ISSN 2303-0135

The Internet as Complex Networks: Understanding Its Structure and Its Dynamical Properties

Ferry Astika Saputra[1]

[1]*Electronic Engineering*
Polytechnic Institute of Surabaya

Abstract— Structures and properties of networks have been attracting research interests in various research fields. The Internet has been investigated as complex networks. The structure of the Internet has been reported to be scale-free. Namely, the degree distribution at a router level obeys a power law. The packet flow in the Internet also has been reported to obey a power law. In this paper, models of networks are investigated theoretically and with simulations. The dynamical properties of those networks are also discussed.

Keywords— Complex networks, structure, dynamical, properties, the Internet.

1 INTRODUCTION

1.1 Brief history of the Internet

The history of the Internet began at the end of the year 1960s. The first concept about “internetworking” was introduced in 1962 by J.C.R. Licklider of MIT[1]. He called his concept “Galactic Network” . This network was envisioned as a set of computers interconnected globally. Everyone could quickly access data and programs from any site through the network. In the same period, July 1961, Leonard Kleinrock of MIT published the first paper about the packet-switching theory. According to his theory, communication between computers could be achieved by packets which are routed between nodes over data links shared with other traffic. In each network node, packets are queued or buffered, resulting in variable delay. Data traffic will be short in duration. Namely a circuit switching approach will require longer overhead for establishing connection than the duration for data transmission. This contrasts with the circuit-switching approach, used in telephone and telegraph network, a pair of users occupy a dedicated connection with a fixed amount of bandwidth between them during their communication.

This approach is efficient for telephone connections. They require a constant transmission rate. The period of their connection is in general longer than the time for establishing the connection. In the late 1965, the first network experiment by using packet-switching technology was directed by Larry Roberts at MIT Lincoln laboratory. He succeeded sending messages between two computers. It confirmed that time-sharing computers could work well together, running programs and retrieving data, but that the circuit-switched telephone-like system was totally inadequate for this task. He concluded that the packet switching approach is an appropriate communications technology for networked computers.

In the end of 1966, the first concept of the first computer network was developed [2]. Researchers worked on the optimization of the network topology. They developed the first host-to-host protocol, called the Network Control Protocol (NCP). The first computer communications network, called the Advanced Research Projects Agency Network (ARPANET), started in 1972. Its first application was e-mail, also developed in 1972.

1.2 Outline of the Internet mechanism

The Internet uses the Transmission Control Protocol/Internet Protocol (TCP/IP) suite as the key protocol for communicating between two hosts. TCP/IP protocol suite consists of four layers: application, transport, network and link layers [3]. As seen in figure 1, each layer defines a function. It can contain multiple protocols. Each provides a service suitable to the function of that layer.

Every protocol communicates with its peer. The peer is an implementation of the same protocol in the corresponding layer on the remote computer. Every layer is responsible for processing the data from its upper layer and its lower layer.

