Examining Graph Properties of Unstructured Peer-to-Peer Overlay Topology

Chao Xie*, Sijie Guo†, Reza Rejaie‡, and Yi Pan*
*Georgia State University, USA ‡University of Oregon, USA †Huazhong Univ of Sci & Tech, China

***Abstract*— During the past few years, unstructured peer-to-peer (P2P) file-sharing systems have witnessed a significant increase in popularity. However, there lacks a systematic study on graph properties of the overlay topology. In this paper, we use accurate snapshots of the Gnutella overlay that span over roughly three years to explore changes in graph properties over long timescale. We investigate the effect of *network address translation* (NAT) on topology analysis. We examine a wide spectrum of graph properties characterizing the Gnutella top-level overlay topology and illustrate some interesting results. We find that the connection limit plays an important role in forming the unstructured overlay topology.**

I. Introduction

During recent years, unstructured peer-to-peer (P2P) file-sharing systems have experienced a dramatic increase in popularity and contributed a significant portion of the total Internet traffic. Examining graph properties of unstructured P2P systems is of great importance. First, we can gain detailed insight into the nature of the underlying system. Moreover, graph properties can affect performance of network algorithms and possibly performance of the networks. For the more, studying graph properties can facilitate generation of accurate artificial topologies for simulation.

The Gnutella network has been extensively examined during the past few years because of its wide deployment and huge user populations. Earlier research [1] [2] has studied characteristics of Gnutella overlay topologies. These previous studies were conducted more than 4 years ago on much smaller user populations. Moreover, they used either partial or distorted snapshots of the Gnutella network and thus the accuracy of their results was significantly affected [3]. While recent work [4] [5] studied the accurate snapshots, they conducted only a limited number of analysis. A complete picture of the Gnutella overlay strongly requires a systematic study on the graph properties. Our paper tries to fill this gap by examining the overlay topology from an array of perspectives.

In this paper, we use three accurate snapshots of the Gnutella overlay that span over roughly three years to explore changes in graph properties over long timescale. We investigate the effect of *network address translation* (NAT) on topology analysis. We examine a wide spectrum of graph properties characterizing the Gnutella top-level overlay topology illustrate some interesting results. To our knowledge, our work is the only study that examines such diverse and comprehensive metrics. Our results show that the connection limit plays an important role in forming the unstructured overlay topology. Due to the limited space, we limit the explanation of the implications of our findings and the potential underlying causes and explore these issues as future work.

TABLE I
Basic Top-level Statistics of Our Gnutella Snapshots

Basic Stat.	10/18/04	02/02/05	05/29/06
n (num. of nodes)	122,776	161,668	409,027
m (num. of edges)	1,317,702	1,940,183	5,264,680
$\langle k \rangle$ (avg. degree)	21.47	24.00	25.74
$\langle c \rangle$ (node clustering coef.)	0.02197	0.01781	0.01613
$\langle \kappa \rangle$ (avg. coreness)	12.89	15.20	16.63
$\langle d \rangle$ (avg. distance)	3.904	4.005	4.263
$\langle \varepsilon \rangle$ (avg. eccentricity)	6.932	7.481	6.950
$\langle b \rangle$ (avg. node betweenness)	301,030	404,542	1,076,450

Our study focuses on the top-level topology of the Gnutella network. Modern Gnutella network adopted a two-tier overlay topology, which divides peers into two groups, *ultrapeers* and *leaf peers*, as shown in Figure 1. There also exist some peers called *legacy peers*, employing the old version of Gnutella network. The ultrapeers and legacy peers form the *top-level* overlay, while the *bottom-level* overlay merely comprises leaf peers. The top-level constitutes the backbone and is the most influential part of the whole network. Therefore, we concentrate on the characteristics of the top-level (ultrapeer) overlay in this paper.

