

On Realistic Network Topologies for Simulation

Oliver Heckmann, Michael Piringer, Jens Schmitt, Ralf Steinmetz

Multimedia Communications (KOM),
Department of Electronic Engineering & Information Technology &
Department of Computer Science
Darmstadt University of Technology, Germany,
Merckstr. 25, 64283 Darmstadt, Germany
Tel.: +49-6151-16-6163
Fax: +49-6151-16-6152
{Heckmann, Piringer, Schmitt, Steinmetz}@kom.tu-darmstadt.de

Abstract — Simulations are an important tool in network research. As the selected topology often influences the outcome of the simulation, realistic topologies are needed to produce realistic simulation results. We first discuss the different types of topologies and present our collection of real-world topologies that can be used for simulation. We then define several similarity metrics to compare artificially generated topologies with real world topologies. We use them to find out what the input parameter range of the topology generators of BRITE, TIERS and GT-ITM are to create realistic topologies. These parameters can act as a valuable starting point for researchers that have to generate artificial topologies.

1 Introduction

1.1 Motivation

Simulations are an important tool in network research. The properties of a selected topology often crucially influence the outcome of the simulation, realistic topologies are therefore highly desirable for realistic simulation results.

There is a lot of work [1, 2, 3, 4, 5, 6, 7, 8, 9] discussing the properties of Internet topologies and there are a number of different topology generators, e.g. [10, 11, 12, 13]. These generators offer a big range of configuration parameters, GT-ITM for example has 16 different configuration parameters (for the transit-stub model).

How realistic a generated topology is depends on the combination of these parameters. How should a researcher set these parameters? With this paper we try to answer that question and furthermore try to make an initial comparison of existing topology generators.

We discuss the different types of topologies depending on the level of abstraction and introduce the term “POP-level” topology arguing that people should carefully distinguish between POP-level and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM SIGCOMM 2003 Workshops, August 25&27, 2003, Karlsruhe, Germany.
Copyright 2003 ACM 1-58113-748-6/03/0008...\$5.00.

router-level topologies. We also point to our library of real-world topologies.

Usually artificially generated topologies are judged realistic or not by pure visual inspection. In this paper, we define objective criteria (similarity metrics) for the similarity between two topologies. Using these metrics we compare artificial topologies with real-world topologies. We present the optimal combinations of input parameters for three different topology generators that yield the highest similarity to two real-world topologies. Researchers can take these parameter combinations as a starting point and vary them to generate topologies that are similar to real-world topologies.

We conclude with a short summary and outlook, our tools are available at: <http://www.kom.tu-darmstadt.de/heckmann/topologies/>.

1.2 Related Work

In [1] powerlaw relationships are found in three inter-domain (AS-level) topologies of the Internet which were constructed from BGP data. This paper opened up a discussion about powerlaw AS-level topologies. [2] investigate based on the work of Barabasi and Albert [14] possible origins of these power laws using topology generators to create artificial topologies, among them are BRITE and GT-ITM which we also use in this paper. More works [6, 5, 8] are based on the powerlaw relationship.

[3] show that during the process of constructing the topologies of [1] from BGP data 20 to 50% of the physical links are missed and that more exact topology graphs do not follow the powerlaw relationship found in [1]. The authors also show that works based on [14], e.g. [2], are not supported by the more exact topologies.

Contrary to the papers mentioned above for the basis of this paper we do not investigate whether the powerlaws hold true or not. In addition, we focus on POP/router-level instead of AS-level topologies. We use the powerlaw metrics in our metrics with low weights because they are under heavy discussion but we include additional metrics that can capture the distribution of, e.g., the outdegree independent of whether it follows powerlaw relationships or not.

Similar work to ours is done in [4] but for AS level graphs with at

least 1000 nodes in order to evaluate topology generators for AS level graphs. We investigate smaller and for POP/router-level graphs.

[9] presents an interesting method of measuring router-level topologies based on traceroutes, BGP and DNS data. The results look very promising and can be used as input for our tool.

2 Types of Topologies

Literature typically distinguishes between AS-level topologies where one node represents one autonomous system (AS) and router-level topologies which implies that one node represents one router. But the latter is not always the case, many topologies that are labeled “router-level” are actually what we call POP-level with one node representing one point of presence (POP) of a provider. Take for example the AT&T topology depicted in Figure 1 which is quite famous and often used for simulations. The New York node in that topology does surely not represent a single router but instead the complete POP of AT&T in New York consisting of a larger number of interconnected backbone and access routers. Using the data from [9] which is based on BGP and DNS data AT&T has at least 61 routers in New York, 24 of them being backbone routers.

