This paper was presented at RNDM 2013 - 5th International Workshop on Reliable Networks Design and Modeling, co-located with ICUMT 2013 Congress (Sept. 10-12, 2013, Almaty, KZ)

On the Fitness of Geographic Graph Generators for Modelling Physical Level Topologies

Egemen K. Çetinkaya*, Mohammed J.F. Alenazi*, Yufei Cheng*, Andrew M. Peck*, and James P.G. Sterbenz*†
*Information and Telecommunication Technology Center
Department of Electrical Engineering and Computer Science
The University of Kansas, Lawrence, KS, 66045, USA
{ekc, malenazi, yfcheng, apeck, jpgs}@ittc.ku.edu
†School of Computing and Communications (SCC) and InfoLab21
Lancaster LA1 4WA, UK
jpgs@comp.lancs.ac.uk
www.ittc.ku.edu/resilinets

***Abstract*—The Internet topology has been studied extensively for decades. However, the emphasis of Internet topology research has been on logical level topologies. On the other hand, physical level topologies are necessary to study the resilience of networks realistically. In this paper, we analyse the structure of synthetic geographic topologies whose node locations are given by those of actual physical level graphs. Our results indicate that the synthetic Gabriel graphs capture the grid-like structure of physical level networks. Moreover, given that the cost of physical level topologies is an important aspect from a design perspective, we also compare the cost of several synthetically generated geographic graphs and find that the synthetic Gabriel graphs achieve the smallest cost among all of the graph models that we consider.**

***Keywords*-Internet, backbone, provider, physical level graphs, modelling; network cost model, connectivity, biconnectivity, optimisation; Gabriel, geometric, geographical threshold, Waxman graph; location constrained graph, population weighted graph**

I. Introduction and Motivation

Internet modelling has been the focus of the research community for decades [1]. The Internet can be examined at the physical, IP, router, PoP (point of presence), and AS (autonomous system) level from a topological point of view [2]. At the lowest level we have the physical topology, which consists of components such as fibre and copper cables, ADMs (add drop multiplexers), cross-connects, and layer-2 switches. The logical level consists of devices operating at the IP-layer. The primary focus of previous studies has been on the logical aspects of the Internet, since tools were developed to collect, measure, and analyse IP-level properties of the Internet (e.g. Rocketfuel [3]). However, given that physical networks provide the means of connecting nodes in the higher levels, the study of physical connectivity is an important area of research [4]–[6]. Moreover, physical topologies are necessary to model area-based challenges on networks, such as power failures and severe weather [7].

Physical level topologies are *necessary* and *important* for studying the structure and evolution of the Internet holistically [8]. Unfortunately, in an effort to maintain security and competitiveness, many providers are unwilling to disclose their physical topologies. We generate adjacency matrices of physical level graphs of four commercial service providers based on a third party map [9], and then make use of the publicly available Internet2 research network and the synthetic CORONET fibre topology. Using the node locations of the physical topologies, we generate synthetic *geographical* graphs of these topologies utilising the Gabriel, geometric, geographical threshold, and Waxman graph models. We analyse the structural properties of the synthetically generated geographical graphs using KU-TopView (KU Topology Viewer) [10] and find that the Gabriel graph model most closely captures the grid-like structure of the physical networks.

Another important aspect of modelling physical graphs is the *cost* of networks, which is particularly important to consider when designing physical level networks. Moreover, from a network design perspective, it is important to design networks that are *resilient* yet *less costly*. Unfortunately, these two objectives fundamentally oppose one another. We compare the synthetically generated geographical graphs based on a cost model and our results indicate that Gabriel graphs are also the best among the ones we consider in terms of minimising cost. Additionally, amongst all of the synthetically generated graphs we find that there are some whose costs are two orders of magnitude greater than their corresponding physical graphs. To the best of our knowledge, there are no other studies that provide structural- and cost-based comparisons of geographic graph models applied to graphs with node locations that are constrained to those of actual physical graphs. Furthermore, we discuss how one might develop a better synthetic graph generator that incorporates the strengths of two of the geographical graph models that we study.

The rest of the paper is organised as follows: The topological dataset we use in this study is presented in Section II. The properties of graphs we analyse are presented in Section III. We describe the synthetic geographical graph models in Section IV. We analyse the cost incurred to design these graphs and present synthetically generated graphs of the Internet2 network visually in Section V. We discuss how one might develop a better alternative geographical graph model to

978-1-4799-1177-6/13/$31.00

capture a graph's structural properties in Section VI. Finally, we summarise our study as well as propose future work in Section VII.

II. Topological Dataset

We study physical level communication networks that are geographically located within the continental United States. Therefore, we only include the 48 contiguous US states, the District of Columbia, and exclude Hawaii, Alaska, and other US territories. We use US *long-haul fibre-optic routes* map data to generate physical topologies for AT&T, Sprint, and Level 3 [9]. In this map, US fibre-optic routes cross cities throughout the US and each ISP has a different coloured link. We project the cities to be physical node locations and connect them based on this map, which is sufficiently accurate on a national scale. We use this data to generate adjacency matrices for each individual ISP. To capture the geographic properties as well as the graph connectivity, cities are included as nodes even if they are merely a location along a link between fibre interconnection. Finally, we also make use of the publicly available TeliaSonera network [11], Internet2 [12], and CORONET [13], [14] topologies. CORONET is a synthetic fibre topology designed to be representative of service provider fibre deployments.