With the Gnutella crawler *Cruiser* [3], we have captured more than 20,000 snapshots of the Gnutella network from 2004 to 2006. In this paper, we examine three snapshots, namely 10/18/04, 02/02/05 and 05/29/06, with each snapshot selected from one different year. Modeling peers by vertices and connections between peers by edges, we treat these snapshots as undirected graphs. In this paper, terms *peer*, *ultrapeer*, *vertex* and *node* are interchangeable. Table I presents basic statistics about the three snapshots.

The rest of this paper is organized as follows. Section II investigates the effect of NAT boxes on overlay topology analysis. In section III, we illustrate the set of graph properties derived from our snapshots. Finally, Section IV concludes our work.

II. Effect of NAT Boxes

Figure 2 plots the *node degree distribution* of the Gnutella network, i.e., the Probability Distribution Function (PDF) of node degree $p(k) = n(k)/n$, where $n(k)$ is the total number of k-degree nodes. The node degree distributions for all three snapshots look similar. The degree distributions have two distinct segments around a spike in degree of 30, which coincides with the results of [4]. This is because the top-level *connection limit* of common Gnutella client softwares, such as Limewire and BearShare, is set to be 30 by the software designers. The nodes with degree lower than the limit are in a

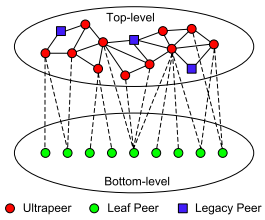


Fig. 1. Gnutella's two-tier overlay topology.

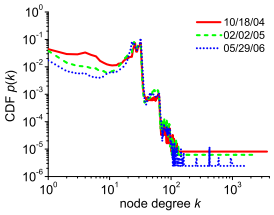


Fig. 2. Node degree distribution.

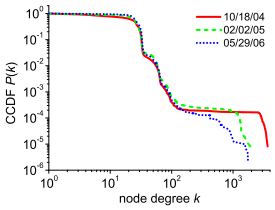


Fig. 3. CCDF of node degree.

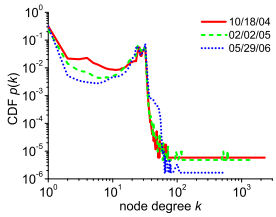


Fig. 4. Node degree distribution, with consideration of NAT.

transitive stage, trying to establish more connections and thus achieve more efficient file-sharing. We say these nodes are in the *flux state*.

The Complementary Cumulative Distribution Function (CCDF) of node degree $P(k) = \int_k^\infty p(k)dk$ in Figure 3 exhibits a sharp drop for node degrees larger than 30, which means that nodes with large connectivity barely exist. When a node reaches the connection limit, it cannot open more connections. We say the nodes that achieve the connection limit are in the *saturation state*.

In both Figure 2 and Figure 3, we can observe a small number of nodes with degree larger than the connection limit. We say these nodes are in the *supersaturation state*. One may instinctively presume that these nodes represent the rare users who had modified their client software and set the top-level connection limit to higher values.

However, the unusual peak in Figure 2 in degree interval [40,60] at the decreasing tail seems odd and requires further explanation. In constructing the snapshot graphs from the crawled data, we merely use IP address to identify different nodes. After a carefully inspection on the crawled data, we find that there exist some response messages from identical IP address but different port numbers. This exhibits the trademark of NAT, a technique that enables a local-area network (LAN) to use one set of IP addresses for internal traffic and a second set of addresses for external traffic. If several peers hide behind a NAT device, they may be treated as one node with the NAT device's external IP. This special node takes the NAT-level degree, which is the sum of the degrees of all peers behind the NAT. With this in mind, we reconstruct the snapshot graphs using IP address plus port number as the node identifier, in order to penetrate the NAT and treat each peer behind the NAT device as an individual node. Figure 4 gives the node degree distribution of these reconstructed graphs. As we expected, the peak in the decreasing tail disappears.

