An Empirical Study of Routing Bias in Variable-Degree Networks

Shudong Jin
Department of Electrical Engineering and Computer Science, Case Western Reserve University
Cleveland, OH 44106
jins@cwru.edu

Azer Bestavros
Computer Science Department, Boston University
Boston, MA 02215
best@cs.bu.edu

Abstract

Recent work on Internet measurement and topology characterization has revealed a feature sharply distinct from the early Erdős-Rényi random graph model: highly variable degree distributions which arguably follow power-laws. Such properties may significantly affect various network algorithms and protocols. This paper explores whether and how degree variability causes routing bias, a phenomenon that routing algorithms may strongly favor nodes with better connectivity. This paper demonstrates that in variable-degree networks, including both synthetic networks and measured Internet topologies, there is a super-linear preference for routing algorithms to choose high-degree nodes in routing decision; in addition, the level of such preference remarkably depends on the degree variability of the network. Empirical study shows that the probability of routing through a node with degree d is approximately proportional to d^θ, where $1 < \theta < 2$ and its value is larger when degree variability is lower. These findings are important to various network practices, including traffic provisioning, routing algorithm design, and network measurement. In particular, this paper illustrates that in traceroute-based topology discovery, routing bias implies that links associated to less well-connected nodes are less likely to be discovered. Hence, routing bias can become an important contributor to the incompleteness of measured router-level topologies.

1 Introduction

Because of its phenomenal growth in size, scope, and complexity, as well as its increasingly central role in society, the Internet has become an important object of study and evaluation. In the last few years, there was a surge in research that attempts to empirically identify invariants about the topological properties of the Internet [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. One topological feature is high variability of node degree distributions, which were often observed to follow approximately a power-law [2]. With such distributions, nodes have non-uniform probabilities of being connected to others, and some of them have extremely large numbers of neighbors. One term to describe this topological feature is *variable-degree networks*, as opposed to the classical Erdős-Rényi model [13]. It is particularly important to understand how variable-degree nature of the Internet topology is relevant to network algorithms and protocols.

This paper focuses on the problem of *routing bias* in variable-degree networks. By routing bias, it is a phenomenon that routing algorithms may favor some nodes and links in a network, but not the others. As a result, different parts of the network, with different levels of connectivity, may have skewed traffi c volume. An example routing protocol is the Open Shortest Path First (OSPF) protocol. OSPF is an Interior Gateway

Protocol used to distribute routing information within a single domain (also called Autonomous System). It uses Dijkstra's algorithm to compute the shortest paths. One key issue is the setting of the cost (also called metric) of an interface in OSPF. The cost could be inversely proportional to the bandwidth of a certain interface, or configured by network administrators in other ways, for example, using a unit cost in order to minimize hop-count of the path. One question is, will this algorithm (and its configurations) cause skewed traffic? This paper explores the inherent relationship between the degree of a node and the probability of routing through this node. More formally, the following quantitative relationship will be empirically studied: let d be the degree of a node, what is the probability p that routing algorithms will choose this node, as a function of d.

This problem is important since understanding routing bias helps us in various practices. For example, routing bias may affect network traffic provisioning. The connectivity of the Internet is constantly changing. A question is, when the degree of a node is increased, how do we expect the change in traffic. Second, understanding routing bias and how such bias arises may also help us in designing new routing algorithms. Third, many measurement activities heavily rely on the routing mechanisms implemented in the Internet. Hence, understanding routing bias may help us conduct more delicate measurements and avoid less appropriate conclusions.

This paper empirically studies how variability of node degree causes routing bias. To derive and validate the results, both synthetic variable-degree networks and real measured Internet topologies are used. The paper reveals that the probability p of routing through a node is often a super-linear function of the degree d of the node. We observed an approximate relationship:

$$p \quad \propto \quad d^{\theta} \tag{1}$$

where $1 < \theta < 2$ and its value depends on the degree distribution of the network. If θ is smaller, then routing bias is weaker. However, it is often that $\theta > 1.5$ for networks with high variability of degree (for example the router-level Internet topology). These findings are important to various practices, including network traffic provisioning, routing algorithm design, and network measurement. In particular, this paper demonstrates that in traceroute-based topology discovery, routing bias implies that links associated to less well-connected nodes are much less likely to be discovered. Hence, routing bias can become an important contributor to the incompleteness of measured router-level topologies, even when measurements are taken from a sufficiently large number of vantage points.

