Analysis and Comparison of Internet Topology Generators

Damien Magoni and Jean-Jacques Pansiot

Université Louis Pasteur – LSIIT
Pôle API, Boulevard Sébastien Brant, 67400 Illkirch, France
{magoni, pansiot}@dpt-info.u-strasbg.fr

Abstract. The modeling of Internet topology is of vital importance to network researchers. Some network protocols, and particularly multicast ones, have performances that depend heavily on the network topology. That is why the topology model used for the simulation of those protocols must be as realistic as possible. In particular a protocol designed for the Internet should be tested upon Internet-like generated topologies. In this paper we provide a comparative study of three topology generators. The first two are among the latest available topology generators and the third is a generator that we have created. All of them try to generate topologies that model the measured Internet topology. We check their efficiency by comparing the produced topologies with the topology of a recently collected Internet map.

1   Introduction

Today simulation tools are widely used to test network protocols. These tools need network topologies as input data. A network topology is usually modeled by an undirected graph where the network devices are modeled by the nodes of the graph and the communication links are modeled by the edges of the graph. A software tool that creates network topologies is usually called a graph generator or a topology generator. The way it builds network topologies (i.e. the set of creation procedures) is called a topology model.

In this paper, we will focus on some of the latest Internet-like topology generators, including one that we have created. We will assess the efficiency of these generators by comparing the graphs that they produce with a graph built from real data and representing a part of the Internet at the router level. The rest of the paper is organized as follows. Section 3 presents the Internet map that we use as a reference, and the generators that we test as well as their settings. Section 4 details the properties studied. Finally, in section 5 we give the results of the analysis of the generated graphs compared to the Internet map analysis.

2   Previous Work

The study of the Internet topology is an area of active research. There is not much information on the topology of Internet at the router level because it is very hard to obtain. An attempt to map the Internet was carried out by Pansiot *et al.* [11] by using

source routing. Their data collect was done during the summer of 1995 and the resulting map contained 3888 routers. They also defined terms that we use in our work. More recently, another collect was undertaken by Govindan *et al.* using a heuristic called hop-limited probes [6]. This heuristic (and many others) has been included in their software called Mercator. Their collect was carried out in 1999 and the resulting map contained 228263 routers. Because it is easier to get exhaustive routing data, the Autonomous System (AS) level topology of Internet has been further investigated. From 1994 to 1995, a study of the Internet inter-domain topology was carried out by Govindan *et al.* [5]. In 1999, Faloutsos *et al.* [4] analyzed the inter-domain routing information provided by the NLANR and the routers' map made by Pansiot *et al.* They found that the Internet topology obey power laws at both the AS level and the router level. In [6], Govindan *et al.* noticed that some of the Faloutsos *et al.* power laws still hold for their router level Internet instance of 1999. Recently, additional power laws were found by Magoni *et al.* at the AS level [8] also by using data provided by the NLANR.

Concerning the topology generators, one of the earliest and most famous models was designed by Waxman in 1988 [12]. This kind of model is usually called flat topology model. The nodes are randomly placed on an euclidean plane irrespective of any hierarchy order among them. This model was later replaced by hierarchical topology models such as the Tiers [3] and the Transit-Stub [13] models. These models try to enforce a multi-level hierarchy that can be found in the Internet (e.g. host-router-AS). The discovery of power laws in the Internet by Faloutsos *et al.* has brought the arrival of a new kind of topology model. We call it the power law topology model because, as the name suggests, it makes use of power laws to generate Internet-like graphs. The BRITE [10] generator as well as our topology generator called network manipulator (*nem*), belong to this category and generate router level graphs. Inet2 [7] also belongs to the power law topology model category but it generates AS level graphs. Finally two models were recently defined to reflect the power laws found in huge network topologies. The first one was defined by Aiello *et al.* in [1] and the second one called Extended Scale-free model was defined by Albert *et al.* in [2].

