Consiglio Nazionale delle Ricerche

Technical Report n. 5/2001

Computer Network Topologies: Models and Generation Tools

Giuseppe Di Fatta[1], Giuseppe Lo Presti[2], Giuseppe Lo Re[1]

[1] CE.R.E. Researcher

[2] CE.R.E., University of Palermo Ph.D. student

{difatta, lopresti, lore}@cere.pa.cnr.it

July 2001

CE.R.E.

Centro di studio sulle Reti di Elaboratori

viale delle Scienze 90128 Palermo - Italy

Computer Network Topologies: Models and Generation Tools

Giuseppe Di Fatta
Giuseppe Lo Presti
Giuseppe Lo Re

*****

Index

1. INTRODUCTION

The explosive growth of networking, and in particular of the Internet, has been accompanied by a wide range of internetworking problems, in particular related to routing, resource reservation, and administration. The topology of a network, or a group of networks such as the Internet, has a strong bearing on many management and performance issues. To effectively engineer the Internet, crucial issues such as the large scale structure of its underlying physical topology, its time evolution and the contribution of its individual components to its overall function need to be well understood.

During the design phase of an Internet-based technology, extensive simulations are usually performed to assess its feasibility, in terms of efficiency and performance, since controlling a real network that is large enough to be interesting is in most cases too difficult and expensive.

In general, Internet studies and simulations assume certain topological properties or use synthetically generated topologies. If such studies are to give accurate guidance as to Internet–wide behavior of the protocols and algorithms being studied, the chosen topologies must exhibit fundamental properties or invariants empirically found in the actual structure of the Internet. As a consequence, good models of the topological structure of a network are essential for developing and analyzing internetworking technologies.

On the other hand, achieving a deep understanding of the Internet topology has proven to be a very challenging task since it involves solving difficult problems such as mapping the actual topology, characterizing it, and developing generation models that capture its fundamental properties. In addition, the topology of the Internet is a target that is constantly evolving, and it is controlled by a set of autonomous authorities that are not often willing to exchange low-level connectivity information.

This report is aimed to investigate the topological characteristics of computer networks, in particular as regards the parameters used to characterize them, and to evaluate some network generation models. Our work is intended to support the selection of the best generation model that fits with a specific task, such as developing and testing routing algorithms performances, and so on.

The rest of this report is structured as follows: in section 2 we present some empirical laws that were observed on real Internet topologies, and we apply these laws to typical computer generated networks from various sources; in particular, we will show how different are some properties of the generated networks; in section 3 we analyze a particular Internet topology instance more deeply, and in section 4 we analyze a method to automatically extract subnets from the Internet topology; we tested these subnets using the same empirical laws and present our results. Finally, in section 5 we conclude our work and discuss future directions.

2. POWER LAWS IN COMPUTER NETWORK TOPOLOGIES

Recently, it has been shown [3] that, despite their apparent randomness, actual Internet topologies exhibit power laws of the form $y \propto x^{\alpha}$, where α is a constant, between some typical

parameters used to characterize computer networks; infact, the authors showed that these laws hold in three Internet snapshots at inter-domain level, with correlation coefficients higher than 96%.

Some of the parameters and graph metrics they used are shown in Table 1.

Metric	Description
δ	Network diameter.
$\|V\|$	Number of nodes or vertices
$\|E\|$	Number of edges
d_c	Connection degree, i.e. fraction of edges with respect to a same-size complete network. It is obtained as $d_c = \frac{2\|E\|}{\|V\|(\|V\|-1)}$; a network is considered *sparse* if $d_c < 5\%$.
d_v	Node outdegree, i.e. the number of edges which exit from the node v.
r_v	Node rank, i.e. its *index* ordering all nodes in decreasing degree. It follows from the definition that if exists two nodes v and w such that $d_v < d_w$, then $r_v > r_w$.
f_d	Degree frequency, i.e. the number of nodes with degree d.
$P(h)$	Number of pairs of nodes within less or equal to h hops, including self-pairs, and counting all other pairs twice.
λ_I	The i-th eigenvalue of the adjacency matrix in decreasing order.

Table 1. Graph metrics used to characterize network topologies.

With these definitions, the power-laws can be stated as follows:

1. The outdegree of a node is proportional to the rank of the node to the power of a constant:

$$d_v \propto r_v^R$$

2. The frequency of an outdegree is proportional to the outdegree to the power of a constant:

$$f_d \propto d^O$$

3. The total number of pairs of nodes within h hops is proportional to the number of hops h to the power of a constant:

$$P(h) \propto h^H$$
$$\text{where} \quad h << \delta$$

4. The eigenvalues of a graph adjacency matrix are proportional to their order to the power of a constant:

$$\lambda_i \propto i^E$$

In this report, we analyze the first three laws, and we verify them on several computer generated networks, by means of a suitable software we developed to calculate these graph metrics. In particular, our software reads the network topology from a file in a standard analysis-suitable

format, and then measures the degree of each node[1], in order to calculate the frequency and the rank through a simple reordering procedure, with a relatively small computational cost; the third law involves the computation of all-pairs distances, so we use an optimized version of the Dijkstra algorithm; however the calculus has a higher computational cost, so we cannot compute it for some big-sized networks.

