

# The Causal Factors Behind Internet Power-Law Connectivity

Sangmin Kim and Khaled Harfoush  
Department of Computer Science  
North Carolina State University  
E-mail: {skim12, harfoush}@ncsu.edu

**Abstract**—In this paper, we investigate the impact of layer-2 devices on our perception of the Internet IP topology and on the performance and security of the Internet. We make the case that a power law connectivity observed in the Internet IP topology is not an illusion as suggested by some researchers. It is mainly manifested due to the blindness of traceroute to layer-2 devices, and this manifestation will persist independent of the nature of the underlying physical topology. Furthermore, we make the case that the Internet physical topology is *not* likely to have a power law connectivity. Our conclusions challenge common wisdom about the Internet topology and highlight the need for more thorough investigations.

## I. INTRODUCTION

Studying the structure and the properties of network topologies has attracted researchers in many disciplines and the biggest human-built network, the Internet, is no exception. Information about the structure of the Internet topology can resolve mysteries about the Internet such as the throughput that it can tolerate, the potential congestion bottlenecks, and the resilience to failures and attacks. Answering these questions is especially important given our growing reliance on the Internet in our everyday lives.

The structure of Internet topology can be abstracted as a graph  $G = (V, E)$ . Vertices,  $v \in V$ , and edges,  $e \in E$ , in a graph take different meanings depending on the level of abstraction. A vertex in an *IP* topology, also known as *router-level* topology, represents a router and an edge connects two vertices if their corresponding routers can exchange traffic without interim routers, as enforced by the underlying routing protocol. A vertex in a *physical* topology represents a network device such as a switch or a router and an edge connects two vertices if their corresponding devices are directly connected with a physical cable. While each level of abstraction provides useful information about the Internet, the physical topology is the most difficult to characterize since the most common probing tool, namely *traceroute*, is blind to layer-2 devices.

One of the most popular but controversial results in Internet topology studies is Faloutsos's power law discovery [1], which tells that the distribution of node degrees in the router-level Internet follows a power law. The implications are (1) the Internet enjoys the *small-world* property delivering

information very efficiently, and (2) it is resilient to *random failures*, but is vulnerable to *targeted attacks*, since some nodes act as heavily connected *hubs* and bringing these hubs down can be disastrous to the Internet [2]. Since then, many studies have been conducted to support this theory using extensive measurements [3] or to reason it with economical and technological constraints [4]. Still, some researchers are questioning the power law conclusion, suggesting that they are a side effect of the sampling bias in the measurement [5].

In this paper, we make the case that our understanding of the Internet properties is misled by missing layer-2 devices. Specifically, we draw the following conclusions: The power law connectivity observed in the Internet IP topology (1) is not an illusion caused by biased data as suggested in [5], (2) It is mainly manifested due to the blindness of traceroute to layer-2 devices, and (3) this manifestation will persist independent of the nature of the underlying physical topology. We also argue that the Internet physical topology does *not* have a power law connectivity. Non-power law physical topology positively affects the Internet performance and security. These conclusions are not in line with the conclusions made by earlier studies and highlight the need for more thorough investigations.

The rest of this paper is organized as follows. In Section II we survey related literature. In Section III analyze the impact of layer-2 devices on our perception of the Internet IP topology. In Section IV we investigate the structure of the Internet physical topology and claim that it is not likely to be a power law. In Section V we revise earlier conclusions about Internet topology generators, and conclude in Section VI.

## II. RELATED WORKS

Internet *topology generators* reflect the evolution of our understanding of the Internet structure and features. Until recently, complex networks were modeled as random graphs introduced by Erdős and Rényi's (ER) [6]. Waxman [7] extended the ER model by factoring the distance between nodes. In [8], Albert and Barabási suggest that most complex networks are *scale-free* with *power law* degree distribution. Faloutsos et al. observed the same power law in the Internet [1]. Lakhina et al., however, argue that the data used in Internet measurement is biased due to the probing methodology and that the power

<sup>0</sup>This work was partially supported by NSF grant CAREER ANIR-0347226.

