Analysis of Large-Scale Topological Properties for Peer-to-Peer Networks*

Hao Chen, Hai Jin, Jianhua Sun, Dafu Deng, Xiaofei Liao
Cluster and Grid Computing Lab
Huazhong University of Science and Technology, Wuhan, 430074, China
Email: {haochen,hjin,jhsun,dfdeng,xfliao}@hust.edu.cn

Abstract

In this paper, we present a thorough study about the topological properties of Gnutella network, which provides us insight into the nature of underlying system, helps us design high performance algorithms and generates more realistic topologies for simulation experiments. We compare two theoretical models of growing networks with the present real data of Gnutella network.

1 Introduction

In the past several years, Peer-to-Peer (P2P) networks have emerged as effective ways for communication and cooperation among geographically distributed computers. P2P systems are often built at the application level and use their own communication protocols to form a virtual network over the underlying physical network. The topology of the virtual network shares some common properties of complex networks in other disciplines of science, and has a significant impact on performance, scalability and robustness of P2P systems.

Recently, a large proportion of research effort has been devoted to the study and modeling of a wide range of natural systems that can be regarded as networks, focusing on large scale statistical properties of networks other than single small networks. Some reviews on complex networks can be found [5, 11]. From biology to social science to computer science, systems such as the Internet [9], the World-Wide-Web [8], social communities, food chain and biological networks can be represented as graphs, where nodes represent individuals and links represent interactions among them. Despite this simple definition, these networks often exhibit high degree of complexity due to the wiring entanglement during their growth. Researches on these networks have revealed some commonalities. Many of these networks have complex topological properties and dynamical features that can not be explained by the classical graph model of random networks, the Erdos-Renyi model [7].

The advances in P2P systems and other complex networks have motivated us to carry out an intensive study of the topology of existing P2P networks, which will provide insight into the nature of underlying system, help us design high performance algorithms according to such topologies and generate realistic topologies in simulation experiments.

In our study, we choose Gnutella as our testbed, due to its large user community and open architecture. Some previous works have been done on the measurement and analysis of Gnutella network [2, 12, 13, 14], but most of them focused on other topics such as user behavior [2], bottleneck bandwidth [13] and search algorithms [1]. In this paper, we put a strong emphasis on the study of topological properties and the model of these properties.

The rest of this paper is organized as follows. Section 2 describes Gnutella protocol and crawler implementation used to discover the topology of Gnutella network. In section 3, we analyze the statistical distributions of Gnutella topology. Section 4 is devoted to a detail discussion of some complex network models compared with the real data analysis. Finally, section 5 ends with our conclusions and future works.

2 Gnutella Protocol and Data Collection

As mentioned in the original Gnutella protocol, individual nodes, also called *servents* perform tasks normally associated with both clients and servers. They provide client-side interfaces through which users can issue queries and view search results. At the same time, they also accept queries from other *servents* and response with applicable results.

Originally, all Gnutella peers connected with each other randomly, which caused scalability problem in Gnutella network. The ultrapeer system has been found an effective way for solving that problem. The scheme organizes the network in a more structured form. All peers are categorized as leaves and ultrapeers. A leaf keeps only a small number of connections to ultrapeers. An ultrapeer behaves as a proxy for the leaves connected to it.

We have developed a *crawler* to collect topology infor-

*This paper is supported by National Natural Science Foundation under grant 60273076.

0-7803-8430-X/04/$20.00 ©2004 IEEE

mation of Gnutella network. According to Gnutella protocol, a ping message with TTL=2 and HOP=0 is regarded as a crawler ping, and peers, upon receiving a crawler ping, would respond with appropriate pong messages. Based on this mechanism, it is possible to discover Gnutella topology by performing a breadth first searching on the network. From our experience and observations, we find that some clients such as *Gnucleus*, *Morpheus* (based on *GnucDNA*) do not respond the crawler ping appropriately. Fortunately, these clients send an information page summarizing servents' status to any web browser trying to connect to it. Motivated by this, we also developed a web spider as a means of collecting topology information from these clients, and integrated the web spider into the crawler, which accelerated the crawling process remarkably. The crawler were written in Java based on *limewire*'s [16] open source client, and ran in parallel using 40 threads. Our crawler can discover more than 50,000 peers within half an hour, which corresponds to 50% of the total number of peers in the system at any time according to the statistical information reported by *limewire*.

3 Topological Properties

In this paper, we use three data sets (collected by our crawler) referenced as V34206, V48134 and V57926 respectively, where the figures represent the number of peers discovered during each crawling.

