1}^{N} k^{-\alpha+1}+\frac{\alpha}{2}\right), \qquad (4)$$

$$\min(\langle l\rangle) = \min\left(\left(1-\frac{1}{\alpha}\right)^{1.7}\cdot\sum_{k=1}^{N} l^{-\frac{0.6\cdot\ln(5.55-\alpha)}{\ln\alpha}}\right). \quad (5)$$

Figure 2 illustrates the relationship between α_c and N. From this figure, we see that α_c decreases as the size of the network N increases, and the range of its values is between 2.1 and 2.3. With N increasing to a large number, for example approximately 10^{10}, α_c approaches 2.1.

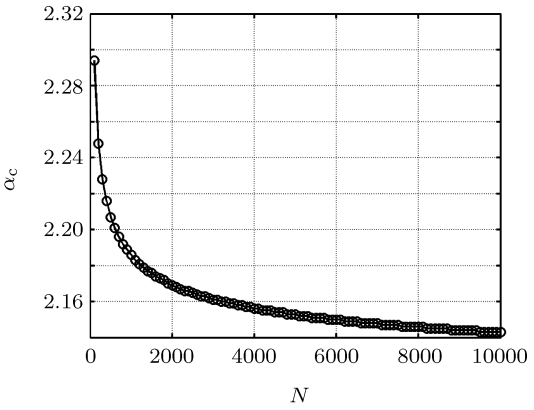


Fig.2. Variation of α_c with N; α_c is a monotonically decreasing function of N. This means that a large network will have a smaller α_c. When the size of the Internet N reaches 10^4, α_c approximately equals 2.14.

The structure of the Internet is optimal when the α value of the network is α_c. With N increasing to 10^{10}, α_c approaches 2.1. That is to say, when the size of the Internet increases to approximately 10^{10}, if the probability distribution function of the node degrees of the Internet satisfies $P(k) \propto k^{-2.1}$, the Internet will show good performance with a low cost. If the power law exponent α was approximately 2.1, the whole Internet would reach an optimal state, and the Internet could achieve good performance with a reasonable cost. This is the desirable result of self-organization optimization. The physical implication of the equality relation of $\langle l\rangle$ and $\langle k\rangle$ is an equilibrium state in which the performance achievable by the network is proportional to the cost of the network. That is to say, the average degree of each node equals the average path length between all pairs of nodes, and thus the network reaches a balanced state.

4. Conclusion

The Internet has been constantly improved. Although it follows a power law degree distribution, the power law exponent keeps changing. Some researchers believe that the Internet has a self-optimizing property on a microcosmic level. However, we wonder if local optimization can benefit global performance. This is what we have studied in this paper. If the number of edges in the network is used as the metric of the cost of the network, and the average path length is used as a measure of the performance of networks, then the statistical properties of the Internet may show a tradeoff between the performance and the cost. With increasing size of the Internet, the critical power law exponent α_c decreases towards 2.1. With this decreased α, the average node degree $\langle k\rangle$ will increase, and the average path length $\langle l\rangle$ will decrease. This shows that as the size of the network increases, more edges will be added to adapt to this enlarged network. But the added edges do not improve the connectivity of the network. As the size of the Internet increases, the impact of the local changes on the global network will be reduced; therefore the change of α_c becomes slower. If the number of nodes of the network is 10^{10}, α_c is approximately 2.1. So we believe that the future Internet will still have a power law degree distribution and the power law exponent is about 2.1. With such an exponent, the Internet has an optimal structure which can obtain the best performance at the lowest cost.

References

[1] Albert R and Barabási A L 2002 *Rev. Mod. Phys.* **74** 47
[2] Albert R, Jeong H and Barabási A L 1999 *Nature* **401** 130
[3] Barabási A L and Albert R 1999 *Science* **286** 509
[4] Faloutsos M, Faloutsos P and Faloutsos C 1999 *Comput. Commun. Rev.* **29** 251
[5] Vázquez A, Pastor-Satorras R and Vespignani A 2002 arXiv:cond-mat/0206084
[6] Zhou S and Mondragón R J 2004 *IEEE Comm. Lett.* **8** 180
[7] Li Y, Shan X M and Ren Y 2004 *Acta Phys. Sin.* **53** 69 (in Chinese)
[8] Li Y, Cao H D, Shan X M and RenY 2008 *Chin. Phys.* B **17** 2327
[9] Li Y, Liu Y, Shan X M, RenY, Jiao J and Qiu B 2005 *Chin. Phys.* **14** 2153