3 Similarity Metrics

To measure the similarity of two network topologies we define a set of metrics that capture the basic connectivity properties of the topology graph. We are interested in graphs with the same connectivity properties but not necessarily in equivalent graphs. We are not interested in any geographical information or whether the graphs look visually similar. We concentrate on the shortest paths, the distance between nodes and the degree of connectivity because these are the properties we deem most important for network simulations.

In the graphs, we can distinguish between edge nodes (which are connected to end-users and other networks) and core/backbone nodes (which are only connected to nodes of the same network).

We define the following similarity metrics to calculate the similarity between a generated topology and a real-world reference topology:

1. The first metric uses the hop-plot of all nodes. For each graph G we look at all n nodes and calculate how many other nodes can be reached on their shortest path within $h=1,2,3\dots$ hops. From this we derive the relative frequency distribution F_h^G with J classes j with frequencies $f_{h,k}^G$. We have to compare the frequency distributions of the generated and the reference topology. We could do this by comparing the mean values but obviously two very different distributions can have the same mean value, so that would not be a good criterion. Comparing the standard-deviation has the same problem. Therefore, we sum up the absolute cumulative difference of each class j :

$$\sum_j \left| \sum_{k=1}^j f_{h,k}^{generated} - \sum_{k=1}^j f_{h,k}^{reference} \right| \quad (1)$$

An example demonstrates the logic behind this:
Consider the two frequency distributions

$$f_{h,j}^1 = \{ 0.5, 0.0, 0.5 \} \text{ and}$$

$$f_{h,j}^2 = \{ 0.0, 1.0, 0.0 \}$$

Both have the same mean value but are obviously very different. Our criterion would return a difference d of

$$d = 0.5 + 0.5 + 0.0 = 1.0$$

Two equal distributions would return a difference of $d = 0.0$.

2. The second metric F_h^{Ge} is similar to the first but only looks at edge nodes. It captures how many other edge nodes can be reached from an edge node within h hops.
3. We derive the relative frequency distribution of the outdegrees d_i of all nodes i of one graph. We use the significance level of a Wald-Wolfowitz test for the similarity of the distributions of the two graphs as third similarity metric.¹
4. While the third criterion tests the complete distribution of outdegrees we also compare with the fourth criterion the rank exponent \mathfrak{R} of the first powerlaw as defined in [1]: The outdegree d_i of a node i is proportional to the rank of the node r_i to the power of a constant \mathfrak{R} :

$$d_i \propto r_i^{\mathfrak{R}} \quad (2)$$

We compare the relative difference of the rank exponent of the artificially generated topology and the real-world reference topology:

$$\left| \frac{\mathfrak{R}_{generated} - \mathfrak{R}_{reference}}{\mathfrak{R}_{reference}} \right| \quad (3)$$

5. Because the powerlaws are a well discussed phenomenon we also include the outdegree exponent O of the second powerlaw [1]. This fifth criterion is quite similar to the third one, we account for that in the weights for this criterion later one. The second powerlaw claims that the frequency f_d of an outdegree d is proportional to the outdegree to the power of a constant O :

$$f_d \propto d^O \quad (4)$$

We use the relative difference of the outdegree exponents as the fifth criterion

$$\left| \frac{O_{generated} - O_{reference}}{O_{reference}} \right| \quad (5)$$

6. As sixth criterion we use the relative difference in the number of nodes n :

$$\left| \frac{n_{generated} - n_{reference}}{n_{reference}} \right| \quad (6)$$

7. For the seventh criterion we use the relative difference in the number of links l :

$$\left| \frac{l_{generated} - l_{reference}}{l_{reference}} \right| \quad (7)$$

1. The Wald-Wolfowitz test did not discriminate well enough for the first criterion but good enough for this criterion.

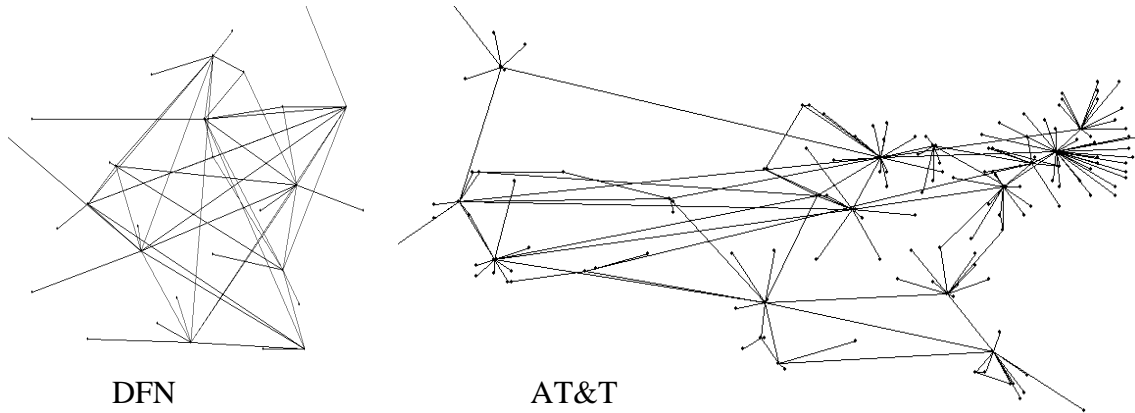


Figure 1: The DFN and AT&T Topologies

All of these criteria were normalized so that 1.0 represents the highest similarity. Some criteria can theoretically return negative values if topologies are extremely different but that was of no significance as we are looking at relatively similar topologies.