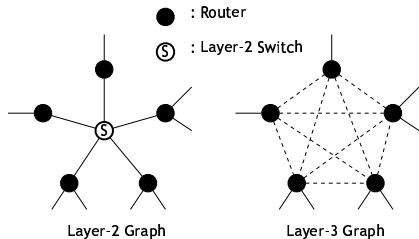


Fig. 1. A portion of a physical topology and the corresponding IP topology. Solid lines are physical links and dotted lines are connections as perceived by a traceroute-like tool.

law observation in the Internet may simply be a side effect of the biased data [5].

Meanwhile, a different research direction has emerged studying the impact of economical and technological drivers on the structure of the Internet [4], and suggest that power laws in complex systems are due to the tradeoffs between yield, cost of resources, and tolerance to risk. Along that line, Li et al. propose a new model to capture Internet topology [4]. They suggest that topology generators need to capture the perceived Internet performance while considering known economical and technological constraints.

### III. INTERNET IP TOPOLOGY

In this section we rely on theoretical proof and simulation results to make the case that the perceived power law Internet IP topology is a side effect of layer-2 devices hidden from probing tools like traceroute, but is not an illusion caused by sampling bias.

#### A. Impact of Layer-2 Devices

Layer-2 devices such as Ethernet or ATM switches are designed to perform transparent bridging between multiple network segments. They are common in today's Internet. Figure 1 (left) provides an example of a switch connecting a set of routers, and Figure 1 (right) shows the corresponding IP graph. The key observation is that in IP graphs, routers connected to switches are perceived as having higher degree than their actual number of physical connections. This simple observation lays the ground for a bolder statement as manifested in Theorem 1.

*Theorem 1:* Consider an arbitrary non-trivial physical topology,  $G$ , in which layer-2 and layer-3 devices are spread uniformly at random. The degree distribution of nodes in the corresponding IP topology,  $G'$ , follows a power law.

*Proof:* Let  $n$  be the number of layer-2 devices in  $G$ . We convert  $G$  into  $G'$  in steps,  $t = 1 \dots n$ . In step 1 we convert  $G$  into  $G_{(1)}$ ; in step 2 we convert  $G_{(1)}$  into  $G_{(2)}$ ;  $\dots$ ; and in step  $n$  we convert  $G_{(n-1)}$  into  $G_{(n)} \equiv G'$ . In each step the physical topology,  $G$ , is brought closer to the IP topology,  $G'$ , by hiding *one* more layer-2 device from  $G$  as explained in Figure 1. Thus after  $n$  steps the resulting graph represents the final IP topology. Consider an arbitrary step,  $t$ , and the layer-2 device,  $s$ , that is to be considered in this step. After hiding  $s$ , the *neighbors* of  $s$  in  $G_{(t-1)}$  will be perceived as

having higher degree in  $G_{(t)}$ . We next show that if layer-2 devices are placed uniformly at random in  $G$  then the resulting  $G'$  will be a power law graph. The proof is by mapping the above construction to Barabási and Albert's (1999) *rich gets richer* generative model [8], the most widely accepted model to generate power law graphs. In our case, we simply need to show that nodes with larger degree are more likely to have even larger degree during the construction of  $G'$ . The key idea is that a vertex that is picked uniformly at random in a graph is more likely to be a *neighbor* of a high degree vertex than a neighbor to a low degree vertex. Since layer-2 devices are placed uniformly at random, then at each step,  $t$ , the newly considered layer-2 device is likely to be a neighbor of a high degree node in  $G_{(t-1)}$ , further increasing its degree in the resulting  $G_{(t)}$ . This concludes the proof. ■

The most common probing approach to reveal Internet IP topology is  $k$ - $m$  traceroute, in which traceroute probes are sent from  $k$  controlled probing sources towards  $m$  arbitrary uncontrolled destinations. Corollary 1 highlights the impact of layer-2 devices on the perceived IP topology revealed by the  $k$ - $m$  traceroute approach.