We define the *size of a NAT box* as the number of peers within the NAT box and analyze the CCDF of NAT box size in Figure 5. Most NAT boxes (more than 90%) have size less than 4. These NAT boxes represent typical home users who own up to 3 computers. There also exist a small number of comparatively huge NAT boxes with size of several hundreds to several thousands, which represent large organizations with many users connecting to the Gnutella network through a single NAT device. We further plot the average NAT-level degree as a function of the NAT size in Figure 6. The average NAT-level degree increases with the augment of the NAT box size. This observation implies that most huge-degree nodes in Figure 2 are actually the huge NAT boxes composed of many peers, while each of these peers maintains only a moderate number of connections. Note that Figure 4 reveals there do exist a few individual peers with huge degree. We speculate that these peers are Internet hubs powered by super-servers.

Hereto we have constructed two different kinds of snapshot graphs from the crawled data: one takes NAT into consideration while the other does not. Which kind of graph is more important? We believe that the latter should deserve more concerns. On the one hand, because the NATed peers are resident in a LAN, routing and resource discovering among them is kind of easy. On the other hand, NATed peers are protected from the external network and LAN connections are comparative stable and robust, while the NAT device is exposed to the wide-area networks and are vulnerable to attacks. If the NAT device is down by failures, all peers within the NAT box will be disconnected from the Gnutella network. Thus it is well-founded to count a NAT box as a single node and study the NAT-level degree rather than the degree of individual NATed peers. Therefore, we will focus on the snapshot graphs without consideration of NAT in the rest of this paper.

III. Graph Properties of the Top-level Overlay

In this section, we quantitatively analyze the top-level topology of the Gnutella network in terms of various graph metrics that have been found important in previous networking literature, covering both theoretic quantities and statistical properties. Although these metrics are not complete, they are sufficiently diverse and comprehensive for understanding and evaluating network topologies [6] [7] [8] [9].

A. Degree Distribution

The node degree distribution may be the most frequently used graph property. The observations [10] [11] that the node

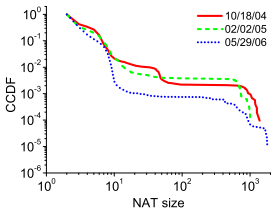


Fig. 5. CCDF of NAT size.

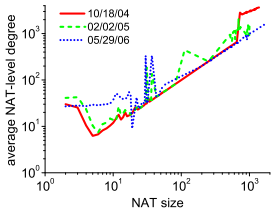


Fig. 6. Average NAT-level degree as a function of NAT size.

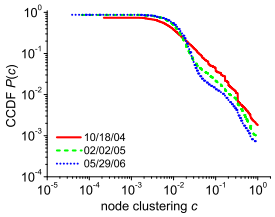


Fig. 7. CCDF of node clustering.

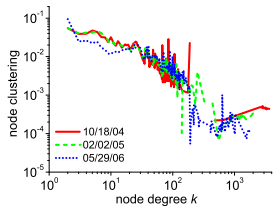


Fig. 8. Average c of k-degree nodes.

degree distributions in the World Wide Web (WWW) induced graph and the AS-level topology of the Internet follow power laws stimulated further interest in network topology research. Earlier studies [1] [2] observed that the degree distribution of the Gnutella network exhibited power law. However, [4] argued that these earlier observations were incorrect characterizations resulted from measurement error and the degree distribution of the Gnutella network does not follow a power law. We confirm the claim as it is shown in Figure 2 and Figure 3.

Comparing the plots in Figure 2 at the flux state, we observe that the degree distributions are decreasing in chronological order (10/18/04>02/02/05>05/29/06). This result is expected since there is a transition underway as Internet households move from narrowband connections to broadband connections [12]. More and more Internet households are shifting from slower dial-up service to faster high-speed connection technologies: DSL and cable modem. With broader bandwidth, a node can establish connections more efficiently and reach saturation state more quickly, thus reduce its stay in the flux state.