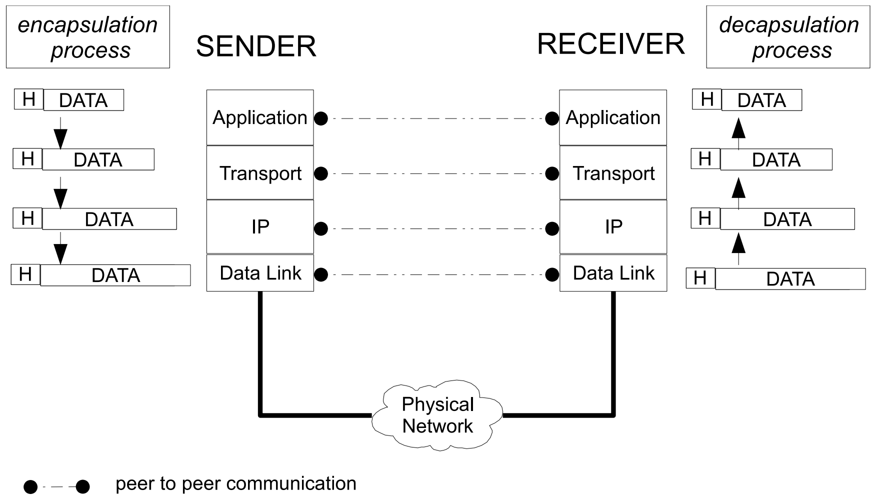

Figure 1. The basic mechanism of sending and receiving data on the Internet

Now, let us focus on the 4th layer (transport layer). Many applications which need reliability for delivering of a stream of bytes such as browser (World Wide Web application), FTP (File Transfer Protocol) and email use TCP. TCP uses mechanisms to achieve high performance and avoid the network congestion. These mechanisms keep the data flow below the capacity of the links. The data transfer of TCP starts from a stage, called slow-start, in which TCP tries to increase its sending rate exponentially, until it encounters the packet loss. At this point, TCP interprets packet loss as an indication of reaching the upper limit of the available bandwidth of the bottleneck link. Thus, from this point (or following another slow-start period, depending on the implementation of TCP), it switches to another stage, called congestion-avoidance, in which TCP employs the Additive Increase, Multiplicative Decrease (AIMD) mechanism to adapt slowly to the available bandwidth.

Another congestion control happens when the network is heavily loaded, during which some of the TCP connections should be closed in order to clear congestion. The algorithm used in this stage is called exponential-backoff [4]. When TCP does not receive any acknowledgments for some timeout period, it assumes that packets are lost. It transmits packets again and doubles its transmission timeout. This process continues until the packets are successfully transmitted and acknowledged, up to some upper limit (usually 64 times the smallest timeout). Essentially, TCP tries to clear congestion by reducing its sending rate in half (or exponentially decreasing its rate). Theoretically, TCP would operate in the congestion-avoidance stage to eciently utilize the network resources.

1.3 Observational fact of the Internet

The Internet is one of the most important communication media in modern societies. Despite to its importance, there is no center for controlling its flow and structure. It evolves without any central controls. In this sense, the Internet is an autonomous evolving network.

Recently, complex properties of autonomous evolving networks, i.e metabolic networks, food webs, Internet web links, academic co authorship relations, actor collaborations, and electric power-grid networks, have been studied from the viewpoints of scale-free properties [5][6]. The structure of the Internet and the flow also has been studied extensively from the viewpoints of statistical physics [7][8][9][10][11][12].

1.4 Scale-free structure of the Internet

The Internet can be described a network whose elements are routers and logical connections.

It is usually viewed as an undirected network, in which nodes represents routers and links represent logical connections between them. By using network program called traceroute, the information of the Internet router connectivity can be obtained. The traceroute command sends out hop-limited IP packets toward a given destination. Each packet contains a specific time-to-live (TTL) that decreases by one at each hop. If the TTL becomes zero, the packet will be discarded out of the network. The network sends the death message to the sender. By incrementing the TTL by one at each successive trial, the intermediate hops in the path to given destination can be explored. It is possible to observe degree k of each router by collecting the routing paths between routers.

Faloutsos et al.[7] have studied the Internet at router level. They concluded that the degree distribution of the Internet routers follows a power law P(k)~k^(-γ),γ=2.2. The distribution is highly variable in the sense that the degree varies over range close to the three orders of magnitude. This behavior is different from the degree distribution of classical random networks where the degree distribution decays exponentially.

Barabási and Albért in their paper in 1999 [5] called the Internet "scale-free" network. Such networks have been specified as scale-free networks because of their degree distribution obey a power law. Consider a power law function.

$$f(x) = \beta x^{-\gamma} \tag{1.1}$$

By changing the scale of the variable x to αx, the function behaves

$$f(\alpha x) = \beta(\alpha x)^{-\gamma} \tag{1.2}$$

$$= \beta \alpha^{-\gamma} x^{-\gamma} \tag{1.3}$$

the factor only changes the value of the constant part. The form does not change. Namely it is "free" from the scale.

1.5 Power-law dynamical properties of the Internet

According to the discussion in Section 1.2, messages in the Internet are delivered as series of packets. The packet flow in the Internet had been modeled as Poisson process, because hosts are assumed to send data packets randomly. The validity of this assumption, however, has been lost completely. Researchers imparted scale-free properties of packet flow in the Internet.

Willinger et al observed Ethernet trace at the packet level [13]. They identified flow between source-destination pairs, and showed that transmission and idle times for those flow obey a power law. Paxson and Floyd [10] traced Internet traffic and observed that burst sizes in FTP transfers, and inter-arrival times of TELNET packets (appearance of keystrokes) obey a power law. In [14], Crovella and Bestavros observed Web traffic. They found that the distributions of file size in Unix systems as well as in Web databases obey a power law. In addition, the distributions of HTTP session times and inter-arrival times of request (time for user's thinking) also obey a power law.