*Corollary 1:* Consider an arbitrary non-trivial physical topology,  $G$ , in which layer-2 and layer-3 devices are spread uniformly at random. The degree distribution of nodes in the corresponding IP topology observed by the  $k$ - $m$  traceroute approach follows a power law.

*Proof:* In the case of one probing source,  $k = 1$ , the graph observed by the source is a *tree* independent of the underlying routing strategy, assuming there are no routing loops. When  $k > 1$ , the overall observed graph from the  $k$  sources is the aggregation of  $k$  trees (a *forest*). Theorem 1 applies to arbitrary graphs, and applies naturally to a forest graph. As a result, the IP topology observed by the  $k$ - $m$  traceroute approach follows a power law. This concludes the proof. ■

Theorem 1 and Corollary 1 prove that the presence of layer-2 devices leads to the power law node degree distribution observed in IP measurement studies. This conclusion is independent of the nature of the underlying physical topology. Note that the assumption that layer-2 devices are spread uniformly at random is reasonable as these devices are quite popular in the Internet.

#### B. Simulation Results

We next verify the results of Theorem 1 and Corollary 1 through simulations. In the simulations, we adopt two well-known random graphs, that were quite popular in representing Internet topologies before the power law models were introduced: (1) Erdős and Rényi (ER) [6], and (2) Waxman [7]. The ER graph has 100,000 nodes and connectivity probability  $p = 0.00015$ . The Waxman graph has 100,000 nodes, that are added incrementally, using  $\alpha = 0.15$ ,  $\beta = 0.2$ ,  $P(i, j) = \alpha e^{-d/(\beta L)}$  where  $0 < \alpha, \beta \leq 1$ ,  $d$  is the Euclidean distance between nodes  $i$  and  $j$ ,  $L$  is the maximum distance between any two nodes. The degree distributions in ER and Waxman

graph follow a Poisson and an exponential distribution, respectively. We pick a fraction of the graph nodes uniformly at random and assume they are layer-2 devices, and all other nodes are assumed to be routers. We also assume shortest path routing between all nodes.

For each of the ER and Waxman models, we study four scenarios: (1) Physical, (2) IP, (3) km-physical, and (4) km-IP. In the first scenario we consider a graph including all nodes, whether routers or layer-2 devices, and all links. In the second scenario we consider a graph in which layer-2 devices are hidden as hinted on in Figure 1. To do that, we identify the shortest path between each pair of nodes in the graph while hiding layer-2 devices, as if they are not revealed by a tool such as traceroute. We then combine all revealed information from the shortest paths to construct the IP graph. This is the graph that would be revealed using traceroute from each node to all other nodes. In the third scenario we consider the graph that would be revealed by running a *hypothetical* version of traceroute capable of revealing layer-2 devices from  $k$  probing sources to  $m$  destinations. The sources and the destinations are picked at random. In the fourth scenario we consider the graph that would be revealed by running the traditional traceroute from  $k$  probing sources to  $m$  destinations, thus missing the layer-2 devices.

