



Technical Report

**KU-LocGen:
Location and Cost-Constrained
Network Topology Generator**

Abdul Jabbar, Qian Shi, Egeman Çetinkaya

and James P. G. Sterbenz

ITTC-FY2009-TR-45030-01

December 2008

Project Sponsor:
National Science Foundation

KU-LocGen: Location and Cost Constrained Network Topology Generator

Abdul Jabbar, Qian Shi, Egemen K. Çetinkaya,
and James P.G. Sterbenz

Abstract

Realistic topology generators are essential to the understanding of network design and survivability analysis. Two important issues that are not sufficiently addressed by current topology generators are node-positioning and cost considerations. We propose that the utility of the existing models could be vastly improved by incorporating these two features. In this paper we introduce a new network topology generator *KU-LocGen*, which enables node positioning in several well-known random graph generation models. We conduct our studies based on two backbone networks. We show that the proposed generator produces graphs that are representative of the real network as well as realistic alternatives. Finally, we present a cost analysis methodology and apply it to our topology generator.

Information and Telecommunication Technology Center
Department of Electrical Engineering and Computer Science
The University of Kansas

31 December 2008

Contents

1	Introduction	1
2	Related Work	2
3	Node Positioning	3
4	Random Link Generation Models	4
4.1	Pure Random	4
4.2	Locality	4
4.3	Waxman	5
5	Network Costs	5
6	Modelled Networks	6
7	Results	7
7.1	Topology Representativeness	7
7.1.1	Node degree distribution	9
7.1.2	Shortest path length	9
7.1.3	Link lengths	10
7.2	Cost Analysis	10
7.3	Comparative Analysis and Model Shortcomings	13
7.3.1	Location constrained pure random	14
7.3.2	Location constrained locality model	16
7.3.3	Location constrained Waxman model	18
8	Conclusions and Future Work	22

1 Introduction

Realistic topology generators are crucial to numerous aspects of networking research. In particular, there are three distinct applications of topology generators [1, 2]: understanding the graphical properties of the network and evaluating the performance of protocols and services over a given topology; resilience and survivability analysis of the network to determine how well the network will react to challenges; and finally, a tool for network architects, providing alternate topologies that meet certain constraints during the design and engineering phase.

The current emphasis of the topological studies primarily focus on the first aspect: descriptive modeling of the graphical properties of a network. There are large number of research efforts that characterize the topology and growth of internetworks using random models like Waxman [3], degree oriented models such as power-law [4] and BA model [5], as well as hierarchy centric models such as Transit-Stub and Tiers [6, 7]. The properties of interest include expansion, distortion, degree distribution, shortest-path distribution amongst a host of others. Several tools such as BRITE [8] and GT-ITM [6, 7] are currently available that can generate one or more different types of network topology graphs.

To some extent, existing studies also address the resilience of a graph [1]. A large section of this area of research is focused on the availability of the alternate paths between nodes and the commonly used challenge model is uncorrelated link failures. However, in practice the challenges at the physical layer tend to be highly correlated to the geographic location. In order to support extensive survivability analysis for location based challenges, it is essential for topology models to consider realistic node positioning. Resilience studies would greatly benefit if established metrics and measures are combined with simulation based studies using realistic topology models.

The emphasis of the topology generators, traditionally, has not been on providing practical tools for network designers of commercial or research networks. The primary reason for this disconnect between the network topology research and real network deployment is the lack of physical location and cost constraints in the topology models, which are two of the most important factors relevant to practical design and deployment. Therefore, design and engineering of commercial networks commonly does not involve rigorous topology analysis using widely researched topology generators.

At a broader level, lack of realistic topologies implies that the perfor-

mance of protocols and mechanisms, especially with respect to resilience, evaluated using simulated topologies does not translate well to actual network deployments. Our objective is to modify the existing topology generators by incorporating location and cost parameters in generation process. To this end, we present a topology generator that uses node positioning to improve the relative representativeness [9] of the network: KuLocGen. We use this generator to investigate the representativeness of topologies generated with random graph models. Given the established shortcomings of random graph models [10, 7], we evaluate the impact of node positioning (with predetermined geographic locations) on their credibility. In addition, we show how cost analysis can be used to derive the range of model parameters that would yield economically feasible topologies. Thus, our contribution is the use of location and cost constraints to derive realistic candidate topologies for practical network design, performance and survivability analysis.

The rest of the paper is organized as follows: The related work is discussed in section 2. Section 3 presents our approach to node positioning followed by its application to random graph models in Section 4. Network costs are discussed in Section 5. We evaluate the proposed models using node locations from two currently deployed networks in Section 6. Finally, an analysis of the generated topologies is presented in Section 7 followed by our conclusions.

2 Related Work

Several research efforts have addressed the aspect of realistic topology generation in the past. A recent comprehensive survey by Haddidi *et al.* [9] on the measurement, inference, modeling, and generation of network topologies highlights the lack of realistic models as a key issue. A majority of network modeling studies are aimed at modeling the Internet as a whole or in parts. Zegura *et al.* [7] studied the relationship between graph models and the real internetwork in terms of topological properties, several of which will be used in this paper. There exists a number of well-established measurement techniques to discover the network topology [11]. The topology maps generated using these techniques are also used in this paper.

A comparison of degree-based and structural network topology generators [1] based on expansion, distortion, and resilience showed that degree based generators are better equipped to capture the hierarchical structure of the Internet at a large scale. Alderson *et al.* [2] present an analysis of the

physical connectivity of single ISP networks and concludes that connectivity patterns of random generated models differs significantly from the inferred router level connectivity.

The need to incorporate realistic design constraints on topology generation was recognized by Alderson *et al.* [12], in which they present a topology generator based on descriptive model called HOT (Highly Optimized Tolerance). Recently, Quinton *et al.* [13] has emphasized the need for incorporation of network design heuristics in the topology models. The IGen tool [13] uses heuristics such as geographic location of the nodes instead of the probabilistic methods to generate topologies. Other models considering the geographic location of nodes include the one presented by Chen *et al.* [14] which aims at finding the optimum location of optical switching nodes and links to minimize the cost in an optical network. Habib [15] investigates the redesigning of an existing network topology with the objective of preserving network elements as well as the original topology, thereby reducing the cost and time to re-engineer the network. To the best of our knowledge, there has not yet been a study that investigates generation of small-scale (e.g. PoP level) network topologies based on geographically fixed located nodes under a cost limitation.

3 Node Positioning

We assert that there are two phases of topology generation: first the positioning of the nodes and second the generation of interconnecting links. Most of the existing graph models, as discussed in Section 2, focus on the latter. The physical location of nodes with respect to each other is commonly a random distribution, however in practice, network designers are almost always constrained on node locations. For example, the node location (point-of-presence) of any major ISP is guided by several economic and policy decisions and plays a dominant role in the resulting network topology. Therefore, we address this issue from a practical perspective. Instead of a random node location, we propose a fixed position for every node in the topology based on pre-determined locations. In case of a commercial network such as an ISP, this information is often a given design constraint. Hence our problem can be formulated as that of topology generation given a set of PoP locations. In this paper, we consider several random graph models to generate edges between a fixed set of nodes as discussed below.