Table I presents the average degree $\langle k \rangle = \sum_k kp(k)$. The average degree increases in chronological order (21.47<24.00<25.74). This resulted from the above reasons as well. We could reasonably conjecture that $\langle k \rangle$ might continue to increase with the technology advancement for the near future. We further observe that the growth speed of $\langle k \rangle$ is declining. From 10/18/04 to 02/02/05 (4 months) $\langle k \rangle$ is increased by 2.53 while from 02/02/05 to 05/29/06 (15 months) it is increased by only 1.74. $\langle k \rangle$ is up-bounded by the top-level connection limit and its increase is hampered by the bound. If the Gnutella software designers increase the connection limit, a prompt rise in $\langle k \rangle$ could be expected.

We can find in Figure 3 that only 5% of nodes have degree higher than 33. This means that only a small amount of Gnutella users have modified their client software and set the top-level connection limit to higher values. Moreover, we find that the 05/29/06 graph has the smallest amount of maximum node degree and the 10/18/04 graph has the largest amount maximum node degree. This implies that the size of the large NAT boxes tends to decrease, which reflects the reconstruction of network infrastructure.

B. Clustering

A common feature of many complex networks is clustering. *Node clustering* is a measure of how well the neighbors of a given node are locally interconnected. Node clustering c_i is defined as the ratio between the number of edges N_i among the neighbors of a node i of degree k_i and the maximum number of possible edges $k_i(k_i-1)/2$. The *node clustering coefficient* of the whole network $\langle c \rangle$ is the average of c_i over all nodes in the network. Clustering expresses local robustness of the network. Newman [13] found that virus outbreaks spread faster in highly clustered networks, although outbreak sizes are smaller.

We first observe in Table I that the node clustering coefficient $\langle c \rangle$ decreases in chronological order. This is very interesting and may be a result of better connectivity heuristics. Therefore, the virus outbreaks, if any, spread slower in the Gnutella network.

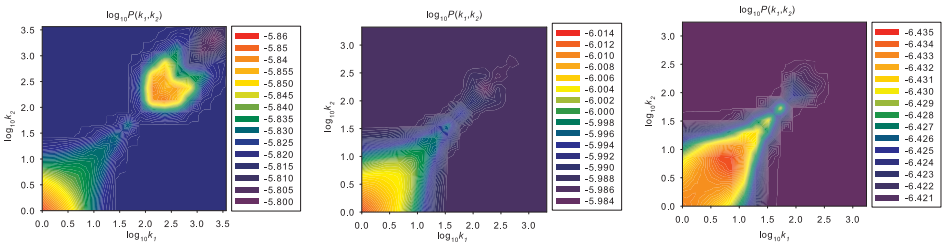
The CCDF plot of node clustering is shown in Figure 7. The tails of the plots decrease more rapidly in chronological order, which is another evidence that the neighborhood of a node becomes less interconnected. Moreover, these tails show scale-free properties, following a power law distribution with exponent 1.5.

Figure 8 shows average node clustering of k-degree nodes. The plots have similar shapes where nodes with lower degrees have higher node clustering than those with higher degrees. We speculate that there are local clusters sparsely interconnected by global backbones. The local clusters are probably formed by nodes with small connectivity but large node clustering. The highly connected nodes act as a bridge and connect to different local clusters. Because these highly connected nodes connect to nodes in different local clusters which are not interconnected, they have small node clustering.

Note that the plots in Figure 8 at supersaturation state decline sharply. This results from the connection limit. Although there are a small number of nodes that modified their client software, most of their neighbors do not and are bounded by the top-level connection limit. Therefore, these neighborhoods are loosely interconnected.

C. Assortativity

The network community has recently begun recognizing the importance of the correlations between the degrees of connected nodes. According to the behavior of the *degree-degree correlation*, the complex networks can be grouped into three types, namely *assortative*, *disassortative*, and *neutral mixing* [14]. A node with large degree in the assortative (disassortative) network tends to connect to nodes with large



(a) 10/18/04 (b) 02/02/05 (c) 05/29/06

Fig. 9. The contour plot of the logarithm of the joint degree distribution $P(k_1, k_2)$.

