Analysis of Topological Properties for IP-Networks in Norway

Tommy Kvalvik
Molde College, 6402 Molde, Norway
Tommy.Kvalvik@hiMolde.no

Judith Molka-Danielsen[1]
Molde College, 6402 Molde, Norway
J.Molka-Danielsen@hiMolde.no

Abstract

We investigate the Norwegian IP-networks connected to the Internet based on recent research in scale free networks and random network topology models. The network has been probed with a simple Traceroute tool and several key metrics have been calculated and presented for making conclusions both about the network and the chosen research methodology. In addition to the established metrics we introduce a new metric to assess the tree-likeness of the explored network.

1. Introduction

The Internet has grown to become the largest computer network in existence today, millions of nodes like personal workstations, laptops, web servers, mobile devices interconnect together through a very complex infrastructure of high-speed links and routers. This infrastructure consists of smaller sub-networks governed by a large amount of independent Internet service providers that use the IP protocols maintained by IETF (Internet Engineering Task Force) to connect their networks together to form a single world-wide IP network. Both new and migrating legacy services and applications are being implemented. Without understanding the topology of the Internet it will become increasingly hard to sustain growth, maintain robustness and availability and create new services and tools that operate effectively as the demand for scalability explodes [1]. This research contributes to such understanding. We explore the topological properties of the Norwegian IP-networks. The IP protocols are engineered with redundancy in mind, but the general perception is that this has not been exploited in the Norwegian networks. Rather sub-networks connect to each other by a single hub, the NIX, Norwegian Internet Exchange, maintained by Uninett. The result is a 'snowflake' topology with the NIX routers in the middle. An exploration of the network should support this perception, but may also reveal alternate paths between ISPs and be helpful when trying to assess the importance of the core routers and other central structures. There has been former research in measuring and modelling the Internet after random processes and models. This research will take a macro view of the Norwegian IP networks; disregarding individual networks, tier based architectures and deterministic processes and seeks if there can be found some order in the chaos.

The report is structured as follows: In section 2 we describe related research, metrics and models. Section 3 describes the research methodology used and the chosen probe

[1] This paper was presented at the NIK-2006 conference. For more information see //www.nik.no/.

method. Section 4 interprets results. Results and problems are discussed in more detail in section 5. Section 6 contains concluding remarks and proposed future work.

2. Related Research, Metrics and Models

CAIDA in their work for promoting Internet research have developed lists of problems that are without good solutions [2]. We focus on some of the problems they have listed. These are: *cannot figure out where an IP address is*; *cannot measure topology effectively in either direction, at any layer; cannot get router to give you all available routes, just best routes and cannot trust whois registry data.* Research with Internet mapping began with, the Internet mapping project started in 1998 has been gathering Internet topology data by a random Traceroute-style method from a single location ever since [3]. Their work is a common data source for much of the theoretical work being done on modelling.

Metrics for Characterizing Networks

The *total number of links and node degree* are the simplest way of describing the connectivity in a network. These are the number of links in the network (*m*). The degree, k_i of a node, v_i is the number of links connected to it, average degree, $\bar{k}$ is the average number of links connecting all the nodes in the graph. Higher values for *m* and $\bar{k}$ indicates a better connected network and is likely more robust. The *degree distribution* used by mathematician Albèrt-Làzlo Barabási and a group of scientists and students developed a web-crawler to map the connectedness of the web [4], they discovered that the degree of the web pages on the web was not as evenly distributed as one would expect from a random network but it had a large number of low degree nodes and a small number of high degree nodes acting as hubs. They found that the probability of a degree, *k*, is; $P(k) \propto k^{-\gamma}$ Where k is the degree and γ is the distribution exponent. This function is known as a power law or Pareto distribution. Other networks that have evolved through somewhat natural growth rather than being engineered will often have this property, the value for γ depends on the network and is usually in the range $2 < \gamma \leq 3$, but can be of any value. The Faloutsos brothers [5] discovered that the Internet, both at IP and AS level displays similar power laws with $\gamma = 2.48$ for the router level and $\gamma = 2.2$ for the AS level. Existence of such a power law would indicate a scale free network. The exponent γ is an important property when describing the connectivity of the network. The discovery, diversity and exploration of such networks are described by Barabási [6].