Assigning positions to several hundreds or thousands of nodes is practical assuming that their coordinates are available in parsable format. However, the objective of our approach is to provide a practical tool that can be used for backbone or nationwide tier-1 networks as opposed to generating the complete Internet structure. Therefore, scalability is not considered to be a major concern in the proposed approach. Moreover, several highly scalable methods as discussed in Section 2 already exists that address generation of scale-free and hierarchical networks.

4 Random Link Generation Models

For the purpose of initial study, we choose three random graph generation models. In each case we use exact node placement instead of a random node distribution while using the original link generation algorithm from the model. The shortcomings of random graph models such as inability to capture all the statistical properties of large-scale graphs are well known. Our objective is to evaluate the impact of node positioning on the generated topologies and their statistical properties.

4.1 Pure Random

In this model, the link between a pair of nodes is generated based on a independent probability P . This is the most simplistic model and only used for reference purposes.

4.2 Locality

The locality model [7] is based on the euclidian distance between a given pair of nodes. First, nodes are placed according to the given data. The probability of a link between a node A and B is then specified as:

$$\begin{aligned} P_{AB} &= \alpha \text{ if } d_{AB} < \tau \\ P_{AB} &= \beta \text{ if } d_{AB} \geq \tau \end{aligned} \tag{1}$$

where τ is a constant threshold and $0 < \alpha < 1$ and $0 < \beta < 1$. The distance based model uses two different probabilities for links shorter and longer than a given threshold.

4.3 Waxman

The Waxman model [3] was first introduced 1988 as a random topology generator in which the nodes are initially placed in plane using with a uniform distribution. The edges between the nodes are added according to the probability that is dependent on the geographical distance between the nodes. The probability of an edge between two nodes A, B is given as:

$$P_{AB} = \alpha e^{\frac{-d}{\beta L}} \quad (2)$$

where d is the distance between the nodes, L is the maximum distance between any two pair of nodes, $0 < \alpha \leq 1$ is a constant, and $0 < \beta \leq 1$ represents the ratio of short distance to long distance links. An increase in α increases the probability of link between a given pair of nodes, whereas an increase in β increases the likelihood of longer links as compared to shorter links. Existing implementations use a Poisson distribution for node placement.

5 Network Costs

The cost analysis of the generated topologies provides valuable perspective on a graph model. In practice, the capital available for network deployment is limited. Therefore, it is essential that if not directly integrated in to the model, the cost should be at least used to obtain a realistic set of model parameters. In this paper we use a realistic cost function with arbitrary units¹ to determine the range of model parameters that provide a feasible solution. Given this range of “affordable” parameters, we can then analyze which topologies optimize the performance. The cost a link between a pair of nodes is defined as:

$$C = f_c + v_c \times d \quad (3)$$

where f_c is the fixed cost associated with deploying a link of any length, v_c is the variable cost expressed per unit distance, and d is the euclidian distance between the nodes.

¹For simplicity, the smallest cost factor is assigned a unit value

6 Modelled Networks

We choose two currently deployed networks to evaluate our topology models. First, we choose the router-level map of Sprint network as shown in figure 1(a) generated by the Rocketfuel project [16]. Secondly, we consider the PoP level graph from the European research network GÉANT2 [17] shown in figure 1(b).

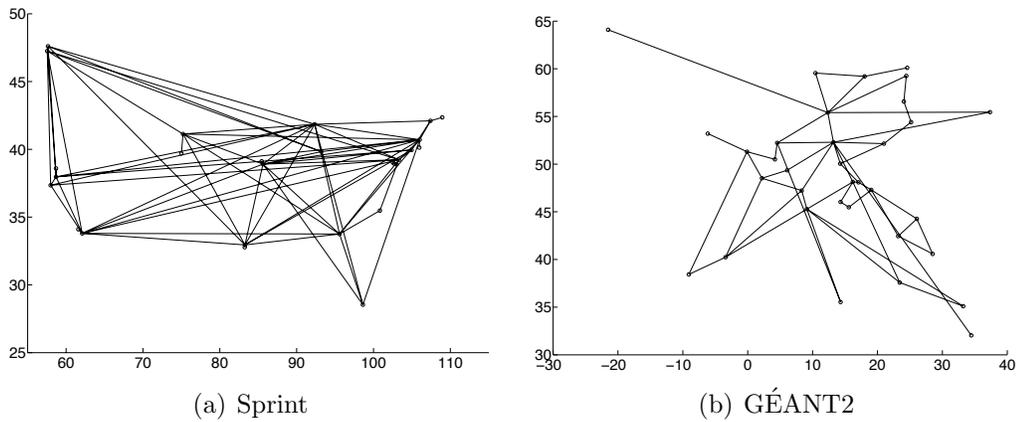


Figure 1: Router-level topology mapped to geographic coordinates

A summary of these networks is give in Table 1. We use the geographic locations from these two graphs to position the nodes in each of the models discussed in Section 4. The results are discussed in the following section.

Table 1: Statistics of modelled network topologies

Network	Number of Nodes	Number of Links	Average Degree
Sprint	27	68	5.04
GÉANT2	34	51	3.06

7 Results

We implemented all three models discussed in Section 4 in MATLAB². In this section, we present a set of topologies generated using the KU-LocGen and evaluate their representativeness. For each model, we generated topologies over the entire parameter (α, β) range with a resolution of 0.01 and repeated each point in the range 100 times. Thus the total number of runs was $(100 \times 100 \times 100) = 1,000,000$. In order to demonstrate the benefits of the proposed approach, we do a comparison between the topologies generated with the original Waxman model and the modified Waxman model in Section 7.1. A more extensive comparison between all the three random models along with the proposed modifications are given in Section 7.3.

7.1 Topology Representativeness

The representativeness of a topology is defined as the qualitative and quantitative measure of the similarity between a real network and the generated topologies. In this section, we show a sample topology generated by the the Waxman model for Sprint and GÉANT2 network. Note that this is a typical topology from range of a over a million topologies generated. Statistical properties of the graphs will be considered in the following sections.

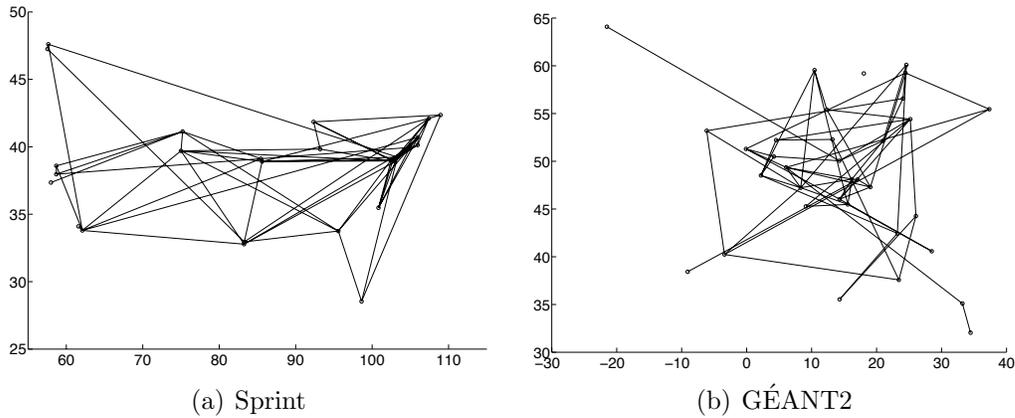


Figure 2: Topologies generated with modified Waxman model

²The topology generators and other code used in this paper is made publicly available for download at [18]

The topologies generated by the Waxman are most representative of the real network as shown in Figures 2(a) and 2(b)³. While the visual plots of the generated models show a fairly representative topology, it is well established that qualitative analysis of topologies is far from accurate [7]. Hence, we consider three metrics: node degree distribution, the shortest path length distribution, and the link length distribution commonly used in the literature for analysing the representativeness of topology generators.