3.1 Small-world and Power-law

Two fundamental topological properties of complex networks are the average shortest path length and the network diameter. We define $\bar{l}$ to be the average shortest path length between vertex pairs in a network:

$$\bar{l} = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij} \qquad (1)$$

where d_{ij} is the shortest path length from node i to node j. Similarly, the network diameter d is defined as the largest among the shortest paths between any node pairs, i.e. d_{ij}, in the network. Many natural networks show surprising small-world effect, i.e. one can go from a node to any other node in the network only through a small number of intermediate nodes averagely. In these networks, the network diameter grows logarithmically with the size n of the network. In Table 1, we give some statistics obtained from our data set and previous work [14] about diameter and average shortest distance of Gnutella network.

In Table 1, the left-hand side is the results computed using our data sets, and the right-hand side is the results presented in [14]. As shown in the table, all snapshots of Gnutella network show small world phenomenon with small average shortest distance and small diameter. These

Table 1: Basic statistics of Gnutella network. (Properties measured are: total number of nodes N; total number of edges E; average shortest distance $\bar{l}$; diameter d; average degree k. Blank entries indicate unavailable data.)

N	34206	48134	57926	992	1008	1007
E	43958	64408	80276	2465	1782	4094
$\bar{l}$	5.4	5.6	5.8	3.7	4.4	3.3
d	16	15	15	9	12	10
k	2.57	2.72	2.83			

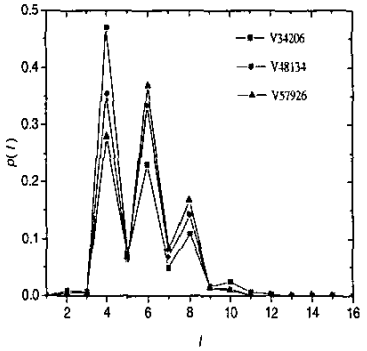

Figure 1: Distribution of the shortest path length for three data sets.

results indicate strong small world properties of the topology of Gnutella network. Both the diameter and the average shortest distance increase compared with previous measurements, that is due to the increasing of the size of the whole network. Furthermore, the network still remains small world property compared with the large size of the network. The distribution $p(l)$ of shortest distances of l between pairs of peers is shown in Figure 1. The distribution is characterized by several peaks around its average value, and its shape remains unchanged from V34206 to V57926 data sets, which also identifies the small average shortest distances.

Another common feature of many complex networks is the property of clustering. In these networks, if node A is connected to B and C, B is likely connected to C. The clustering can be quantified by clustering coefficient, which is defined as the ratio between the number of edges E_i among the k_i neighbors of a given node i and its maximum possible value $k_i(k_i-1)/2$, i.e.

$$C_i = \frac{2E_i}{k_i(k_i-1)}. \qquad (2)$$

The clustering coefficient of the whole network $\langle C \rangle$ is the average of C_i over all nodes in the network. It provides a measurement of how well the neighbors of a node are locally interconnected.The maximum value of $\langle C \rangle$ is 1, corresponding to a fully connected network.

In Figure 2, we plot the distribution of $\langle C \rangle_k$ as a function of the node degree k for the three data sets. As shown in Figure 2, we can not find obvious evidence for a power-law behavior of this distribution. But some useful information

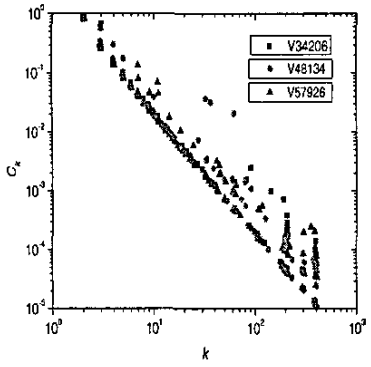

Figure 2: Log-log plot of clustering coefficient as a function of the node degree for three data sets.

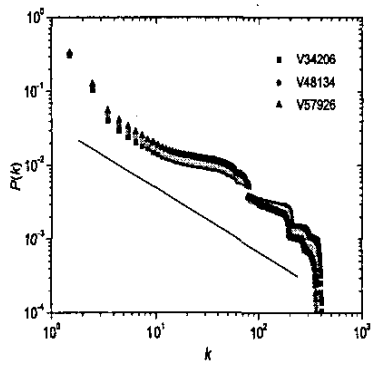

Figure 3: Log-log plot of degree vs. cumulative degree distribution for three data sets. The power-law behavior is characterized by a slope -1.0, which yields a degree exponent $\gamma = 2.0$.

can be obtained from analyzing the clustering coefficient c_k as a function of the node degree k. According to this measurement, we see that the clustering coefficient c_k follows an evident decay with the increasing of the node degree k, which indicates that nodes with small degree have larger local clustering coefficient than those with large degree. Low degree nodes form well connected local subgraph by connecting to high degree nodes (called hubs), and at the same time local subgraphs are connected to each other by hubs. That fits into the pattern in Gnutella network, where leaves (having small degree) are always connected to several ultrapeers (hubs) with each ultrapeer connecting to several other ultrapeers.