Other metrics are *clustering, local and mean local* [7]. A cluster in a graph is a group of nodes that are more highly connected to each other than to the rest of the graph. Clustering is important in several ways, firstly it is a good measure for validating models [8], and it is a practical expression for local robustness in the graph as a high local clustering indicates more links and high path diversity around that node. $\overline{m}_{nn}(k)$ is the clustering coefficient as defined in [9].

Definition: Let $\overline{m}_{nn}(k)$ *be the average number of links between the neighbors of k-degree nodes. The local clustering ratio is the ratio of this number to the maximum possible such links.*

The Internet AS level network has been shown to exhibit a rich club phenomenon in [2]. The rich-club is described as the connectivity between rich nodes. The interactive growth model [10] is designed to mimic the rich club found in the AS network. The rich club connectivity is considered an important metric for capturing network topology. This measures how close p induced sub graphs are to complete graphs. The most successful AS level topology model, the positive feedback preference (PFP) model [11], tries only to mimic three characteristics; (1) degree distribution, maximum degree and rich club connectivity. Capturing the RCC of the network provides perhaps the most useful metric for comparing it to the models. This metric is also an indication of how centralized the network is by indicating if the redundant links are concentrated around a few rich nodes or spread out across the network. Next, there is more than one definition of coreness, we use [12]. Relevant coreness related metrics are, firstly the core at each level and $k_{\max}$, the maximum k where the k-core is not empty. This metric should, like rich club, provide an indication of centralized the network is. High coreness is an indication of a high number of redundant links interconnecting a subset of the network, leaving the average degree of the rest of the nodes lower.

Several measures of *distance* exist, but we use two common ones: The average distance between nodes and the diameter. The Diameter is the maximum distance between any two nodes in the network.

Finally in this research we offer a new definition and use of the *tangential link ratio (TLR)*. This measure attempts to measure how close a network resembles a star or a tree. In a star or a tree each link will either take you closer to a destination or further away from it. One way of doing this will be to divide the links into radial links (connecting "vertically" between nodes of different distance to the destination) and tangential links (connecting "horizontally" between nodes of equal distance to the destination). In [7] the authors claim that their traceroute-like skitter program discover more radial links than tangential links since tangential links do not lie on any shortest path [13]. Their definition of radial and tangential links does however assume that high degree nodes populate the center of the network, while low degree nodes are on the circumference. We use a slightly different definition that does not make this assumption.

Definition: Let e_{ij} *be a link connecting node i and j, and* v_s *a start node. Then* d_{is} *is the distance from node i to the start node and* d_{js} *is the distance node j to the start node. If* $d_{is} = d_{js}$ *then* e_{ij} *is tangential to* v_s *, otherwise the link is radial.*

Counting the number of radial vs. tangential links should provide an indication of how tree-like the network is; the metric used is: *TLR*

Definition: Let T_s *be the subset of tangential links in E to a given* v_s*, the ratio of tangential links in the network will then be:*

$$TLR = \frac{\sum_{s=0}^{n} |T_s|}{n(m-(n-1))}$$

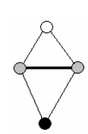

Figure 1 TLR ratio and Graph with a start node and one highlighted tangential link

Here $m-(n-1)$ is the maximum number of tangential links as there will always be $n-1$ radial links connecting the start node to every other node. The final measure is the ratio of tangential links compared to the maximum possible amount of tangential links. If a link is tangential or not will depend on our start node, so the measure is the average of all nodes as starting point.