In the following sections, we evaluate the modified Waxman model using the above mentioned metrics and compare it with the original Waxman model as well as the real network. Since the generated topology can vary significantly based on the model parameters (α, β in case of Waxman), we choose only those parameters that on average result in an average node degree similar to that of the real network. We call this the *feasible parameter set*. In other words, the number of edges in the real network and the simulated network are within a small tolerance. We generate 100 runs at each point within the feasible parameter set using a resolution of 0.01. The statistics (mean μ and standard deviation σ) of a sample run are shown in Table 2 for both Sprint and GÉANT networks.

Table 2: Statistical distributions of modelled network topologies

Topology	Degree (μ/σ)	Shortest Paths (μ/σ)	Link lengths (10^3 km) (μ/σ)
Sprint real network	5.04/3.88	2.44/1.10	1.41/1.18
Sprint original Waxman	4.74/1.75	2.25/0.87	1.15/0.63
Sprint modified Waxman	5.04/2.36	2.51/1.17	0.81/0.74
GÉANT2 real network	3.00/1.74	3.47/1.55	0.74/0.62
GÉANT2 original Waxman	3.06/1.48	3.03/1.19	1.63/0.84
GÉANT2 modified Waxman	3.06/1.76	3.34/1.42	1.20/0.71

In the following sections, we discuss specific representativeness metrics and compare their probability distributions. The plots are intended to give an idea of the shape of the distribution curves for a sample run. The exact

³Given the only constraint of node location, the generated topologies may suffer from key vulnerabilities such as cut vertices and nodes. Addition of further practical constraints such as k -connectivity and path diversity [19] remains a part of our future work.

difference between the models is shown quantitatively in Table 2.

7.1.1 Node degree distribution

The node degree distribution is defined as the probability that a randomly selected node is of degree k [9]. The average node degree for the Sprint and GÉANT2 network are plotted in Figure 3 for the real network, original Waxman, and modified Waxman. We see that modified Waxman reproduces the degree distribution more faithfully as compared to random node distributed Waxman model especially at the higher degrees. Table 2 shows that standard deviation of the modified Waxman model is much closer to the the real network.

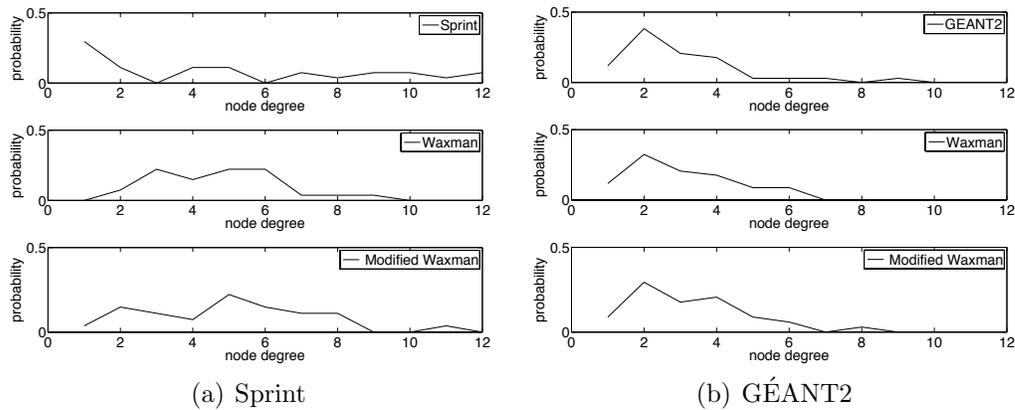


Figure 3: Node degree distributions

7.1.2 Shortest path length

The *shortest path length distribution* is defined as the distribution of the probability of the two randomly selected nodes being at minimum distance of k hops from each other [9]. We calculated this distribution for all feasible values of α and β , with 100 repeats per point. The distributions from one specific set ($\alpha = 0.95, \beta = 0.20$) are shown in Figure 4.

We observe that to some extent both the original Waxman and modified Waxman generate similar hop counts. This is expected since hop count depends primarily on the number of nodes and links which are equal by design for both the models. However, the difference is evident towards the tail of the distribution and this fact is captured in the statistics that show

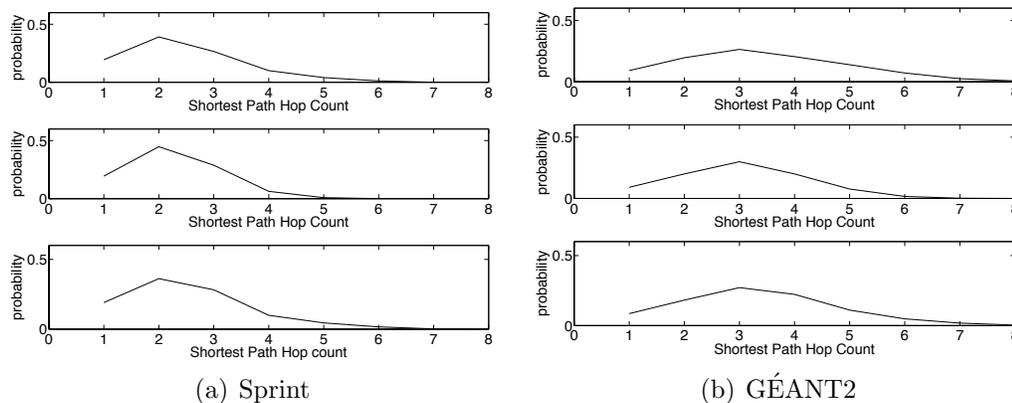


Figure 4: Shortest path length distributions

that the mean and standard deviation of the modified Waxman model is closest to the real network.

7.1.3 Link lengths

We define the *link length distribution* as the probability that a random edge in the network will be of length k such that $m < k < n$ where $n - m = \delta$ is a fixed length increment. Figure 5 shows that the modified Waxman model closely duplicates the link length distribution of the original network. The statistics show that over the entire range of the model parameters, modified Waxman is particularly representative of the real network. Link length distribution is of particular significance to economic feasibility as the overall cost of the network is directly proportional to the sum of all the link lengths in the deployed network.