It can be noted that in cases where these laws hold, the exponents are always negative, so that the relationships are a kind of inverse proportionality; moreover, each one captures a different part of the network: in particular, the first one says that there are very few nodes with a very high degree, where the correlation is typically less and less strong when rank increase and consequently degree decrease; so this law captures the core of the network where nodes are usually clusters and links are high bandwidth backbones; on the other hand, the second one states that there are a lot of nodes with very small degree, where again the shape of the function is typically more and more noisy when degree increase and frequency decrease; so this law well describes the peripheral zones of the network, where there are a lot of leaves and links are low bandwidth connection to the backbones; finally, the third one provides information about the local connectivity of the nodes, on average, in the whole network we considered.

In the following, we present results about four groups of computer generated networks: the former was taken from the *SteinLib* Steiner tree problems public library [8], the others were generated using public-domain software packages.

SteinLib test set

SteinLib is a collection of Steiner tree problems in graphs [8]. The aim of this library is to collect freely available instances of Steiner tree problems in graphs and provide some information about their origins, solvability and characteristics; the library is composed by several group of graphs with homogeneous characteristics, and include graphs from VLSI applications, genetic contexts and computer networks applications.

We analyzed some sparse networks with random-weights chosen from several groups among those available on the web site, ranging in size from 1,000 to 2,500 nodes, and we found that the connection degree is relatively high, from 0.1% to 10% for graphs with the highest degree of redundancy. We calculated the parameters explained above in order to verify if these networks exhibit some power laws; the results are shown in figures 1, 2, 3, where two networks with 2000 edges and 25000 edges (0.4% and 5% connection degree respectively) from the “E” group are analyzed.

[1] Since we used symmetrical networks, the degree and the outdegree for each node are the same.

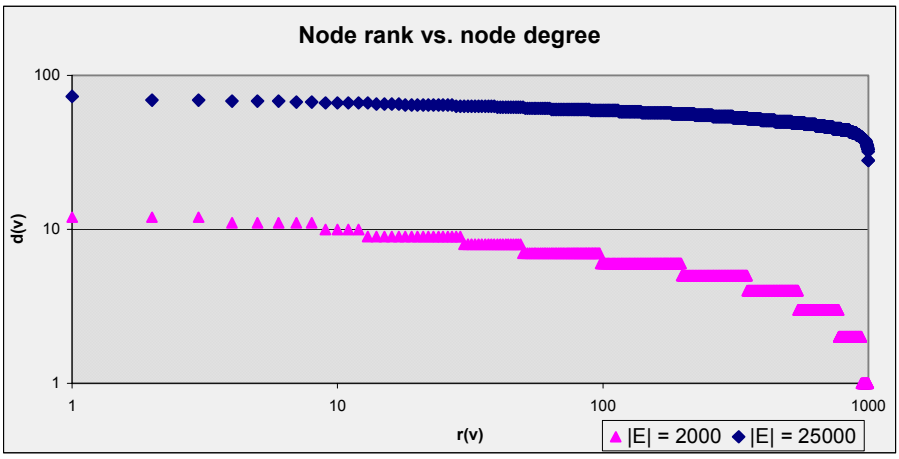

Fig. 1 – Node rank vs. node degree for two 1000-node networks from SteinLib.

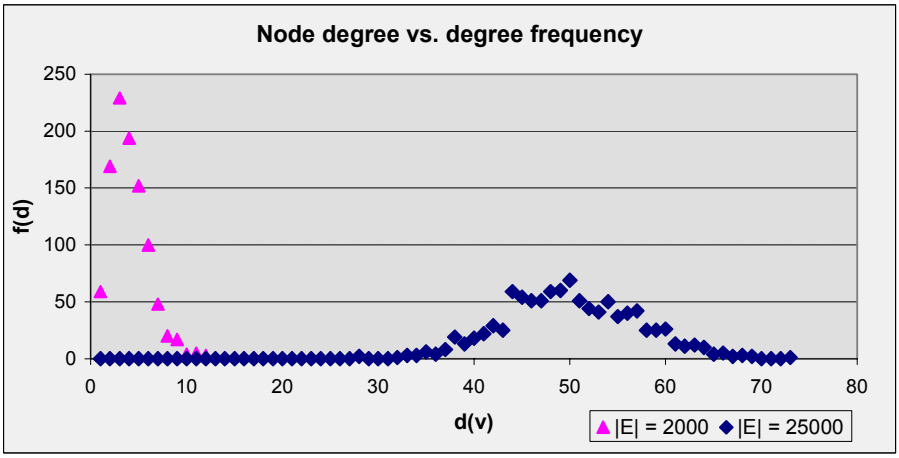

Fig. 2 – Node degree vs. degree frequency for the same 1000-node networks from SteinLib.