III. Properties of Networks

We investigate a number of other quantities that reveal a graph's structural properties using the Python NetworkX library [15]. In Table I we list a number of relevant quantities for each of the provider networks. A detailed analysis of graph metrics for the given physical networks was presented in our earlier work [16]. We observe from the node and link counts that AT&T, Level 3, and Sprint are the larger among the networks. Moreover, all of the physical topologies have an average degree between 2 and 3. In our previous work, we noted that the average degree of these physical topologies was much smaller than the average degree of their corresponding logical topologies due to the difficulty involved in connecting nodes in a physical topology, where one must physically lay down fibre between nodes.

TABLE I
Topological characteristics of physical level networks

Network	Geographical				Full mesh	
	Nodes	Links	Avg. Node Degree	Tot. l [km]	Links	Tot. l $\times 10^6$ [km]
AT&T	383	488	2.55	50,026	73,153	116.8
Level 3	99	130	2.63	28,538	4,851	7.5
Sprint	264	312	2.36	33,627	34,716	57.8
TeliaSonera	21	25	2.38	14,190	210	0.4
Internet2	57	65	2.28	19,050	1,596	2.7
CORONET	75	99	2.64	28,325	2,775	4.6

A. Network Cost Model

Structural properties impact the connectivity and *cost* of building networks. While at the logical level the cost is captured by the number of nodes and the capacity of each node (i.e. the bandwidth and number of ports available in a router [17], [18]), at the physical level, the length of the fibre is a major determinant of the cost. After all, logical level links are arbitrarily overlaid links on top of the underlying physical links. Previously, we provided a network cost model as:

$$C_{i,j} = f + v \times d_{i,j} \tag{1}$$

where f is the fixed cost associated with the link (including termination), v is the variable cost per unit distance for the link, and $d_{i,j}$ is the length of the link [19], [20]. Moreover, in a modest attempt to capture the total cost of fibre topologies, if we assume that the fibre length dominates wide-area network cost and ignore the fixed cost associated with each link, the network cost can be written as:

$$C = \sum_i l_i \tag{2}$$

where l_i is the length of the i-th link [21], [22]. We calculate the total link length for each provider with this simplified network cost model as shown in 5th column in Table I. The total link length of each physical topology is somewhere between 14,000 to 50,000 km. For these topologies, the smaller the size of the network, the smaller the total length link of the fibre.

Next, for each physical level topology, we consider as an upper baseline the full-mesh topology whose vertex set is identical to that of the original topology. We then calculate the total link count and length of each full-mesh topology as shown in column 6 and 7 in Table I, respectively. Note that the total link lengths are given in millions of km for a hypothetical full-mesh physical level topology, emphasising that real networks cannot have unlimited resilience due to cost constraints.

B. Structure of Physical Level Graphs

The physical level topologies consist of a number of degree two intermediate nodes for accurate geographic representation that are necessary for modelling area-based challenges on the network, such as power failures and severe weather. However, these intermediate nodes artificially change the graph theoretic properties of the networks, in particular artificially skewing the degree distribution toward degree-2 nodes. Therefore, we modify the existing physical level graphs by removing nodes with a degree of two, if there is not a logical level node at that location that is serviced by the physical node. The number of nodes, links, and average degree of the *structural* graphs are shown in Table II. Each structural graph has fewer nodes and links than its corresponding physical level graph. However, with the exception of TeliaSonera, each structural graph has a larger average degree than its corresponding physical level graph. For example, the structural graph of Internet2 has 16 nodes, 24 links, and an average degree of 3 whereas the original Internet2 physical graph has 57 nodes, 65 links, and an average degree of 2.28. We believe that the structural graph of TeliaSonera has a smaller average degree than the original

graph of TeliaSonera due to the latter's small order and size. However, we note that total fibre length of the structural graph (14,040 km) is close to that of the original physical graph (14,190 km).

TABLE II
FIBRE LINK LENGTHS OF STRUCTURAL GRAPHS

Network	Nodes	Links	Avg. Deg.	Tot. l [km]
AT&T	130	191	2.94	37,489
Level 3	48	71	2.96	25,390
Sprint	52	73	2.81	25,190
TeliaSonera	18	21	2.33	14,040
Internet2	16	24	3.00	18,146
CORONET	39	63	3.23	27,579

IV. GRAPH MODELS FOR PHYSICAL LEVEL NETWORKS

In the following section we present four different geographic graph models. The Gabriel graph model is a parameterless model that uses only node locations as input, while the geometric, geographical threshold, and Waxman models all require at least one parameter. The geometric graph model uses a single threshold parameter, while the geographic threshold model and the probabilistic Waxman model use two parameters. We apply each of these graph models to graphs with node locations constrained to those of actual physical topologies. Given the diverse nature of these models, we believe the following sections represent a fairly comprehensive analysis of geographic graph models applied to physical topologies.