In Figure 2 we plot the degree distribution of nodes in the four graphs corresponding to the four scenarios in the ER and Waxman models. We plot the curves of scenarios 3 and 4 only for  $k = 1$  and  $m = 10,000$  since larger values of  $k$  lead to similar results. As shown in the figure the presence of layer-2 devices result in power law IP topologies (Scenarios 2 and 4), whether  $k$ - $m$  traceroute is used or not, even though the underlying physical topologies are not power laws. The graphs for the scenario 2 and 4 confirm the conclusions from Theorem 1 and Corollary 1. The IP topology (scenario 2) shows exponential cut-off at the end of the curve, also revealed in measurement studies. As seen in Table I, the fraction of layer-2 devices increases, the exponent tends to be smaller, which means that more nodes are perceived as highly connected. Also, as expected, the physical curves (Scenario 1) in Figure 2 reflect the actual node degree distributions of the ER and the Waxman graphs. What is *not* expected though is that the km-physical curve (Scenario 3) is *not* a power law. According to [5],  $k$ - $m$  traceroute should have led the captured graph to a power law distribution due to the sampling bias. We clarify this seemingly contradictory result in the Section III-C. It is worth mentioning that missing layer-2 devices affects our perception of *topological* metrics other than the node degree distribution as well. For example, the *clustering coefficient* [9],  $C$ , is slowly inflated, which means more IP nodes are perceived as highly connected, as more layer-2 devices are included.

### C. Sampling bias

In [5], the authors suggest that nodes that are close to the probing sources reveal their connectivity better than nodes that are further away, and as a result the measurements are

TABLE I  
CHARACTERISTIC PARAMETERS BY HIDING LAYER-2 VERTICES: POWER LAW EXPONENT  $\gamma$ , AVERAGE PATH LENGTH  $l$ , AND CLUSTERING COEFFICIENT,  $C$ . *Switch %* IS THE PERCENTAGE OF LAYER-2 DEVICES.

Switch %	ER			Waxman		
	$\gamma$	$l$	$C$	$\gamma$	$l$	$C$
0%	-	4.5	0.00004	-	21.8	0.00004
20%	2.41	3.9	0.00005	3.16	17.2	0.00005
40%	2.35	3.3	0.00006	3.10	10.1	0.00006
60%	2.37	2.9	0.00007	2.85	7.1	0.00007

distorted leading to power law node degree observations. One can highlight their conclusions through our km-physical curves in Figure 2 (a-c) (Scenario 3). Interestingly, the km-physical curves in Figure 2 do not reveal power law trends. The reason for this seemingly contradicting results is that the authors of [5] plot the degree distribution only up-to a degree of 25, while we plot all revealed node degrees. Notice in Figure 2 that if we draw the km-physical curves up-to a degree of 25 for the ER setup, we get a seemingly power law curve. This does not mean that there is no sampling bias associated with the  $k$ - $m$  traceroute approach; it simply means that with thorough probing, whether by increasing  $k$  and/or  $m$ , the *nature* of the distribution is revealed. In addition, the  $k$ - $m$  traceroute itself is *not* able to produce a long heavy-tailed power law graph, because it can *not* inflate the degree that each node actually has.

Measurement studies targeting Internet IP topologies relying on  $k$ - $m$  traceroute have been quite thorough, revealing node degrees of up-to 1500 [5]. The degrees certainly exceed the maximum possible physical connectivity of any router. For example, the popular high-end Cisco 7600 series and Juniper’s T640 brands offer routers with a maximum of 128 physical ports. Furthermore, measurement results that are not based on  $k$ - $m$  traceroute have revealed also power law traces. Measurement result in [10], which limits the bias toward nodes close to the probing host, reveals node degrees up-to 4000. As a result, the sampling bias itself cannot explain the power law prevalent in Internet measurements.

## IV. INTERNET PHYSICAL TOPOLOGY

Understandably, there is no measurement study about the Internet physical topology. Layer-2 devices can not respond to IP probe packets, and network administrators do not publish the structure of their networks. Still, there is standing evidence that the structure of the Internet physical topology is *not* a power law. Researchers have invested their effort in classifying large scale networks based on their features. Watt’s taxonomy classifies networks into *relational* and *spatial* networks [11]. In relational networks, new connections between vertices are governed by a certain probability that is a function of existing network connectivity. Web sites linkage [12] and airline systems connecting cities [13] are examples of relational networks. In spatial networks, connections between vertices are governed by Euclidean distance between the vertices. The power grid networks [11] and the U.S. highway system [14] are examples