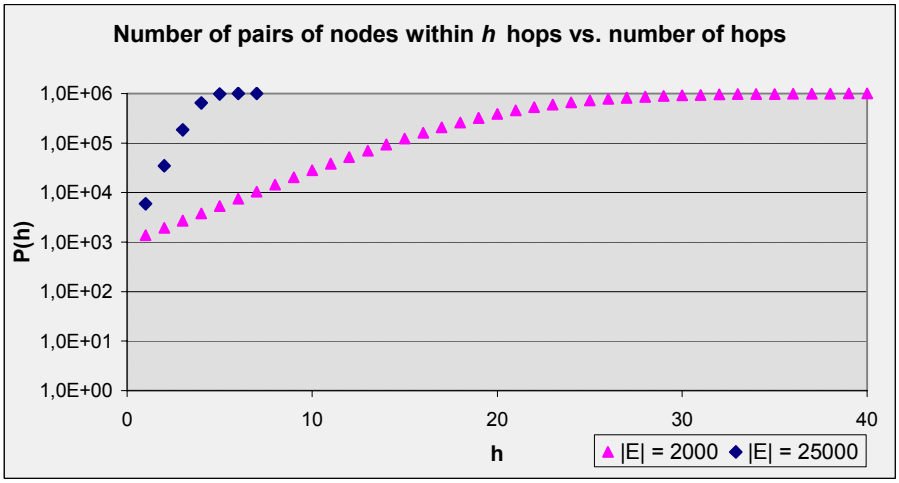

Fig. 3 – Hop-plot for the same 1000-node networks from SteinLib.

As it can be seen, the networks do not exhibit any power law correlation; they show a Gaussian-shape function as regards node degree distribution, and no particular well-known shape as regards node rank distribution, so it can be argued that they were generated randomly using for instance the simple Waxman probability distribution [2] or some other non-hierarchical model. However, it should be noted that the third graph does exhibit a power law correlation, with a typical asymptote when $P(h)$ reaches the maximum number of pairs of nodes in a network, i.e. n^2; nevertheless one should analyze all plots in a whole in order to state that a network is similar to the Internet, so this type of plot should be read carefully before correct conclusions can be drawn.

TIERS

TIERS is a software package for the generation of hierarchic multi-tier networks [1], [2]; the generated networks can have a user-definable redundancy degree and could be very sparse: we generated some networks with a connection degree less than 1%. In particular, the network model is based on a three level hierarchy that represents WAN, MAN and LAN, where the LAN are represented as star topologies instead of as complete subgraph.

We generated some networks ranging in size from 500 to 2,000 nodes, and found the results shown in figures 4, 5.

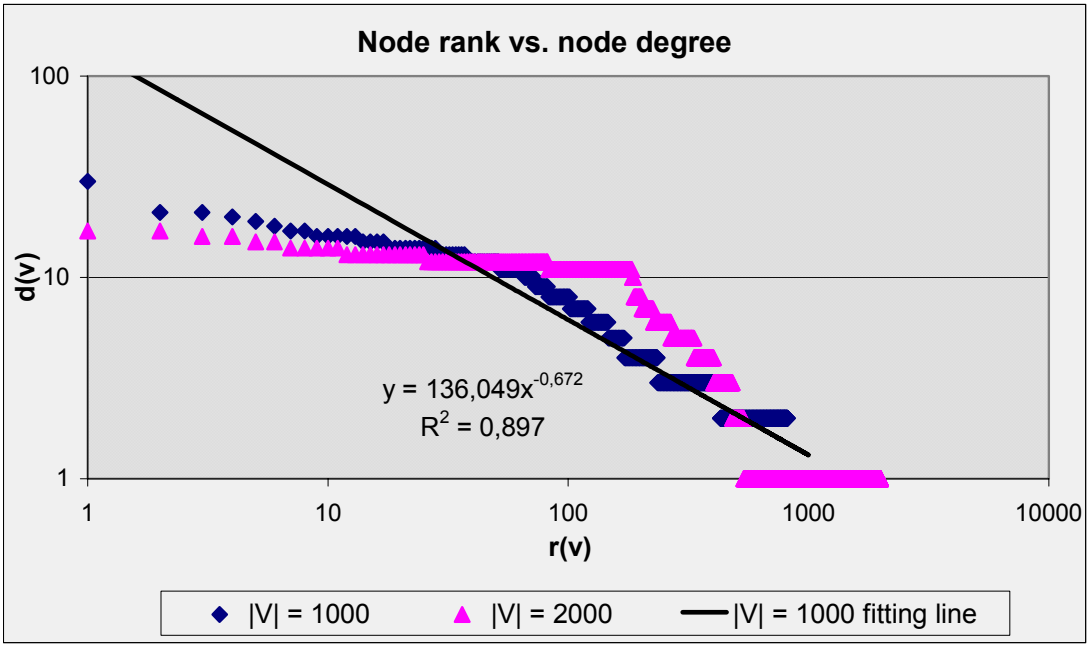

Fig. 4 – Node rank vs. node degree for a 1000-node and a 2000-node TIERS-generated network.