Models

In this research we compare our explored network to several know network models. The first is the *random model (ER)* [14]. The simplest random model called $G(n, p)$ is simply generating *n* nodes and linking each pair together with probability *p*. This produces a graph with a set number of links (*n*) and $m = p(n*(n-1))$ links and the resulting degree distribution will still follow a Poisson distribution. The next examined model, is the *scale free model (SF)*. The scale free model was discovered and formulated by [15] and is a product of the power law discovered in the WWW network. While the vertices in the basic random network are generated all at once and wired up afterwards the basic scale free models add one vertex at a time and add a fixed number of random edges to wire it in with the rest of the network. [16] add one vertex at a time and wire it up with two random links to the existing network. Continued studies of scale free networks revealed that in many cases hubs have a larger proportion of the total links because when a new node is introduced to a network it will often prefer to connect to already high degree nodes. This is known as the "rich-get-richer" phenomenon. According to [4] prior models failed to account for growth and preferential attachment. The *preferential attachment (BA)* (Barabási-Albert) model introduces a simple probability scheme to accommodate this, links are not formed uniformly but the probability of a new node forming a link to node i is: $\prod(i) = \frac{k_i}{\sum_j k_j}$ is called the preference function of the model. The basic BA model generates networks with $\gamma = 3$. The same adjustments are made to this model as to the general scale free model to satisfy the average degree of the explored network. There is also a tendency among rich nodes to be connected to other rich nodes with a higher probability than then lower degree nodes. This causes the rich nodes to connect and form a highly connected network; the previously mentioned club phenomenon. An actual measurement of the AS-level shows that it has this feature. So does the BA model, but a lot less clearly so additional modifications are needed. *The general linear preference model (GLP)* [8] is a recently introduced extension to the BA model. This model grows the components in a slightly different order. The assumption is that a network does not evolve by only adding nodes, but new links between existing nodes as well. The model starts with a network with m_0 nodes and $m_0 - 1$ links. Each step one of two operations is performed: 1. $m < m_0$ new links are formed between chosen existing nodes. 2. One new node is added and connected to m existing nodes. New links are formed with a generalized linear preference function: $\prod(i) = \frac{k_i - \beta}{\sum_j (k_j - \beta)}$, $\beta < 1$ The parameter β adjusts how strongly new links will prefer high degree nodes, if set smaller high degree nodes will be preferred less, adjusting β and the probability of executing step 1 or 2 each iteration will modify γ for the network. This model fits the actual AS level network with regards to clustering coefficients, characteristic path length and small worlds. Real networks does however not grow randomly, but when new nodes are added traffic increases around the nodes it connects and new links are more likely to form in that area of the network. The *interactive growth model (IG)* [11] is designed to take this into account, it adds one new node to the network each step and connects it to existing nodes (hosts) using the BA preference function, then new links are added to the host nodes to balance the additional traffic simulating how a real network would grow.

3. Research Methodology

There are generally two sources of IP connectivity data. (1) The border gateway protocol tables obtained from a BGP (Border Gateway Protocol) server reveal routing tables from individual routers and (2) Traceroute like probing of individual nodes. Method 2 is the one used here and is the same one used in the Internet Mapping Project [3]and CAIDA's skitter program [7][13]. The method is chosen for its simplicity and ease of use as BGP lookups require more complex software. We extend our former research in Traceroute exploration [17].

The resulting output of Traceroute is a series of paths the echo requests took; the output is stored in a database with a timestamp. Each line in each trace is translated into a host and consecutive lines are links. All traces are then merged into one undirected, un-weighted graph that will represent the explored network. Traceroute tools need a destination host and since the scope of this work is limited to the Norwegian IP networks the addresses need to be in Norway. Finding the actual location of an IP resource is hard and not possible to do in an efficient way. The only feasible way of doing it for the volume of data needed here is to trust whois lookups and trace addresses registered to Norwegian registrars. The IP-to-Country database [18] is a compilation of all address blocks found in APNIC, ARIN, RIPE and LACNIC databases mapped to the country of their registrar. The database can be downloaded as an excel spreadsheet. The spreadsheet lists the addresses as a decimal encoded IP range with a corresponding country code. The list contained 649 ranges belonging to Norwegian registrars; the version of the database used was last updated 15. Dec 2005. Unfortunately an address is not necessarily located in the country its registered, so there will be traces that leave the country and end up in foreign networks, this requires us to filter results.

