Analysis of Network Heterogeneity by Using Entropy of the Remaining Degree Distribution

Lu Chen, Shin'ichi Arakawa, and Masayuki Murata
Graduate School of Information Science and Technology
Osaka University
Osaka, Japan
$\{$*l-chen, arakawa, murata*$\}$*@ist.osaka-u.ac.jp*

***Abstract*—As the Internet becomes the social infrastructure, a network design method that has the adaptability against the failure of network equipment and has the sustainability against changes of traffic demand is becoming important. Since we do not know in advance when the environmental changes occur and how large the changes are, it is preferable to have heterogeneity in topological structures so that the network can evolve more easily. In this paper, we investigate the heterogeneity of topological structures by using mutual information of remaining degree. Our results show that the mutual information is high at the most of router-level topologies, which indicate that the route-level topologies are highly designed by, e.g., the network operators. We then discuss and show that the mutual information represents the heterogeneity of topological structure through illustrative examples.**

Keywords-power-law network; router-level topology; topological structure; mutual information; network heterogeneity; degree distribution.

I. Introduction

As the Internet becomes the social infrastructure, it is important to design the Internet that has adaptability and sustainability against environmental changes. However, dynamic interactions of various network-related protocols make the Internet into a complicated system. For example, it is shown that interactions between routing at the network layer and overlay routing at the application layer degrade the network performance [1]. Therefore, a new network design method which has the adaptability against the failure of network equipment and has the sustainability against changes of traffic demand is becoming important. Since complex networks display heterogeneous structures that result from different mechanisms of evolution [2], one of the key properties to focus on is the network heterogeneity where, for example, the network is structured heterogeneous rather than homogeneous by some design principles of information networks.

Recent measurement studies on Internet topology show that the degree distribution exhibits a power-law attribute [3]. That is, the probability P_x, that a node is connected to x other nodes, follows $P_x \propto x^{-\gamma}$, where γ is a constant value called scaling exponent. Generating methods of models which obey power-law degree distribution are studied widely, and Barabáshi-Albert (BA) model is one of it [4]. In BA model, the topology increases incrementally and links are placed based on the connectivity of topologies in order to form power-law networks. The resulting topology has a large number of links connected with a few nodes, while a small number of links connected with numerous nodes. Topologies generated by BA model are used to evaluate various kind of network performances [5], [6].

However, it is not easy to explain topology characteristics of router-level topology by such models because topology characteristics are hardly determined only by degree distribution [7], [8]. Li et al. [7] enumerated several different topologies with power-law, but identical degree distribution, and showed the relation between their structural properties and performance. They pointed out that, even though topologies have a same degree distribution, the network throughput highly depends on the structure of topologies. The lessons from this work suggest us that the heterogeneity of the degree distribution is insufficient to discuss the topological characteristics and the network performance of router-level topologies.

In this paper, we investigate the diversity of router-level topologies by using mutual information of remaining degree. Here, the diversity of topology means how diverse the interconnections are in any sub graphs chosen from the topology. Mutual information yields the amount of information that can obtain about one random variable X by observing another variable Y. The diversity of topology can be measured by considering Y as some random variable of a part of the topology and X as the rest of it. Solé et al. [2] studied complex networks by using remaining degree distribution as the random variable. They calculated the mutual information of remaining degree of biological networks and artificial networks such as software networks and electronic networks, and shown that both of them have higher mutual information than randomly connected networks. In this paper, we use this mutual information to evaluate the diversity of topology.

Milo et al. [9] have introduced a concept called Network Motif. The basic idea is to find several simple sub graphs in complex networks. Arakawa et al. [10] shows the characteristic of router-level topologies by counting the

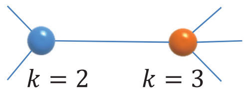

Figure 1. Remaining degree

number of each kind of sub graph which consists of 4 nodes respectively. They conclude that router-level topology has more sub graphs called "sector", that is removing one link from 4 nodes complete graph, than other networks. However, Network Motif is expected to evaluate the frequency of appearance of simple structure in a topology, and is not expected to measure the diversity of topology.

The rest of this paper is organized as follows. The definition of remaining degree and mutual information is explained in Section II. Mutual information of several router-level topologies are calculated, and shown in Section III. In Section IV, we investigate the topological characteristic by changing the mutual information through a rewiring process. Finally, we conclude this paper in Section V.