3   Source and Tools

3.1   Internet Map

There are basically two levels in Internet topology. The router level and the AS level. Although AS level maps are easier to make, we will focus our attention on the router level maps of the Internet. The main reason for this choice is that a router level map provides a higher accuracy for IP layer protocol simulations. A simulation of a protocol designed for the IP environment should use such a map because it displays the IP connection topology.

As we said in the previous section, one of the most recent router level Internet maps was constructed by Govindan *et al.* with their software tool Mercator [6]. Their map is called **scan**. They also recovered another map built by researchers at the *Lucent* laboratories. This map is called **lucent**. Both maps were merged to create one of the most recent and complete Internet map ever built. This map is called **scan+lucent**. Govindan *et al.* have made the anonymized version of this map freely available for download.

It is this map that we have used in our study. We call it an Internet reference map. It is a huge map containing 284772 nodes and 449228 edges. When we show a property value of the **scan+lucent** map in a figure, we label it "Internet" for simplicity instead of **scan+lucent**. We are aware that the map is probably not exhaustive.

3.2   Topology Models

The three topology generators studied in this paper are BRITE, Inet2 and *nem*. All of them belong to the power law topology model (i.e. the latest and most accurate model). The graph sizes we have chosen to study are 500, 1000, 2000, 4000, 8000 and 16000 nodes. This should give a good view of the scaling effect on the properties of the generated graphs. Although researchers may need graphs smaller than 500 nodes (e.g. for resource consuming simulations), it is difficult for these generators to create very small graphs as they are based on power laws that arise only with big numbers. As Inet2 is, by design, not able to generate graphs of size below 3037, we have only created Inet2 graphs of sizes 4000, 8000 and 16000. We have generated **20** graphs of each chosen size for each topology generator.

We explain here the parameter settings that we used to generate the graphs. We have chosen to test BRITE with $m = 1$, 2 and 3 (it is the number of links added per new node). As we use incremental growth, we obtain graphs with an average node degree of $2m$. The Internet reference map has an average node degree of 3.15, so taking m above 3 would have given graphs with too high an average node degree (i.e. it means too many edges compared with the number of nodes). Furthermore we found different results for $m = 1$, 2 and 3, despite what Medina *et al.* found in [10]. That is why in the result section we consider three scenarios for the use of BRITE (i.e. for $m = 1$, 2 and 3) as if we had three different graph generators. Given the results of the authors of BRITE, we use a random node placement because a Pareto node placement gives similar results. Also we use preferential connectivity and incremental growth both turned on because only these settings generate graphs that obey the outdegree and rank power laws as shown by Medina *et al.* in [10]. The generation method (i.e. how graphs are created) of BRITE is fully explained in [10].

Concerning Inet2, a first very important remark is that it generates **AS level** graphs. To compare it with the other generators, we simply consider Inet2 output to be router level graphs. We note that Jin *et al.* did the same in [7] when they compare Inet2 with Waxman, Tiers, Transit-Stub and BRITE that they also consider as AS level generators. We will refer back to this point when it is relevant, and sometimes compare Inet2 graphs with an AS level map of May 2000 analyzed by Magoni *et al.* in [8]. The generation method of Inet2 is fully explained in [7].

Concerning *nem*, it is worth noticing that it creates graphs by extracting a subgraph from a real Internet map. Thus we usually call it a topology modeler rather than a topology generator. We use the **scan+lucent** map as its input real Internet map. Of course it can be argued that if we compare the graphs generated by *nem* with a reference map that is its input map, the results will automatically be matching what is desired but this is not true. The process of extracting a graph of a few hundred or thousand nodes from a map having nearly 300,000 nodes can make this graph have a completely different topology than the originating map. The generation method of *nem* is fully explained in [9].

4   Properties of Interest

In this section we describe which topological properties we have chosen to study. We use the regular terminology set forth in previous papers by Pansiot *et al.* [11], Faloutsos *et al.* [4] and Magoni *et al.* [8]. From now on and for the sake of simplicity, we will talk about the property values of the *Internet* instead of the property values measured in the **scan+lucent** map (i.e. the map that we use as our reference map). (e.g. we write "the diameter of the Internet is . . . " instead of "the diameter of the **scan+lucent** reference map is . . . ".)