Csabai [15] also observed that series of round trip time (RTT) for two weeks. He found the existence of the $1/f$ type fluctuations. Takayashu et al. [11] observed the packet flow behavior by using a simplified Ethernet simulation. He found that the competition of sending packets among nodes and the binary exponential back-off algorithm are revealed to play important roles in producing $1/f$ fluctuations.

Recent interesting study by Tadaki [12], the Detrended Fluctuations Analysis (DFA) is applied on MRTG data. He reported that packet flow on the Internet has one-day periodic behavior with power law behavior.

1.6 The Outline of this paper

The objective of this paper is to give understanding to the reader of evolving processes of the structure of the Internet. For reaching the objective, at the beginning, we introduce the observational results of the Internet in [5][6]. Then, we analyze three known network models: the Erdös-Rényi random network, the random growing network and the Bárabasi-Albért scale-free network theoretically. By simulating those three network models, we give a basic concept to the reader by choosing the most suitable network model which can display similar properties as the Internet has.

The rest of this paper is organized as follows: Section 2 discusses the introduction to dynamic network models. Chapter 3 contains the summary and suggests future directions for research.

ISSN 2303-0135

2 NETWORK MODELS

2.1 Definition of networks

A network consists of a set of N nodes connected by L edges. As shown in Figure 2.1 computers in the system are drawn by circles called nodes and the file sharing relations are represented by lines called edges. Edges can be directed or undirected. A network whose edges are undirected is called an undirected network. It is represented by $g = (V, E)$. A set of nodes is denoted by $V = \{v_1, v_2, v_3, \dots, v_N\}$. A set of edges is denoted by $E = \{e_{i_1 j_1}, e_{i_2 j_2}, e_{i_3 j_3}, \dots, e_{i_L j_L}\}$. Edge e_{ij} is identical to the e_{ji}.

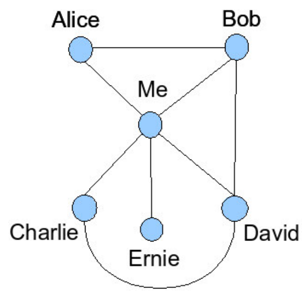

Figure 1.1. **A file sharing relation of the network for computers**

An undirected network g can be represented by its *adjacency matrix* A. An $N \times N$matrix whose element a_{ij} represents the existence of e_{ij} between v_i and v_j. An element a_{ij} has only two possible integer values: 0 and 1. If the value of a_{ij} is equal to 0, the edge e_{ij} is absent. Otherwise, the edge e_{ij} exist. An adjacency matrix of an undirected network is symmetric: $a_{ij} = a_{ji}$. Diagonal elements are equal to 0: $a_{ii=0}$. The matrix A in eq. (2.1) represents the adjacency matrix of the network shown in the Figure 2.1 :

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad (2.1)$$

2.2 Degree of node

The degree k_i is the number of edges attached to v_i. It is equal to the number of adjacent nodes. In literatures, sometimes it is called as node connectivity. The degree k_i is calculated from the adjacency matrix of the network:

$$k_i = \sum_{j=1}^{N} a_{ij} \qquad (2.2)$$

The total degree of the network, K is defined as the sum of degrees of all nodes. It is equal to the sum of the elements of A. It is also equal to two times the number of edges $2L$, because elements which represent ee_{ij} are counted twice in A.

$$K = \sum_{j=1}^{N} k_i \qquad (2.2)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} \qquad (2.2)$$

The degrees of nodes in the example are: $k_{me} = 5$, $k_{Alice} = 2$, $k_{Bob} = 3$, $k_{Charlie} = 2$, $k_{David} = 3$, $k_{Ernie} = 1$ and $K = 5 + 2 + 3 + 3 + 3 + 1 = 16$.

2.3 Network model

2.3.1 Random growing networks (RGN) model

The random growing network model is inspired from the facts that the natural and artificial networks are growing. Figure 2.3 gives an illustration of the random growing network model. Here we observe the simplest form of the model. It is initialized with two nodes, $v_s s$ $(s = 1;\ 2)$. These nodes connect each other with two links. Namely, the degree k of both nodes is 2. At every time step t, a new node v_s $(s = t)$ is created. The new node is connected to a randomly selected node $v_{s'}$ $(0 < s' < t)$. The probability that the new node in the network selects an existing node $v_{s'}$ is uniformly $p(v_{s'}) = 1/_{t-1}$ In other words, all nodes with any value of degree k have the same probability to be chosen. The number of nodes is kept equal to t.