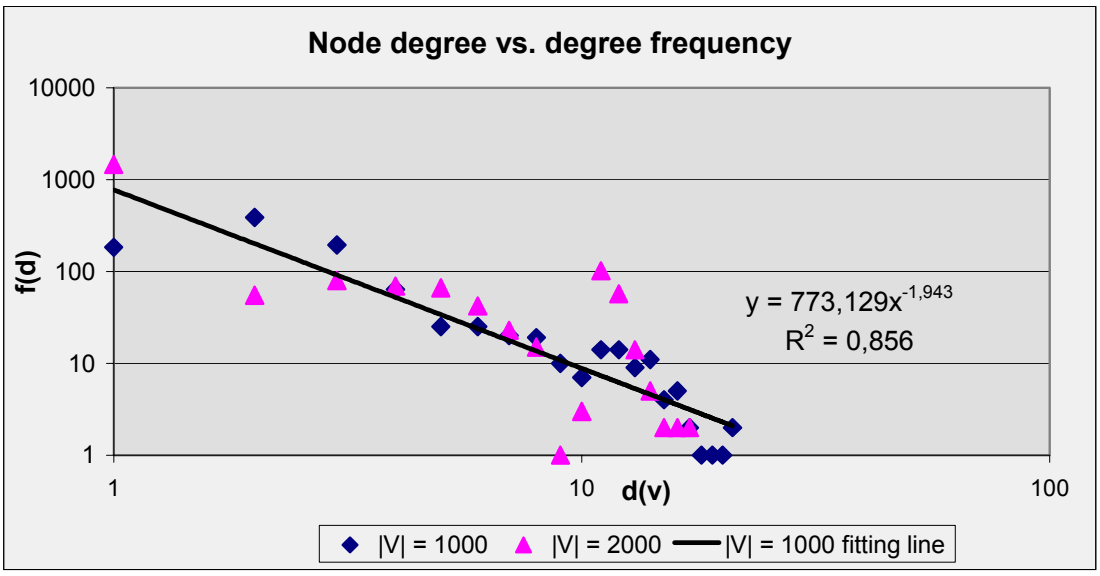

Fig. 5 – Node degree vs. degree frequency for the same networks.

In this case, we found some kind of correlation between the graph metrics in both graphs, and it could be explained since the topologies are hierarchical, but there are still some points out of fitting: for example, the node with the highest degree does not respect the fitting for the degree graph (fig. 4), and the local maximum in the same graph located near $d = 11$ is due to the relatively high number of LAN stars centers, which are artifacts of the model; moreover, the rank graph (fig. 5) doesn't exhibit a power fitting and the correlation coefficient for a power law is quite low (less than 90%).

BRITE

BRITE (for Boston university Representative Internet Topology gEnerator) was developed to investigate the growth of large computer networks [4]. It was recently updated [5] in order to build a framework to test various topology generation models and analyze their characteristics with the graph metrics we presented above.

The key characteristic of this generator is the incremental growth and the preferential connectivity used during the generation process; the authors claim that these are the primary reasons for power-laws on the Internet, since the generated topologies exhibit the power-laws with a high correlation. As an example, the plot in fig. 6 shows that for a typical BRITE-generated network the correlation is strong and the topology match the power laws very well.

It should be noted however that BRITE is not based on a hierarchical model but on a incremental growing strategy, where the previous considered model, TIERS, is hierarchical instead: as was noted in [9], though is a common idea that the Internet is hierarchical, it is a fact that a degree based generator fits better the real networks than a hierarchy based generator. Studies are going on to investigate this strange behavior.

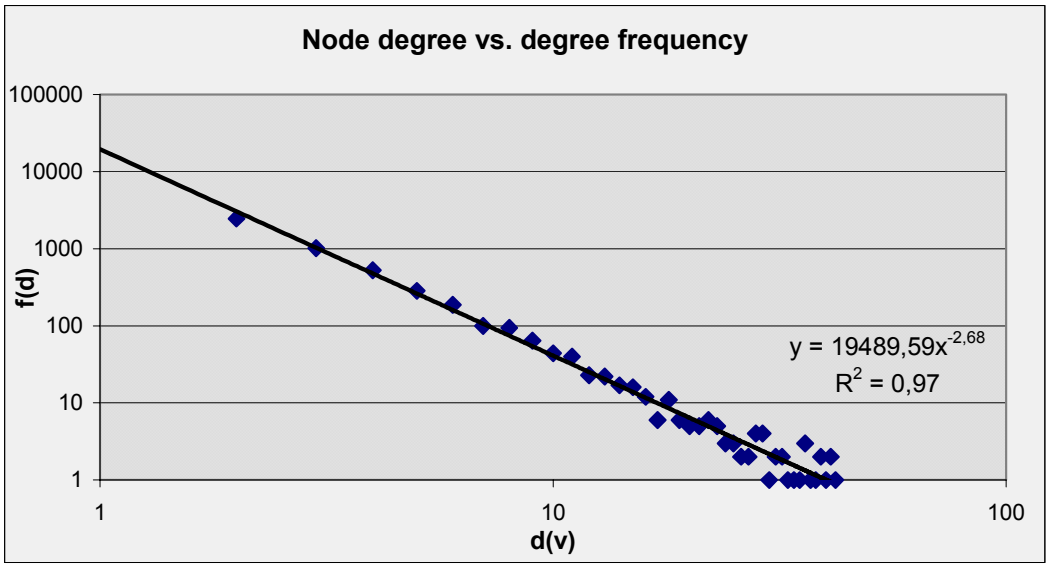

Fig. 6 - Node degree vs. degree frequency for a 5000-node BRITE generated topology.