Technical problems with Traceroute exploration

Unfortunately finding a path with Traceroute is not completely reliable or accurate, a number of technical problems may result in incomplete or no path information at all, or in worst case data with errors. These problems can be caused by a number of factors like strict security policies, bad configurations and well known bugs in routing software. Some of the more common ones are mentioned in the BSD Traceroute manual [19] and our handling of them described in Kvalvik [20]. The biggest problem with Traceroute from a topology standpoint is that it is impossible to capture it all. Tangential links are especially hard to find and we rely on variations in routing tables to find them. A high number of traces will pass through the center of the network and fewer traces will venture along the fringes, this will skew the measured metrics. By intuition we can tell that the number of links and average degree will be lower than the real values. In the worst case the explored network will be a tree spanning from the start node, in which case $m = n - 1$, the minimum required links in a connected network. Higher number of links indicates more alternate paths and a higher detection rate of the network. Incomplete detection of links will also skew the other metrics, but how and how much is hard to tell. In [21] the authors argue that degree distribution is a very important measure in assessing the statistical accuracy of the sampled network. They conclude that the shape of the degree distribution function is fairly well sampled for tail heavy distributions although deviations should be expected, especially at lower degrees. This sounds like good news, the authors of [22] also claim that Traceroute sampling of scale free models reproduce degree distribution fairly well. They also conclude that Traceroute exploration of random networks (with Poisson degree distribution) also results in a measured power law, this is supported by [23]. Accurate calculations on what happens in a BA network when explored with Traceroute can be found in [24]. It supports the notion of accurate reproduction among high degree nodes, but low degrees are skewed and the measured γ is lower then the actual value.

If traces were stopped too early, it will probably make the discovered network too sparse. Both networks size and average degree should be stable towards the end. One should expect a reduction in detection rate once a decent portion of the network has been explored.

Running the Traceroute Program

The trace program itself is implemented as a java program using the Traceroute program shipped with all recent versions of MS Windows for tracing. Addresses are picked randomly from one of the 649 ranges in the IP-to-Country database [18]. Each trace is decomposed into nodes and links and is stored in a database with a timestamp and trace number for when it was first discovered. Traceroute was run with a timeout of 200ms, meaning that if a host does not reply within 200ms the trace is aborted. This is to limit the time spent on each individual trace so that if a host does not reply we abort and move on to hopefully more responsive hosts. Traces were also limited to a maximum of 30 hops to avoid flooding badly configured networks with echo requests when probing addresses not in use. To maximize the number of completed traces we need to run more than one trace at a time. 40 traces were run concurrently from January 21 to February 28 2006. Running more traces at the same time seemed to disturb the networking subsystem on the computer running the probe and cause Traceroute to crash. Total number of traces run by the end was 2 089 015.

4. Interpretation of Results

There were no expectations to the size of the network except the assumption that the growth would be fairly high early on and decline as more nodes and links will already be discovered. Number of alternate paths found and average degree is also expected to grow slightly over time. Figure 2 and 3 shows the explored network size during the trace.

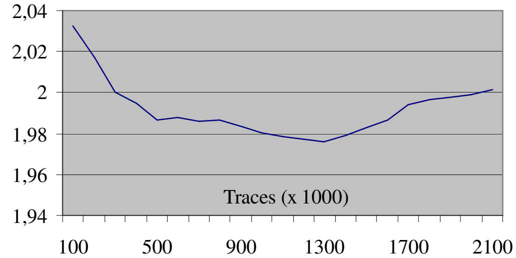

Figure 2 Discovered average degree as a function of number of traces.

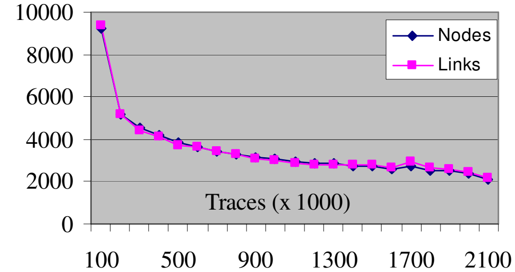

Figure 3 Newly discovered each 100 000 traces

Figure 3 shows the growth of the network every 100 000 traces. Note that these are unfiltered results with foreign links and unconnected fragments still in the network.

