Research of Modeling Complex Topology on Internet
MIFUBA BABA SERGE and Ye XU

1College of Information Science and Engineering, Shenyang Ligong University, Shenyang, China
2xuy.mail@gmail.com
*Corresponding author

Keywords: BA model, Degree-rank power-law, Frequency-degree power-law, Genetic Algorithm, Internet topology modeling

Abstract. Method of modeling complex networks with large scale topology is studied and an example of modeling Internet topology is followed. First, mathematical tools like frequency-degree power-law, degree-rank power-law and CCDF (d)-degree power-law is studied to outline the network power-law properties. Then, models like BA model are improved to set up an model for Internet topology by means of Genetic Algorithm. Algorithm was given at last.

Introduction

The node degree distribution in a network topology follows the power-law distribution if the network has uneven topology structure where most of nodes have a small degree whereas a few nodes have very large degrees. In such a network, the Max degree, Min degree or Average degree, however, can not properly capture topology properties. Recently, the power-law distribution has been introduced into this field [1, 2].

The Internet is an example of such network and power-law tools are used for studying Internet topology [1, 2, 4]. In 1999, for the first time, Faloutsos et al. [1] consider the notion of frequency-degree power-law to character the topology at the AS-level and router-level Internet. Thereafter, similar approaches such as degree-rank power-law, eigenvalue-rank power-law and others are discussed to evaluate the network topology [1, 2]. In 2003, Siganos et al. [3] found that the frequency-degree power-law distribution is quite similar to but better than the probability density function (PDF) where degree (d) is independent variable and frequency (f) is dependent variable. Then, the complementary cumulative distribution function (CCDF)-degree power-law distribution was proposed [3]. Therefore, we develop a power-law based approach to study the Internet topology modeling, based on a giant set of measured samples of router-level Internet topology.

The Measured Samples of the Router-level Internet

In this section, we discuss the measuring approaches and select the appropriate approaches for us to measure the collected data.

1) Measuring approaches

Static methods based on the BGP route table and the dynamic methods based on the active probing are two ways to measure the router-level Internet topology [16]. The static methods, however, are incapable of effectively measuring of the redundant routers.

The dynamic methods can be divided into three categories [19]: (1) Single-monitor-measurement is able to record the source routers in the route path. The Internet Mapping Project (IMP) in Bell Lab is an example of this approach to map and visualize the Internet [20]. The Mercator[21] project is another approach to discover the Internet map; (2) Active measuring based on the Public Traceroute Server (PTrS) is the second type of the dynamic approach. ISP topology measurement project by Boston University is able to measure the network topology with rocket fuel [22] to achieve dynamic measurement; (3) The third type is called as multi-monitor- measurement or
measurement-from-multiple-vantage-points. These approaches include the CAIDA projects [17, 18] and the Active Measuring Project by Harbin Institute of Technology [19].

**Power-law Analysis**

**Frequency-degree power-law**

We calculate the frequency and degree from one-monitor sample, two-monitor sample, five-monitor sample and twenty-one-monitor sample and plot the result in Fig.3. The power-law curve fitting results were also illustrated in Fig.1.

![Figure 1. The frequency-degree power-law relationship on the router-level Internet topology. Both axis x and axis y are in logarithm. Monitors of the four sub-graphs are separately one monitor (arin), two monitors (arin, b-root), five monitors (arin, b-root, cam, cdg-rssac and champagne) and twenty-one monitors (including all monitors).](image)

We can clearly see the power-law relationship between the variable of frequency and degree since the straight line shown in the Figure. Moreover, we can see that the curve fitting results (the straight line) are close to the sample, and all four ACCs (Absolute value of the correlation coefficient) are greater than 0.95, meaning that the fitting results are acceptable.

**Internet Topology Modeling**

**BA Model.** We began to construct an Internet topology model according to the power-law analyses results.

The power exponent of frequency-degree power-law is $|R|=2.1406$. We will try to construct a model that could generate a network whose frequency-degree power exponent is close to this value.