II. Definitions

Mutual information of remaining degree is defined by Solé et al. [2]. Remaining degree k is the number of edges leaving the vertex other than the one we arrived along. The example is shown in Figure 1, where the remaining degree is set to two for the left node and three for the right node. This distribution $q(k)$ is obtained from:

$$q(k) = \frac{(k+1)P_{k+1}}{\Sigma_k k P_k}, \quad (1)$$

where $P(P_1, \ldots, P_x, \ldots, P_K)$ is the degree distribution, and K is the maximum degree.

The distribution of mutual information of remaining degree, $I(q)$, is

$$I(\text{q}) = H(\text{q}) - H_c(\text{q}|\text{q'}), \quad (2)$$

where q=($q(1), \ldots, q(i), \ldots, q(N)$) is the remaining degree distribution.

The first term $H(\text{q})$ is entropy of remaining degree distribution:

$$H(\text{q}) = -\sum_{k=1}^{N} q(k)\log(q(k)). \quad (3)$$

Within the context of complex networks, it provides an average measure of network's heterogeneity, since it measures the diversity of the link distribution. $H = 0$ in a homogeneous networks such as ring topology. As network become more heterogeneous, the entropy H gets higher. For example, Abilene inspired topology [7] shown in Figure 2 is

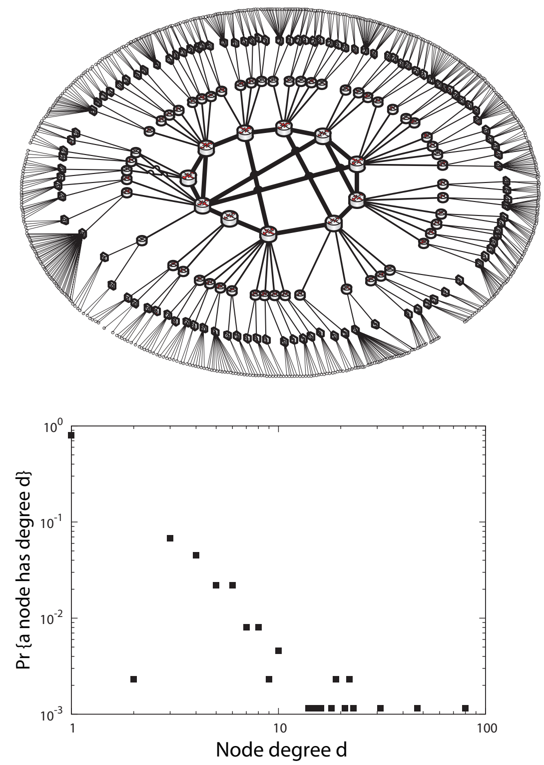

Figure 2. Abilene ($H = 3.27$,$H_c = 2.25$)

Table I
MUTUAL INFORMATION OF ROUTER-LEVEL TOPOLOGIES

Topology	Nodes	Links	$H(G)$	$H_c(G)$	$I(G)$
Level3	623	5298	6.04	5.42	0.61
Verio	839	1885	4.65	4.32	0.33
ATT	523	1304	4.46	3.58	0.88
Sprint	467	1280	4.74	3.84	0.90
Telstra	329	615	4.24	3.11	1.13
BA	523	1304	4.24	3.98	0.26
Random	523	1304	3.22	3.15	0.07

heterogeneous in the degree distribution, thus it has higher entropy.

The second term $H_c(\text{q}|\text{q'})$ is the conditional entropy of the remaining degree distribution,

$$H_c(\text{q}|\text{q'}) = -\sum_{k=1}^{N}\sum_{k'=1}^{N} q(k')\pi(k|k')\log\pi(k|k'), \quad (4)$$

where $\pi(k|k')$ are conditional probability. They give the probability of observing a vertex with k' edges leaving it provided that the vertex at the other end of the chosen edge has k leaving edges. For Abilene inspired topology, combinations of remaining degrees which are the ones of a pair of linked nodes are biased; therefore, the conditional entropy H_c is low.