7.2 Cost Analysis

In this section, we conduct a cost analysis of the topologies generated by the proposed models. The objective is to derive a range of economically feasible model parameters for a given value of total capital available. In order to generate topologies that cover the entire model range, we increment the parameters in steps of 0.01 and generate 100 runs for each set. For every single topology generated, we calculate the total cost of the network using Equation 7.2, which is repeated here:

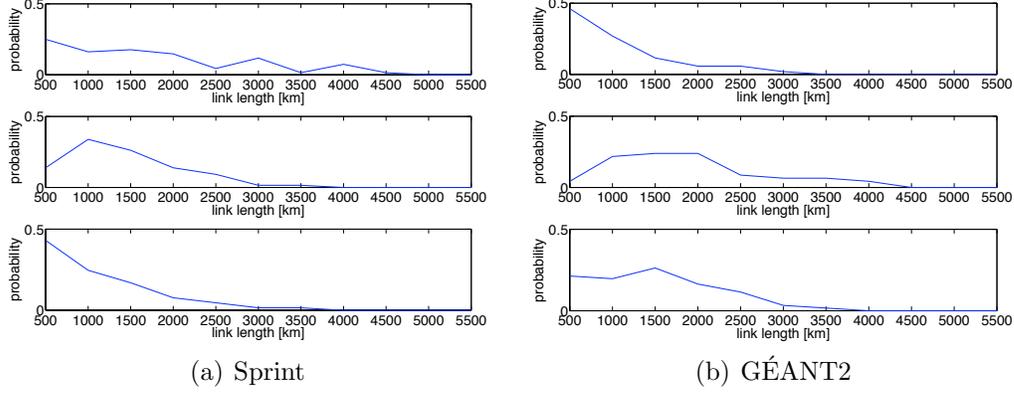


Figure 5: Link length distributions

$$C = f_c + v_c \times d$$

In order to calculate cost, we need a relationship between the fixed costs f_c and variable costs per unit distance v_c . This relationship is dependent on a number of market factors. For the purpose of analysis, we assume the fixed costs to be defined by the following equation:

$$f_c = \bar{d} \times v_c \tag{4}$$

where \bar{d} is the average distance between all node pairs. In other words, the fixed cost is assumed to be the average of variable costs of all the links. While the actual relationship may vary depending upon the geographic location of the network (urban vs. rural centers), this simplistic assumption demonstrates the method without loss of generality. Figures 6 show the cost function versus the model parameters for the locality and modified Waxman model respectively.

The next step is to find the model parameters such that the total network cost is within the maximum allowable cost given the economic constraints. For example, the range of feasible parameters α, β using the modified Waxman model and cost constraint of 5 million units is plotted in Figure 7.

Given the α, β pairs feasible within the cost constraint, the network designer can evaluate the relative merits of different generated topologies. These include optimizing connectivity, capacity, and resilience.

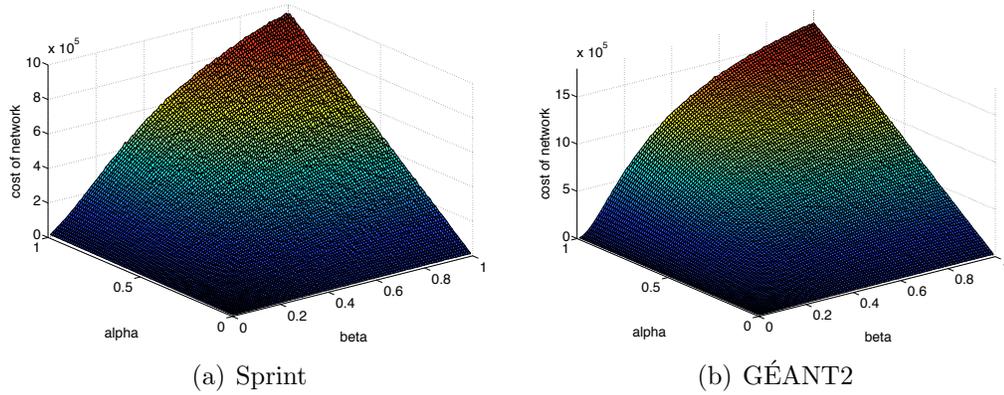


Figure 6: Cost function for modified Waxman model

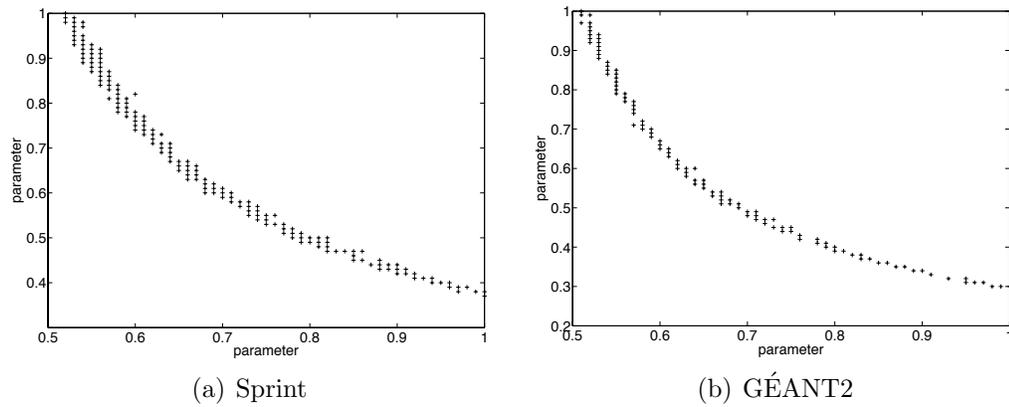


Figure 7: Parameters α and β that upper-bound the cost to 5M units

7.3 Comparative Analysis and Model Shortcomings

In this section, we present a detailed comparison of topologies generated using the proposed location constrained models with link probabilities generated from pure random, locality, and Waxman models. We conduct a qualitative analysis to evaluate the advantages as well as the shortcomings of the location constraints on the above mentioned models. We will use the node locations of two currently deployed networks for the purpose of our analysis. First we look at the router-level map of the tier 1 service provider Sprint, followed by a much sparser map of the GÉANT2 research network.

The objective of this section is to get a qualitative measure of the representativeness as well as resilience of graphs generated using the proposed model. This allows us to evaluate the potential shortcomings of the proposed approach and motivates future refinement to KU-LocGen that instruments further constraints on the generated topologies.

7.3.1 Location constrained pure random

Figure 8 shows three different topologies generated using the pure random model with the location constraints similar to that of nodes in the Sprint network shown in Figure 8(a). The resilience of graphs generated with pure random model is relatively high because the link probabilities that are equal independent of the distance. In order to generate graphs that match the node degree of the actual topology, we use a link probability α , where $\alpha = 0.19$.

It can be seen that the topologies generated can vary significantly between runs. While the topology of Figure 8(b) shows a resilient graph that closely approximates the real network, multiple runs yield graphs that lack resilience (Figure 8(c)) or results in sub-optimal paths leading excessive delays and bottlenecks in the network (Figure 8(d)).