These numbers seem a bit strange at first, average degree seems below 2 for a long duration indicating that a lot of traces are incomplete leaving degree 0 nodes in the network. It increases a little during the end of the trace indicating more alternate paths are found form trace 1 500 000 and until the end. The final size of the unfiltered network is 72759 nodes and 72809 links resulting in a average degree of 2.003, just slightly more than a pure tree topology. The addresses we traced are registered to a Norwegian network, but not necessarily utilized there. A filtering strategy is needed to remove foreign nodes from the network. These nodes might belong to networks far away causing everything in between to appear as tendrils or very sparse pieces of network. The strategy for removing these nodes is described fully in [20]. The low degree is also an indication of high network stability.

In summary, target nodes from international backbone providers where chosen as destinations. Finding candidates for removal required some manual work and some of the nodes removed where identified as foreign based simply on their URL string. Others where looked up with a whois tool (http://www.radb.net/) and identified as foreign registrars. This method is by no means exact, but it did serve its purpose as we are trying to get rid of the sparse fringes not belonging in Norway rather than remove every single router not in the country. Routers belonging to the tier1 provider Level 3 were chosen as targets, one target was picked at a time and the first probable foreign node in its path is removed. The start node was molde-gw2.uninett.no. When a path is no longer found a new destination was chosen and the procedure repeated. The target was changed three times and a total of 23 nodes removed before all Level 3 routers where cut off. Manual inspection did not reveal any identifiable foreign nodes. Final size for the filtered network is 62804 nodes and 65768 links with an average degree of 2.1

Metrics for the network of the described size were recorded based on the developed network models and on the collected data of the explored network. These are summarized in Table 1 and described in the following section.

Metrics	Explored	ER	SF	SF-scaled	BA	BA-Scaled	GLP	IG
γ	1.82	N/A	4	4	2.6	2.6	3.4	2.5
C(k)	.0062	$.202\text{x}10^{-4}$	$.119\text{x}10^{-3}$	$.936\text{x}10^{-6}$	.0011	$.773\text{x}10^{-5}$	$.476\text{x}10^{-4}$	$.439\text{x}10^{-4}$
k-max	5	1	1	1	1	1	2	2
D	35	34	12	34	10	25	22	24
Avg-d	10.3	14.2	7.5	15	5.7	9.3	9.1	9.5
TLR	0.39	0.64	0.52	0.54	0.51	0.52	0.48	0.51

Table 1 Compared Metrics: Degree distribution exponent γ, mean local clustering C(k), k-max coreness for each network, diameter D, average distances Avg-d, and tangential link ratios (TLR).

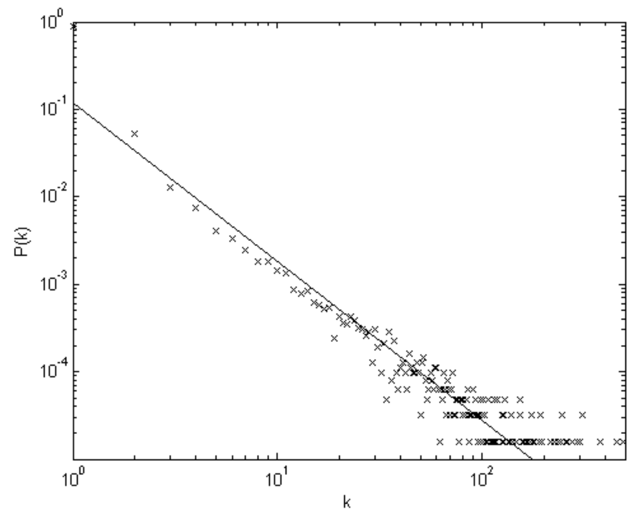

Figure 4 Degree distribution. The line is $\gamma = 1.82$ plotted over the observed *P(k)*

Degree distribution: The measured degree exponent is low as expected. There is a higher probability for high degree nodes than even the measured low power law allows. This supports the assumption that high degree nodes are favored by Traceroute. The calculated exponent is based on degrees with more than 4 observations. All observations are plotted in Figure 4. Distribution exponents for generated networks are listed in Table 1. The ER model does not produce a power law. Note that scaling the number of links in the network for the SF and BA models does not affect the degree distribution. Degree distribution is only determined by the growth algorithm and not by density.