The degree of a node in a network is the number of edges connected to that node. Some important information about a network can be extracted from its degree distribution $p(k)$, which is the fraction of nodes that have degree k. The degree distribution of random graphs obeys binomial distribution, or Possion distribution with a peak at $p(\langle k \rangle)$. However, the degree distribution of many real networks always has a power-law tail [11]

$$p(k) \sim k^{-\gamma} \tag{3}$$

We also use cumulative probability distribution to express the probability that a node has degree larger than or equal to k. The degree distribution is presented as:

$$P(k) = \sum_{k'=k}^{\infty} p_{k'} \sim \sum_{k'=k}^{\infty} k'^{-\gamma} \sim k^{-(\gamma-1)} \tag{4}$$

It has the advantage of being less noisy than the original distribution. Cumulative distributions also have power-law tails, but with exponent $\gamma - 1$ rather than γ.

In Figure 3, we plot the cumulative probability distribution of the node degree corresponding to three data sets on a log-log scale, and use a linear regression to fit a line for the plot. As shown in Figure 3, it is obvious that the plots of all data sets follow power-law distributions with slope close to -1.0, yielding a degree exponent $\gamma = 2.0$, which is in good agreement with previous results [1]. The average degree of the network is shown in Table 1, which indicates that most peers (leaves) keep only a small number of connections except for ultrapeers.

3.2 Centrality

One of the most important measurements in network analysis is to identify the most 'central' or 'influential' nodes in a network. To locate these central nodes is to evaluate the centralities [10] of these nodes. Measuring of centrality can be defined in two different ways: node centrality (one value per node) and network centralization (one value for the whole network). This three cases correspond to three centrality measurements–degree centrality, closeness centrality, and betweenness centrality. Their formal definitions are as follows.

(1). Degree centrality is define as the ratio between the degree of a node and the highest possible degree.

$$C_d(x) = \frac{degree\ of\ node\ x}{N-1} \tag{5}$$

where N is the number of nodes in a network.

(2). Closeness centrality is defined as

$$C_c(x) = \frac{1}{\sum_{y \subseteq U} d(x,y)} \tag{6}$$

where $d(x,y)$ is the shortest path length between node x and y, U is the set of all nodes.

(3). Betweenness centrality is defined as

$$C_b(x) = \sum_{\{(y,z)\}} \frac{n_x(y,z)}{n(y,z)} \tag{7}$$

where $n_x(y,z)$ is the number of the shortest paths running through node x, and $n(y,z)$ is the number of the shortest paths between node y and z.

(4). Freeman [10] defined general network centralization index

$$C_n = \frac{\sum_{x \in U}(C_n^* - C_n(x))}{max \sum_{x \in U}(C_n^* - C_n(x))} \tag{8}$$

where C_n^* is the highest value of selected node centrality $C_n(x)$ in the set of nodes of a network.

Network centralization is a number between 0 and 1. The value 0 indicates that if all nodes have equal centrality (ring topology), and 1 if one node completely dominates all other nodes (star topology). There are also three measurements about network centralization according to the three definitions of node centrality.

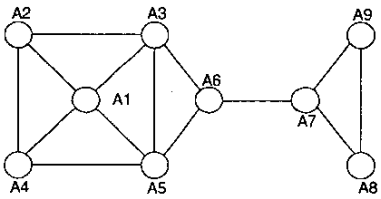

Figure 4: A sample network.

Figure 4 shows a simple network. In this network, A1 has the most connections, making it the most 'cental' node in the network with the highest degree centrality. A1 corresponds to the ultrapeer in Gnutella network that connects to a large number of leaves. Ultrapeers are always more active in routing *ping* and *query* messages. The failures of these nodes have a negative impact on the performance of the whole network.