 ISBN: 978-1-61208-226-4

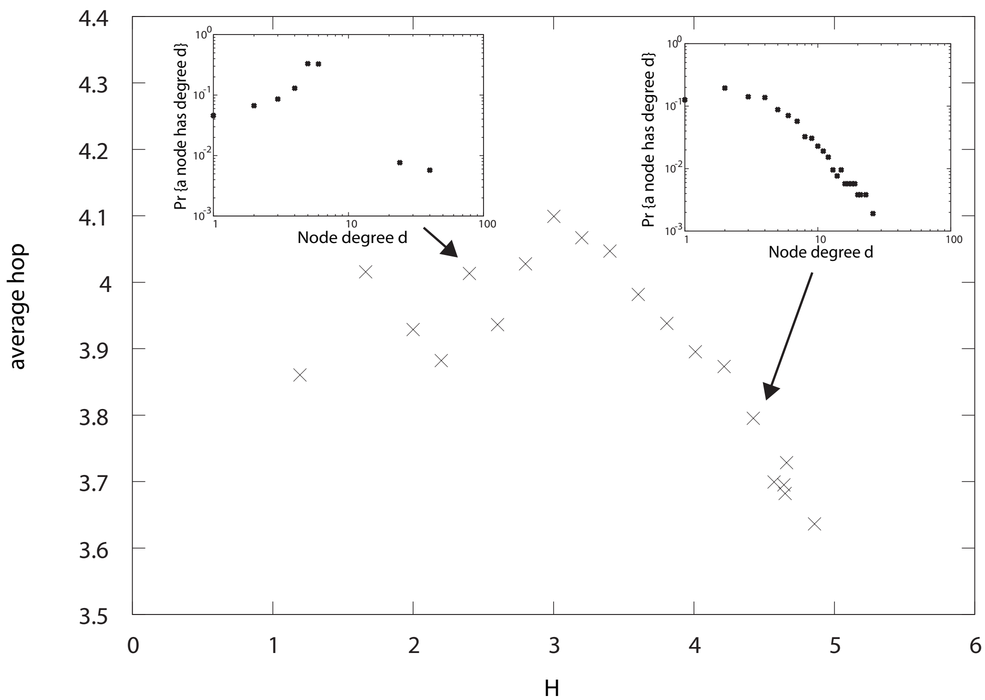

Figure 3. Average hop distance

III. Diversity of router-level topology

In this section, we show the mutual information of some router-level topologies: Level3, Verio, AT&T, Sprint and Telstra. The results are summarized in Table I. The router-level topologies are measured by Rocketfuel tool [11]. To compare with those router-level topologies, topologies made by BA model [4] and ER model [12] which has the same number of nodes and links with AT&T are also shown. From Table I, we can see that, except Verio, the mutual information of router-level topologies are high, and that of model-based topologies, such as the ones generated by BA model and ER model, are low. This can be explained by a design principle of router-level topologies. Because router-level topology is designed under the physical and technological constraints such as the number of switching ports and/or maximum switching capacity of routers, there are some restrictions and a kind of regulations on constructing the topologies, so that they are less diverse. Note, however, that the mutual information of Verio is low. This can be explained by its growing history. Because Verio grows big with small ISPs [13], it contains various kinds of design principles conducted in each ISP. Therefore, Verio is more diverse than other router-level topologies.

IV. Mutual information and the characteristic of topologies

As we mentioned in Section II, mutual information is defined by entropy and conditional entropy. In this section, we explore the relationship between entropy, conditional entropy and the characteristic of topologies respectively.

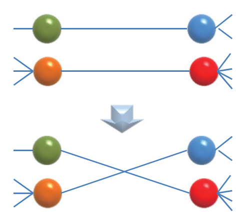

Figure 4. Rewiring method to leave the degree distribution unchanged

Table II
TOPOLOGIES OBTAINED BY SIMULATED ANNEALING

Topology	Nodes	Links	$H(G)$	$H_c(G)$	$I(G)$
BA	523	1304	4.24	3.98	0.26
T_{Imin}	523	1304	4.24	4.13	0.12
T_{Imax}	523	1304	4.24	1.54	2.70

A. Entropy $H(q)$ and the characteristic

To show the relationship between degree distribution and the characteristic of topologies, we generate topologies having different entropy, and compared their average hop distance and degree distribution.