A. Gabriel Graphs

Next, we generate Gabriel graphs of the six service provider networks. Gabriel graphs are useful in modelling graphs with geographic connectivity that resemble grids [23], [24]. We would expect the Gabriel graph to be one of the best ways to model physical topologies for this reason. In a Gabriel graph, two nodes are connected directly if and only if there are no other nodes that fall inside the circle whose diameter is given by the line segment joining the two nodes. The number of links and the total link length of Gabriel graphs of six networks are shown in Table III.

TABLE III
FIBRE LINK LENGTHS OF GABRIEL GRAPHS

Network	Links	Tot. l [km]
AT&T	686	66,157
Level 3	170	33,991
Sprint	474	57,104
TeliaSonera	26	12,111
Internet2	94	27,786
CORONET	127	33,265

B. Geometric Graphs

A 2-dimensional geometric graph is a graph in which nodes are placed on a plane or surface and any pair of nodes is connected if and only if:

$$d(u, v) \leq d_\theta \tag{3}$$

where $d(u, v)$ is the Euclidean distance between the two nodes $\{u, v\}$, and d_θ is a distance threshold parameter [25]. In the conventional random 2-dimensional geometric graph model, nodes are distributed randomly on a plane.

Using the physical level node locations of six provider networks, we generate four different geometric graphs based on four different d_θ distance threshold values. For the first set of graphs, we use the maximum link length of the actual physical graph as the d_θ value. For the second set of graphs we select the largest possible values of d_θ such that the total link lengths of these graphs are less than the total link lengths of the original physical level graphs. Using this methodology, we find that all of the synthetically generated graphs are disconnected. For the third set of graphs, we select the smallest value of d_θ such that the graphs are connected. It turns out that none of these graphs are biconnected. For the fourth set of graphs we select the smallest values of d_θ such that the graphs are biconnected: that is, such that the graphs will remain connected after the failure of any one node or link. This is a basic requirement for basic network resilience and survivability [26], [27]. The link lengths l of the actual graphs as well as the synthetically generated geometric graphs are shown in Table IV.

To further explain the data in Table IV, consider the AT&T physical graph with the given node locations. The number of links, total link length, and maximum link length of the actual AT&T physical graph are shown in columns 2, 3, and 4, respectively. For the case of AT&T, when we assign $d_\theta = \max(l_i)$ (where $\max(l_i) = 629$ km in this case), the synthetically generated geometric graph has 15,062 links and the total length of the graph is approximately 5.7×10^6 km. Using this *threshold optimised* methodology we obtain the number of links, total link length, and d_θ as shown in columns 5, 6, and 7, respectively. With the second *cost optimised* methodology we generate synthetic geometric graphs such that the total link length is less than that of the actual physical topology. In the case of AT&T, the generated graph has a total link length of 49,937 km, which is less than that of the actual AT&T graph whose total link length is 50,026 km. We note that the cost optimised geometric graphs of all service providers are disconnected graphs. The number of links, total link length, and d_θ for cost-optimised graphs are shown in columns 8, 9, and 10, respectively. Since the cost-optimised geometric graphs are disconnected graphs, we increase the value of d_θ until we obtain connected graphs. Applying this *cost and connectivity optimised* methodology to AT&T, the total number of links is 4,916, the total length of the links is 918,353 km, and d_θ = 302 km, as shown in columns 11, 12, and 13, respectively. While cost and connectivity optimised graphs are connected, none of them are biconnected. Therefore, we increase d_θ so that the resulting geometric graphs are biconnected. Applying this *cost and biconnectivity optimised* methodology to AT&T, we obtain a synthetically

TABLE IV
COST OF GEOMETRIC GRAPHS BASED ON A THRESHOLD VALUE

Network	Actual			Threshold Optimised			Cost Optimised			Cost & Con. Optimised			Cost & Bicon. Optimised		
	Links	Tot. l [km]	Max. l [km]	Links	Tot. l [km]	d_θ [km]	Links	Tot. l [km]	d_θ [km]	Links	Tot. l [km]	d_θ [km]	Links	Tot. l [km]	d_θ [km]
AT&T	488	50,026	629	15,062	5,719,021	629	783	49,937	99	4,916	918,353	302	8,343	2,169,572	424
Level 3	130	28,538	1,063	2,107	1,326,422	1,063	209	28,358	226	749	234,721	528	1,104	449,360	683
Sprint	312	33,627	602	6,478	2,327,659	602	466	33,573	112	3,417	804,197	390	4,261	1,159,340	452
TeliaSonera	25	14,190	1,592	106	88,151	1,592	37	13,757	614	56	27,842	859	93	68,635	1,425
Internet2	65	19,049	910	442	246,259	910	83	18,997	334	131	37,532	424	258	104,793	616
CORONET	99	28,325	943	922	506,209	943	156	28,144	280	512	188,663	604	613	253,812	691

generated geometric graph with 8,343 links, 2.2×10^6 km of total link length, and a d_θ value of 424 km, as shown in columns 14, 15, and 16, respectively. The rest of the service provider data is shown in the consecutive rows in Table IV.