The node *outdegree* or *degree* (i.e. as we consider undirected graphs) distribution is one of the fundamental properties of a graph. The degree distribution of the Internet is a skewed distribution. From a graph's degree distribution we infer the average node degree, power law 1 (rank exponent) and power law 2 (outdegree exponent). Both power laws have been found by Faloutsos *et al.* in [4]. Medina *et al.* have shown in [10] that Waxman graphs and Transit-Stub graphs do not comply with power laws 1 and 2.

Another important property is the distance distribution. We define the distance between two nodes as being the hop count between the two (i.e. the minimum number of edges to cross to get from one node to the other). The distance is also called shortest path length (defined by a number of hops). It is worth noticing that the distance distribution of the Internet is not skewed. It seems to be Gaussian. This means that the average distance inferred from the distance distribution is a good indicator to study. The biggest distance of a given node to any other node is called the *eccentricity* of the node. The eccentricity distribution of the Internet also seems to be Gaussian and thus we study the average eccentricity.

Furthermore we will not study power law 3 (i.e. eigen exponent) found by Faloutsos *et al.* because Medina *et al.* have shown in their paper [10] that power law 3 holds for large Waxman graphs, Transit-Stub graphs and nearly all BRITE configurations. As Waxman and Transit-Stub graphs do not model the Internet topology accurately, we think that power law 3 is not a primary indicator.

From the definition given by Magoni *et al.* in [8], a node belonging to a cycle or lying on a path connecting two cycles is called an *in-mesh* node. The *mesh* is the set of all in-mesh nodes of the graph. We examine the mesh size and we give results about the mesh connectivity such as the number of cutpoints and the biggest bicomponent size. The study of the mesh gives information about the amount of reliability vs connection failures and the possibility of load balancing by using alternate paths.

Finally we examine the forest part of the graphs. We look at the number of trees and at the tree size distribution. In the Internet, we have found that power laws 6 and 7 (found by Magoni *et al.* at the AS level in [8]) can be inferred from this distribution. We examine the generated graphs to see if they also comply with these power laws. The properties concerning the trees are interesting for studying the network reliability and connectivity as each node belonging to a tree is a cutpoint (excepted the leaf nodes) and thus it can make the graph disconnected if it fails.

5   Results

This section contains the results of the analysis of all the graphs generated by the three topology generators. These results are compared with the results of the analysis that we made on the **scan+lucent** map. *In all the following figures, the value measured in the scan+lucent map of the given property is plotted as a dashed horizontal line. Of course it is not corresponding to the network size coordinate axis. It only serves as a reference value that can be easily compared to the values measured in the graphs.* Any property value for a generator for a given size is the average of the values measured for each of the 20 generated graphs. For instance, the average node degree of the 500-node graphs generated by BRITE with m = 2 is 3.6. This means that this value is the average of the average node degree of each of the 20 graphs of size 500 generated with BRITE (m = 2). In what follows we will write BRITE x instead of BRITE with m = x.

5.1   Degree Properties

Figure 1 shows the plots of the average node degree. The first striking observation is that it is just below 2 for the BRITE 1 graphs. In fact we checked that each graph of size n generated with BRITE 1 has exactly $n - 1$ edges and is connected. This means that these graphs are merely trees. This is confirmed by the authors of BRITE who state that for a newly considered node they connect it with only one link [10] when m = 1. BRITE 3 graphs have an average degree between 5.5 and 6, which is nearly the double of the Internet average degree value. However, as the degree distribution is skewed, this may not be of great importance. Finally, *nem* has exactly the same average degree as the Internet. This is because, as we saw in section 3, it generates the amount of edges needed to match the Internet average degree. It is worth noticing that the average node degree of the graphs from all the generators does not depend much on the size of the graph (excepted for 2000-node or less BRITE 3 graphs).