The paper is organized as follows. Section 2 describes the Internet routing topologies under study and illustrates their variable-degree nature. Section 3 demonstrates routing bias in variable-degree networks. Section 4 demonstrates how routing bias may affect the accuracy of traceroute-based topology discovery. Section 5 reviews recent work on topology measurement and characterization. The findings from this research are summarized in Section 6.

2 Variable-Degree Networks

This section describes Internet topologies and shows that they exhibit high variability of node degree. This section also describes the power-law random graph (PLRG) model [14] for generating synthetic variable-degree networks.

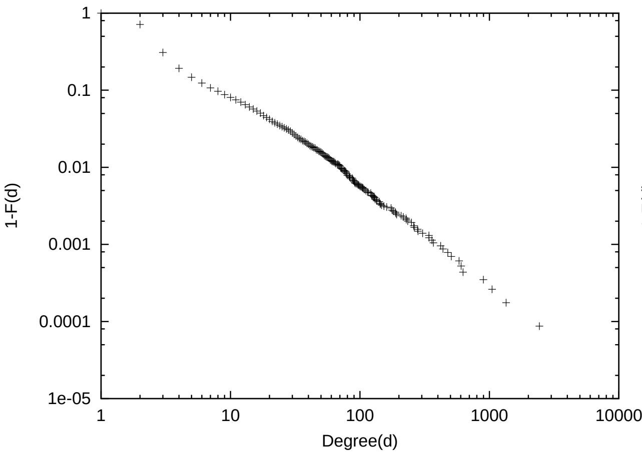
2.1 Internet Topologies under Study

We consider both autonomous system (AS) level and router level Internet topologies. Table 1 summarizes an AS-level topology and a router-level topology.[1] We have also used other Internet topologies, such as the

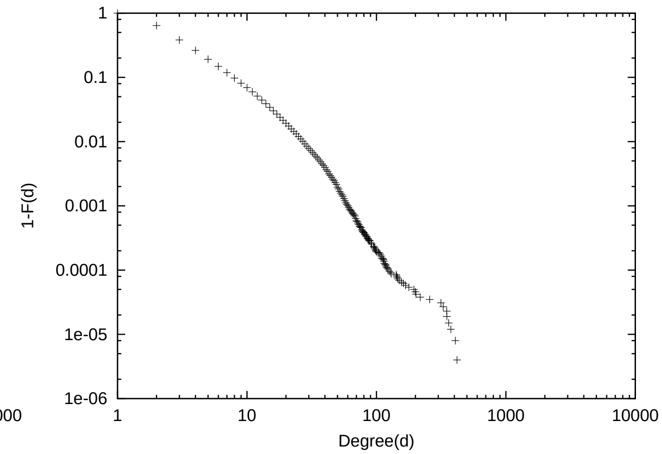
[1]These two topologies may not represent the complete pictures of the Internet and it could be very difficult to find the complete AS-level or router-level Internet topologies. We hope such incompleteness will not change the inherent topological properties.

Table 1: Internet topologies used in this paper.

Topologies	AS-2001+	Skitter
Number of nodes	11461	260080
Mean degree	5.71	3.39
Maximum degree	2432	412
Degree distribution	Power law, $\alpha \approx 1.22$	Weibull, $\beta \approx 0.49$



(a) AS-level topology



(b) Router-level topology

Figure 1: Vertex degree distributions of the AS-level topology and the router-level topology under study. In both topologies, degree is highly variable, but the AS-level topology has higher level of variability.

newer data-sets from CAIDA and from [15]. The results are either similar, or the topologies are not large enough to provide good estimation.

AS-level topology reflects the peering relationship between different routing domains. Many studies used the routing tables at route-views.oregon-ix.net. Since 1997, the routing tables have been collected once a day by the National Laboratory for Applied Network Research (NLANR) [16]. However, it was found that the Oregon route-views is incomplete [6]. Hence, for this study an AS-level topology is obtained from [17], which incorporates not only Oregon route-views, but also Looking glass data and Routing registry data. This topology is dated May 26, 2001. Hereafter, it is called "AS-2001+" topology.

Router-level topology reflects the lower-level physical connectivity of the Internet. For this study a snapshot of the router-level Internet topology is obtained from CAIDA in early 2003. This topology was recorded by 13 of the 19 Skitter monitors worldwide. There are many incomplete paths and non-responding routers in the Skitter logs. During preprocessing, incomplete paths are still used in generating the topologies. Some non-responding routers do not provide their IP addresses. These routers and their associated links are simply ignored. The author has also used old router-level topologies from [18], including the SCAN topology obtained around October/November 1999 using the Mercator software [3]. For the purpose of this study, it suffices to consider a representative topology. Hence, this paper does not present results from using other router-level topologies.