C. Population-weighted Geographical Threshold Graphs

A threshold graph is a type of graph in which links are formed based on node weights [28]. Two nodes $\{u, v\}$ with node weights $\{w_u, w_v\}$ are connected if and only if:

$$w_u + w_v \geq t \tag{4}$$

in which t is a threshold value that is a non-negative real number. A modified version of a threshold graph is a *geographical* threshold graph that includes geometric information about the nodes [29]. In this case, two nodes $\{u, v\}$ with node weights $\{w_u, w_v\}$ are connected if and only if:

$$w_u + w_v \geq \psi d(u, v)^\phi \tag{5}$$

where ψ and ϕ are model parameters and $d(u, v)$ is the distance between nodes $\{u, v\}$. In our study, we assign the node weights to be the population estimates of cities for year 2011, which are taken from the US Census Bureau [30]. The population statistics for each provider are given in Table V. For the AT&T physical graph, the total of population of all of the cities (e.g. 383 cities) is about 76 million, and the average city population is about 197,000. The most populous city (NYC for all networks) has about 8.2 million people, and the least populated city has 182 people. These statistics are shown in columns 2, 3, 4, and 5 in Table V respectively for each provider network.

TABLE V
POPULATION STATISTICS OF CITIES AS NODE WEIGHTS

Network	Total	Average	Maximum	Minimum
AT&T	75,753,034	197,789	8,244,910	182
Level 3	53,221,035	537,586	8,244,910	12,695
Sprint	67,794,208	256,796	8,244,910	448
TeliaSonera	27,944,279	1,330,680	8,244,910	65,397
Internet2	40,980,611	718,958	8,244,910	8,438
CORONET	49,559,726	660,796	8,244,910	33,395

Using city populations as node weights, we generate synthetic graphs for each provider network. We choose $\phi = 1$ so that we can manipulate only ψ. Moreover, by choosing $\phi = 1$, we find that the righthand side of inequality (5) varies linearly with distance. Hence, as the distance increases between two nodes they are less likely to be connected. Having fixed $\phi = 1$, we first choose ψ so as to minimise cost while ensuring connectivity, and then choose ψ so as to minimise cost while ensuring biconnectivity. More specifically, for each network, we select the largest value of ψ rounded to the nearest tenth such that the graph is connected, and then select the largest value of ψ rounded to the nearest tenth such that the graph is biconnected. The results of both methodologies are shown in Table VI. For AT&T, we find that the largest value of ψ such that AT&T is connected is 3.1, yielding a link number of 1670 and a total link length of 690,941 km. Additionally, the largest value of ψ such that AT&T is biconnected is 2.4, which yields a link number of 2,336 and a total link length of 1,036,747 km.

TABLE VI
POPULATION-WEIGHTED GEOGRAPHIC THRESHOLD GRAPHS FOR $\phi = 1$

Network	Connectivity Optimised			Biconnectivity Optimised		
	ψ	Links	Tot. l [km]	ψ	Links	Tot. l [km]
AT&T	3.1	1,670	690,941	2.4	2,336	1,036,747
Level 3	3.4	324	158,316	2.4	526	304,696
Sprint	3.0	1,164	500,678	2.4	1,532	717,311
TeliaSonera	3.4	43	31,099	2.3	62	58,492
Internet2	3.2	151	98,733	2.3	233	194,938
CORONET	3.3	244	127,387	2.4	374	233,360

D. Location-constrained Waxman Graphs

The Waxman model provides a probabilistic way of connecting nodes in a graph [31]. Given two nodes $\{u, v\}$ with a Euclidean distance $d(u, v)$ between them, the probability of connecting these two nodes is:

$$P(u, v) = \beta e^{\frac{-d(u,v)}{L\alpha}} \tag{6}$$

where $\beta, \alpha \in (0, 1]$ and L is the maximum distance between any two nodes. Increasing β increases the link density and a large value of α corresponds to a high ratio of long links to short links.

In the Waxman model nodes are uniformly distributed in the plane. We modify the Waxman model so that it is constrained by the node locations. The resulting link properties of the location-constrained Waxman model, along with the β and α parameters, are shown in Table VII. For each network, we

choose β and α such that the resulting graph is a connected graph with the smallest possible total link length. For example, in the AT&T graph, using the node geographic locations we use β and α values of 0.1 and run the experiments 10 times, which results graphs that are disconnected. Then, we keep β at a value of 0.1 and increase α to a value of 0.2, which results in connected graphs but with a mean of 1.6 million km total link length. We calculate total link length by averaging 10 runs with increments of 0.1 for β and α parameters until we find connected graphs that result in least total length. The β and α parameters for each provider are shown in columns 2 and 3 in Table VII. The average number of links for each topology resulting from 10 runs is shown in column 4, whereas the standard deviation σ of the number of links resulting from 10 runs is shown in column 5. The average total link length of 10 runs is shown in column 6, and the standard deviation σ of the total link of length resulting from 10 runs is shown in column 7.