For theoretical discussion, let us consider a set of random growing networks G where each network has the same initial condition and follows the rules mentioned above.

$$G = \{g_m | 0 \leq m \leq M\} \tag{2.3}$$

The number of nodes of degree k at time t in a network g_m is

$$N(k,t) = \sum_{s=1}^{t} \delta_{k(s,t),k}. \tag{2.4}$$

Where $k(s,t)$ is defined as the degree of s-th node at time t. The number of nodes of degree k in G is

$$n(k,t) = \sum_{m=0}^{M-1} N_m(k,t). \tag{2.5}$$

The average number of nodes of degree k in the set G is

$$\langle N(k,t)\rangle = \frac{1}{M} n(k,t) \tag{2.6}$$

The degree distribution is defined by

$$P(k,t) = \frac{1}{t}\langle N(k,t)\rangle. \tag{2.7}$$

Figure 2.2, 2.3, and 2.4 represent a set of RGNs model.

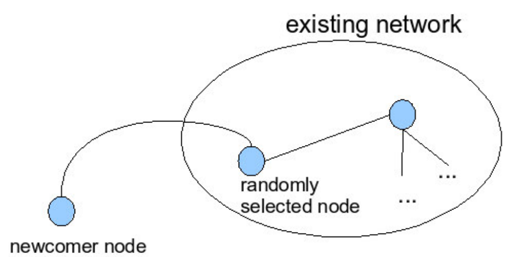

Figure 2.2. Basic idea of the RGN model.

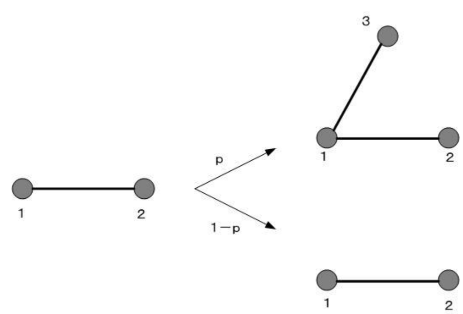

Figure 2.3. The propbability of node s to gets an edge is p, which is equal to $\frac{1}{t-1}$.

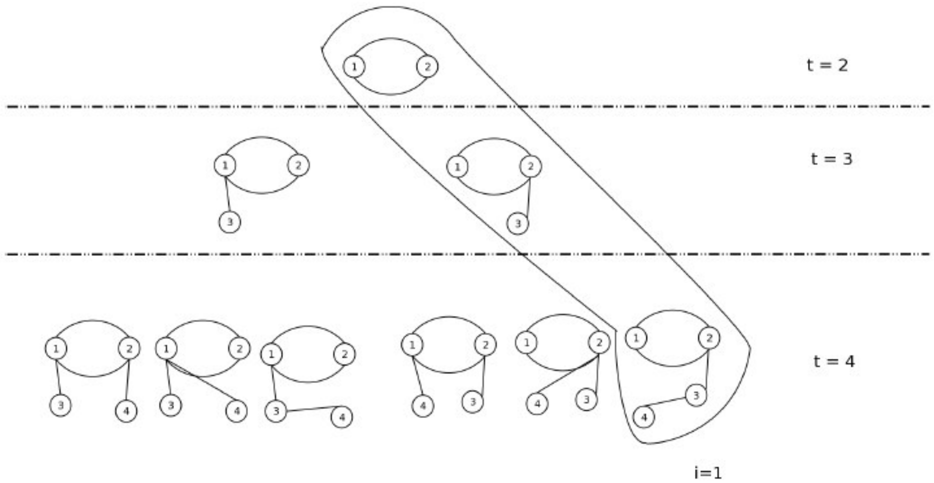

Figure 2.4. Simulating random growing networks by hand for t=4.
The networks in the set G can be categorized into six networks.

We introduce a master equation that can describe as the change of the average number of nodes of degree k in G for every time step. Before we discuss it in detail, lets us describe three possible conditions at every time step t as the changes of the number of nodes of degree k in G:

1. The number of nodes of degree k decreases, because in some networks in G the nodes of degree k are chosen by new nodes for attaching their ends of edges with probability $\frac{1}{t-1}$
2. The number of nodes of degree k increases, because in some networks in G the nodes of degree $k-1$ are chosen by new nodes for attaching their ends of edges with probability $\frac{1}{t-1}$
3. The number of nodes with $k = 1$ in G increase by one. This is because in G a new nodes have to create edges to randomly chosen nodes.