Some researches[4][14] indicate that, the network having frequency-degree power-law properties is a kind of scale-free network, and the traditional model - BA model[29] can be used to generate such scale-free networks. Thus, we use BA model as a base to form the Internet topology model.

The basic of the algorithm of BA model can be described as below. It generates $m_n(n_0 > 1)$ nodes, and links them randomly. The algorithm repeats the following step: for network $G(t-1)$ at time $t-1$, it add one new node with $n$ links to $G(t-1)$ and form a new network $G(t)$ for time $t$. The $n$ links could be connected between the new added node and any selected current node in the network if the selected node $i$’s $\sum_j k_j/k_i$ is greater than a given threshold, where $i, j$ are nodes existed in $G(t-1)$ and $k_i, k_j$ are degree value of corresponding nodes.

The network generated by the this algorithm conforms to a frequency-degree power-law distribution $p(k) \sim k^{-\alpha}$, where the power exponent $\alpha$ is irrelevant to $m_0$ and $n$. 

222
Researches in [4] [14] show that the power exponent of the network generated by BA model is usually 3, which is different from 2.1406 in this paper. Therefore, we need to improve the algorithm as illustrated in the next subsection.

**Improvement of BA Model**

1) **Improvement approaches**

How to improve the power exponent of BA model is still a open issue. Reference [15] gives an algorithm using limit calculation and is too complicated to fit for the improvement requirement in this paper. Reference [7] gave another way of improvement for AS level Internet topology. In this approach, the probability of linking nodes (as mentioned in the upper BA algorithm description) is:

\[ \Pi_i = k_i / \sum_j k_j \]  

where \( k_i, k_j \) are degree value of node \( i \) and \( j \). It can be rewritten as:

\[ \Pi_i = k_i^{1+\varepsilon} / \sum_j k_j^{1+\varepsilon} \]  

Then the power exponent of BA model would be modulated to be around 2.2 when parameter \( \varepsilon \) is set in an interval \([0.1, 0.3]\) [7]. Since value 2.2 is close to value 2.1406 as we calculated, this method should be effective for our requirement. In the next step, we determinate the parameter \( \varepsilon \).

2) **Optimize parameter \( \varepsilon \) by Genetic Algorithm**

One option to find an appropriate parameter \( \varepsilon \) is to search it in the interval of \([0.1, 0.3]\), or \((0, 0.6]\), enlarged the domain to ensure we can find the optimal parameter \( \varepsilon \) by steps. This method is called step-based searching. If the network generated by the improved BA model with parameter \( \varepsilon \) could produce a power exponent close to 2.1406, i.e., the difference between the two values is less than a predefined threshold, we say the parameter \( \varepsilon \) is found. Otherwise, we continue the algorithm by adding a step for parameter \( \varepsilon \).

The step-based searching is straightforward and lacks efficiency. Genetic Algorithm (GA)[30][31] is a searching approach directed by past searching experience. Instead of searching in a straightforward way step by step, GA generates a random \( \varepsilon \) value in the interval and evaluates this value through a GA’s evaluation function. GA algorithm repeats these operations till the generated network by the selected \( \varepsilon \) could produce power exponent close to 2.1406. Experiments showed that GA could converge in comparatively shorter time than the straightforward searching method. The steps of GA for optimization of parameter \( \varepsilon \) are:

i) **Gene code**: We define a gene code \( x \) as a vector comprising primary parameters to be optimized. Only one parameter \( \varepsilon \) is to be optimized:

\[ x = (\varepsilon) \]  

ii) **Random initialization of gene group**: Assuming the size of the gene group is \( N \) (e.g., \( N =100 \)), we randomly initialize a gene group having \( N \) genes, i.e., 100 copies of randomly selected parameter \( \varepsilon \).

iii) **Evaluation function**: The choice of \( \varepsilon \) should minimize the difference of the power exponents of generated network and 2.1406. The evaluation function should be:

\[ f(x) = |P_x(n) - 2.1406| \]  

where \( P_x(n) \) is the power exponent of the generated network with parameter \( \varepsilon \), and \( n \) is the number of the nodes in the network. The evaluation function is expected to score the parameter \( \varepsilon \).
\( n \) is an important parameter because it’s closely related to the calculation efficiency of the target network’s power exponent. It’s easy to know that the bigger \( n \) is, the longer time is needed to calculate the power exponent. A good choice of \( n \) would quickly produce a better outcome.