TABLE VII
LOCATION-CONSTRAINED WAXMAN GRAPHS

Network	β	α	Avg. No. of Links	σ Links	Avg. Tot. l [km]	σ Tot. l
AT&T	0.2	0.1	1,981	54	1,044,856	29,509
Level 3	0.6	0.1	392	14	205,036	7,896
Sprint	0.2	0.1	904	43	475,943	24,271
TeliaSonera	0.6	0.2	31	3	24,498	4,743
Internet2	0.6	0.1	102	10	62,100	7,723
CORONET	0.5	0.1	174	15	91,002	10,062

V. ANALYSIS OF PHYSICAL-LEVEL GRAPHS

In this section we present the cost incurred using different graph models, as well as show the structure of the synthetic models for the Internet2 network.

A. Cost Analysis of Graphs

We presented the total link lengths of the synthetically generated graphs in the previous section. However, in order to see the big picture we summarise them again in Table VIII. The first column shows the providers, and the second column gives the types of graphs generated for each provider, which we discussed in the previous section. For each type of graph, if any optimisation technique was used, this is indicated in column 3, where optimisation techniques include CO (connectivity optimised), CCO (cost and connectivity optimised), and CBO (cost and biconnectivity optimised). The total number of links, average node degree, and total link lengths are shown in columns 4, 5, and 6, respectively. For the Waxman graph (as discussed in Section IV-D), among the set of ten connected graphs we generated, we choose the graph with the smallest total link length to present in Table VIII.

The results are not surprising: when networks are required to be biconnected, additional costs are incurred. Furthermore, for geometric graphs, as the threshold distance is increased, the graph becomes more connected, which increases the total cost. Moreover, geometric graphs incur the highest cost among all of the graph models considered. The PWGTGs (population weighted geographical threshold graphs) also incur a high cost due to the fact that nodes with heavy weights are connected even if they are far apart. For our future work, we will generate additional PWGTGs by further varying the ψ and ϕ parameters. Depending on the physical topology used, PWGTGs may incur a higher cost than Waxman graphs or vice

TABLE VIII
STRUCTURAL PROPERTIES OF FIBRE TOPOLOGIES

Provider	Graph	Opt.	Links	Avg. Degree	Tot. l [km]
AT&T	Physical	N/A	488	2.55	50,026
	Structural	N/A	191	2.94	37,489
	Gabriel	N/A	686	3.58	66,157
	Geometric	TO	15,062	78.65	5,719,021
		CCO	4,916	25.67	918,353
		CBO	8,343	43.57	2,169,572
	PWGTG	CCO	1,670	8.72	690,941
		CBO	2,336	12.19	1,036,747
	Waxman	CCO	1,896	9.90	991,037
Level 3	Physical	N/A	130	2.63	28,538
	Structural	N/A	71	2.96	25,390
	Gabriel	N/A	170	3.43	33,991
	Geometric	TO	2,107	42.57	1,326,422
		CCO	749	15.13	234,721
		CBO	1,104	22.30	449,360
	PWGTG	CCO	324	6.55	158,316
		CBO	526	10.63	304,696
	Waxman	CCO	364	7.35	192,546
Sprint	Physical	N/A	312	2.36	33,627
	Structural	N/A	73	2.81	25,190
	Gabriel	N/A	474	3.59	57,104
	Geometric	TO	6,478	49.08	2,327,659
		CCO	3,417	25.89	804,197
		CBO	4,261	32.28	1,159,340
	PWGTG	CCO	1,164	8.82	500,678
		CBO	1,532	11.61	717,311
	Waxman	CCO	826	6.26	429,285
TeliaSonera	Physical	N/A	25	2.38	14,190
	Structural	N/A	21	2.33	14,040
	Gabriel	N/A	26	2.48	12,111
	Geometric	TO	106	10.09	88,151
		CCO	56	5.33	27,842
		CBO	93	8.86	68,635
	PWGTG	CCO	43	4.09	31,099
		CBO	62	5.90	58,492
	Waxman	CCO	25	2.38	15,937
Internet2	Physical	N/A	65	2.28	19,050
	Structural	N/A	24	3.00	18,146
	Gabriel	N/A	94	3.29	27,786
	Geometric	TO	442	15.51	246,259
		CCO	131	4.59	37,532
		CBO	258	9.05	104,793
	PWGTG	CCO	151	5.29	98,733
		CBO	233	8.18	194,938
	Waxman	CCO	85	2.98	48,252
CORONET	Physical	N/A	99	2.64	28,325
	Structural	N/A	63	3.23	27,579
	Gabriel	N/A	127	3.39	33,265
	Geometric	TO	922	24.59	506,209
		CCO	512	13.65	188,663
		CBO	613	16.35	253,812
	PWGTG	CCO	244	6.51	127,387
		CBO	374	9.97	233,360
	Waxman	CCO	154	4.11	73,498

versa. On the other hand, the probabilistic nature of Waxman can result in graphs that are not structurally optimal.

We presented the total link lengths of the synthetically generated graphs in Table VIII. However, in order to better illustrate the results we summarise them again in Figure 1. The y-axis shows the cost incurred in terms of total link length in units of m for each graph and x-axis shows six provider networks for different graph models from left to right. We use the graphs that provide minimal connectivity with the least cost. For the Waxman graph (as discussed in Section IV-D), among the set of ten connected graphs we generated, we choose the graph with the smallest total link length to present in Figure 1.