Clustering: Given such a sparse network there were not high expectations of any clustering. Given that most traces will take shortest paths no local clustering will be discovered. However, this is not the case. The explored network actually shows larger clustering values all over the range than any of the models are producing. The ER and SF models where formation of links takes place randomly only between new nodes and existing nodes show hardly any clustering while the IG and GLP models where links also grow in between existing nodes have slightly higher clustering. The unscaled BA model is much denser than the real network and do produce some clustering. The average local clustering is *C(k)* [9], for k-degree nodes.

Rich club: All networks display power law RCC though the real network has a falloff among the highest degree nodes. Lower γ models seem to have overall higher RCC than the others while the SF network has more moderate values. The explored network has the lowest γ, but do however not have a significant rich club.

Coreness: Perhaps the most interesting and surprising measurement done is coreness. Considering the average degree of only 2.1 one should not expect to find any fully connected core. Nodes with degree 1 will not be in any core in the network. All redundant links in the network will form cycles and all nodes contained in those cycles are in the 1-core of the network. Any network with $m > n - 1$ will therefore have a non-empty 1-core. The models fit the expectations well, all but the two last models where unable to produce cores connected by more than two links. In the two models where links can form within the existing network (GLP and IG) we find an emerging core, but still only connected with three links. Maximum coreness, *k-max*, of the explored network is 5, meaning there is still nodes present after recursively removing all nodes with degree less or equal to 5. This must be considered a significant and highly connected core. Average degree in the *5-core* was 7.57 with 42 nodes.

Distances: Measured diameter and average distance between node pairs are listed in Table 1. The diameter is largest of the measured ones. The average distance is however not far off the models.

Tangential link ratio: As mentioned earlier, this metric is defined for the purpose of this research and is designed to measure the tree-likeness of a network. The assumption is that most of the links discovered by Traceroute all “takes you somewhere” in the network (a.k.a. radial links), while few links interconnect node pairs of equal distance to the core and fringes (a.k.a. tangential links). Table 1 lists the measured tangential link ratios. The models that produce power laws also all produce a TLR of approximately 0.5, meaning that half of the redundant links are interconnecting nodes of equal distance from an origin while the other half are connecting between the fringe and the origin. Note that both scaled and un-scaled versions of the SF and BA networks are about the same and does not seem dependant on density. The explored network exhibits a significantly lower value while the ER network produces more tangential than radial links.

5. Discussion

The network grew at a declining rate as expected with a growing average degree. The number of links discovered is very low indicating that packets do not travel much along alternate paths. This might suggest that few alternate paths exist, especially along the fringes of the network. The average degree after filtering is only 2.1 and should be considered extremely low. Unfortunately, the low detection rate for links means that values for the rest of the values will also be significantly skewed. This should however provide the opportunity to validate the assumptions made regarding problems with Traceroute.

Degree distribution is a power law as expected; this is also supported by others using BGP table probing [5], though they found $\gamma = 2.48$ and this is assumed to be correct though no proof exists. The measured $\gamma = 1.82$ is probably a lot lower than the real, meaning that the trace largely favored high degree nodes. Looking at the plot in Figure 4 we also see that there is no cut off at the tail of the distribution and nodes are scattered at much higher degrees than the power law should allow. Judging from this we can see that there probably are a significant number of higher degree nodes acting as important hubs. A lot of packets are routed through these and detection rate of links around these hubs is a lot higher than for the rest of the network. *Local clustering* values are somewhat surprising considering the overall extremely low link count. This is confirmed by the fact that none of the models produce mean local clustering anywhere close the values of the explored network. ER, SA and BA models are simply unable to produce clustering in networks this sparse. Kvalvik [20] produced un-scaled BA and SF plots and the BA model does produce significantly more clustered network than the SF at the same density. The only factor different in these two models is the preferential attachment. This can indicate that preferential attachment leads to clustering, or that the lower γ , in this case produced by the preferential attachment, induces higher clustering. Models that grow links between existing nodes independently from new nodes (GLP and IG) naturally produce the most clustered networks. The trend of higher clustering among lower degree nodes in the explored network is not shared with any models, this is an indication that high degree nodes will most often act as a "snowflake" hub connecting to mostly low degree nodes.