Although A3 and A5 have fewer connections than A1, their locations in the network allow them to communicate with all the nodes more quickly than any others. They are close to everyone else with higher closeness centrality. One property strongly correlated with closeness centrality is the average shortest path length. A network with more short paths is more efficient in transferring data, and more flexible with the change of the network topology. Hence, in a P2P network, maximizing the closeness centrality of all nodes will decrease the average path length and improve the performance of communication among nodes correspondingly.

As shown in Figure 4, A6 has fewer connections than A1, A3, and A5 (also fewer than the average degree), but it is located at such an important point that many communication paths in the network must across it. It acts as an agent between two communities, and plays a central role in the network with larger betweenness centrality. Intended attacks on such node would break down the functionality of the whole network.

As centralities of individual nodes provide position information about these nodes in the network, the network centralization obtained from individual centralities can reveal much about the overall topology of the network. A very centralized network is controlled by one or a few central nodes, and networks of this type characterizing with larger network centralization values are vulnerable by removing the central nodes. However, networks with low centralization values, such as those decentralized P2P networks, are always resilient under random failures of nodes. In next section, we will present a detail study about the resilience of Gnutella network.

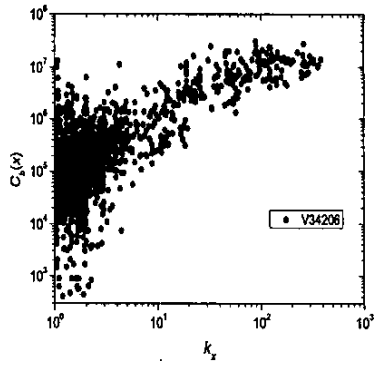

Figure 5: Log-log plot of betweenness distribution as a function of node degree for data set V34206.

From the above analysis about centrality, we see that centralities have great impact on optimizing topologies of networks. In the following, we give our measurements about centrality in Gnutella network.

Table 2: Network centralization of Gnutella network.

N	34206	48134	57926
E	43958	64408	80276
degree	0.01177	0.00844	0.007
closeness	0.16742	0.18938	0.19221
betweenness	0.05131	0.04256	0.02634

Three measurements of network centralization are shown in Table 2. The small values in Table 2 imply that Gnutella network is not dominated by one or a few central nodes, hence it is fault tolerant under random failures of nodes. Figure 5 and Figure 6 depict the distributions of betweenness and closeness as a function of node degree respectively. As shown in Figure 5, there are also some nodes with low degrees having high values of betweenness centrality, which indicates that attacks on such low degree nodes would also influence the performance of the whole network compared with attacks on high degree nodes as described earlier. Furthermore, nodes with large degree (with fewer vertically plotted scatters) obviously have large values of betweenness centrality because they are often more active or central than others. Figure 6 shows that closeness centrality is distributed uniformly within the range [0.1, 0.2], which in another way illustrates the fact that all nodes with different degrees in the network can reach others easily.

3.3 Network Resilience

Recent interest about network resilience has been sparked by the work [3], which studied the effect of random removal (called random failure) and intentional removal (called attack) of the nodes in networks.

With the removing of nodes from a network, some paths between pairs of nodes is broken. The average length of these paths increases (the closeness centrality decreases accordingly). Eventually nodes are isolated in different

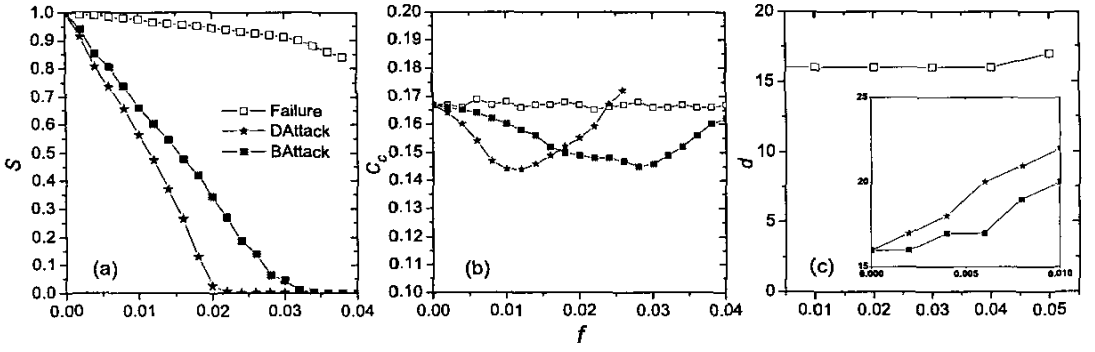

Figure 7: Results for random failures (open square), degree-based (star), and betweeness-based (filled square) attacks of nodes measured by the relative size of largest cluster S, the closeness centralization C_c, and diameter d as functions of the fraction of removed nodes f in Gnutella network. Inset in (c) is magnification in the early stage of attack.