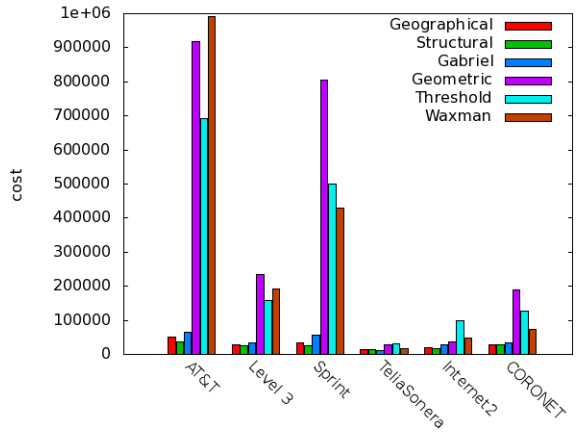


Fig. 1. Cost analysis of physical graph models

Cost analyses of the graphs indicate that the cost of each synthetically generated graph depends on the order (number of nodes) of the network as expected. For example, the cost incurred for TeliaSonera is the smallest and TeliaSonera also has the lowest number of nodes. Second, we can infer that geographical, structural, and Gabriel graphs incur about the same cost for all providers. The cost of geometric, population-weighted geographical threshold, and Waxman graphs are much higher than the previous three models. However, the cost difference between different graph models for TeliaSonera is not as drastic as larger size networks due to its smaller order. In other words, the difference between the first three and last three graph models differs more as the number of nodes increase. The location-constrained Waxman model is probabilistic in nature and the cost values are shown for a sample generated graph using this model with $\beta = 0.6$ and $\alpha = 0.1$. The cost incurred with the Waxman model is generally higher than that of the original geographic physical level graphs across all providers, because the links are added probabilistically.

A graph's connectivity can be improved by adding links; however this adds additional cost to achieve resilience [22]. By examining the synthetically generated topologies using the geometric graph model and geographical threshold graph model in Tables IV and VI respectively, we observe that it incurs about 90% or more additional cost to result in biconnected graphs. For example, applying the geometric graph model on Internet2 topology yields a total link length of about 37,000 km for a minimal connected graph. However, for the same node locations of Internet2, when we generate a biconnected synthetic graph, the total link length is about 105,000 km, which is more than double the cost of the uniconnected version. Similar conclusions can be also observed for the geographic threshold model. When we compare these cost values against the upper bound of the Internet2 graph, which is 2.7 million km, we observe that they are far less than the upper bound. From these results, we conclude that all synthetic graph models discussed in this paper—with the exception of the Gabriel graph model—result in a total link length that is not feasible to model physical level topologies.

B. Visual Analysis of Graphs

We inspect all the synthetically generated topologies of all the providers using KU-TopView (KU Topology Map Viewer) [10], [20]. We find the results to be similar across all providers. We discuss the Internet2 graph here because its smaller order makes it easier to visualise, and thus more informative for demonstrating the fitness of each synthetic graph model on this topology. The geographical, structural, Gabriel, geometric, population weighted geographical threshold, and location-constrained Waxman model of the Internet2 physical-level graphs are shown in Figure 2.

The geographic physical-level Internet2 topology with 57 nodes and 65 links is shown in Figure 2a. In our earlier work we showed using graph spectra that geographic physical-level graphs resemble a grid-like structure [16]. The structural physical-level Internet2 topology in which degree-2 nodes are removed is shown in Figure 2b. The synthetically generated Gabriel graph of the geographic Internet2 graph is shown in Figure 2c. While the Gabriel graph preserves the grid-like structure of the geographic physical-level topology, it omits some of the links at the periphery of the actual geographic physical-level graph (e.g. link between Baton Rouge, LA and Jacksonville, FL) and adds links that are infeasible to deploy due to terrain. The synthetically generated geometric graph based on a distance threshold value that incurs minimal cost to obtain a connected graph is shown in Figure 2d. In this case, while islands of nodes that are close to each other are richly connected, overall the graph is far from being biconnected. The geographical threshold graph of the Internet2 topology using population of cities as node weights is shown in Figure 2e. This synthetic graph resembles multiple star-like structures, because highly-populated cities become central nodes and connect to nodes that are far away. In this connected graph, there is only one link that connects east and west portions of the US. Finally, a location-constrained Waxman graph with $\beta = 0.6$ and $\alpha = 0.1$ values is shown in Figure 2f. Because of the probabilistic nature of this graph model, the links between nodes are established randomly. In conclusion, Gabriel graphs are the closest to model physical level topologies with some caveats which we discuss in the next section.

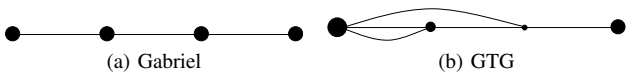
(a) Geographical (b) Structural (c) Gabriel

(d) Geometric (e) Geographical threshold (f) Waxman

Fig. 2. Visual representation of Internet2 physical-level topologies in KU-TopView [10]

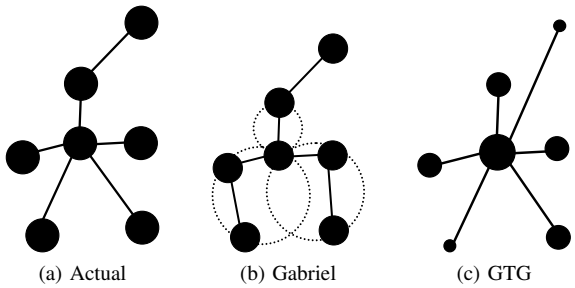
VI. Discussion

In Section V we demonstrated that none of the synthetic geographical graph models we study capture the cost and structural properties perfectly. Based on our observations we present some ideas about how to develop a new geographic graph model that more closely captures the cost and structural behavior of physical topologies. First, we observe that the presence of parameters within a graph model gives the user more control with regards to optimising the graph based on an objective function. Second, we note that while Gabriel graphs capture linear topologies that are horizontally aligned, they fall short in capturing star-like structures.