Topologies are generated by simulated annealing that

 ISBN: 978-1-61208-226-4

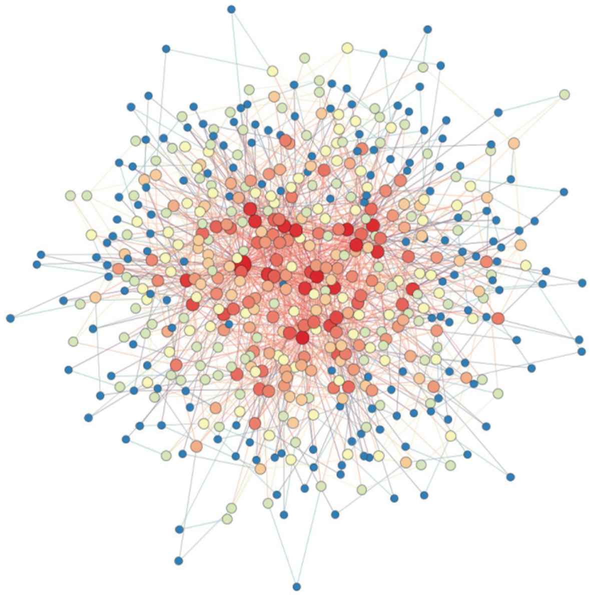

Figure 5. T_{Imin} with minimum mutual information

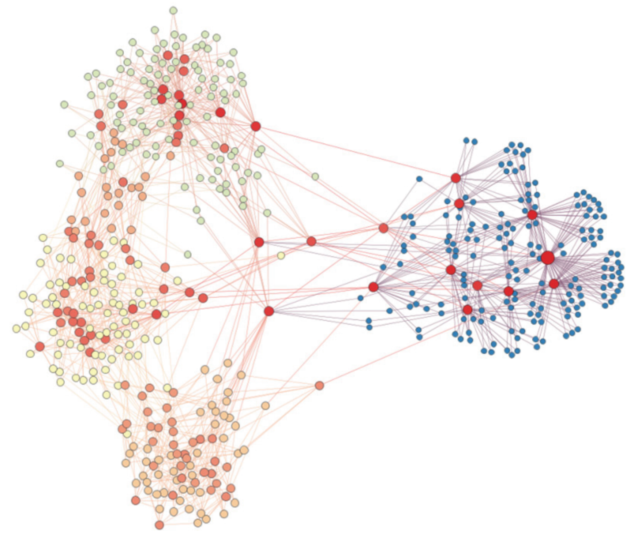

Figure 6. T_{Imax} with maximum mutual information

looks for candidate networks that minimize the potential function $U(G)$. Here, the temperature is set to 0.01, and the cooling rate is set to 0.0001. The simulation searched 450000 steps. The initial topology is set to the topology obtained by BA model which has the same number of nodes and links with AT&T. Topologies are changed by random rewiring, and try to minimize the following potential function:

$$U(G) = \sqrt{(H - H(G))^2 + (H_c - H_c(G))^2}. \quad (5)$$

Here H and H_c are pre-specified value of entropy and conditional entropy respectively. $H(G)$ and $H_c(G)$ are entropy and conditional entropy calculated by the topology G generated in the optimizing search process. We generated topologies by setting H, H_c as $H = H_c$ from 1 to 5. Every time in the search process, $U(G)$ converge to approximately zero. Therefore, entropy and conditional entropy of the generated topologies are almost equal.

Figure 3 shows the average hop distance of topologies we generated. It can be seen that, when H increases higher than 3, the average hop distance decreases. This is because, as H increases, the degree distribution become biased, and it gets close to power-law around $H = 4$.

B. Conditional entropy $H_c(q|q')$ and characteristic

Next, we show the relationship between mutual information and the characteristic of topologies. Because router-level topologies obey power-law, we compare topologies having high $H(\mathrm{q})$.

Topologies are again generated by the simulated annealing. We set the same parameter and the same initial topology as we have used in the previous section. The different points are the way to rewire the topology and the potential function $U^I(G)$. For the first point, topology is changed by a rewiring method [14] that leaves the degree distribution unchanged, i.e., by exchanging the nodes attached to any randomly selected two links (Figure 4). For the second point, the potential function we used to minimize is $U^I(G)$ defined as,

$$U^I(G) = |I - I(G)|, \quad (6)$$

where I is pre-specified mutual information, and $I(G)$ is mutual information calculated by the topology G generated in the optimizing search process. Note that looking for a pre-specified mutual information I is as the same as looking for a pre-specified conditional entropy H_c under the same entropy H. Because the entropy is same when the degree distribution unchanged, minimizing mutual entropy is identical to maximize conditional entropy.