Figure 1 shows the complementary cumulative distribution function $1 - F(d)$ of node degree for the two topologies under study. This distribution function quantifies the probability that a node has a degree larger than a certain value d. A common property of these topologies is the long tails of the distributions, *i.e.*, node degree exhibits high variability. This property departs sharply from the classical Erdős-Rényi model [13].

Erdős-Rényi model assumes nodes have uniform probability of connecting to others and the tail of degree distribution decays exponentially fast.

Previous work [2] has shown that node degree distributions in Internet topologies follow a power-law. This is best illustrated by the AS-level topology. With a power-law distribution, the complementary cumulative distribution function $1 - F(d) = cd^{-\alpha}$, where c and α are constants, and its logarithmic scale plot is a straight line, as shown in Figure 1(a). Estimation using a linear regression shows that for the AS-level topology, the power-law exponent α is close to 1.22. For the router-level topology, it appears that the power-law distribution for the router-level topology does not fit the empirical dataset. Figure 1(b) shows that the tail of the distribution drops faster than any power-law. In [5], Weibull distribution was found to provide a good fit to many Internet object size distributions. Its tail takes on the form $\exp(-(x/\eta)^{\beta})$, where η is the scale parameter and β is the shape parameter. Estimated using rank regression on y-axis, Weibull distribution fits the empirical distribution better, though not perfectly.

Several studies [6, 5] discussed whether power-laws are good fit to the empirical degree distributions, and whether power-laws emerge as the consequence of a particular construction [19]. This paper avoids such debates since the main objective is routing bias caused by degree variability. It is less important to determine if node degree follows a particular distribution, a power-law or other.

2.2 Synthetic Variable-Degree Networks

There are many models for generating variable-degree networks, including those in [20, 7, 21, 14]. This study uses the random graph model described in [14]. For self-inclusion, this model is shown as follows.

(1) Form a set S containing d_v distinct copies of each node v.

(2) Choose a random matching of the elements of S.

(3) For two nodes u and v, the number of edges joining u and v is equal to the number of edges in the matching of S joining copies of u and v.

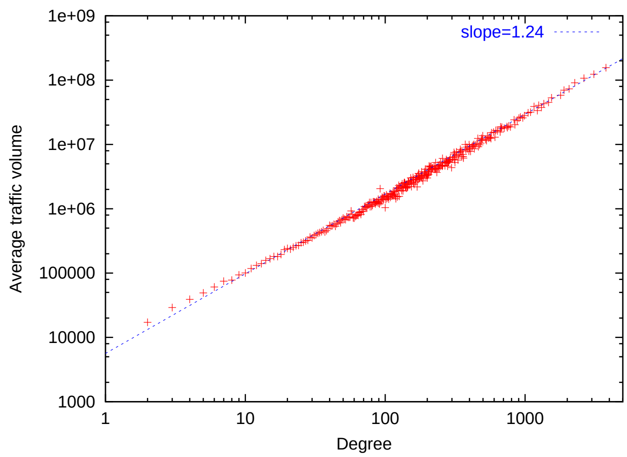
Briefly, given a sequence of nodes and the power-law exponent, the model first assigns degrees to the nodes and then connects the nodes using a random matching. The resulting network is in fact a multi-graph with duplicated edges and self-loops. In all experiments in this study, duplicated edges and self-loops are simply removed. This model is often successful in generating strict power-law distributions. The resulting network might be disconnected. In such cases, the largest connected component is used. The largest connected component typically contains majority of the nodes (its size depends on the parameters of the input, especially the power-law exponent [14]). Note, this model may also be used to generate networks with other degree distributions.

3 Evidence of Routing Bias

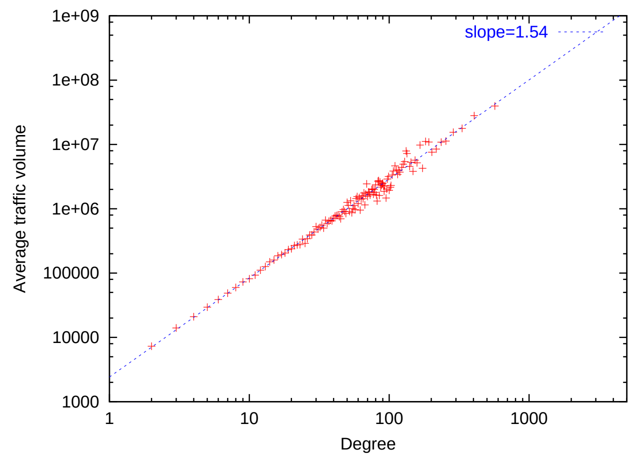
This section studies routing bias when the routing algorithm uses hop-count as the routing metric. The section first examines routing bias in synthetic variable-degree networks, and then extends the main findings to real Internet topologies.