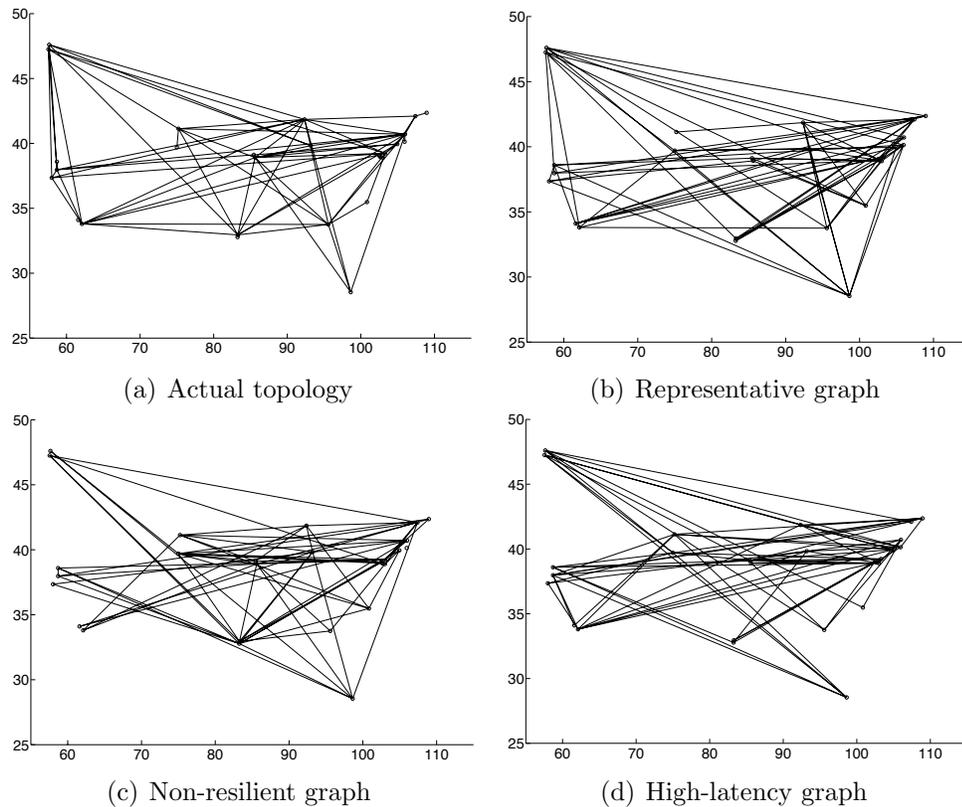


Figure 8: Location constrained pure random graphs for Sprint Network

The second set of graphs in Figure 9 show the topologies generated by the pure random model when the locations are constrained based on the GÉANT2 research network nodes of Figure 9(a). The link probability for this network is calculated to be 0.09.

Again, we observe a significant variance in the the representativeness of the graphs between multiple runs. Some of the graphs such as the one in Figure 9(b) are fairly representative of the actual topology and possess resilience characteristics, while others resulted in less than optimal graphs in term of resilience and latency (Figures 9(c) and 9(d)).

Equal link probabilities correspond to many long distance links, thereby reducing the probability of generating graphs with cut vertices and edges. Further, the graph is often bi-connected. However, the cost of such a network is very high.

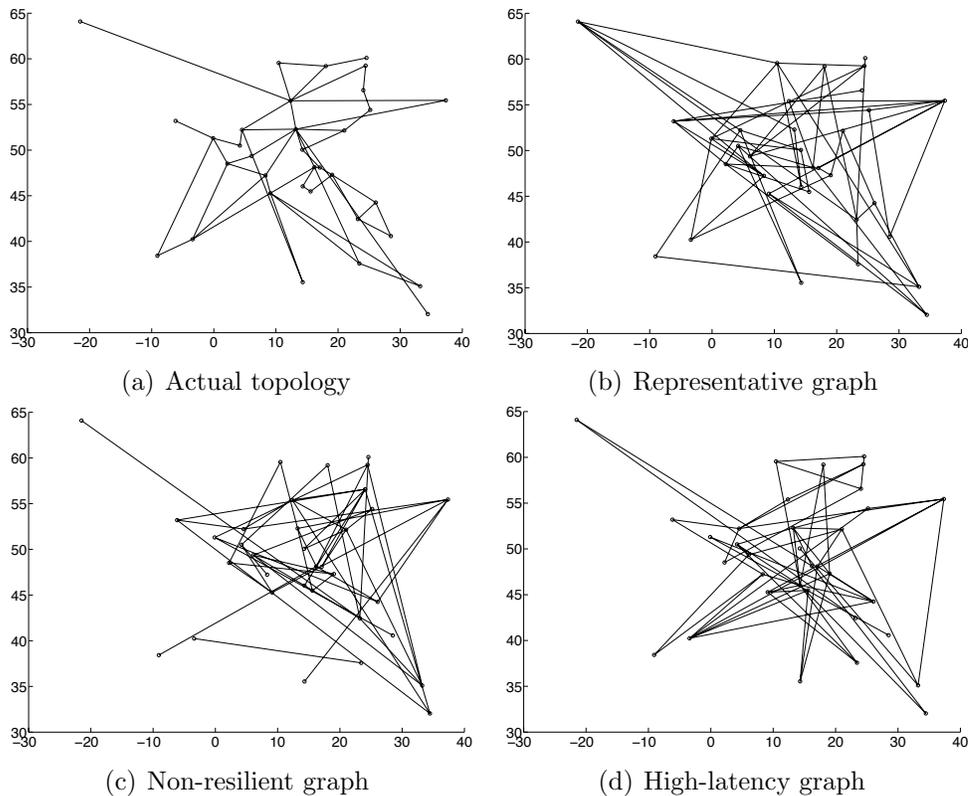


Figure 9: Location constrained pure random graphs for GÉANT2 Network

7.3.2 Location constrained locality model

In this section, we look at the differences in topologies generated using locality model as shown in Figure 10. Recall that locality model uses two different link probabilities based on a distance threshold. The value of this threshold and the link probabilities is calculated such that the resulting plots have, on average, the same node degree distribution as that of the original topology. For the Sprint topology, the model parameters are $\alpha = 0.4$, $\beta = 0.1$, and the distance threshold is 1000 km.

We observe that the locality model can generate graphs (see Figure 10(b)) that are more representative of the actual topology as compared to the pure random model. However, there are cases where the generated graph may lack key resilience features (Figure 10(c)), or have sub-optimal link placement (Figure 10(d)) resulting in unrealistic topologies.

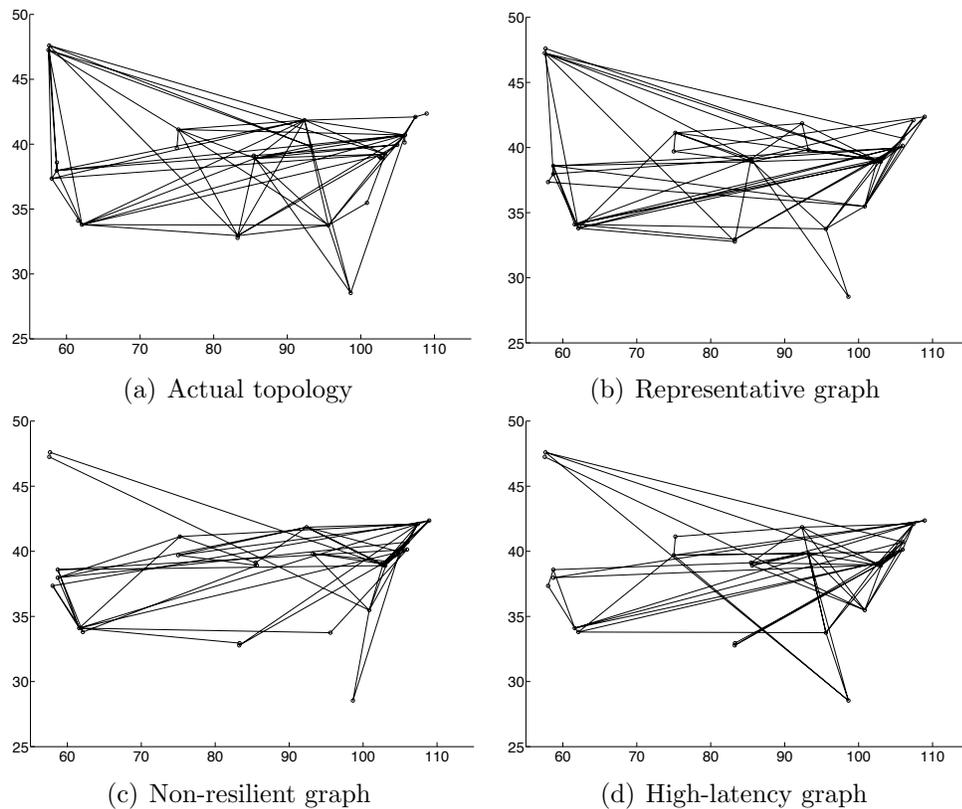


Figure 10: Location constrained locality graphs for Sprint Network

Figure 11 shows sample topologies based on the node locations of the GÉANT2 network using the locality model. Here, we use $\alpha = 0.2$, $\beta = 0.07$, and a distance threshold is 1000 km. Given the low node degree of the network, it is likely that the generated graphs may have a disconnected node. Further, the network is subjected to partitions as well as the occurrence of cut nodes and vertices.