To explain the relationship between mutual information and the characteristic of topologies, we use two topologies: topology T_{Imin} with minimum mutual information and topology T_{Imax} with maximum mutual information. T_{Imin} is generated by setting $I = 0.0$ for simulated annealing, and the resulting mutual information is 0.12. The topology is shown in Figure 5. T_{Imax} is generated by setting $I = 3.0$ for simulated annealing, and the resulting mutual information is 2.70. The topology is shown in Figure 6. In both figures,

 ISBN: 978-1-61208-226-4

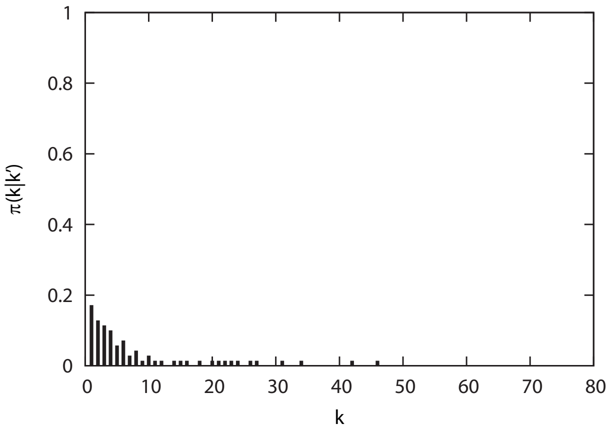

Figure 7. $\pi(k|k')$ of nodes with the largest remaining degree in T_{Imin}

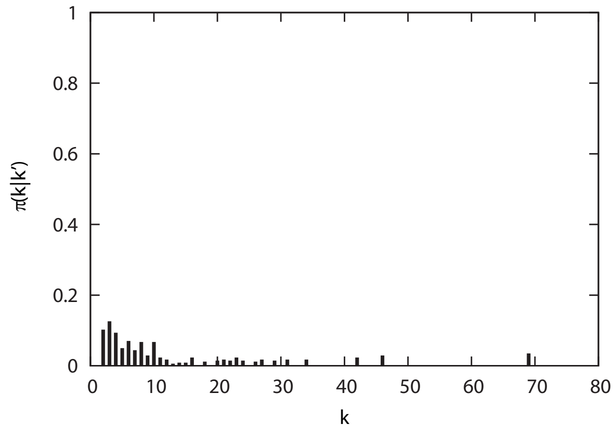

Figure 8. $\pi(k|k')$ of nodes with the smallest remaining degree in T_{Imin}

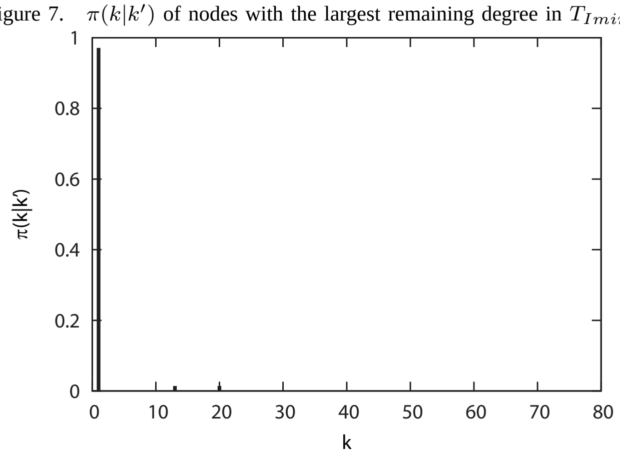

Figure 9. $\pi(k|k')$ of nodes with the largest remaining degree in T_{Imax}

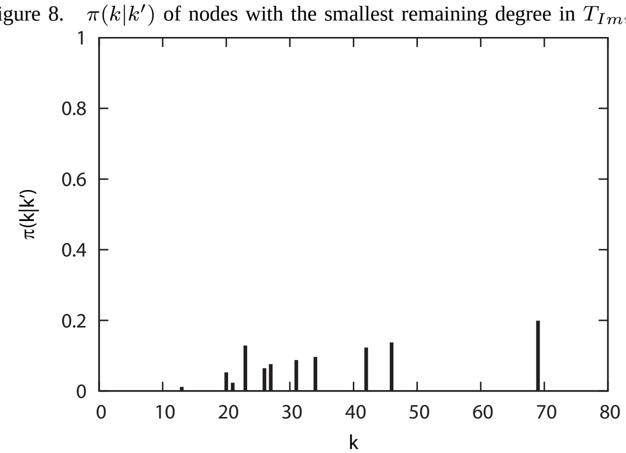

Figure 10. $\pi(k|k')$ of nodes with the smallest remaining degree in T_{Imax}

colors represent node degrees. Nodes which have the same color have the same node degree. Topological characteristics of the initial topology, T_{Imin} and T_{Imax} are summarized in Table II.

From Figure 5 and Figure 6, we can see that topology with high mutual information is less diverse, and have more regularity than the one with low mutual information. From Figure 7 to Figure 10, we show $\pi(k|k')$ dependent on remaining degree k. $\pi(k|k')$ is defined as the probability that observing a vertex with k' edges leaving it provided that the vertex at the other end of the chosen edge has k leaving edges. Figure 7 and Figure 8 show $\pi(k|k')$ of nodes with the largest remaining degree and nodes with the smallest remaining degree in T_{Imin}, respectively. Figure 9 and Figure 10 show $\pi(k|k')$ of nodes with the largest remaining degree and nodes with the smallest remaining degree in T_{Imax}, respectively. We can see that $\pi(k|k')$ of T_{Imax} is more biased than that of T_{Imin}. This also represents that the topology with high mutual information is less diverse than the one with low mutual information.

V. Conclusion and Future Work

In this paper, we investigated the network heterogeneity of router-level topologies by using mutual information. From calculating mutual information of some router-level topologies, we found that router-level topologies have higher mutual information than model-based topologies. We also generated topologies with different mutual information, and showed that the topology is diverse when mutual information is high, and the topology has regularity when mutual information is low.