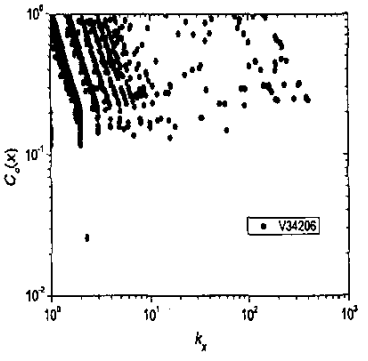

Figure 6: Log-log plot of closeness distribution as a function of node degree for data set V34206.

clusters (the network is fragmented into many small clusters), and communications between them becomes impossible. Some real networks display high degree of robustness against random failures of nodes, but they are also very vulnerable under attacks to the highest degree nodes. As described in the above section, betweenness centrality is an important concept that captures the prominence of a node in the network. Thus, it is natural to expect that removal of nodes with high betweenness centrality also degrades the functionality of networks. In this paper, we take into account both degree and centrality to see how differently these two properties affect the network resilience of Gnutella network. To explain the damages caused by attacks and random failures, we measure three parameters: the relative size of the largest cluster S (defined as the ratio between the size of the largest cluster and the size of original network), the average closeness centrality C_c (defined as the average of the closeness centralities of all nodes in the largest cluster) and network diameter d.

Figure 7 summarizes the results for random failures (Failure), degree-based (DAttack) and betweeness-based (BAttack) attacks of nodes measured by S, C_c and d as functions of the fraction of removed nodes f. According to our observations, we find that the percentage of ultrapeers in Gnutella network is about 3.3%. Due to this, the fraction of removed nodes f in simulating attacks is limited in the range $[0, 0.05]$.

As shown in Figure 7, Gnutella network shows high degree of tolerance against random failures. However, the fault tolerance comes at the expense of attack vulnerability: rapid increasing of the network diameter, rapid decreasing of the relative size of the largest cluster and the closeness centrality in early stage. Both S and C_c show threshold phenomenons: $S \simeq 0$ when $f > 0.022$ and $f > 0.032$ under DAttack and BAttack, respectively. C_c reaches the lowest point when $f = 0.01$ and $f = 0.028$ under DAttack and BAttack, respectively. With the increasing of f, C_c decreases accordingly because the removal of some important nodes lengthen the average path distance between node pairs. However, after some critical points, the largest cluster becomes much smaller than the initial size of the network, which causes the fallback of average path length in such clusters and the increasing of C_c correspondingly. Specially, C_c even exceeds the initial value when f reaches around 0.025. Another interesting characteristic shown in Figure 7 is that degree-based attacks affect the network topology more significantly than betweenness-based attacks (the curves change faster). The reason will be described in the next section, where we compare that with other two theoretical models. In Figure 7(c), we only show the results of DAttack and BAttack in the inset in the range $[0, 0.01]$, which is sufficient to demonstrate the impact caused by attacks.

4 Modeling P2P Networks

In the above section, we presented a thorough study of the topology of Gnutella network. Applications running on top of a P2P network rely heavily on the topology of the network. Thus designing accurate network models of P2P network is of importance to simulate these applications on top

of P2P networks. First, we introduce three general models for complex networks. The models are defined as follows:

ER (Erdos-Renyi) model: This model defines a random graph of n nodes with each pair of nodes being connected with probability p. The probability space of random graphs $G(n,p)$ is a finite probability space whose elementary events are all graphs on a fixed set of n nodes and where the probability of a graph with m edges is $p(G) = p^m(1-p)^{\binom{n}{2}-m}$.

BA (Barabasi-Albert) model: Two main features presented by this model [6] are the growing nature of the network and a preferential attachment rule, in which the probability of adding new connections to a given node grows linearly with its degree. The algorithm of BA model is described as follows: Staring with a small number (m_0) of nodes, at every step we add a new node with m edges that link the new node to m different nodes already in the system. The probability that a new node is connected to node i depends on the degree k_i of node i, such that

$$\prod_{BA}(k_i) = \frac{k_i}{\sum_j k_j} \tag{9}$$

After n steps, we obtain a network with degree distribution $p_k(k) \sim k^{-3}$.