3.1 Routing Bias in Synthetic Topologies

Two variable-degree networks are generated using the PLRG model described in the last section. These two networks are indeed the largest connected components of two outputs of the PLRG model. The first network has 99778 nodes and the average degree is 5.2. The degree distribution follows a power-law with



(a) Network generated by PLRG model, $\alpha = 1.25$.



(b) Network generated by PLRG model, $\alpha = 2.1$.

Figure 2: Traffic volume versus node degree relationship. In both synthetic networks, there is super-linear relationship, but for networks with different level of degree variability, routing bias is also different.

the exponent $\alpha = 1.25$. The second network has 100031 nodes and the average degree is 3.4. The degree distribution also follows a power-law but its exponent $\alpha = 2.1$. Thus the first network has a more heavy-tailed degree distribution, and its power-law exponent is close to that of the AS-level Internet topology under study. The second network is less heavy-tailed and its degree variability is close to that of the router-level topology under study. In addition, we have used other topology models [20, 21] to generate networks with the same levels of degree variability. The use of those models will not change the conclusions reached here.

With these networks, we can examine traffic volume through the nodes as a function of degree d. The following experiment is conducted. Traffic originates randomly in the network, travels through the network, and finally reaches other nodes at the edge. The source and destination are picked randomly (they are edge nodes who have degree one). The routing algorithm is shortest path first. In addition, since the routing algorithm use hop-count as the metric, there are possible ties, *i.e.*, there can be multiple shortest paths from a source to a destination. A tie-breaker is introduced, similar to the one in [12]. Each link is assigned a weight which is $1 + \frac{\epsilon}{|V|}$ where $|V|$ is the number of nodes in the network and ϵ is a uniform random number in interval (0,1). This experiment is repeated for different source-destination pairs. Finally, we can calculate the frequency at which the routing algorithm chooses the nodes with a particular degree value. For this experiment, an all-pair shortest path algorithm is necessary which is computationally rather expensive for large networks.

The result is shown in Figure 2. Each point in the figure represents one degree value. The y-axis shows an average quantity, *i.e.* traffic volume averaged over all nodes with the same degree. Figure 2(a) shows the result from using the synthetic network with higher variability, and Figure 2(b) shows the result from using the other synthetic network. Notice that both x-axis and y-axis are in logarithmic scale.

We have the following observations from this figure. First, the points are closely located on straight lines. It indicates that the probability of routing through a node p is approximately a power-law function of degree d, noticing again the logarithmic scale of the plots. It means the following relationship $p \propto d^{\theta}$, where θ is called the routing bias index. Using a least-square fit, we can estimate the slopes of these two plots, *i.e.*, the values of θ. The figure shows the fits. Second, for the two networks, there are different values of the routing bias index. For the network with lower degree variability, see Figure 2(b), the routing bias index is larger.

The second observation stimulates more interests. It suggests that routing bias may largely depend on the degree variability of the underlying network. To validate this, the following experiment is conducted. We

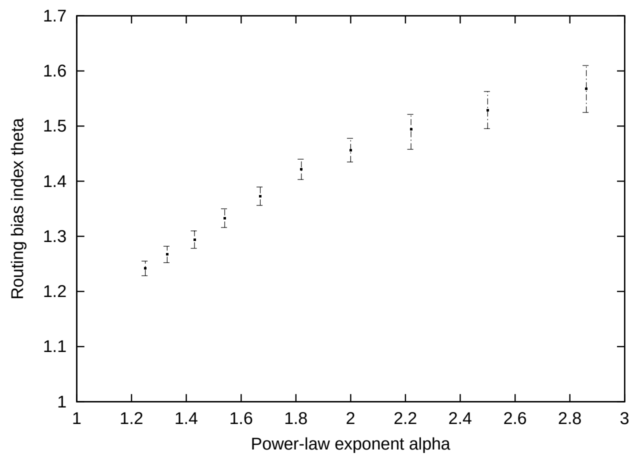


Figure 3: Dependence of the routing bias index on the variability of node degree. Lower variability is consistent with stronger bias.

generate networks using the PLRG model and vary the value of the power-law exponent α. The networks have approximately 20000 nodes. Then traffic volume is calculated as before and the routing bias index is estimated. Finally, we can examine the routing bias index as a function of the power-law exponent α. Figure 3 shows the result. For each value of α, a number of power-law networks have been generated and the figure shows the 90% confidence interval of 100 runs.