Inet

INET is an Internet topology generator based on Autonomous System (AS) connectivity in the Internet [7]. The purpose of this generator is similar to BRITE, and the authors based their model over the power laws we exposed above, though they used a different growth model and also included time in their system in order to capture the network growing during time.

We generated some networks ranging from 3,000 to 5,000 nodes and found similar results as with BRITE: the generated topologies exhibit the power laws with a high correlation coefficient. On the other hand, the model is constructed over these laws, so we could expect this kind of results; finally, this tools is relatively simple for the final user since it needs only the target size, and calculates the topology in some steps according to the degree power law.

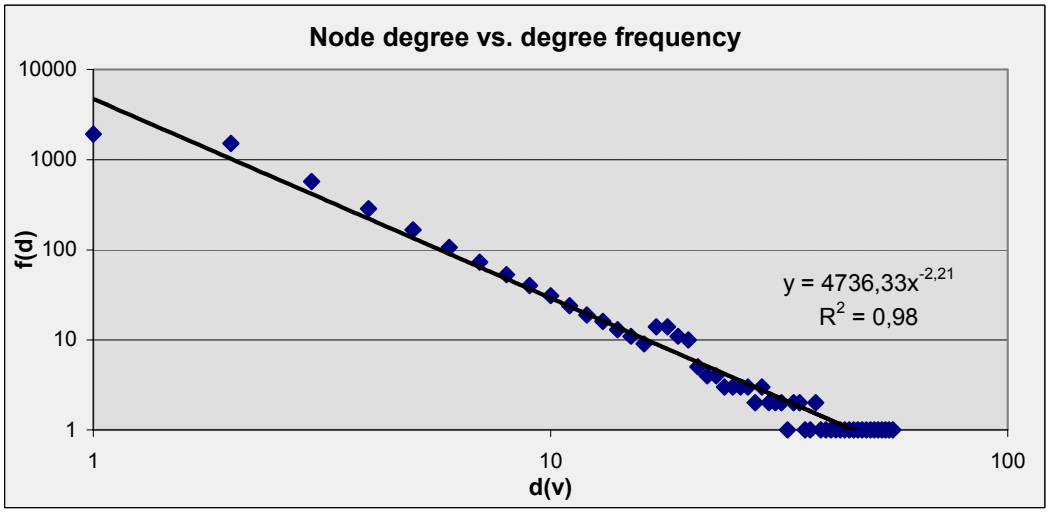

Fig. 7 - Node degree vs. degree frequency for a 5000-node Inet generated topology.

3. INTERNET ANALYSIS

To further investigate the Internet topology structure, the *Bell Labs* [6] – among others – have developed a software system in order to obtain the real Internet map. Their system *Mercator* is based on several traceroute started from a dedicated server and gathered into a database; then all the traces were carefully merged in order to obtain a single topology, and the IP addresses were deleted.

The resulting network, obtained in November 1999, includes 284806 nodes and 449306 edges, with a connection degree lower than 0,001%; a low-resolution graphical representation of this network from [6] is presented.