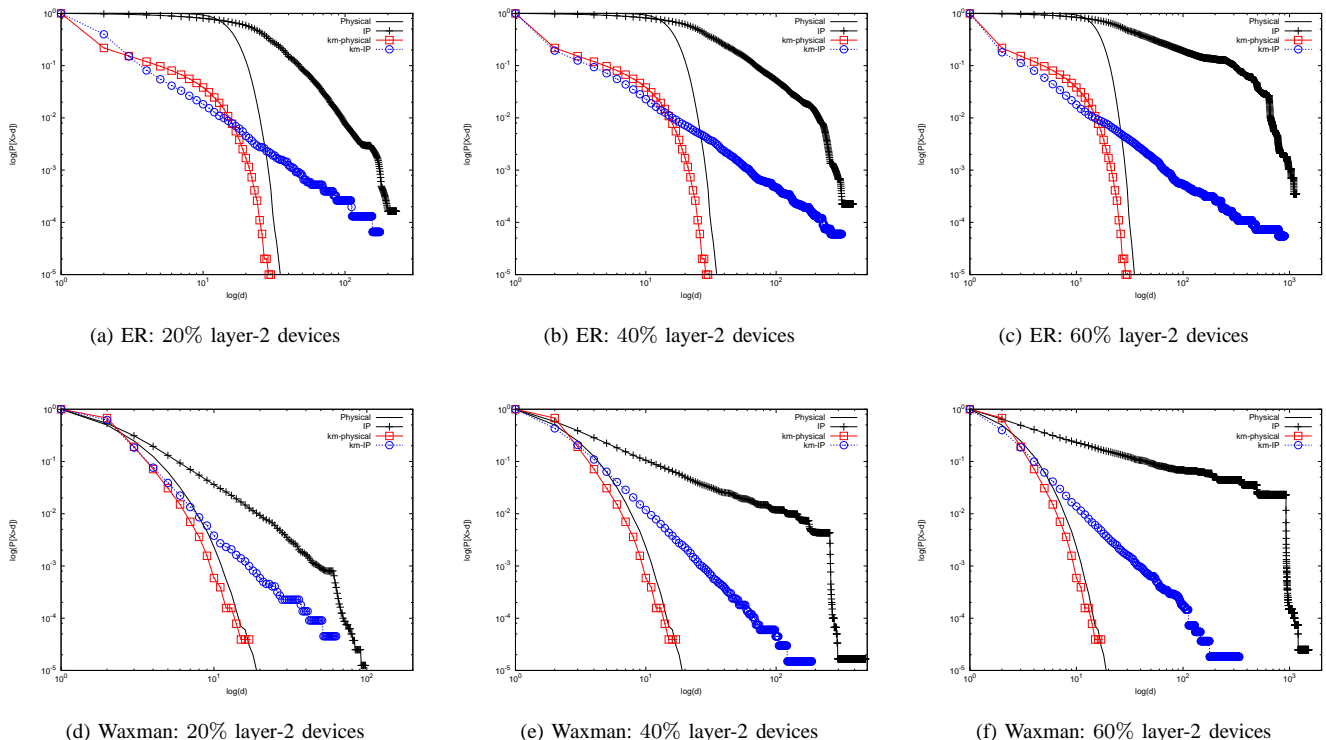


Fig. 2. Node degree distribution of ER and Waxman graphs with 100,000 nodes, when the fraction of layer-2 devices is 20% (a,d), 40% (b,e), 60% (c,f). Each figure include four curves representing the (1) Physical, (2) IP, (3) km-physical, and (4) km-IP scenarios. For scenarios 3 and 4, we use  $k = 1$ , and  $m = 10,000$ .

of spatial networks. Research studies on spatial networks confirm that they have exponential-like degree distributions [13]. The Internet physical topology can be classified as a spatial network and it is likely that it also has exponential connectivity. While there is no evidence of this conjecture, it is worth highlighting that none of the existing and well-documented spatial networks has a power law connectivity.