EBA (Extended BA) Model: The Extended BA model [4] describes a more realistic description of network formation by incorporating additional local events that are known to appear in real networks. In this model, starting with m_0 isolated nodes, and at each step we perform one of the following operations:

(1). *With probability p we add $m(m \leq m_0)$ new links.* For each of these links, the starting point of a new link is selected randomly, and the other end of the link is selected with probability

$$\prod_{EBA}(k_i) = \frac{k_i + 1}{\sum_j (k_j + 1)} \tag{10}$$

(2). *With probability q we rewire m links.* We randomly select a node i and a link l_{ij} connected to it, and then remove this link and replace it with a new link $l_{ij'}$ that connects i with node j' chosen with probability $\prod(k'_j)$ given by (10). This process is repeated m times.

(3). *With probability 1-p-q we add a new node.* The new node has m new links that with probability $\prod(k_i)$ are connected to node i already present in the system. Equation (10) leads to a power-law degree distribution with an exponent given by [4]

$$\gamma = \frac{2m(1-q) + 1 - p - q}{m} + 1 \tag{11}$$

Changing the values of m, p, and q, we get the desired degree distribution exponent γ. In the simulations, we use the values $m = 2$, $p = 0.3$, and $q = 0.54$, which yield the exponent $\gamma = 2.0$.

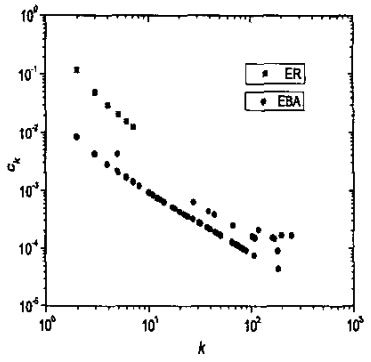

Figure 8: Log-log plot of clustering coefficient as a function of the node degree for ER and EBA models.

As described above, the BA model produces a power-law exponent $\gamma = 3.0$, which does not accord with our result shown in Figure 3. However, EBA model is more flexible than BA model so that we can obtain desired power-law exponent by choosing different parameters. Thus, we perform comparison among our empirical data, ER model and EBA model (although ER model does not show a power-law degree distribution, we still compare it with our results to see how different it behaves against real networks). In ER model, we choose the probability $p = 0.2$.

We perform simulations of the two models with parameters mentioned above and the number of nodes $N = 30246$ according to the size of data set V30246.

Table 3: Properties of the ER and EBA models, compared with the values of data set V34206.

N	ER model	EBA model	V34206
$\bar{l}$	10.8	2.6	5.4
d	26	8	16
closeness	0.04	0.269	0.167

In table 3, we report the values of average shortest path length, diameter, and closeness centralization for the two models, compared with the data set V34206. From the table, we find that the values of V34206 are all between those of ER and EBA models. The EBA model provides a too small value for the average shortest path length $\bar{l}$, and the diameter of ER model is obviously much larger than the other two counterparts.

In Figure 8, we plot clustering coefficient as a function of node degree of the two models. For ER model, C_k does not exhibit a strong dependency with the degree k. However, for EBA model, it is similar to Gnutella network (shown in Figure 2) that C_k follows an evident decay with the increasing of the node degree k except for a little more slower decaying speed.

In Figure 9, we plot the degree distributions of the two models. Obviously, the degree distribution of ER model does not show a power-law tail. On the other hand, a power-law distribution of EBA model is characterized by the exponent $\gamma = 3.0$ in the figure. The exponent in EBA model is

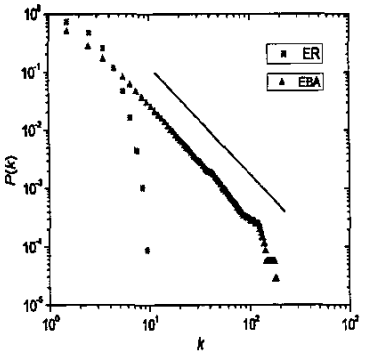

Figure 9: Log-log plot of degree vs cumulative degree distribution for ER and EBA models.

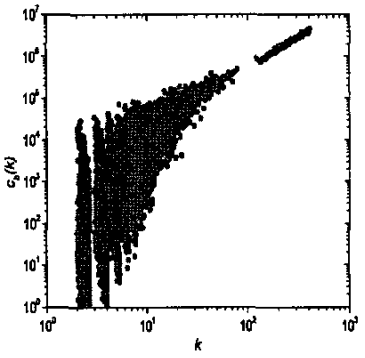

Figure 11: Log-log plot of betweenness distribution as a function of node degree for EBA model.