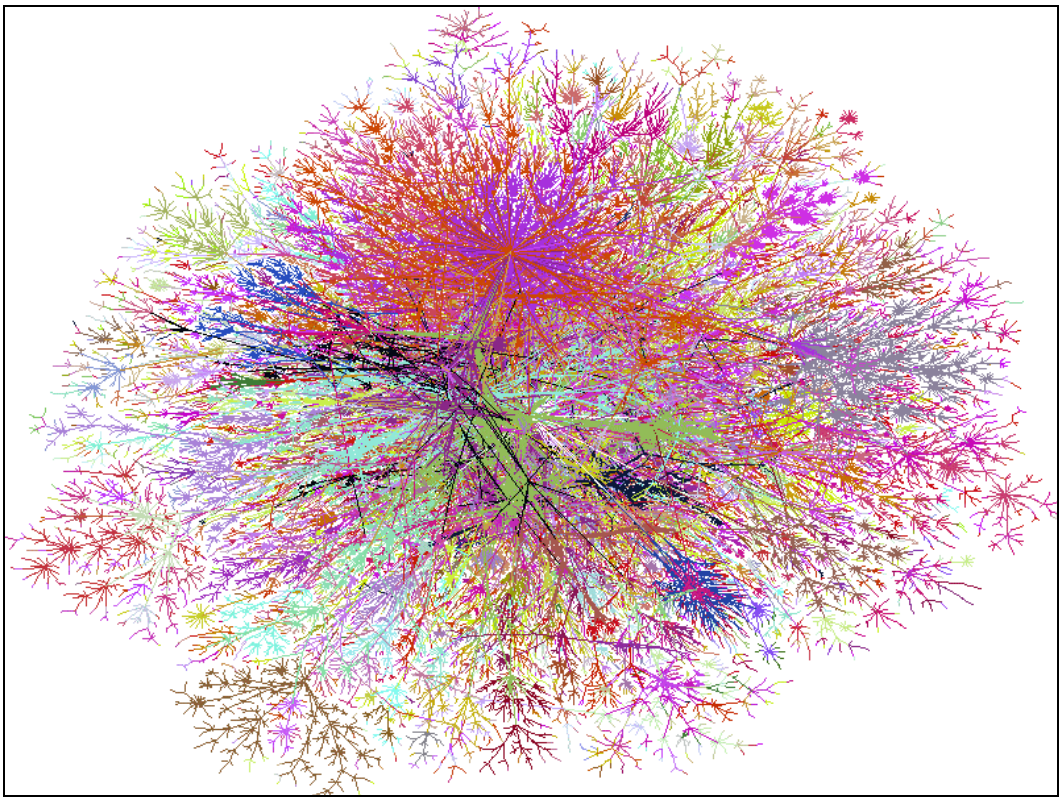

Fig. 8 – A map of the Internet as discovered by the Bell-labs.
Links painted with different colors represent different geographical locations.

It should be noted that this map was obtained at a router-level, so it doesn't include any terminal host (which are about 100 millions); as a comparison, the inter-domain level Internet maps used in [3] contain only from 3000 to 4500 nodes.

We converted the entire topology to our standard format and then we analyzed it using our software; we calculated only the graph metrics for the first two laws, since the others would involve too high computational cost due to the size of the network.

The plots in fig. 9 and 10 show that Internet does exhibit the power laws with a high correlation: as it could be seen, we calculated a correlation coefficient greater than 97% for both graphs.

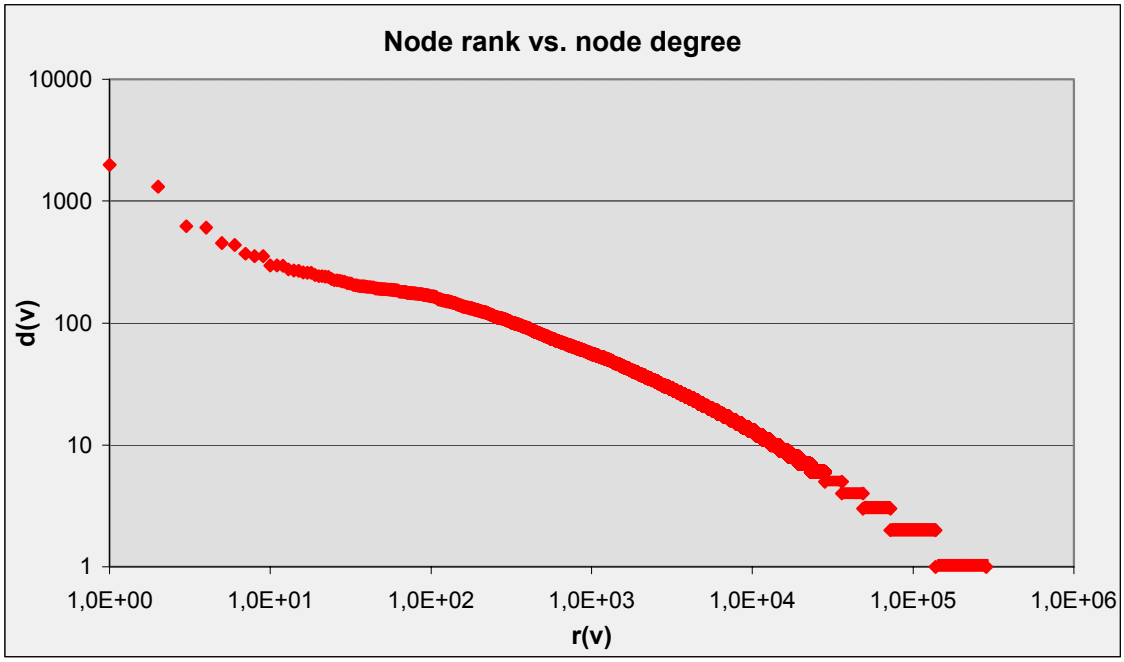

Fig. 9 – Node rank vs. node degree for the Internet.