We note that representativeness of the graphs when compared to the actual topology is low for the locality model, particularly when the underlying graph is sparse. For example, the graph shown in 11(c) is partitioned and the failure of node located at (-3,40) can further partition the network.

When compared to pure random graphs, the locality model yields graphs that are more representative of the actual topology, but are less resilient. However the cost of the generated topology is significantly less due to the reduced number of long distance links.

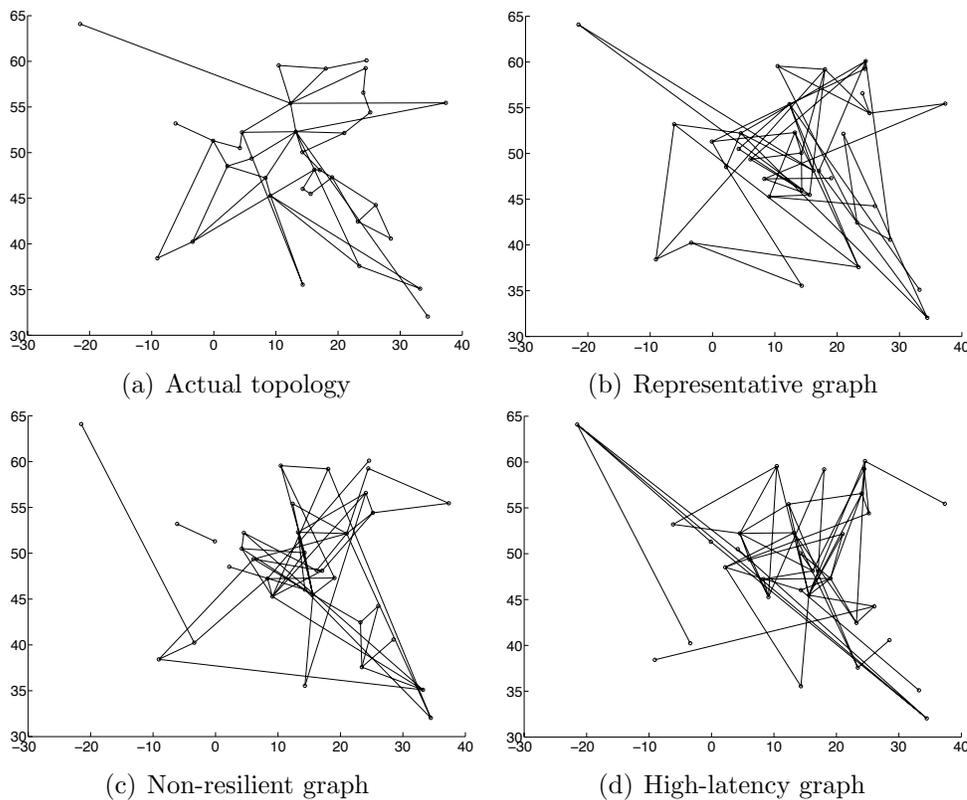


Figure 11: Location constrained locality model for GÉANT2 Network

7.3.3 Location constrained Waxman model

We consider different topologies generated when using Waxman model to calculate link probabilities while constraining the node locations to those of the Sprint network. In order to maintain the average node degree similar to that of actual topology, we use the following values for the model parameters: $\alpha = 0.95$, $\beta = 0.18$. The various graphs shown in Figure 12 indicate a wide variance in the characteristics of the generated plots.

Figure 12(b) shows a sample graph that is fairly representative of the original topology. The graph in Figure 12(c) is an example of a topology that is generated with the same model parameters but has significantly different characteristics in terms of resilience. There are several one-degree nodes and the failure of a cut vertex at (75,41) would partition the network in to two. Similarly, the delay characteristics of Figure 12(d) are suboptimal.

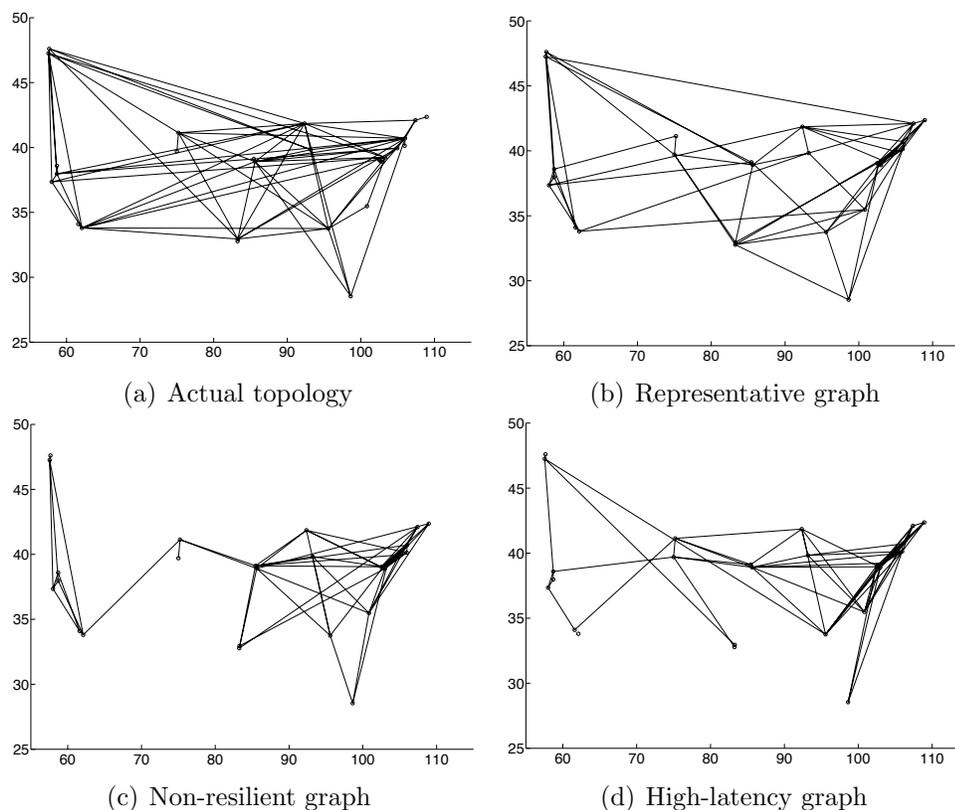


Figure 12: Location constrained Waxman - Sprint network ($\alpha = 0.95$, $\beta = 0.18$)

For a given an number of nodes and average degree, the Waxman model equation (Eq. 2), yields a range of feasible values for the model parameters α and β . In Figure 12, we used a particular set of values that results in topologies which are representative of the actual Sprint network. In this section, we use a different set of values: $\alpha = 0.55$, $\beta = 0.35$ to generate multiple graphs as shown in Figure 13.