Figure 2 shows the plots of the absolute correlation coefficient (ACC) of the average degree distribution with respect to power law 2 (i.e. outdegree exponent). We clearly see that BRITE 2 and BRITE 3 graphs do not comply with power law 2. We also see that Inet2 graph ACCs tend to decrease when the size increases. In particular the graphs of size 16000 have an average ACC a little under 0.95. Finally, *nem* and BRITE 1 graphs have very good ACCs that decrease a little when the graph size is small (e.g. 500).

We have examined some samples of the degree distributions of BRITE 2 and 3 graphs. Here is an example of the beginning of a degree distribution of a 4000-node BRITE 3 graph:

```
Degree    Frequency    Frequency in %
  1           5             0.125
  2           4             0.1
  3         1601           40.025
  4          766           19.15
  5          450           11.25
 ..          ..             ..
```
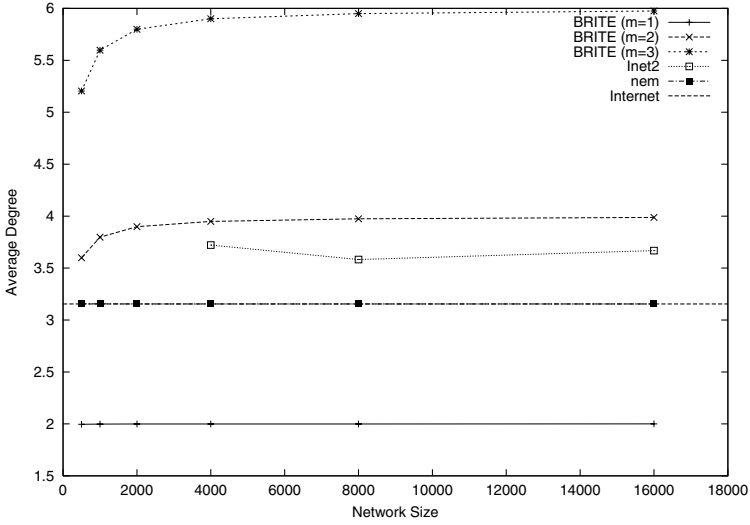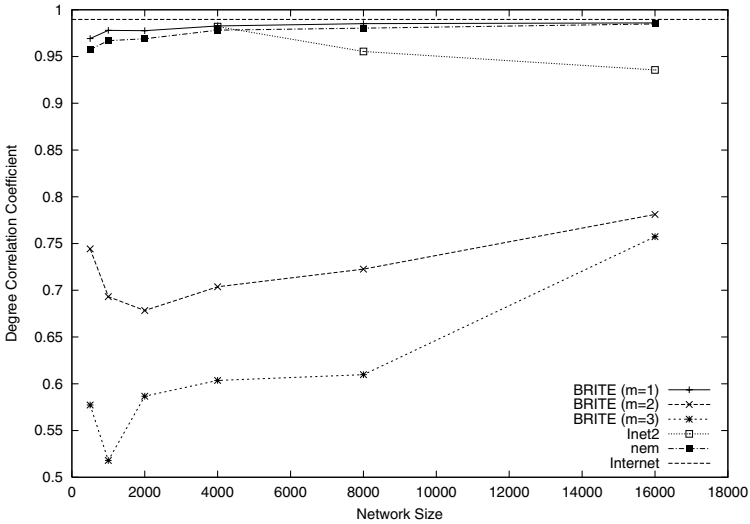
Fig. 1. Average Degree