(a) Gabriel (b) GTG

Fig. 3. Graph models under linear geography



(a) Actual (b) Gabriel (c) GTG

Fig. 4. Graph models under star geography

GTGs (geographical threshold graphs), on the other hand, generate star-like structures *aggressively* around heavily weighted nodes. For example, in Figure 3, we show the behavior of the Gabriel model and GTG model applied to a linear topology consisting of nodes horizontally aligned. In Figure 3a, we see that the Gabriel model perfectly captures the linear topology, while in Figure 3b we see that the GTG model aggressively adds more links.

Next, consider a star-like graph as shown in Figure 4. While the Gabriel model aggressively changes it to a grid-like structure as depicted in Figure 4b showing the circles for determining the links, the GTG model can capture this star-like structure better than the Gabriel model depending on the node weight distribution, which we represent by giving different node sizes as shown in Figure 4c. While each of these two models captures different structures better than the other, a better model would be able to select either the Gabriel model or the GTG model based on local structural criteria.

Finally, a detailed examination of Gabriel graphs show that they have two undesirable properties as compared to highly-engineered physical graphs. First, they add unnecessary ladder cross-connections between parallel linear segments in an attempt to increase the grid-like structure, and second they leave stub links that do not biconnect nodes on the edge into the rest of the graph. We will explore heuristics for a modified-Gabriel graph to address these issues in future work.

VII. Conclusions and Future Work

Modelling the Internet at the logical level has been the focus of the research community. On the other hand, physical level topologies are *necessary* to study the resilience of networks more realistically. In this paper, we discuss the *fitness* of four geographical graph models applied to graphs with node locations given by those of six actual networks. We evaluate the cost of these synthetically generated graphs based on a cost model, and we find that among the synthetic graph models we studied, the Gabriel model yields topologies with the smallest

cost. Furthermore, the cost incurred using synthetic models depends on the number of nodes and the geographic distribution of these nodes. We analyse the topologies generated by the synthetic geographic graph models, and visual inspection of these topologies shows that the Gabriel graphs best capture the grid-like structure of physical level topologies. Based on our observations we present some ideas about how to develop a new geographic graph model that more closely captures the structural behaviour of physical topologies.

For our future work, we intend to generate synthetic graphs based on the structural physical-level topologies. Moreover, we will investigate heuristics that increase connectivity and biconnectivity while representing grid-like structure of physical-level topologies.

ACKNOWLEDGMENTS

This research was supported in part by NSF grant CNS-1219028 (Resilient Network Design for Massive Failures and Attacks) and by NSF grant CNS-1050226 (Multilayer Network Resilience Analysis and Experimentation on GENI).

REFERENCES

[1] J. M. McQuillan, "Graph theory applied to optimal connectivity in computer networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 7, pp. 13–41, Apr. 1977.

[2] B. Donnet and T. Friedman, "Internet topology discovery: a survey," *IEEE Communications Surveys & Tutorials*, vol. 9, no. 4, pp. 56–69, 2007.

[3] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson, "Measuring ISP topologies with Rocketfuel," *IEEE/ACM Transactions on Networking*, vol. 12, no. 1, pp. 2–16, 2004.

[4] H. Haddadi, M. Rio, G. Iannaccone, A. Moore, and R. Mortier, "Network Topologies: Inference, Modeling, and Generation," *IEEE Communications Surveys & Tutorials*, vol. 10, no. 2, pp. 48–69, 2008.

[5] D. Krioukov, k. claffy, M. Fomenkov, F. Chung, A. Vespignani, and W. Willinger, "The Workshop on Internet Topology (WIT) Report," *ACM Comput. Commun. Rev.*, vol. 37, no. 1, pp. 69–73, 2007.

[6] R. Durairajan, S. Ghosh, X. Tang, P. Barford, and B. Eriksson, "Internet Atlas: A Geographic Database of the Internet," in *Proceedings of the 5th ACM HotPlanet Workshop*, (Hong Kong), August 2013.

[7] E. K. Çetinkaya, D. Broyles, A. Dandekar, S. Srinivasan, and J. P. G. Sterbenz, "Modelling Communication Network Challenges for Future Internet Resilience, Survivability, and Disruption Tolerance: A Simulation-Based Approach," *Telecommunication Systems*, vol. 52, no. 2, pp. 751–766, 2013.

[8] E. K. Çetinkaya, A. M. Peck, and J. P. G. Sterbenz, "Flow Robustness of Multilevel Networks," in *Proceedings of the 9th IEEE/IFIP International Conference on the Design of Reliable Communication Networks (DRCN)*, (Budapest), pp. 274–281, March 2013.