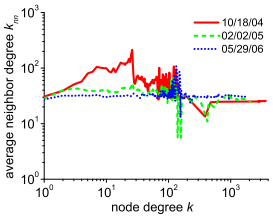


Fig. 10. k_{nn} of k-degree nodes.

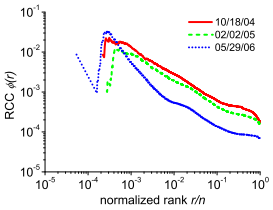


Fig. 11. Rich club connectivity.

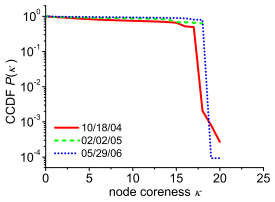


Fig. 12. CCDF of node coreness.

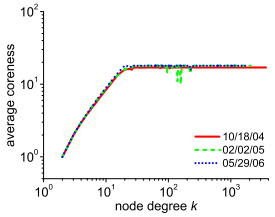


Fig. 13. Average coreness of k-degree nodes.

(small) degree, while for the network of neutral mixing, there are no such clear tendencies.

One metric of degree-degree correlation is *joint degree distribution* [7]. Joint degree distribution (JDD) is the probability that a randomly selected edge connects nodes of degree k_1 and k_2 respectively, $P(k_1, k_2) = m(k_1, k_2)/m$, where $m(k_1, k_2)$ is the total number of edges between k_1- and k_2-degree nodes.

The degree-degree correlation can also be investigated in terms of the average degree of the neighbors of k-degree nodes, $k_{nn} = \sum_{k'} k' p(k'|k)$, where $p(k'|k)$ is the conditional probability that a k-degree node is connected to a k'-degree node [15]. For the assortive (disassortative) networks, k_{nn} increases (decreases) with increasing k, while the neutral networks, k_{nn} is independent of k.

Disassortative networks are vulnerable to both random failures and targeted attacks [14]. Other metrics, such as *likelihood* [16], *radial*, and *tangential* [17], are also directly related to assortativity. In this paper, we use *tangential links* to refer the links connecting nodes of similar degrees and use *radial links* to refer the links connecting nodes of different degrees.

We plot the joint degree distribution in Figure 9. Obviously, all three graphs are disassortative because they exhibit small JDD values along the anti-diagonal. All graphs have small frequency of tangential links interconnecting low-degree nodes (the bottom-left corner). This finding is intuitively reasonable. These nodes are either in the flux state, working on opening more connections, or in the saturation state, restricted by the connection limit. The most frequent links are either radial (bottom-right and top-left corners) or high-degree tangential (the top-right corner) because the connection limit do not apply to high-degree nodes in the saturation state. Furthermore, we observe that the area constituted by tangential links interconnecting low-degree nodes expands in chronological order. With more widely adopting of broadband technologies, nodes can establish connections more efficiently and reach the connection limit more quickly. Consequently the boundary of the above area extends.

Note, however, that there is one exceptional area (the upper-right area) in Figure 9(a) which exhibits a small frequency of tangential links. If we look back at Figure 3, we can observe that, compared with the other two plots, the 10/18/04 plot has a visible gap between degree 100 and degree 2000. For the 10/18/04 snapshots, there are few nodes in this range, which certainly leads to the sparsity of links in the above area.

Figure 10 plots the average nearest neighbor connectivity k_{nn} as a function of node degree k. If we focus on the nodes in the flux state and the saturation state, we observe that these three graphs are getting less assortative in chronological order and the 05/29/06 graph is almost neutral. Therefore, the Gnutella network may become less resilient to random failures and targeted attacks.

D. Rich Club Connectivity

Rich club connectivity reveals information about how tightly the high-degree nodes are interconnected to each other. Many networks contain a small number of high-degree nodes, which are called "rich" nodes. The set of these nodes are called the *rich club* [18]. Let node rank r denote the position of a node in a list sorted by decreasing degree. The rich club is defined as the set of the first r nodes in the ranked list. The *rich club connectivity* (RCC) $\phi(r/n)$ is the ratio of the number of links interconnecting the club numbers over the maximum possible number of links $r(r-1)/2$, where r/n is the rank normalized by the total number of nodes in the network.