Fig. 2. Degree Correlation Coefficient

It is the few degree 1 and degree 2 nodes that cause these graphs to *not* comply with power law 2. We verified for BRITE 2 and 3 graphs that the BRITE algorithm generates a degree distribution that starts at m instead of one (when m is greater than 1) and a few outliers of degree less than m cause BRITE graphs to *not* follow power law 2. We do not show the plots of the absolute correlation coefficient (ACC) of the rank distribution because all generators comply with power law 1 (i.e. rank exponent).
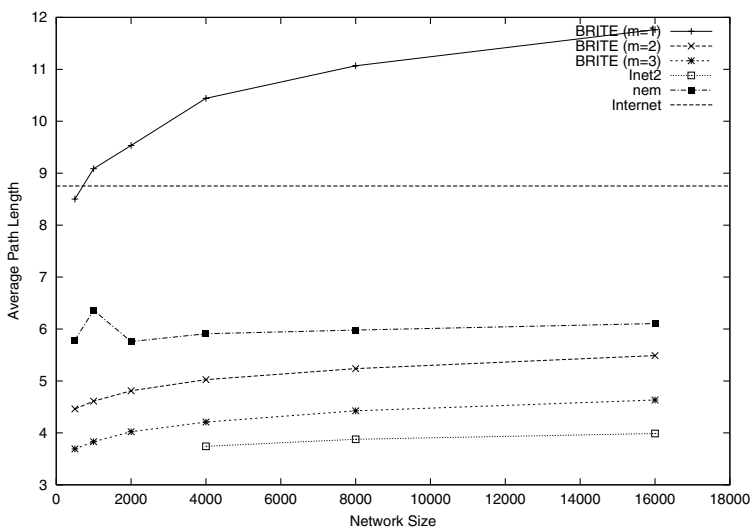
Fig. 3. Average Path Length

To conclude the study on the degree distribution of the graphs, we can already see that many factors contribute to the realism of a graph generator. We saw here that BRITE 1 complies with power laws 1 and 2, but its graphs are trees and thus do not match the Internet topology at all (particularly in terms of redundant links). BRITE 2 and 3 do not comply with power law 2 because of the way they shift the degree distribution (i.e. the majority of the nodes should have degree 1 and not degree m). Inet2 and *nem* generate graphs that bear a closer similarity to the Internet degree properties.

5.2   Distance Properties

Figure 3 shows the plots of the average distance (i.e. path length). Except for BRITE 1 graphs, all graphs have an average distance below the Internet average distance which is 8.75. BRITE 1 average distance increases when the graph size increases presumably because of its tree structure. The average distance of the others does not seem to vary with changes in size. Inet2 graphs have the lowest values (all below 4) but this is surely due to Inet2 design (i.e. an AS level generator). Magoni *et al.* found that the average distance of the AS level Internet in May 2000 was 3.65 which is very close to the average distance values of Inet2 graphs. As distances are an important factor in simulations (particularly for determining delays) it would be desirable to have average distances as close as possible to the Internet average distance of 8.75. *nem* graphs with average distance values around 6 and BRITE 1 graphs with values around 11 (although much influenced by the graph size) are the most realistic ones.

Figure 4 shows the plots of the average node eccentricity. Here BRITE 2 and 3 graphs have values not influenced by the graph size and below 8, which is far from the Internet eccentricity value of nearly 20. BRITE 1, Inet2 and *nem* values seem to depend
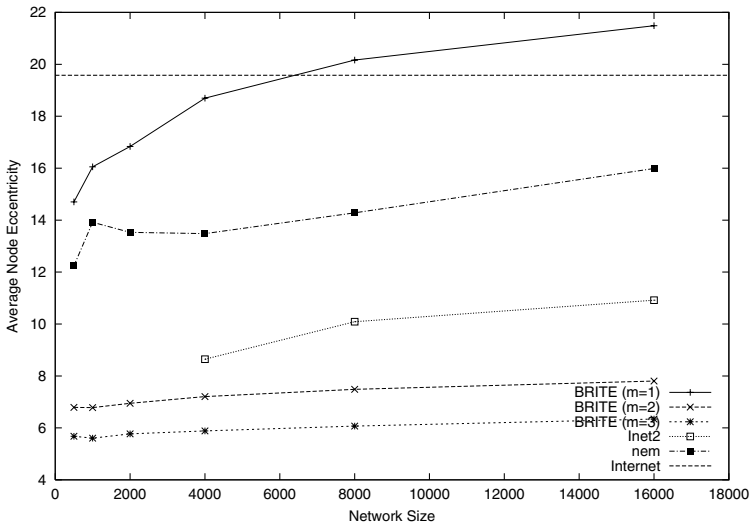
Fig. 4. Average Node Eccentricity