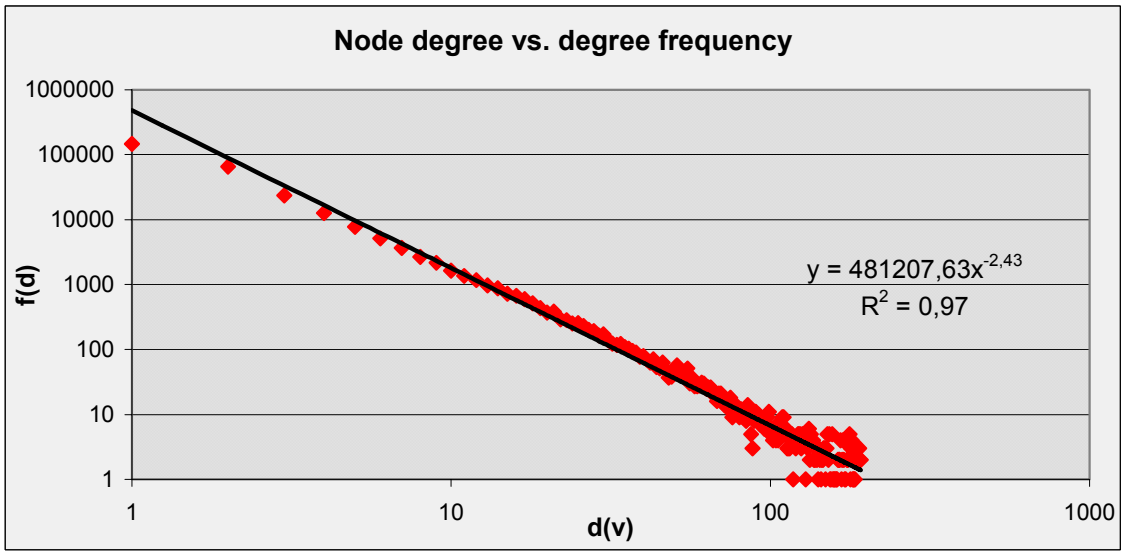

Fig. 10 – Node degree vs. degree frequency for the Internet.

4. SUBNETS EXTRACTED FROM THE INTERNET

Extraction methods

In order to generate realistic Internet topologies we are studying some extraction methods which start from a complete Internet topology – such as the one examined in the previous section – and cut off a subnet with a user definable size.

We implemented a simple method that extracts a network as a growing oil-spot; it starts from a randomly selected node and iteratively adds the nodes in its neighborhood until the growing network reaches the target size. In particular, the algorithm could be stated in a C++ like language as follows:

```
function OilSpotSubnet(Network net, int size) returns a network  {
  Network target = ∅;
  Queue queue = ∅;
  int count = 0;
  int start = random(net.size);
  for each node in net do
        node.status = available;
  queue.put(start);
  while(queue.isFull() && count < size) {
        start = queue.get();
        for each node in net.neighborhood(start) do
              if (node.status == available) {
                    node.status = picked_up;
                    target.add(node);
                    queue.put(node);
                    count++;
                    }
        }
  return target;
  }
```

This algorithm, although very simple, preserves the local structure of the network, so we could expect good results as regards the topology properties of the generated subnetworks.

Another method could search the minimum number of cuts needed to extract a subnet with a target size; we would extend it further in the future.

Analysis and results

We tested the oil-spot subnet algorithm generating several 5000-node networks from the Mercator Internet topology and obtained the results shown in fig. 11, 12, 13.

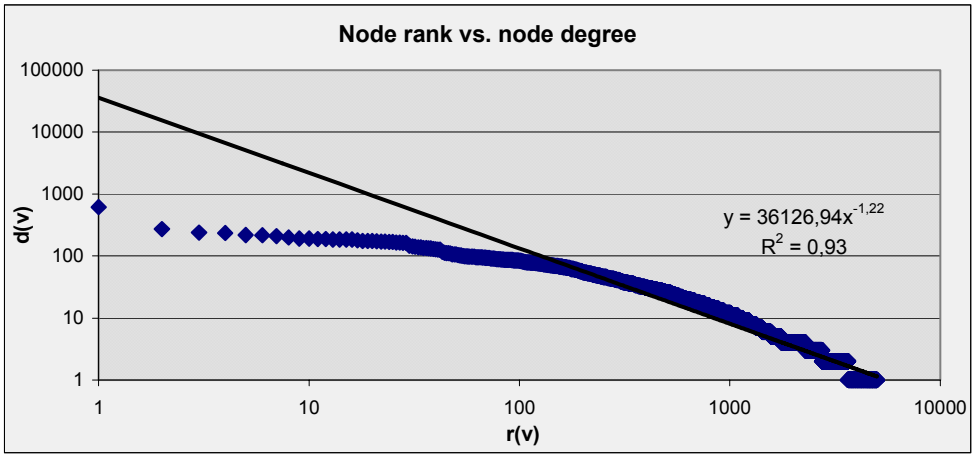

Fig. 11 – Node rank vs. node degree for a typical 5000-node Internet subnetwork.

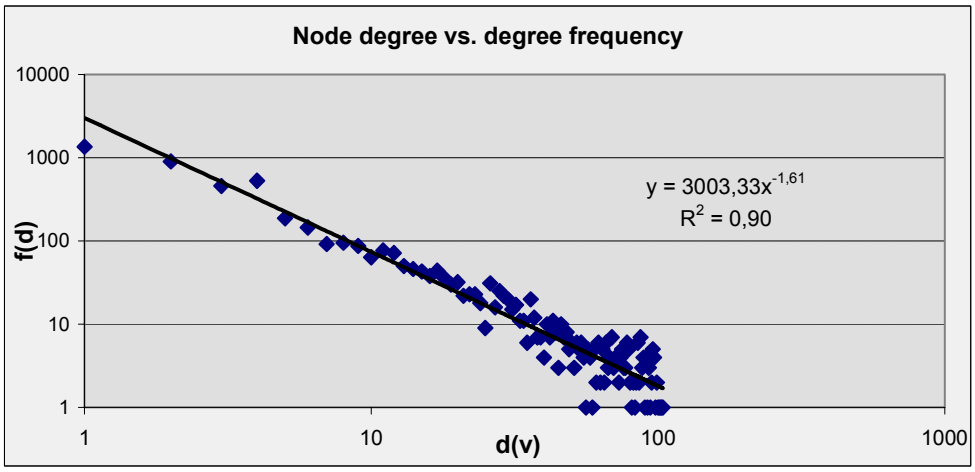

Fig. 12 – Node degree vs. degree frequency for the same Internet subnetwork.