We can write these conditions as the change of the number of nodes of degree k in G at time t:

$$\langle N(k,t+1)\rangle - \langle N(k,t)\rangle = \frac{1}{t}\langle N(k+1,t)\rangle - \frac{1}{t}\langle N(k,t)\rangle + \delta_{k,1}. \quad (2.8)$$

By using eq. (2.7) we rewrite eq. (2.8) to

$$(t+1)P(k,t) - tP(k,t) = P(k-1,t) - P(k,t) + \delta_{k,1} \quad (2.9)$$

In eq. (2.9) the birth of new nodes in the RGNs is $\delta_{k,1}$.

The stationary solution of the master equation eq. (2.9) is obtained by setting $P(k) = P(k,t)$. The equation for stationary degree distributions is

$$2P(k) - P(k-1) = \delta_{k,1} \quad (2.10)$$

with boundary condition $P(0) = 0$. The stationary solution of RGNs is

$$P(k) = 2^{-k} \quad (2.11)$$

The degree distribution of the random growing network obeys an exponential distribution. The number of nodes of degree k decays exponentially. The RGN is not a model for explaining a power law degree distribution which is seen in the Internet and other natural and artificial networks.

We have performed simulations of constructing random growing networks. We generate 100 net-works. Each network has evolved up to 1000 nodes. Figure 2.5 shows the average degree distribution. The average degree distribution is well fit by the theoretical curve (2.1).

We found that the number of nodes with large values of k decays exponentially. The random network model fails to capture the properties of the Internet topology whose degree distribution $P(k)$ follows a power law as seen in Section 1.4.

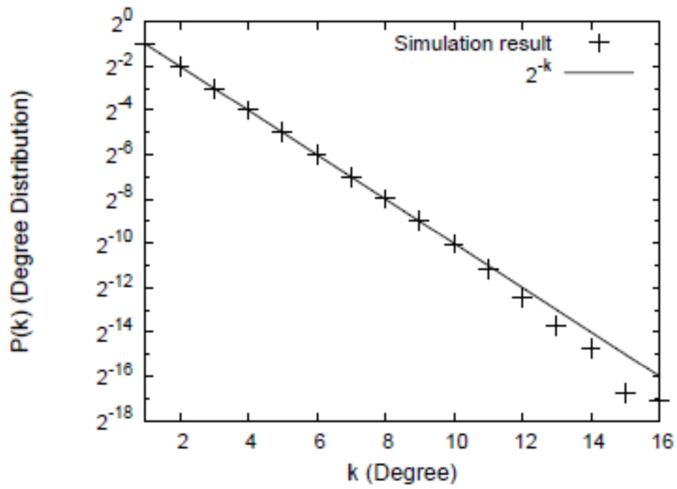

Figure 2.5. The average degree distribution of generated random growing networks (1000 networks) and the theoretical curve $P(k) = 2^{-k}$.

2.3.2 Barabási-Albért network model

We have discussed network models in which nodes connect each other in a random fashion and independently from their properties. Barabasi and Albert [5] introduced a different network model, where new nodes tend to make links to existing nodes with a large degree. This linking mechanism is called preferential attachment. Figure 2.4 gives the illustration of the basic idea of the Barabási-Albért network model. As in the random growing network, the initial state of the BA network consists of two nodes v_s $(s = 1, 2)$ connected with two links. At every time step t, a new node v_s $(s = t)$ is created. It is connected to a randomly selected node $v_{s'}$ $(0 < s' < t)$ with the preferential probability, which is proportional to its degree $k_{s'}$. The number of nodes is kept equal to t.

As for random growing networks, a set of BA networks is introduced for discussing the master equation. We define G as a set of BA networks where each network has the same initial condition and obeys the rules mentioned above :

$$G = \{g_r | 0 \leq r \leq M\}. \quad (2.12)$$

The number of nodes of degree k at time t in the graph g_r as

$$N(k,t) = \sum_{s=1}^{t} \delta_{k(s,t),k}, \quad (2.13)$$

Where $k(s,t)$ is the degree of s-th node at time t.