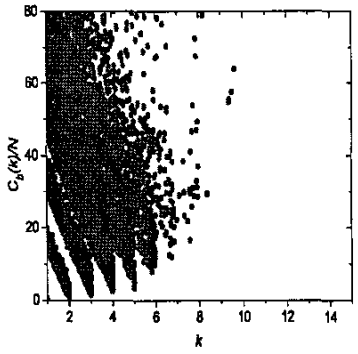

Figure 10: Betweenness distribution as a function of node degree for ER model.

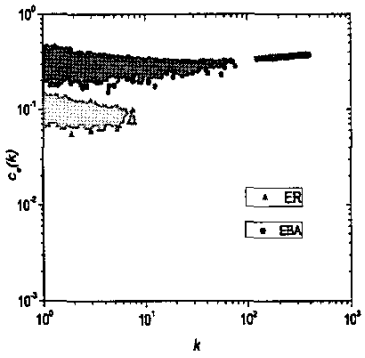

Figure 12: Log-log plot of closeness distribution as a function of node degree for ER and EBA models.

not in good agreement with our data sets that have a smaller value $\gamma = 2.0$ (Figure 2).

We show respectively in Figure 10 and 11 the node betweenness versus the node degree of ER and EBA models. In ER model, despite the absence of high degree nodes, correlations between betweenness and degree are still evident for ER model with an exponential cut-off in the distribution. The EBA model shows similar behavior as compared with Gnutella network. In both EBA model and Gnutella, low degree nodes have large values of betweenness, differing from ER model, where low degree nodes always have smaller values of betweenness than high degree ones.

The distribution of closeness against node degrees of ER and EBA models (plotted in Figure 12) has analogical characteristic with Gnutella network (Figure 6) except for the relatively smaller region of distribution of ER model.

In analogy with Gnutella network, in Figure 13, measurements are plotted with the same parameters as used in Figure 7. The figures (a), (b) and (c) in Figure 13 are the measurements about ER model, and (d), (e) and (f) illustrate EBA model. As shown in the figure, the response of EBA model under failures and attacks is similar to Gnutella network, except for larger values of critical point, such as in Figure 13(d), $S \simeq 0$ when $f > 0.24$ and $f > 0.22$, and in Figure 13(e),C_c reaches the lowest point when $f = 0.11$ and $f = 0.09$ under DAttack and BAttack, respectively. These values are all larger than the counterparts in Figure 7. However, the response of ER model to failures and attacks is rather different. Failures and attacks have almost equal impact on the network structure (the overlapping curves as shown in Figure 13 (a),(b),(c)), since all nodes in ER model have approximately the same number of connections. Comparing (b) and (e), (a) and (d) in Figure 13, we find that the response to attacks in EBA model is faster than the response in ER model: with smaller critical points than those for ER model. The insets in Figure 13 are magnifications in the early stages of attacks, which are plotted using comparable range of x axis with Figure 7. As shown in these insets, all curves decay much slower than those plotted in Figure 7 within the same interval. A remained problem in the previous section is that the response to degree-based attack and betweenness-based attack in Gnutella network is much different from those in ER and EBA models. In ER and EBA models, betweenness-based attacks are more harmful than degree-based attacks (curves change faster), while it is contrary in Gnutella network. The main reason for this discrepancy is that in Gnutella network, leaf peers are most dependent on ultrapeers, which causes the formation of many structured subgraphs in the network, and these subgraphs at last compose a more structured topology of the whole net-