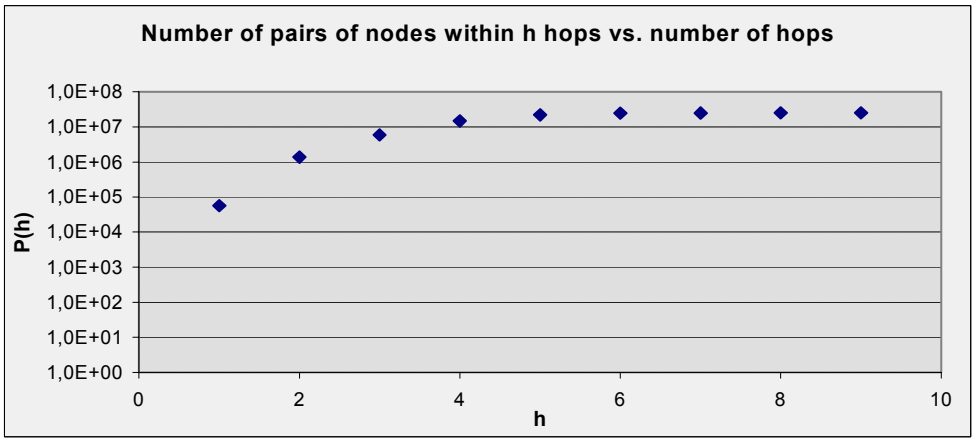

Fig. 13 – Hop-plot for the same Internet subnetwork.

As we stated in the previous paragraph, the topology properties of the generated subnetworks are very similar to those of the entire topology, and they exhibit the power laws with good correlation coefficents.

5. CONCLUSIONS

In this report the topological characteristics of computer network were studied and several computer-generated topologies were analyzed. The analysis done show that the Internet has some particular topological characteristics, and a random generated network does not match in general these characteristics. Among others, *Brite* and *Inet* have a network generation model that simulates the Internet structure with the best accuracy, and they are suitable for testing network algorithms and protocols prior to release them.

On the other hand, a hierarchical model such as *TIERS* is less accurate on simulating the Internet, and it is still an open question why a degree based generator fits better the real networks than a hierarchy based generator: authors found [9] some evidence that in degree based generators hierarchy arises from the degree distribution, but currently this is a guess which should be more widely validated.

Finally, subnets extracted from the Internet shows the same topological characteristics of the whole network: this fact can be explained with the intrinsic auto-similarity of the entire Internet, which grows in a fractal-like way, as was already stated in [4]: it's well known that most non-linear systems grows as fractals. On the other hand, the exhibited power laws – typical of fractal autosimilar systems – confirm this view and could help us in developing more and more accurate models.

REFERENCES

[1] M. Doar, *TIERS Network Generator*, Proc. of IEEE Globecom, November 1996, www.geocities.com/ResearchTriangle/3867/sourcecode.html.

[2] K. L. Calvert, M. B. Doar, E. W. Zegura, *Modeling Internet Topology*, IEEE Trans. on Comm., December 1997, pages 160 - 163.

[3] M. Faloutsos, P. Faloutsos, C. Faloutsos, *On Power-Law Relationships of the Internet Topology*, ACM SIGCOMM 1999.

[4] A. Medina, I. Matta, J. Byers, *On the Origin of Power Laws in Internet Topologies*, ACM SIGCOMM 2000, **30**, 2, April 2000.

[5] A. Medina, A. Lakhina, I. Matta, J. Byers, *BRITE Universal Topology Generator*, January 2000 - April 2001, cs-pub.bu.edu/brite.

[6] R. Govindan, H. Tangmunarunkit, *Heuristics for Internet Map Discovery*, Proc IEEE Infocom 2000, Tel Aviv, Israel, www.isi.edu/scan/mercator/mercator.html.

[7] C. Jin, Q. Chen, S. Jamin, *Inet Internet Topology Generator*, University of Michigan Technical Report, 2000, topology.eecs.umich.edu/inet.

[8] S. Voss, A. Martin, T. Koch, *SteinLib Testdata Library*, February 2001, elib.zib.de/steinlib/steinlib.php.

[9] H. Tangmunarunkit, R. Govindan, S. Jamin, *et al.*, *Network Topologies, Power Laws, and Hierarchy*, SIGCOMM 2001 poster, June 2001. To be appeared on Computer Comm. Rev., October 2001.