[9] KMI Corporation, "North American Fiberoptic Long-haul Routes Planned and in Place," 1999.

[10] "ResiliNets Topology Map Viewer." http://www.ittc.ku.edu/resilinets/maps/, January 2011.

[11] "TeliaSonera." http://www.teliasoneraic.com.

[12] "Internet2." http://www.internet2.edu.

[13] "The Next Generation Core Optical Networks (CORONET)." http://www.darpa.mil/Our_Work/STO/Programs/Dynamic_Multi-Terabit_Core_Optical_Networks_%28CORONET%29.aspx.

[14] G. Clapp, R. A. Skoog, A. C. Von Lehmen, and B. Wilson, "Management of Switched Systems at 100 Tbps: the DARPA CORONET Program," in *International Conference on Photonics in Switching (PS)*, pp. 1–4, September 2009.

[15] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring Network Structure, Dynamics, and Function using NetworkX," in *7th Python in Science Conference (SciPy)*, (Pasadena, CA), pp. 11–15, August 2008.

[16] E. K. Çetinkaya, M. J. F. Alenazi, J. P. Rohrer, and J. P. G. Sterbenz, "Topology Connectivity Analysis of Internet Infrastructure Using Graph Spectra," in *Proceedings of the 4th IEEE/IFIP International Workshop on Reliable Networks Design and Modeling (RNDM)*, (St. Petersburg), pp. 752–758, October 2012.

[17] D. Alderson, L. Li, W. Willinger, and J. C. Doyle, "Understanding Internet Topology: Principles, Models, and Validation," *IEEE/ACM Transactions on Networking*, vol. 13, no. 6, pp. 1205–1218, 2005.

[18] J. C. Doyle, D. L. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger, "The 'robust yet fragile' nature of the Internet," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 41, pp. 14497–14502, 2005.

[19] M. A. Hameed, A. Jabbar, E. K. Çetinkaya, and J. P. Sterbenz, "Deriving Network Topologies from Real World Constraints," in *Proceedings of IEEE GLOBECOM Workshop on Complex and Communication Networks (CCNet)*, (Miami, FL), pp. 400–404, December 2010.

[20] J. P. Sterbenz, E. K. Çetinkaya, M. A. Hameed, A. Jabbar, Q. Shi, and J. P. Rohrer, "Evaluation of Network Resilience, Survivability, and Disruption Tolerance: Analysis, Topology Generation, Simulation, and Experimentation (invited paper)," *Telecommunication Systems*, vol. 52, no. 2, pp. 705–736, 2013.

[21] E. K. Çetinkaya, M. J. F. Alenazi, A. M. Peck, J. P. Rohrer, and J. P. G. Sterbenz, "Multilevel Resilience Analysis of Transportation and Communication Networks," *Springer Telecommunication Systems Journal*, 2013. accepted on July 2013.

[22] M. J. F. Alenazi, E. K. Çetinkaya, and J. P. G. Sterbenz, "Network Design and Optimisation Based on Cost and Algebraic Connectivity," in *Proceedings of the 5th IEEE/IFIP International Workshop on Reliable Networks Design and Modeling (RNDM)*, (Almaty), September 2013.

[23] K. R. Gabriel and R. R. Sokal, "A New Statistical Approach to Geographic Variation Analysis," *Systematic Zoology*, vol. 18, no. 3, pp. 259–278, 1969.

[24] D. W. Matula and R. R. Sokal, "Properties of Gabriel Graphs Relevant to Geographic Variation Research and the Clustering of Points in the Plane," *Geographical Analysis*, vol. 12, no. 3, pp. 205–222, 1980.

[25] M. Penrose, *Random Geometric Graphs*. Oxford Studies in Probability 5, 2003.

[26] J. P. G. Sterbenz, R. Krishnan, R. R. Hain, A. W. Jackson, D. Levin, R. Ramanathan, and J. Zao, "Survivable mobile wireless networks: issues, challenges, and research directions," in *Proceedings of the 3rd ACM workshop on Wireless Security (WiSE)*, (Atlanta, GA), pp. 31–40, 2002.

[27] J. P. G. Sterbenz, D. Hutchison, E. K. Çetinkaya, A. Jabbar, J. P. Rohrer, M. Schöller, and P. Smith, "Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines," *Computer Networks*, vol. 54, no. 8, pp. 1245–1265, 2010.

[28] N. Mahadev and U. Peled, *Threshold Graphs and Related Topics*, vol. 56 of *Annals of Discrete Mathematics*. Elsevier North-Holland, Inc., 1995.

[29] M. Bradonjić, A. Hagberg, and A. G. Percus, "The Structure of Geographical Threshold Graphs," *Internet Mathematics*, vol. 5, no. 1-2, pp. 113–139, 2008.

[30] "US Census Bureau Population Estimates." http://www.census.gov/popest/data/cities/totals/2011/files/SUB-EST2011_ALL.csv, 2013.

[31] B. M. Waxman, "Routing of Multipoint Connections," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 9, pp. 1617–1622, 1988.