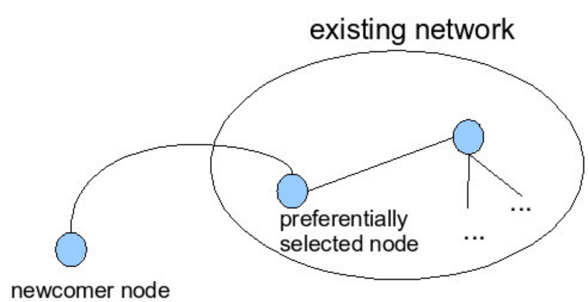

Figure 2.6. The newcomer node selects a previously existing node with probability proportional to its degree.

The number of nodes of degree k in G is defined as:

$$n(k,t) = \sum_{r=0}^{M-1} N_r\,(k,t). \tag{2.14}$$

The average number of nodes of degree k in the set G :

$$\langle N(k,t)\rangle = \frac{1}{M} n(k,t). \tag{2.15}$$

The degree distribution is defined by

$$P(k,t) = \frac{1}{t}\langle N(k,t)\rangle. \tag{2.16}$$

We construct a master equation that can describe the change of the average number of nodes of degree k in G for every time step. Lets us describe three possible conditions at every time step t of the change of the number of nodes of degree k in G:

1. The number of nodes of degree k decreases. This condition happens when the nodes of degree k are chosen by new nodes with probability $p = k/2t$
2. The number of nodes of degree k increases. Because in the networks in G the nodes of degree $k-1$ are chosen by new nodes with probability $q = k - 1/2t$
3. The number of nodes with $k = 1$ in G increase by one. This is because in G thenew nodes have to create edges to the chosen nodes.

We can describe these conditions above as the change of the number of nodes of degree k in G at time t:

$$\langle N(k,t+1)\rangle - \langle N(k,t)\rangle = \frac{k-1}{2t}\langle N(k-1,t)\rangle - \frac{k}{2t}\langle N(k,t)\rangle + \delta_{k,1} \tag{2.17}$$

By following the eq. (2.16), we rewrite eq.(2.17) to

$$(t+1)P(k,t+1) - tP(k,t) = \frac{k-1}{2}P(k-1,t) - \frac{k}{2}P(k,t) + \delta_{k,1} \tag{2.18}$$

The stationary solution of the master equation eq. (2.2.45) is obtained by setting $P(k) = P(k,t)$. The equation for stationary degree distributions is

$$0 = \frac{k-1}{2}P(k-1) - \frac{k}{2}P(k) - P(k) + \delta_{k,1} \tag{2.19}$$

$$\frac{1}{2}\left(kP(k) - (k-1)P(k-1)\right) + P(k) = \delta_{k,1} \tag{2.20}$$

The degree distribution of BA scale-free networks is

$$P(k) = \frac{4}{(k+2)(k+1)k}\,. \tag{2.21}$$

The stationary degree distribution of Barabasi-Albert networks is shown by eq. (2.21). If k is large, then the degree distribution behaves as power law $P(k) \sim k^{-3}$. This result indicates the preferential attachment mechanism seems to be the origin of power law degree distribution as seen in the Internet and wide range of natural and artificial networks.

We have performed simulations of constructing BA networks. We generate 100 networks. Each network has evolved up to 1000 nodes. Figure 2.7 shows the average degree distribution. The average degree distribution is well fit by the theoretical curve (2.21).

An interesting fact about the degree distribution of this model is that there is a finite probability of finding nodes with degree much larger than the average $\langle k\rangle$. In other words, the consequence of a power-law is that the average behavior of the system is not typical. The characteristic degree is the one that, picking up a node at random, should be encountered most of the time. In this distribution, most of of nodes will have small degree, but there is a probability of finding nodes with large degree values. This condition is totally different with Poisson distributions where the average value is very close to the maximum of the distribution. It is also different with exponentially decaying distributions observed in random growing network where the exponential function decays at constant rate. The scale-free structure of the Internet as described in Section 1.4 seems to be explained by the BA scale-free network model.