Two scale-free networks with 100 and 500 nodes respectively are illustrated in Fig.2.

From the Fig.2, there is already a sign of scale-free property with 100 nodes, and a perfect scale-free property in Fig.(b) with 500 nodes. The average, e.g., 300, is taken in this paper, to ensure that the 300-node network generated by improved BA model could show both clear scale-free property and its simplicity in calculating its power exponent.

iv) Selection: Genes are sorted in descending order by scores in the gene group, and the first \( m*N \) genes, \( m \) is a random number \((0<m<1)\), are selected for the next round of calculation by GA. Then, we duplicate the best \( m*N \) genes, and together with the genes that were not selected, i.e., \( N(1-m) \) genes. We get the gene group with size of \( 2*m*N + N(1-m) = N+mN \).

In order to keep the size of group remaining to be same, i.e., \( N \), we remove the last (i.e., worst) \( m*N \) genes in the sorted group, and then the size of group is \( N \). This group is applied to GA for the next round of calculation.

v) Crossover: Crossover operation is:

\[
\begin{align*}
\varepsilon'_i &= \varepsilon_i (1 - \alpha) + \beta \varepsilon_j, \\
\varepsilon'_j &= \varepsilon_j (1 - \alpha) + \beta \varepsilon_i,
\end{align*}
\]

where \( \alpha, \beta \) are random numbers, and \( 0 < \alpha < 1, 0 < \beta < 1 \).

vi) Mutation: Mutation operation is:

\[
\varepsilon'_i = \varepsilon_i (1 + \alpha) \quad \text{if} \quad \gamma \geq 0.5
\]

\[
\varepsilon'_i = \varepsilon_i (1 - \alpha) \quad \text{if} \quad \gamma < 0.5
\]

where \( \alpha, \gamma \) are random numbers, and \( 0 < \alpha < 1, 0 < \gamma < 1 \).

Unlike crossover operations, not all genes are selected to perform mutation. We set up a threshold of 0.3 in the algorithm, which means only 30\% genes would be performed by mutation.

vii) Termination conditions: There are two termination conditions in GA. The first condition is when evaluation function outcome of the best gene in the group is less than a threshold \( s \), \( s \) is set to be 0.01 in the algorithm. The other condition is when GA have repeated for more than 1000 times before finding the best gene (parameter \( \varepsilon \)). If so, we terminate the algorithm.

Construct Internet Topology Model Based on the Improved BA Model

With the improved BA model for frequency-degree power exponent, we then construct an Internet topology model. Before this, the model, however, would have to execute another process as below in the two phases of degree-rank power-law analysis outcomes.
Studies on AS-level Internet topology in [32] indicated that nodes in a network would not definitely conform to a power-law distribution with only one power exponent, especially the CCDF(d)-degree power-law and degree-rank power-law distribution.

**Test and Evaluation of IBA Model**

1) **By power-law analysis**

Frequency-degree power-law analyses for two networks generated by IBA model are illustrated in Fig. 3.

First, obvious power-law features are found in both groups. Then SSSE of fitting the power-law exponents of two groups are 388.4035 and 273.9731, respectively, indicating that the fitting results is acceptable.

Finally, the power-law exponents from the figure are 1.5518 and 1.8191, different from the value 2.1406 in this paper. The difference occurs here might originate from the procedure of optimizing parameter $\varepsilon$ in GA. And finding a better way to do this would be our future job.

However, the power-law exponents of the two tests are gained from two IBA network with rather small size. According to the principle of power-law, properties of it would be getting more obvious with increasing size. Besides, the power-exponents are not far different from 2.1406, so, the tests results could be used to prove that IBA model is acceptable.

![Fig. 3](image1.png)

**Fig. 3** Results of frequency-degree power law distributions and their fitting curves.

2) **By Normalized Laplacian spectra (NLS)**

NLS analyses for two IBA networks are illustrated in Fig. 4.