#### A. Economical Drivers

The cost of digging and laying out network cables outweighs the cost of adding routers and switches. Interconnecting local networks together before interconnecting them to distant networks is thus more economical. This cost efficient recursive clustering helps reduce the degree of each vertex in the Internet structure, making it unlikely that some vertices would have a degree much larger than the average vertex degree. Notice that while recent edge router technology can accommodate thousands of DS0 channels or hundreds of DS1 channels, physical ports on routers are more or less around 100, which means that switch devices should be placed to multiplex hundreds of ports prior to interconnecting them to the core of the Internet.

#### B. Internet Performance

We consider two performance metrics: (1) The *maximum throughput*, and (2) the *average path length*. We use the former to make the case that the maximum throughput that

can be achieved by the Internet is improved with non-power law physical topology. We use the latter to show that the small-world phenomenon observed in the Internet can be explained even if the underlying Internet physical topology has exponential degree distribution.

**Maximum Throughput.** Routers and switches have a limited number of physical ports, and utilizing a large number of ports degrades the throughput [4]. Furthermore, the typical number of physical ports on a device is not large making it again unlikely that some nodes would have much larger degree than the average node degree. In [4], the authors argue that highly connected node in the Internet should be located at the edge of the network to obtain the maximum possible throughput while satisfying the router degree-bandwidth constraint. Their model, however, does not lead to the maximum *possible* throughput as the router degree-bandwidth constraint limits the capacity of highly connected routers at the edge of the network as well. This itself is unlikely in practice as it is hard to conceive that an administrator would intentionally use a heavily connected router and degrade the potential capacity that would be otherwise available by the use of two less connected routers.

**Average Path Length.** Researchers have suggested that the Internet exhibit the *small world* property, which is inherent in power law graphs and Poisson graphs like ER, but *not* in exponential graphs like Waxman. In general, the *small world*

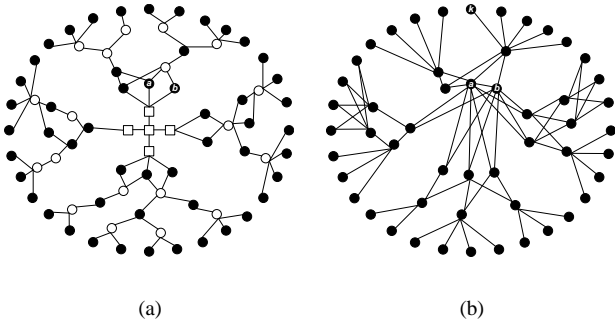


Fig. 3. (A) Degree distribution for physical connectivity. There is no node highly connected in this level. (B) Degree distribution for IP-level topology based on  $k$ - $m$  traceroute. Traceroute begins from node ‘ $k$ ’. The result topology of the IP-level is exaggerated by layer-2 devices, and has few highly connected hub nodes. ●: router, ○: Ethernet switch, ◻: ATM switch,  $k$ : traceroute source node

property exists in relational networks, not in spatial networks. In Table I we show the average path length of the ER and Waxman graphs using the setup described in Section III-B. It is clear from the figures that when no layer-2 devices are included, the ER graph has the small world property while Waxman does not. However, as layer-2 devices are introduced, the observed average path length in the resulting IP topologies decreases significantly both in the ER and the Waxman, satisfying the small world property. As a result, the small world phenomenon observed in the Internet may be a side effect of the presence of layer-2 devices.

### C. Internet Security

In [2], the authors suggest that the Internet is resilient to random failures but vulnerable to targeted attacks to its highly connected hubs. Transparent layer-2 devices however suggest that the physical structure of the Internet may be resilient to both failures and attacks. For illustration, nodes  $a$  and  $b$  in Figure 3(b) appear as heavily connected hubs in the perceived IP topology and it seems that attacking and removing these nodes isolates a large portion of the network. However, by observing the underlying physical topology in Figure 3(a), removing nodes  $a$  and  $b$  affects only a smaller fraction of the network. Overall, security favors an Internet physical topology without heavily connected nodes.