Figure 11 shows RCC as a function of the normalized node rank. It is interesting that highest ranks do not result in largest RCCs. This observation implies that a few top ranked nodes, which have exceptional high degrees, do not tightly interconnect with one another. Due to the lack of global

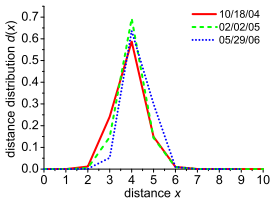


Fig. 14. Distance distribution.

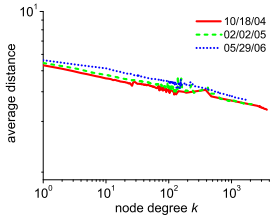


Fig. 15. Average distance from k-degree nodes.

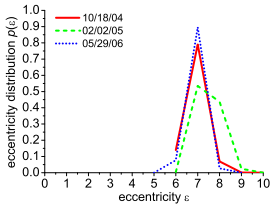


Fig. 16. Eccentricity distribution.

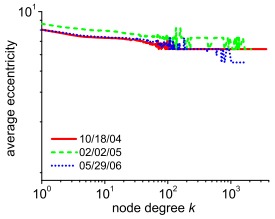


Fig. 17. Average eccentricity from k-degree nodes.

information of the whole Gnutella network, these nodes are not acquainted with each other and thus not interconnected. RCC increases with the growth of rank r and approaches a peak. In this stage, the addition of more club members enriches the interconnections in the club. These new club members play a pivotal role in keeping the whole network connected. After the peak, RCC declines with the growth of rank r. In this stage, those lately enrolled club members fail to bring significant amount of interconnections and the denominator $r(r-1)/2$ tends to dominate the RCC value.

E. Coreness

The k-core of a graph is the subgraph obtained by recursively removing all nodes of degree less than k from the original graph [19]. If a node belongs to the k-core but not to the $(k+1)$-core, we say that the *node coreness* κ of this node is k. Nodes with degree 1 have $\kappa = 0$. The node coreness contains information about how "deep" in the core the node is. It is related to the degree of nodes but is a more sophisticated measure than the node degree. A node with small coreness is not well connected and can be disconnected easily by removing its poorly connected neighbors, even if its degree is high.

Figure 12 plots the CCDF of node coreness. It shows that more than 70% of nodes have coreness larger than 13. Most nodes in the Gnutella network tend to dive deeply into the network core and thus are well connected. This implies that the Gnutella network successfully attain a well-connected structure by itself. In Table I, the average coreness increases in chronological order, which is the inevitable result of the growth of average degree $\langle k \rangle$.

Figure 13 plots the average node coreness of k-degree nodes. In the flux state, the node coreness increases with the growth of node degree. In this stage, a node pushes into the core while it opens more connections. Interestingly, when the node reaches the saturation state, its coreness seems saturated as well. For nodes that successfully attain the supersaturation state, increasing node degree does not increase coreness. We claim that the coreness is also confined by the connection limit potentially.

F. Distance

The *distance distribution* or the shortest path length distribution $d(x)$ is the probability that a random pair of nodes are at a distance of x hops from each other, i.e., the number of pairs of nodes at a distance x divided by the total number of pairs n^2 of the graph. The average distance $\langle d \rangle$ is the shortest path length between two nodes, averaged over every pair of nodes in the network. Another metric *eccentricity* is an extreme form of distance. Eccentricity of a node is the maximum distance from the node to other nodes. Eccentricity ε_i of node i is defined as $\varepsilon_i = max(d_{ij}), \forall j$, where d_{ij} is the distance between nodes i and j. Distance distribution is important for evaluating the routing algorithms, whose performance parameters depend mostly on the distance distribution [20].