on the graph size. When the graph size increases, the eccentricity increases. However only BRITE 1 seems to match the Internet eccentricity. Inet2 has values that are too low (i.e. below 11) and *nem* values although reaching 16 for 16000-node graphs are still 25% under the Internet eccentricity value. Eccentricity gives an idea of the spreading of the graph. If it is low it means that there are only a few nodes that are far removed from the others, most of them being concentrated in a small area. On the other hand, a high eccentricity (as in the Internet) means that the nodes are spread over a wide area. It is a measure of how far a node is from all the other nodes. In the Internet a node is, on average, at most 20 hops from all the others. At the AS level, the eccentricity of the Internet is 7 which is well under the Inet2 graph values (especially for 16000-node graphs with a fraction of 0.35, equal to the AS level fraction, where the eccentricity reaches 11).

To conclude, we can say that distance properties are hard to model with accuracy. Average values such as distance and eccentricity are usually too low in the generated graphs compared to the Internet values. The increase in the values of BRITE 1, Inet2 and *nem* when the graph size increases is a good sign that taking bigger graphs will give adequate values. However we think that it is important to try to do better because these two properties are based on Gaussian distributions and thus are good indicators of graph realism. In fact we are currently creating a filter in our *nem* generator that allows us to generate graphs with average distances nearly equal to those measured in the Internet (i.e. between 9 and 11). It works when generating graphs having 500 nodes or more. This filter will be described in a future work.
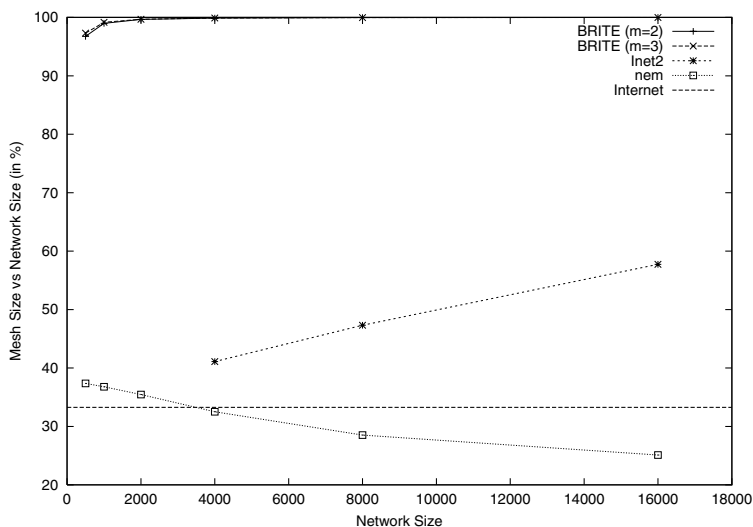
Fig. 5. Mesh Size vs Network Size (in %)

5.3   Mesh and Trees Properties

We have given the definition of the mesh in section 4. As BRITE 1 graphs are trees, they do not have a mesh and thus they will not be studied in this section. Figure 5 shows the plots of the mesh size in percentage of the graph size. The Internet has a mesh whose size represents 33% of the size of the whole Internet graph. This means that one node out of three in the Internet belongs to the mesh. We can already make a striking observation. BRITE 2 and 3 graphs have an average mesh proportion of nearly 100% ! (the lowest value is 96.74% for 500-node BRITE 2 graphs). This does not coincide with the Internet mesh proportion of 33%. Inet2 has values starting at 41% and increasing accordingly with the graph size. *nem* on the other hand has values starting at 37% and decreasing when the graph size increases. Nevertheless *nem* graphs seem to have the closest mesh proportion values to the Internet mesh proportion. It is worth noticing that Inet2 mesh proportions are accurate (particularly for 16000-node graphs) when compared with the AS level mesh proportion which equals 63%.