The metrics mentioned above and some additional metrics are implemented in our Topology Comparison Tool. It is written in Java and can be easily extended with additional metrics. Every metric can be given a weight and based on those the tool calculates the difference between two topologies normalized to a number between 0.0 and 1.0.

4 Fitting based on Similarity Metrics

Based on the similarity metrics of the last section we try to determine the input parameters of the most often used and referenced topology generators BRITE [10], TIERS [11] and GT-ITM [12] for two well-known real-world topologies (see Figure 1). The 154 node AT&T topology is quite famous and often used for simulations. The 30 nodes DFN (German Research Network) topology is also often used in German projects. Both topologies are POP-level topologies. We did not have any router level topologies at the time of our experiments. It is quite hard to get real-world topologies, ISPs are somewhat reluctant to reveal information about their topologies. Because of this we started to collect available real-world topologies, they are publicly available at our topology site: <http://www.kom.tu-darms-tadt.de/~heckmann/topologies/>.

4.1 Fractional Factorial Analysis

We started with a fractional factorial analysis to determine the influence of the different input parameters of the topology generators on the different similarity metrics. The results were very disappointing because of the strong non-linear influence of practically all input parameters on the similarity metrics. Only for small parameter variations the linear model which the factorial analysis is based upon can be assumed. Because of these results another search heuristic to search the parameter space was implemented.

4.2 Systematic Search

The topology generators use random numbers in their topology cre-

Criteria	Weight
1 with h = 3 with h = 5 with h = 10	2 2 2
2 with h = 3 with h = 5 with h = 10	2 2 2
3	5
4	2
5	2
6	5
7	5

Table 1: Weights

ation algorithm. We first tested how similar topologies created with the same parameter set are. For all topology generators the combined similarity metric varies less than 2% if the same parameters set is used. So the topology generators return stable results if the parameters are unchanged, the influence of the random number generator on the similarity metrics is not strong.

We then used a heuristic similar to Hooke & Jeeves [15] to search for the parameter combination that yields the maximum combined similarity metric. The weights chosen for the different metrics from Section 3 are summarized in Table 1. We are aware of the fact that these weights are subjectively chosen. We put our emphasis on the number of nodes and links and the distributional criteria and not on power-laws. Our tool allows to add further metrics and change the weights

to what the researcher deems important for his experiment.

4.3 Results

The parameters of Table 2 were found for Brite and the DFN and AT&T topologies with a high and satisfying combined and normalized similarity of 0.972 resp. 0.951. Please note that the values α and β do not seem to significantly influence the outcome of the measurements when the parameter links/node is set to 2.

The parameters for Tiers result in a similarity of 0.998 and 0.995, the

highest similarities found in our experiments, they are given Table 3.

The results for GT-ITM are displayed in Table 4 and Table 5 and have a similarity of 0.966 resp. 0.879.

To conclude, Tiers was able in both cases to produce topologies that had the highest similarity to the real world ISP topologies, GT-ITM produced the least similarities. The level of similarity that could be reached is quite high and indicates that hierarchical topology generators are able to produce realistic POP level topologies. This is contrary to the findings of [1] for AS level topologies.

topology	type	method	AS	nodes	router model	α	β	links/node
DFN	Bottom up	random pick	17	30	GLP	0.42-0.46	0.62-0.68	3
AT&T	Bottom up	random pick	31	154	GLP or BA	[0, 1.0]	[0, 1.0]	2

Table 2: Parameters of Brite for DFN / AT&T-like Topologies

DFN	WAN	MAN	LAN	nodes/WAN	nodes/MAN	nodes/LAN
	1 ^a	1	1	9	4	17
	redundancy for WAN	redundancy for MAN	redundancy for LAN	redundancy for MAN to WAN	redundancy for LAN to MAN	
	6	4	1 ^a	7	2	
AT&T	WAN	MAN	LAN	nodes/WAN	nodes/MAN	nodes/LAN
	1 ^a	1	3	25	6	41
	redundancy for WAN	redundancy for MAN	redundancy for LAN	redundancy for MAN to WAN	redundancy for LAN to MAN	
	3	4	1 ^a	3 - 4	4	