Figure 14 shows the distance distribution. The distance distributions of all three snapshots are almost identical. The majority (60%) of shortest paths are 4-hop paths. We also observe in Table I that the average distance slightly increases during the three year period over which the network size has more than tripled. This means that the Gnutella network maintains the small-world property and is highly scalable.

Figure 15 plots the average distance from k-degree nodes. One important observation is that the average distance as a function of the node degree shows an evidently stable power law with exponent 0.07 in the three year period. Moreover, these plots increase in chronological order (10/18/04<02/02/05<05/29/06) in the full range of node degrees. This result is expected because the size of the Gnutella network keeps increasing.

Figure 16 shows the eccentricity distribution. The largest eccentricity of all graphs is 10 hops, while the smallest eccentricity of the 05/29/06 graph is one hop shorter than that of the other two graphs. More than half (53%) of the nodes have eccentricity 7. An interesting observation in Table I is that the 02/02/05 graph has the largest average eccentricity.

Figure 17 plots the average eccentricity from k-degree nodes. For nodes in the flux state, the average eccentricity decreases slightly as the node degree increases. For nodes in the supersaturation state, the average eccentricity remains nearly stable.

G. Betweenness

Betweenness is the most frequently used measure of centrality. Betweenness of a node i, b_i, is defined as the total number of shortest paths between pairs of nodes that pass through node i. Betweenness is an indicator of who the most influential nodes in the network are, the ones who control the flow of traffic between most node pairs. The nodes with highest betweenness also result in the largest increase in typical distance between others when they are removed. Metrics such

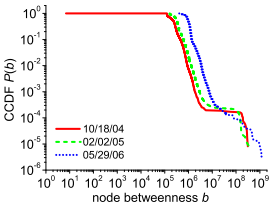


Fig. 18. CCDF of node betweenness.

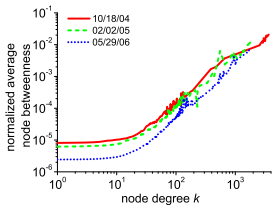


Fig. 19. Normalized average node betweenness of k-degree nodes.

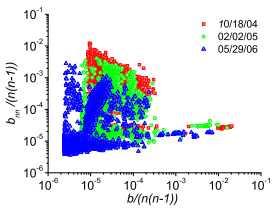


Fig. 20. Normalized b_{nn} of nodes with normalized betweenness b.

as *router utilization* [16] and *link value* [21] are directly related to betweenness.

Figure 18 shows the CCDF of node betweenness. It follows a power law distribution of exponent 2 with an exponential cutoff in the tail. From Table I, we observe that the average node betweenness increases considerably. This is not surprising because of the significant augmentation of the network size over the three years.

As shown in Figure 19, the normalized average node betweenness of k-degree nodes is nearly identical for the three years and follows a growing power law function with exponent -1.4. Higher node degree leads to greater path diversity and, hence, aggrandizes node betweenness for these nodes.

One may think that the *betweenness-betweenness correlation* [22] would also be worthy of investigation. A simple and clear measurement is to compute the average betweenness b_{nn} of the neighbors of the nodes with a given betweenness b. We plot normalized $b_{nn}/(n(n-1))$ as a function of normalized node betweenness $b/(n(n-1))$ in Figure 20. It is interesting that the distribution of each snapshot graph is confined to a triangle area with a descending hypotenuse and an ascending base. There are some low betweenness nodes with high betweenness neighborhood. It implies that the mean influence of neighbors of a node is almost independent of the influence of the centered node.

IV. Conclusions

In this paper, we examine the Gnutella top-level overlay topology over roughly three years. In particular, we investigate the effect of NAT boxes on graph analysis and study an array of interesting graph properties of the Gnutella top-level overlay topology. We illustrate the importance of the connection limit in shaping the unstructured overlay topology. Due to the limited space, we limit the explanation of the implications of our findings and the potential underlying causes and explore these issues as future work.