Rich club measures how close rich nodes are to forming complete sub networks. According to [2] the BA network does not produce a distinctive rich club. The measured RCC yields similar results to them for the BA networks and slightly lower for the other models. Overall, no rich clubs can be found in any of the networks.

Coreness values are probably the most noteworthy of the metrics used. The expected values for *k-max* are strongly connected to the density of the network, but despite a very sparse network there is a significant highly connected core. However, none of the models produce any core at all at this low number of links. All models will produce loops due to $m > n-1$, but only the two models that grows links independently from nodes have a $k_{\max} = 2$, meaning nodes connected with more than 3 links. The explored networks $k_{\max} = 5$ indicate a very highly centralized core with mostly tree-like fringes.

Diameter of the explored network was as expected higher than for any of the models, though only slightly larger than for the ER network. Average distance is shorter for the for the explored network than for the ER and SF networks. Considering the low detection rate of links one can assume that the real value is shorter than for all the other networks. The reason for this large diameter but relatively short average distance can be explained by the distance distribution in [20]. *Tangential link ratio* is lower than the models as expected; this is mainly a product of too stable routing in the network and most traces taking a direct path to the target. Considering this the assumption of Traceroute not discovering tangential links is correct and the resulting network looks like a tree. None of the models produced networks that resembled the real very well. None of the models produce anywhere near the clustering and coreness of the explored network suggesting that the real network has significantly different and more complex preference functions. The scaled BA network has diameter and average distance close to the explored but does not fit well on any other metric. Overall, none of the models captures the processes involved in the growth of the real network.

Problems

It is safe to assume that the real network is far bigger than the one explored here and the tracing was not allowed to continue for long enough. Judging from the falloff in newly discovered network elements in Figure 3 we can carefully estimate at least 3 times as many traces were needed. Second, only one Traceroute probe was set up at one location, the data would have been more complete if more probes were set up at several locations in the network. The finished result would then be a union of several tree like topologies instead of just this one. This will discover more links and increase TLR as probes send packets in different directions within the network. Third, no steps have been taken to validate these data statistically or otherwise. The data gathering method used is well known and some of its problems are described in [21-24]. Claims in these sources have not been challenged and assumed to be true for this network as well. Other secondary data sources were not explored due to time restrictions. We suggest Internet Mapping Project or BGP table lookups.

6. Conclusions

In this research we have explored the Internet in Norway by a well known method. The result has been compared to several well known models and measured with established metrics for structural analysis of networks. We have also introduced a new metric that defines and measures tangential links in the network. This could help us understand the importance of such links and its role in network robustness and redundancy. The frequently used exploration tool, Traceroute, seem to miss this metric and the findings could indicate that findings based on this tool alone are misleading. Existing publications [7] are aware of this but does not quantify it. The tangential link ratio metric is defined for as an aid in validating the results of tree-like explorations, primarily for link discovery rate. Theoretical work on the role of tangential links and ways to measure them in networks is needed to understand its impact on robustness and scalability. This underlines the need for additional methods for collecting a more complete description of such large IP networks as the Internet.