Compared to the graphs generated using the previous set of model parameters ($\alpha = 0.95$, $\beta = 0.18$), these graphs are in general more resilient. This is due to the higher value of β , which is the ratio of long to short links. Since there are more long links, the network is likely to be bi-connected. Further, cut vertices and links are avoided because of links spanning across geographically distant nodes. However, we still see some topological differences between multiple runs.

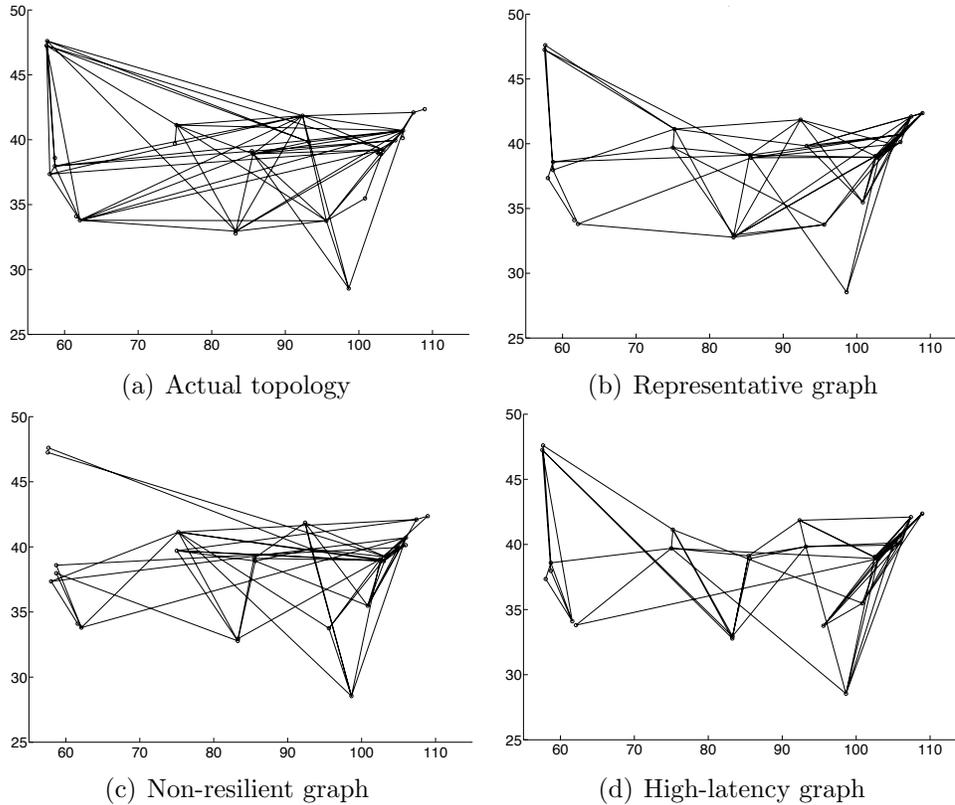


Figure 13: Location constrained Waxman - Sprint network ($\alpha = 0.55$, $\beta = 0.35$)

The topologies generated with the Waxman model when constrained with location of nodes similar to that of GÉANT2 network are shown in Figure 14. Here the model parameters are: $\alpha = 0.35$, $\beta = 0.3$.

An example of a fairly representative graph is given in Figure 14(b). On the other hand, Figure 14(c) shows a uncharacteristic graph generated with the same model parameters. In this particular case, there are several zero and one-degree nodes as well as cut edges. Figure 14(d) shows a case, in which the link locations are such that the shortest distance paths between geographically adjacent nodes is several long length hops away.

When compared to the pure random and locality models, the Waxman model can produce the most representative graphs. Furthermore, the link length and the shortest path distribution also matches with the actual topology. However, like all the other models, graphs generated with Waxman for sparse networks are less representative when compared to dense networks.

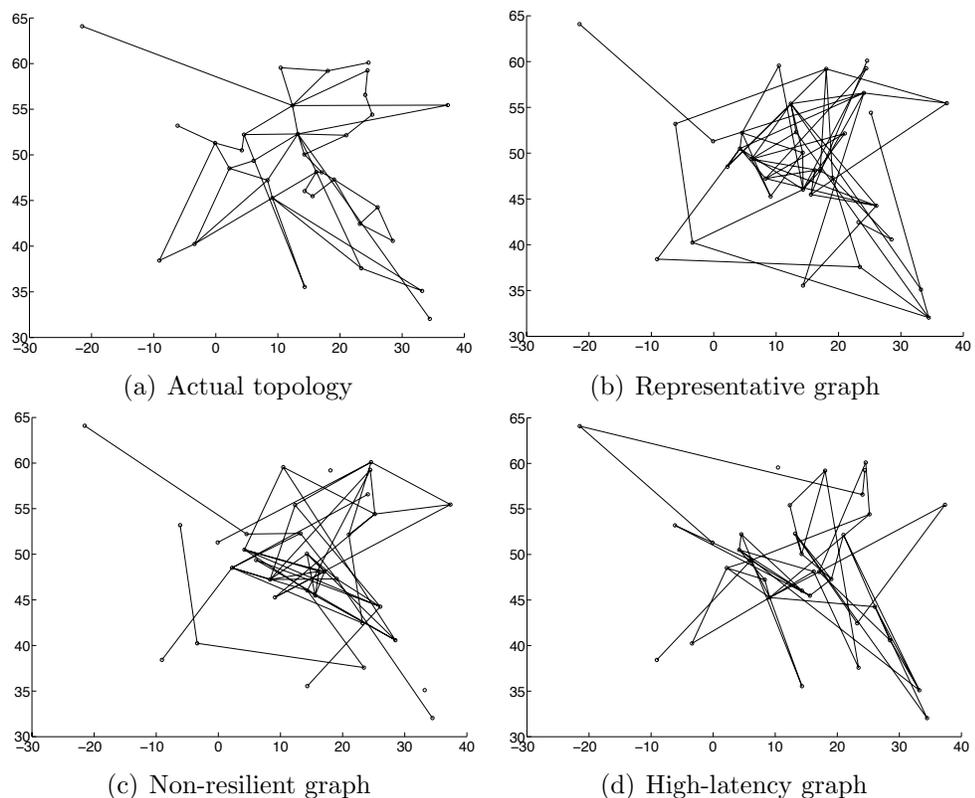


Figure 14: Location constrained Waxman model for GÉANT2 Network

It was observed that all three models can fail to produce realistic topologies on a run-by-run basis. We attribute this to the lack of additional parameters and constraints such as the realization of a minimum spanning tree, bi-connectedness, and path-diversity in the topological models. The representativeness metrics for the topologies based on Sprint and GÉANT2 node locations discussed in this section are given in Table 3 and Table 4 respectively.