References

[1] M. A. Jovanovic, "Modelling large-scale peer-to-peer networks and a case study of gnutella," Master's thesis, University of Cincinnati, Cambridge, June 2000.

[2] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, "Search in power-law networks," *Physical Review E*, vol. 64, pp. 46 135–46 143, 2001.

[3] D. Stutzbach and R. Rejaie, "Capturing accurate snapshots of the gnutella network," in *Proc. IEEE Global Internet'05*, Mar. 2005, pp. 2825 – 2830.

[4] D. Stutzbach, R. Rejaie, and S. Sen, "Characterizing unstructured overlay topologies in modern p2p file-sharing systems," in *Proc. ACM/USENIX IMC'05*, 2005, p. 49C62.

[5] A. H. Rasti, D. Stutzbach, and R. Rejaie, "On the long-term evolution of the two-tier gnutella overlay," in *Proc. IEEE Global Internet'06*, Apr. 2006.

[6] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, "Large-scale topological and dynamical properties of the internet," *Phys. Rev. E*, vol. 65, no. 6, p. 066130, Jun 2002.

[7] P. Mahadevan, D. Krioukov, M. Fomenkov, X. Dimitropoulos, k c claffy, and A. Vahdat, "The internet as-level topology: three data sources and one definitive metric," *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 1, pp. 17–26, 2006.

[8] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, "Systematic topology analysis and generation using degree correlations," in *Proc. ACM SIGCOMM'06*, New York, NY, USA, 2006, pp. 135–146.

[9] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas. (2006) Characterization of complex networks: A survey of measurements. [Online]. Available: http://arxiv.org/abs/cond-mat/0505185

[10] R. Albert, H. Jeong, and A.-L. Barabasi, "The diameter of the world wide web," *Nature*, vol. 401, pp. 130–131, 1999.

[11] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," in *Proc. ACM SIGCOMM'99*, New York, NY, 1999, pp. 251–262.

[12] (2006) United states of america internet usage and broadband usage report. [Online]. Available: http://www.internetworldstats.com/am/us.htm

[13] M. E. J. Newman, "Properties of highly clustered networks," *Phys. Rev. E*, vol. 68, no. 2, p. 026121, Aug 2003.

[14] ——, "Mixing patterns in networks," *Phys. Rev. E*, vol. 67, no. 2, p. 026126, Feb 2003.

[15] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, "Dynamical and correlation properties of the internet," *Phys. Rev. Lett.*, vol. 87, no. 25, p. 258701, Nov 2001.

[16] L. Li, D. Alderson, W. Willinger, and J. Doyle, "A first-principles approach to understanding the internet's router-level topology," in *Proc. ACM SIGCOMM'04*. New York, NY, USA: ACM Press, 2004, pp. 3–14.

[17] L. Tauro, C. Palmer, G. Siganos, and M. Faloutsos, "A simple conceptual model for the internet topology," in *Proc. Global Internet*, San Antonio, Texas, USA, November 2001.

[18] S. Zhou and R. J. Mondragn, "The rich-club phenomenon in the internet topology," *IEEE Communication Letters*, vol. 8, pp. 180–182, 2004.

[19] M. Gaertler and M. Patrignani, "Dynamic analysis of the autonomous system graph," in *2nd International Workshop on Inter-Domain Performance and Simulation (IPS 2004)*, 2004, pp. 13–24.

[20] D. Krioukov, K. Fall, and X. Yang, "Compact routing on internet-like graphs," in *Proc. IEEE INFOCOM'04*, 2004.

[21] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "Network topology generators: degree-based vs. structural," in *Proc. ACM SIGCOMM'02*. New York, NY, USA: ACM Press, 2002, pp. 147–159.

[22] K.-I. Goh, E. Oh, B. Kahng, and D. Kim, "Betweenness centrality correlation in social networks," *Phys. Rev. E*, vol. 67, no. 1, p. 017101, Jan 2003.