Table 3: Parameters of Tiers for DFN / AT&T-like Topologies

a. Parameter cannot be changed in TIERS 1.2

method	avg stubs/transit	extra t-s links	extra s-s links
transit-stub	1	10	6
top nodes	edge method	alpha	beta
1	3	0.99	-
transit nodes	edge method	alpha	beta
5	4	0.35	100
stub nodes	edge method	alpha	beta
5	2	0.5	100

Table 4: Parameters of GT-ITM for DFN-like Topologies

method	avg stubs/transit	extra t-s links	extra s-s links
transit-stub	3	12	12
top nodes	edge method	alpha	beta
3	3	0.3	-
transit nodes	edge method	alpha	beta
4	3	0.5	-
stub nodes	edge method	alpha	beta
4	3	0.2	-

Table 5: Parameters of GT-ITM for AT&T-like Topologies

Further experiments showed that the similarity with regard to most metrics (except of course the number of links and nodes metrics) remains roughly equal if the number of nodes and links are increased proportionally for all topology generators. The parameters found seem to be scalable to larger topology graphs and thus are a valuable hint for researchers.

5 Summary and Outlook

This work should help the researcher in finding or creating realistic topologies. We first introduced the term POP-level topologies, many topologies that are called “router-level” are in fact POP-level topologies and no true router-level topologies.

We also introduced our collection of real-world topologies that can be used for simulations. They are available at our topology page (<http://www.kom.tu-darmstadt.de/~heckmann/topologies/>) in human-readable GML [16] file format and as an NS2 OTcl script that can directly be used for NS2 simulations. On that page we also offer our software, for example a topology converter that can read the most common topology file formats.

In this work, we also presented similarity metrics for comparing network topologies and based on these we derived the combination of input parameters for 3 topology generators that lead to the highest similarity with 2 real world ISP-level topologies. The results show ranges of parameter combinations that generate realistic topologies and can act as a starting point for anybody who wants to do realistic ISP level simulations.

We are aware that our results are only estimations for a limited number of topologies and metrics and plan to continue it using more topologies as well as more and different combinations of similarity metrics.

Acknowledgments

This work is partly sponsored by the german research network provider DFN (<http://www.dfn.de>) as part of the LETSQoS project (<http://www.letsqos.de/>).

References

- [1] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On Power-law Relationships of the Internet Topology. In *Proceedings of SIGCOMM 1999*, pages 251–262. IEEE, 1999.
- [2] A. Medina, I. Matta, and J. Byers. On the Origin of Power Laws in Internet Topologies. *Computer Communications Review*, April 2000. <http://www.cs.bu.edu/brite/publications/ccr00.pdf>.
- [3] Q. Chen, H. Chang, R. Govindan, S.Jamin, S. Shenker, and W. Willinger. The Origin of Power Laws in Internet Topologies Revisited. In *IEEE Infocom 2002*. IEEE, 2002.
- [4] H. Tangmunarunkit, R. Govindan, S.Jamin, S. Shenker, and W. Willinger. Network Topologies, Power Laws, and Hierarchy. Technical Report Technical Report USC-CS-01-746, University of Michigan, USA, 6 2001.
- [5] T. Bu and D. Towsley. On Distinguishing between Internet Power Law Topology Generators. In *INFOCOM '2002*. IEEE, 2002.
- [6] W. Aiello, F. Chung, and L. Lu. A Random Graph Model for Power Law Graphs. *Experimental Mathematics Volume 10*, pages 53–66, 2001.
- [7] E. W. Zegura, K. Calvert, and M. J. Donahoo. A Quantitative Comparison of Graph-based Models for Internet Topology. *IEEE/ACM Transactions on Networking*, December 1997.
- [8] C. R. Palmer and J. G. Steffan. Generating network topologies that obey power laws. In *GLOBECOM '2000*, November 2000.
- [9] N. Spring, R. Mahajan, and D. Wetherall. Measuring ISP Topologies with Rocketfuel. In *SIGCOMM '2002*. ACM, 2002.
- [10] BRITE. Boston University Representative Internet Topology Generator. <http://www.cs.bu.edu/brite/>.
- [11] TIERS. Tiers Topology Generator. <http://www.isi.edu/nsnam/ns/ns-topogen.html#tiers>.
- [12] GT-ITM. Georgia Tech Internetwork Topology Models. <http://www.cc.gatech.edu/projects/gtitm/>.
- [13] Inet Topology Generator. <http://topology.eecs.umich.edu/inet/>.
- [14] A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, October 1999.
- [15] R. Hooke and T. A. Jeeves. Direct search solution of numerical and statistical problems. *Journal of the Association for Computer Machinery*, pages 212–229, 8 1961.
- [16] O. Heckmann, K. Pandit, J. Schmitt, M. Hoffmann, and M. Jobmann. LETSQoS Milestone 2. <http://www.letsqos.de>, June 2002.