We have also studied the connectivity properties of the graph meshes and the Internet mesh. The number of cutpoints in the mesh of the Internet is equal to 3.7% of the total number of nodes in the mesh. In all the graphs that we studied this ratio was at most 0.39% which is much less than the Internet value. Cutpoints are important because the failure of a cutpoint router leads the network to be disconnected. Although we saw that cutpoints are rare, it is interesting to see how they partition the mesh. To examine this point, we calculated the sizes of the biggest biconnected component of the meshes. In the Internet, the biggest bicomponent contains 87% of the nodes of the mesh. This means that although the Internet mesh contains 3.7% cutpoints, these nodes only fraction a small part of the mesh (i.e. 13%), as the larger part is biconnected. Concerning the graphs, we saw that except for 500-node *nem* graphs whose average biggest bicomponent size

is 95.8% of the mesh, all the other graphs have a ratio above 97.8%. To conclude we can say that the generated graphs have almost entirely biconnected meshes, which is not really the case of the Internet mesh that still contains a few cutpoint nodes and bridge nodes (i.e. nodes not belonging to a cycle but on a path linking two cycles).

The forest is simply the set of nodes of the graph that do not belong to the mesh. These nodes are located in trees and the union of these trees form the forest. The trees are connected to the mesh by special nodes called roots. We consider the roots as nodes belonging to the mesh. As we said before, BRITE 1 graphs are trees. Any one BRITE 1 graph of size n has exactly $n - 1$ edges and is connected hence it is a unique tree. This implies that BRITE 1 graphs are always composed of one tree and thus the assumption that the graphs have forests with multiple trees is false for BRITE 1 graphs. Hence we will not study BRITE 1 graphs in the rest of this section.

Figure 6 shows the plots of the number of trees (i.e. the number of roots) *vs* the number of nodes in the graph. This is an interesting measure as root nodes represent the connection points to the in-tree nodes of the graph. We can see that BRITE 2 and 3 graphs have nearly 0% trees except for small sized ones. This is a consequence of what we saw in the previous section: nearly all the nodes of BRITE 2 and 3 graphs belong to the mesh part. When we look closer at these graphs we see that they only have a handful of trees, most of them having depth one (i.e. nodes are directly connected to the roots). Hence we have not been able to calculate tree power laws for BRITE 2 and 3 graphs and thus we do not study them in the rest of this section. In the Internet the ratio of the trees *vs* the Internet size is 11.7% (more than one in-mesh node out of three is a root node). We can see that Inet2 and *nem* graphs have roughly similar values. For small sizes, *nem* graphs have a higher percentage because their mesh is smaller and thus we find more trees. It is worth noticing that Inet2 values, already above *nem* values, are far from the AS level tree percentage which equals 7.7%.

In the following of this section we examine the presence of tree power laws [8] defined by Magoni *et al.* in Inet2 and *nem* graphs. We found that the Internet complies with power laws 6 and 7 (i.e. tree rank exponent and tree size exponent) with a high degree of accuracy. It is worth noticing that Magoni *et al.* have already found that the AS level of the Internet complies with these tree power laws.

We do not show the plots of the tree size ACCs of *nem* and Inet2 graphs because we found that the values are close to or above 0.95. Thus all these graphs comply with power law 7 (tree size exponent).We also do not show here the plots of the tree rank ACCs of *nem* and Inet2 graphs because they are all above 0.97. Thus they all closely comply with power law 6 (tree rank exponent).

To summarize this section, we can notice that only Inet2 and *nem* graphs comply with the power laws concerning the trees. Their proportion of trees matches the Internet tree proportion. Because they do not have a sizeable number of trees, the tree size distributions of the BRITE 2 and 3 graphs cannot be used to calculate the ACCs and thus they do not comply with tree power laws 6 and 7. As both the AS level and the router level of the Internet comply with the tree power laws, we think that these laws are valid indicators of a graph topology similarity to the Internet topology.
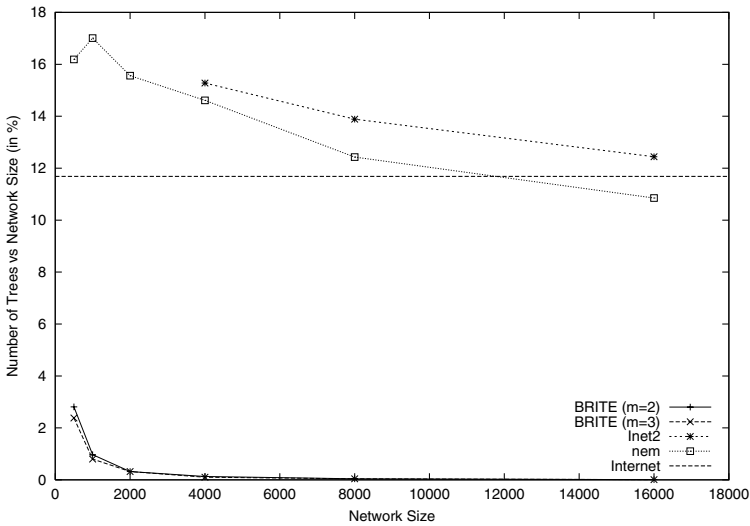
Fig. 6. Number of Trees *vs* Network Size (in %)