## V. REVISITING INTERNET TOPOLOGY GENERATORS

Internet topology generators fall into two main categories: (1) structure-based [15] and (2) degree-based [16]. Structure-based topology generators model the hierarchical structure of the Internet, while degree-based topology generators model power law node degree distributions in measurement study. Degree-based topology generators in general are perceived as superior to structure-based ones since the structure-based generators do not exhibit power law node degree distributions [3]. The conclusions that we draw in this paper, however, sheds some doubt into this perception. Structure-based topology generators may indeed reflect a power law node degree

distribution if layer-2 switches are modeled. This does not mean that structure-based generators produce more realistic Internet topologies. It is rather that earlier conclusions need to be revisited.

## VI. CONCLUSIONS

In this paper, we made the case that (1) layer-2 devices are the reason for the perception of a power law node degree distribution in the Internet’s IP topology, not sampling bias nor node’s tendency of preferential attachment. (2) The Internet physical topology should not be a power law to enjoy better performance and security. Our conclusions help to explain many of the Internet measurement results, to question some of the earlier conclusions, and to call for more thorough investigation of the Internet properties.

## REFERENCES

- [1] M. Faloutsos, P. Faloutsos, and C. Faloutsos, “On power-law relationships of the internet topology,” in *SIGCOMM*, 1999, pp. 251–262.
- [2] R. Albert, H. Jeong, and A.-L. Barabási, “Error and attack tolerance of complex networks,” *Nature*, vol. 406, p. 378, 2000.
- [3] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. 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*, 2002, pp. 147–159.
- [4] L. Li, D. Alderson, W. Willinger, and J. Doyle, “A first-principles approach to understanding the internet’s router-level topology.” Portland, OR: ACM SIGCOMM, 2004.
- [5] A. Lakhina, J. Byers, M. Crovella, and P. Xie, “Sampling biases in ip topology measurements,” in *IEEE INFOCOM*, San Francisco, CA, March 2003.
- [6] P. Erdős and A. Rényi, “On the evolution of random graphs,” *Publ. Math. Inst. Hung. Acad. Sci.*, 1960.
- [7] B. Waxman, “Routing of multipoint connections,” *IEEE J. Sel. Areas Communications*, vol. 6, no. 9, pp. 1617–1622, December 1988.
- [8] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 15, pp. 509–512, October 1999.
- [9] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature*, vol. 393, pp. 440–442, 1998, characteristic path length and clustering coefficient.
- [10] S. Kim and K. Harfoush, “Efficient estimation of more detailed internet ip maps,” in *IEEE International Conference on Communications 2007*, Scotland, June 2007.
- [11] D. Watts, *Small worlds: the dynamics of networks between order and randomness*, ser. Princeton studies in complexity. Princeton, N.J.: Princeton University Press, 1999.
- [12] R. Albert, H. Jeong, and A.-L. Barabási, “Diameter of the world-wide web,” *Nature*, vol. 401, p. 130, September 1999.
- [13] L. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, “Classes of small-world networks,” *Proc Natl Acad Sci U S A*, vol. 97, no. 21, pp. 11 149–11 152, October 2000.
- [14] A.-L. Barabási and E. Bonabeau, “Scale-free networks,” *Scientific American*, pp. 50–59, May 2003.
- [15] K. Calvert, M. Doar, and E. Zagura, in *IEEE Communications Magazine*, no. 35, June, pp. 160–163.
- [16] A. Medina, A. Lakhina, I. Matta, and J. Byers, “BRITE: Universal topology generation from a user’s perspective, Tech. Rep. 2001-003, Jan 2001.