Table 3: Statistical distributions of graphs based on Sprint node constraints

Topology	Degree (μ/σ)	Shortest Paths (μ/σ)	Link lengths (10^3 km) (μ/σ)
Actual topology	5.04/3.88	2.44/1.10	1.41/1.18
<i>Modified Random Model</i>			
Representative graph	5.04/1.53	2.11/0.71	2.02/1.35
Non-resilient graph	4.66/2.35	2.08/0.69	1.87/1.07
High-Latency graph	5.04/2.17	2.15/0.76	2.00/1.26
<i>Modified Locality Model</i>			
Representative graph	5.40/2.29	2.07/0.71	1.45/1.29
Non-resilient graph	4.07/2.40	2.37/0.86	1.42/1.29
High-Latency graph	4.44/1.84	2.28/0.80	1.44/1.32
<i>Modified Waxman Model ($\alpha = 0.95, \beta = 0.18$)</i>			
Representative graph	5.48/1.60	2.20/0.85	0.86/0.87
Non-resilient graph	5.04/2.53	2.90/1.52	0.65/0.55
High-Latency graph	5.41/2.92	2.42/1.06	0.70/0.63
<i>Modified Waxman Model ($\alpha = 0.55, \beta = 0.35$)</i>			
Representative graph	5.33/2.32	2.34/1.03	0.95/0.65
Non-resilient graph	4.59/2.18	2.26/0.83	1.17/1.33
High-Latency graph	4.96/2.08	2.46/1.04	0.90/0.72

Table 4: Statistical distributions of graphs based on GÉANT2 node constraints

Topology	Degree (μ/σ)	Shortest Paths (μ/σ)	Link lengths (10^3 km) (μ/σ)
Actual topology	3.00/1.74	3.46/1.54	0.741/0.615
<i>Modified Random Model</i>			
Representative graph	3.23/1.34	2.91/1.02	1.59/0.84
Non-resilient graph	2.70/1.40	3.46/1.38	1.55/0.917
High-Latency graph	3.23/1.68	3.412/1.527	1.40/0.82
<i>Modified Locality Model</i>			
Representative graph	3.0/1.28	3.25/1.29	1.31/0.87
Non-resilient graph	3.23/1.72	2.87/1.12	1.07/0.83
High-Latency graph	3.23/1.85	3.15/1.34	1.24/0.96
<i>Modified Waxman Model ($\alpha = 0.35, \beta = 0.30$)</i>			
Representative graph	3.41/1.87	2.95/1.16	1.21/0.788
Non-resilient graph	2.88/1.70	3.33/1.50	1.16/0.59
High-Latency graph	2.88/1.66	2.84/1.10	1.14/0.69

8 Conclusions and Future Work

It has been argued in the past that realistic topology generation is an important step both for network research as well as practical design of real networks. Prevailing methods focus primarily on improving the structural representativeness of the network and to some extent the on path survivability to link failures. We argue that comprehensive resilience analysis requires tighter integration of geographic constraints in to the graph models. Pre-determined node positioning along with cost considerations will enable the use of topology generators in the network design phase of real network.

We showed that given a set of node locations, the location and cost constrained models in KU-LocGen not only produce realistic topologies, but also provide a significant improvement over the random graph models in terms of the representativeness of the topology. While to some extent it is intuitive that using node positions of the real network is bound to produce realistic

topologies, a contribution of our work is to demonstrate this improvement in quantitative terms for random graph models and provide a usable topology generator for research purposes.

A logical extension of the cost analysis in this paper is cost limited topology generation, in which cost constraints are implicit in graph models as opposed to its use after generation. Given that random graph models are overly simplistic for certain types of networks, our future work includes applying the proposed ideas to other graph models. Future work also involves addition of further constraints that follow the practical network design process such as forcing minimum link degree, k -connectedness, and path diversity. The results in this report show that incorporating location and cost constraints in graph models can improve the credibility of generated topologies and deserves further research.

Acknowledgments

We would like to acknowledge Justin P. Rohrer for his technical and editorial contributions. This work is supported in part by the National Science Foundation Future Internet Design program.

References

- [1] Hongsuda Tangmunarunkit, Ramesh Govindan, Sugih Jamin, Scott Shenker, and Walter Willinger. Network topology generators: degree-based vs. structural. In *SIGCOMM '02: Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 147–159, New York, NY, USA, 2002. ACM.
- [2] D. Alderson, Lun Li, W. Willinger, and J.C. Doyle. Understanding internet topology: principles, models, and validation. *Networking, IEEE/ACM Transactions on*, 13(6):1205–1218, Dec. 2005.
- [3] B.M. Waxman. Routing of multipoint connections. *Selected Areas in Communications, IEEE Journal on*, 6(9):1617–1622, Dec 1988.
- [4] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, pages 251–262, New York, NY, USA, 1999. ACM.
- [5] A.L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509 – 512, October 1999.
- [6] K.I. Calvert, M.B. Doar, and E.W. Zegura. Modeling internet topology. *Communications Magazine, IEEE*, 35(6):160–163, Jun 1997.
- [7] E.W. Zegura, K.L. Calvert, and S. Bhattacharjee. How to model an internetwork. *INFOCOM '96. Fifteenth Annual Joint Conference of the IEEE Computer Societies. Networking the Next Generation. Proceedings IEEE*, 2:594–602 vol.2, Mar 1996.
- [8] Alberto Medina, Ibrahim Matta, and John Byers. On the origin of power laws in internet topologies. *SIGCOMM Comput. Commun. Rev.*, 30(2):18–28, 2000.
- [9] H. Haddadi, M. Rio, G. Iannaccone, A. Moore, and R. Mortier. Network topologies: inference, modeling, and generation. *Communications Surveys & Tutorials, IEEE*, 10(2):48–69, Quarter 2008.

-
- [10] Deepayan Chakrabarti and Christos Faloutsos. Graph mining: Laws, generators, and algorithms. *ACM Comput. Surv.*, 38(1):2, 2006.
- [11] B. Donnet and T. Friedman. Internet topology discovery: A survey. *Communications Surveys & Tutorials, IEEE*, 9(4):56–69, Quarter 2007.
- [12] David Alderson, John Doyle, Ramesh Govindan, and Walter Willinger. Toward an optimization-driven framework for designing and generating realistic internet topologies. *SIGCOMM Comput. Commun. Rev.*, 33(1):41–46, 2003.
- [13] Bruno Quoitin. Topology generation based on network design heuristics. In *CoNEXT '05: Proceedings of the 2005 ACM conference on Emerging network experiment and technology*, pages 278–279, New York, NY, USA, 2005. ACM.
- [14] B.K. Chen and F.A. Tobagi. Network topology design to optimize link and switching costs. *Communications, 2007. ICC '07. IEEE International Conference on*, pages 2450–2456, June 2007.
- [15] Sami J. Habib. Redesigning network topology with technology considerations. *Int. J. Netw. Manag.*, 18(1):1–13, 2008.
- [16] Neil Spring, Ratul Mahajan, David Wetherall, and Thomas Anderson. Measuring ISP topologies with Rocketfuel. *IEEE/ACM Trans. Netw.*, 12(1):2–16, 2004.
- [17] GÉANT2. European research network. <http://www.geant2.net>, Dec 2008.
- [18] James P. G. Sterbenz and David Hutchison. ResiliNets: Multilevel resilient and survivable networking initiative. https://wiki.ittc.ku.edu/resilinetns_wiki/index.php/Topology_and_Challenge_Tolerance#Location_and_Cost_Constrained_Network_Topologies, December 2008.
- [19] Justin P. Rohrer, Ramya Naidu, and James P.G. Sterbenz. Multipath at the transport layer: An end-to-end resilience mechanism. In *Proceedings of the IEEE International Conference on Communications*, Dresden, Germany, June 14–18 2009. Submitted.