6   Conclusions

Designing Internet-like topology generators is not an easy task. The generated graphs have to comply with many laws and their topological properties must have suitable values that match the Internet ones. Ensuring that important power laws and Gaussian distribution averages are accurately reproduced is already a big step towards making a graph topology similar to the Internet one. In particular we think that the tree proportion and the tree power laws are interesting indicators for assessing the reliability of an Internet-like generated graph. The aims of our paper can be summarized as follows:

- Compare some of the latest generators that belong to the power law topology model, as a means to evaluating their performance.
- Analyze the generated graphs to examine how well their properties compare with the ones measured in the **scan+lucent** router level Internet map, as a means to evaluating their accuracy.
- Give, at the same time, results on the Internet topology properties inferred from the **scan+lucent** map (one of the biggest available Internet maps).

Our topology generator (*nem*) holds positive results and we hope that it will be helpful to the research community. The complexity of the Internet topology opens up the prospect of finding new properties and creating new generators in the near future.

References

1. William Aiello, Fan Chung, and Linyuan Lu. A random graph model for massive graphs. In *Proceedings of ACM STOC'00*, pages 171–180, 2000.

2. Réka Albert and Albert-László Barabási. Topology of evolving networks: local events and universality. *Physical Review Letters*, (85):5234, 2000.
3. Matthew Doar. A better model for generating test networks. In *Proceedings of IEEE GLOBE-COM'96*, November 1996.
4. Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *Proceedings of ACM SIGCOMM'99*, Cambridge, Massachusetts, USA, September 1999.
5. Ramesh Govindan and Anoop Reddy. An analysis of internet inter-domain topology and route stability. In *Proceedings of IEEE INFOCOM'97*, Kobe, Japan, April 1997.
6. Ramesh Govindan and Hongsuda Tangmunarunkit. Heuristics for internet map discovery. In *Proceedings of IEEE INFOCOM'00*, Tel Aviv, Israël, March 2000.
7. Cheng Jin, Qian Chen, and Sugih Jamin. Inet: Internet topology generator. Technical Report CSE-TR-433-00, University of Michigan, 2000.
8. Damien Magoni and Jean-Jacques Pansiot. Analysis of the autonomous system network topology. *Computer Communication Review*, 31(3):26–37, July 2001.
9. Damien Magoni and Jean-Jacques Pansiot. Internet topology analysis and modeling. In *Proceedings of IEEE Computer Communications Workshop*, Charlottesville, Virginia, U.S.A., October 2001.
10. Alberto Medina, Ibrahim Matta, and John Byers. On the origin of power laws in internet topologies. *ACM Computer Communication Review*, 30(2), April 2000.
11. Jean-Jacques Pansiot and Dominique Grad. On routes and multicast trees in the internet. *ACM Computer Communication Review*, 28(1):41–50, January 1998.
12. Bernard Waxman. Routing of multipoint connections. *IEEE Journal on Selected Areas in Communications*, 6(9):1617–1622, December 1988.
13. Ellen Zegura, Kenneth Calvert, and Michael Donahoo. A quantitative comparison of graph-based models for internetworks. *IEEE / ACM Transactions on Networking*, 5(6):770–783, December 1997.