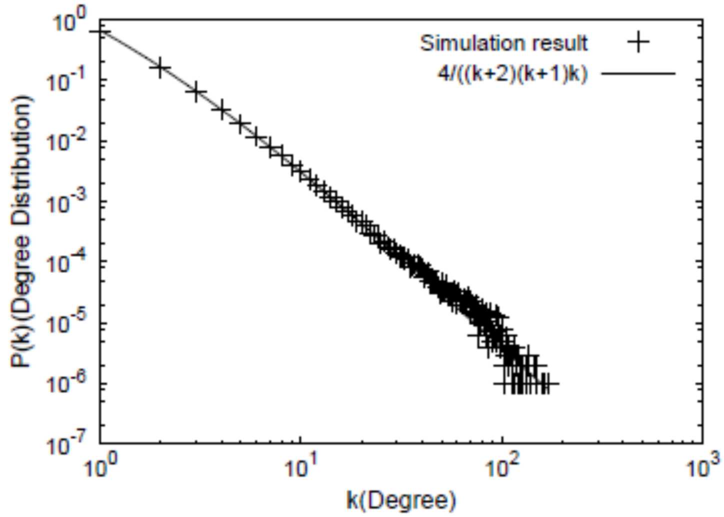

Figure 2.7. The average degree distribution of generated BA scale-free networks (1000 networks) and the theoritical curve $P(k) = \frac{4}{(k+2)(k+1)k}$.

3 Summary and Discussion

3.1 Summary

The Internet is one of the most important communication media in modern societies. Despite its importance, the Internet has no center for controlling its flow and structure. As other autonomous evolving networks, the Internet has been observed in the viewpoint of scale-free properties [5][6][9].

In this paper we have analyzed and simulated the degree distribution of two known network models: the random growing network and the BA scale-free network. The scale-free structure of the Internet seems to be explained by the random growth with preferential attachment mechanisms which belongs to BA scale-free network. The BA scale-free network has a capability for displaying the same properties as the Internet, where its degree distribution obeys a power law. The important characteristic shape of the degree distributions of BA scale-free networks is that most nodes have small number of connections while a few nodes have many connections to others.

3.2 Discussion

Studying the topology of the Internet is just a part of understanding the overall characteristics and dynamical behavior of the Internet. There might be step costs, for instance, understanding the dynamical properties of packet flow of the Internet. Does the scale-free structure affect the dynamical properties of the Internet?

The packet flow in the Internet has been reported to bear power-law fluctuations [12][15][9]. There are two types of candidates for origins of power-law behavior. One is internal and the other is external. As internal mechanisms, the Internet has various types of queuing and retardation mechanisms. Examples are CSMA/CD (Carrier Sense Multiple Access/ Collision Detection) mechanisms in the Ethernet, queuing at IP routers, and re-sending mechanisms in TCP. If the amount of packets which flows over the Internet is very small, the correlations between packets generated by internal mechanisms will be lost. In this case, the powerlaw behavior will be affected by external sources. The external sources are the structure of the Internet and the demand of communication in the Internet. The relation between the structure and the dynamical properties of the Internet is in our future research plan.

4 References

[1] J. Licklider, Spring Joint Computer Conference 21, 113, 1962.
[2] D. C. B.Leiner, V.Cerf, *Communications of the ACM* 40, 102, 1997.

ISSN 2303-0135

[3] W. R. Stevens, “TCP/IP Illustrated”, Volume 1: The Protocols, Addison-Wesley, 1994.
[4] P. arn and C. Partridge, *ACM Transactions on Computer Systems* 9, 364, 1991.
[5] A. Barabási and R. Albért, “Reviews of Modern Physics”, 74, 47, 2002.
[6] S. Dorogovtsev and J. Mendes, “Advances in Physics”, 51, 1079, 2002.
[7] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *ACM SIGCOMM Computer Communication Review*, 29, 251, 1999.
[8] R. Govindan and H. Tangmunarunkit, *IEEE INFOCOM 2000* , 1371, 2000.
[9] R. Pastor-Satoras and A. Vespignani, “Evolution and Structure of the Internet”, Cambridge University Press, 2004.
[10] V. Paxson and S. Floyd, *IEEE/ACM Transactions on Networking* 3, 226, 1995.
[11] M. Takayasu, H. Takayasu, and T. Sato, *Physica A* 233, 824, 1996.
[12] S. Tadaki, *Journal of the Physicsl Society of Japan*, 76, 044001, 2007.
[13] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, *IEEE/ACM Transactions on Networking,* 5, 71, 1997.
[14] M. Crovella and A. Bestavros, *The ACM International Conference on Measurement and Modeling of Computer Systems,* 24, 160, 1996.
[15] I. Csabai, *Journal of Physics A: Mathematical and General*, 27, L417, 1994.
[16] P. Erdos and A. Rényi, Publ. Math. Debrecen 6, 290, 1959.