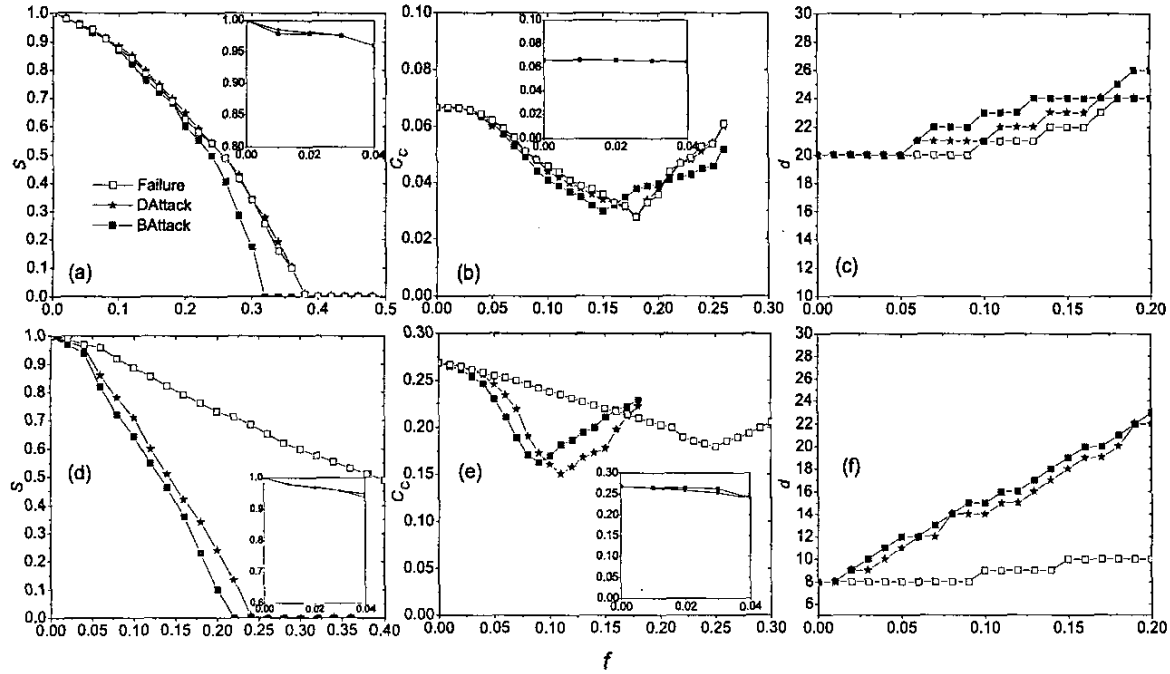

Figure 13: Results for random failures (open square), degree-based (star) and betweeness-based (filled square) attacks of nodes measured by the relative size of largest cluster S, the closeness centralization C_c and diameter d as functions of the fraction of removed nodes f in ER and EBA models. Insets are magnifications in the early stage of attack.

work.

5 Conclusions

In this paper, we have introduced some topological properties of Gnutella network. The investigation about these properties is of great importance to the applications running on top of it. We have shown that Gnutella network exhibits small-world effect and power-law degree distribution. Furthermore, centrality and network resilience are also studied in detail. The results of comparison between Gnutella network and theoretical models show that the understanding of P2P network still need more real data analysis and theoretical modeling.

References

[1] L. Adamic, R. Lukose, A. Puniyani, and B. Huberman, "Search in Power-Law Networks", *Phys. Rev. E*, Vol.64, 2001.

[2] E. Adar and B. Huberman, "Free riding on Gnutella", *First Monday*, Vol.5-10, 2000.

[3] R. Albert, H. Jeong and A. L. Barabasi, "Attack and error tolerance of complex networks", *Nature*, Vol.406, pp.378-382, 2000.

[4] R. Albert and A. L. Barabasi, "Topology of evolving networks: local events and universality", *Phys. Rev. Lett*, Vol.85, pp.5234-5237, 2000.

[5] R. Albert and A. L. Barabasi, "Statistical mechanics of complex networks", *Rev. Mod. Phys*, Vol.74, pp.47-97, 2002.

[6] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks", *Science*, Vol.286, pp.509, 1999.

[7] B. Bollobas, *Random Graphs*, Academic Press, New York, 2nd ed, 2001.

[8] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, and R. Stata, "Graph structure in the web", *Computer Networks*, Vol.33, pp.309-320, 2000.

[9] M. Faloutsos, P. Faloutos, and C. Faloutsos, "On Power-law Relationships of the Internet Topology", *Computer Communications Review*, Vol.29, pp.251-262, 1999.

[10] L. Freeman, "Centrality in Social Networks: A Conceptual Clarification," *Social Networks*, Vol.1, 1979.

[11] M. E. J. Newman, "The structure and function of complex networks", *SIAM Review* , Vol.45, pp.167-256, 2003.

[12] M. Ripeanu, I. Foster, and A. Lamnitchi, "Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System", *J. Internet Computing*, 2002.

[13] S. Saroiu, K. P. Gummadi, and S. D. Gribble, "Measuring and analyzing the characteristics of Napster and Gnutella hosts", *Multimedia Systems*, Vol.9, pp.170-184, 2003.

[14] M. A. Jovanovic, "Modeling Large-scale Peer-to-Peer Networks and a Case Study of Gnutella", Master thesis, Department of Electrical and Computer Engineering, University of Cincinnati, 2000.

[15] S. Milgram, "The small-world problem", *Psychology Today*, Vol.1, pp.62-67, 1967.

[16] http://www.limewire.org/