From the figure, the NLS distributions of three topologies are consistent by showing step styles. But there is only one step in the figure, different from three steps gained by a NLS research on the China side Internet (CERNet) in [2]. And this might due to the difference of the size of the samples.

The increasing parts before and after the steps are of highly similarities among three groups. Especially the group (2), indicating that the group (2) is better than group (1), closer to real Internet. And this is quite the same as what was found in power-law tests, power-exponent of group (2) is closer to real Internet, i.e., better than group (1).

With these, IBA model is regarded to be accepted.

**Conclusions**

In this paper, we study the Frequency-degree power-law, degree-rank power-law and CCDF(d)-degree power-law distributions on the router-level Internet topology. We look at the frequency-degree power-law relationship and find that the power-exponent is 2.1406 from the experiments. For the degree-rank power-law, two phases of power-law relationships are identified.
with power-exponents of 0.29981 and 0.84639 respectively. However, the CCDF (d)-degree power-law relationships are not quantified in the research.

With the power-law relationships and power-exponents found in the experiments, we construct a mathematical model—IBA model to improve the traditional BA model. In the IBA model, we use the GA algorithm to search the optimal parameter $\epsilon$. Thereafter, modulation by two-phase degree-rank power-law distributions is designed to enhance the accuracy of IBA model.

Acknowledgements.

This work is financially supported by the National Natural Science Foundation of China (No. 61373159), the program for Liaoning Excellent Talents in University (LJQ2015095), the Shenyang Key Laboratory Project (No. 170434(17-118-1-00)) and the Open foundation of Key lab of Information Networking and Confrontation of Shenyang Ligong University (No. 4771004kfs18).

References

[1] Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the Internet topology [J]. ACM SIGCOMM Computer Communication Review, 1999, 29 (4):251-262.

[2] Jiang Y, Fang B.X., Hu M.Z. An Example of Analyzing the Characteristics of a Large Scale ISP Topology Measured from Multiple Vantage Points [J]. Journal of Software, 2005, 16 (5):846-856.

[3] Siganos G, Faloutsos M, Faloutsos P, Faloutsos C. Power laws and the AS-level Internet topology[J]. IEEE/ACM Trans. on Networking, 2003, 11(4):514-524.

[4] Wang X.F., Li X., Chen G.R., Complex networks theory and its application [M]. Beijing: QingsHua Press, 2006, 49-70.

[5] Dam E, Haemers WH. Which graphs are determined by their spectrum? [J]. Linear Algebra and its Applications, 2003, 373:241-272.

[6] Lakhina A, Byers JW, Crovella M, Xie P. Sampling biases in IP topology measurements[C]. In: Proc. of the IEEE INFOCOM 2003, Vol 1. San Francisco: IEEE, 2003. 332-341.

[7] Sagy B, Mira G, Avishai W. An incremental super-linear preferential Internet topology model[C]. Proc. 5th Annual Passive and Active Measurement Workshop, LNCS 3015, 2004, 53-62.

[8] Sagy B, Mira G. Avishai W. A geographic directed preferential Internet topology model [C]. Arxiv: CS, 2005, NI/0502061.

[9] Cao L.B., Dai R.W., The intelligent Information System—Internet[M]. Beijing: Science Press, 2001, 121-130.

[10] Broido A, Claffy KC. Internet topology: Connectivity of IP graphs[C]. In: Fahmy S, Park K, eds. Scalability and Traffic Control in IP Networks (Proc. of the SPIE ITCom Vol. #4526). Washington: SPIE Press, 2001. 172-187.

[11] Spring N, Mahajan R, Wetherall D. Measuring ISP topologies with rocket fuel[J]. ACM SIGCOMM Computer Communication Review, 2002, 32(4):133-145.

[12] Waxman BM. Routing of multipoint connections [J]. IEEE Journal on Selected Areas in Communications, 1988, 6 (9):1617~1622.

[13] Zhang W.B. Research on the Life Characteristic and Evolution of Internet macroscopic Topology [D]. Shenyang: Northeastern University, 2005, 6-23, 49-67.