Finally, the findings the findings for this research are relevant to the explored network in Norway and it has significance for explaining topology characteristics for other IP sub-networks of the Internet. It is clear that the gathering of data using Traceroute requires more probes than used here and should have been run for much longer. Therefore data gathered should be considered incomplete. Nevertheless some interesting metrics were discovered. The most surprising is the high $k_{\max} = 5$ and the high average degree of 7.57 in the *5-core*. This suggests a highly centralized network with a high concentration of interconnected mid-high degree nodes in the core. This does however mean that the rest of the network outside the small core is even sparser than the 2.1 average degree, approaching that of a tree topology. The large core combined with higher than expected clustering is an indication that Traceroute probing does capture some central structures, but fails at discovering more than spanning trees at distant parts of the network.

References

[1] Beiberich, M. (2004) “Service Providers Define Requirements for Next-Generation IP/MPLS Core Routers”, *Yankee Group Report*. Accessed from CISCO: http://www.cisco.com/warp/public/779/largeent/landingPage/pdf/yankee.pdf

[2] Zhou, S. and R.J. Mondragòn (2002) “The 'rich-club' phenomenon in Internet topology”, IEEE Communication Letters, Vol. 8, No. 3, March, 180-182.

[3] Burch, H. (2005) “Measuring an IP Network” in *situ*, Carnegie Mellon University.

[4] Albert, R. and A. Barabási (1999) “Emergence of scaling in random networks” *Science*, October, 509-512.

[5] Faloutos, M., P. Faloutos, and C. Faloutos (1999) “On power-law relationships of the Internet topology”, *ACM SIGCOMM*, Vol. 29, 251-262.

[6] Barabási, A.L., (2003) *Linked*, Plume Press.

[7] Mahadevan, P., Krioukov, D. and et al. (2005) “Lessons from three views of internet topology” *CAIDA* Technical Report.

[8] Bu, T. and D. Towsley (2002) “Distinguishing between Internet Power Law Topology Generators” IEEE INFOCOM 2002 Conference Proceedings.

[9] Watts, D.J. and S.H. Strogatz (1998) “Collective dynamics of "small world" networks” *Nature*, 393, 440-442.

[10] Zhou, S. and R.J. Mondragòn (2003) “Towards Modelling The Internet Topology – The Interactive Growth Model”, *ITC18 Conference Proceedings*, Berlin.

[11] Zhou, S. and R.J. (2004) Mondragòn, Accuraltely moddeling the Internet topology. *Physical Review*, Vol. 70.

[12] Gaertler, M. and M. Patrignani (2004) “Dynamic Analysis of the Autonomous System Graph” *IPS Conference Proceedings*.

[13] Skitter. (1999) CAIDA, TR: http://www.caida.org/tools/measurement/skitter/.

[14] Erdõs, P. and A. Rèny (1959) “On random graphs” *Publicationes Mathematique,* Vol. 6, 290-297.

[15]Albert, R., H. Joeng, and A. Barabási (1999) “Diameter of the World Wide Web” *Nature*, 401, 130-131.

[16] Albert, R and A.L. Barabási (2002) “Statistical mechanics of complex networks.” *Review of Modern Physics* 74, 47.

[17]Kvalvik, T. (2004) “Tree-Like Exploration of IP-Networks” *eNytt*, Nr 1, Høgskolen I Molde, online at http://kursinfo.himolde.no/in-kurs/felles/eNytt/enytt-1204.pdf .

[18] IP-to-Country database. 2005, WebHosting.info.

[19] BSD traceroute manual. 2000.

[20] Kvalvik, T. (2006) “Measuring the Internet in Norway” Master Thesis, Molde University College, Dept. of Economy, Informatics and Social Sciences.

[21] Asta, L.D., et al. (2005) “Exploring networks with Traceroute like probes” Theoretical *Computer Science, Special Issue on Complex Networks*.

[22] Lakhina, A., J. Byers, and M.C.P. Xie (2005) “Sampling Biases in IP Topology Measurements”, IEEE INFOCOM 2003 Conference Proceedings.

[23] Clauset, A. and C. Moore (2003) “Traceroute sampling makes random graphs appear to have power law degree distribution”, arXiv:cond-mat/0312674 .

[24] Peterman, T. and P.D.L. Rios (2004) “Exploration of Scale-Free networks, Do we measure real exponents?” arXiv:cond-mat/0401065.