Our next work is to evaluate network performance of topologies with different mutual information, and to apply this measure to designing information network that has adaptability and sustainability against environment changes.

Acknowledgment

This research was supported in part by Grant-in-Aid for Scientific Research (A) 24240010 of the Japan Society for the Promotion of Science (JSPS) in Japan.

REFERENCES

[1] Y. Koizumi, T. Miyamura, S. Arakawa, E. Oki, K. Shiomoto, and M. Murata, "Stability of virtual network topology control for overlay routing services," *OSA Journal of Optical Networking*, pp. 704–719, Jul. 2008.

[2] R. Solé and S. Valverde, "Information theory of complex networks: On evolution and architectural constraints," *Complex networks*, vol. 650, pp. 189–207, Aug. 2004.

[3] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," *ACM SIGCOMM Computer Communication Review*, vol. 29, pp. 251–262, Oct. 1999.

[4] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, Oct. 1999.

[5] R. Albert, H. Jeong, and A. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, 2000.

[6] K. L. Goh, B. Kahng, and D. Kim, "Universal behavior of load distribution in scale–free networks," *Physical Review Letters*, vol. 87, no. 27, Dec. 2001.

[7] L. Li, D. Alderson, W. Willinger, and J. Doyle, "A first-principles approach to understanding the Internet's router-level topology," *ACM SIGCOMM Computer Communication Review*, vol. 34, pp. 3–14, Oct. 2004.

[8] R. Fukumoto, S. Arakawa, and M. Murata, "On routing controls in ISP topologies: A structural perspective," in *Communications and Networking in China, 2006. ChinaCom'06. First International Conference on*. IEEE, Oct. 2006, pp. 1–5.

[9] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, Oct. 2002.

[10] S. Arakawa, T. Takine, and M. Murata, "Analyzing and modeling router-level Internet topology and application to routing control," *Computer Communications*, vol. 35, pp. 980–992, May 2012.

[11] 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, Feb. 2004.

[12] P. Erdős and A. Rényi, "On random graphs," *Publicationes Mathematicae Debrecen*, vol. 6, pp. 290–297, 1959.

[13] M. Pentz, "Verio grows big with small clients," *Business Journals*, Feb. 1999.

[14] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, "Systematic topology analysis and generation using degree correlations," *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 4, pp. 135–146, Oct. 2006.