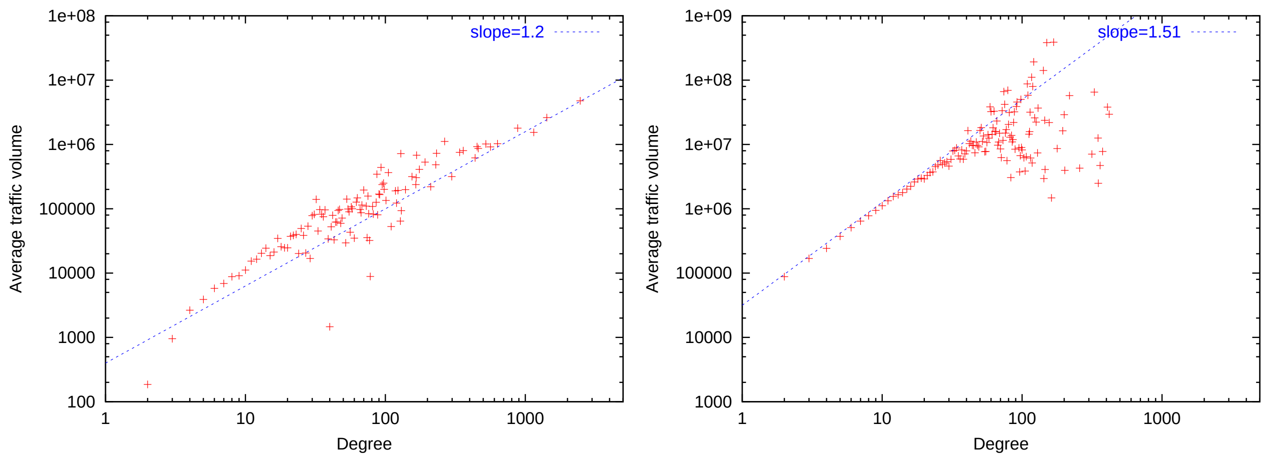
This figure clearly demonstrates that within a wide range of α, degree variability leads to routing bias. The value of $\theta < 2$ but it consistently larger than unity. An explanation is as follows. If the network has a small number of very well-connected nodes. then these well-connected nodes are good navigators and they are very much favored by the routing algorithms. Hence, traffic through these nodes are super-linearly higher. Meanwhile, Figure 3 also shows that higher variability (*i.e.* smaller value of α) leads to weaker routing bias. This is a little surprising. One plausible explanation is as follows. When the degree is extremely variable, there are so many good navigators in the network that routing algorithm may no longer favor individual well-connected nodes much. In other words, only if the network has few but rather well-connected nodes, they become very important in routing.

3.2 Routing Bias in Internet Topologies

Previous experiments and analysis have drawn a conclusion that in variable-degree networks, in particular synthetic power-law networks, routing bias exists due to degree variability. An immediate question is, does routing bias also exist in real Internet topologies, since the Internet topologies exhibit richer and possibly much different properties. To answer this question, let us consider the AS-level and router-level topologies summarized in Table 1. As in the last subsection, same experiments are conducted. The result is shown in Figure 4.

From this figure, there is an indication of routing bias, although unlike the results using synthetic networks, the plots do not fit a straight line perfectly. A possible reason is that in real Internet topologies, there are other structural properties, such as hierarchy [8] and clustering behavior [7, 22]. Random graph models such as the PLRG model may not capture such properties very well. For example, in real Internet it is likely a well-connected regional node serves as the navigators for a small number of end-hosts. This is contrast to a less well-connected core node which is on the paths between many end-host pairs.

For the two plots in Figure 4, we have also used a least-square fit to estimate their slopes, *i.e.* the values of the routing bias index θ. For the AS-level topology, the value of θ is close to 1.2. Since this topology



(a) AS-level topology

(b) Router-level topology

Figure 4: Routing bias also exists in Internet topologies. The level of bias depends on the variability of degree.

has the power-law exponent $\alpha \approx 1.22$, this result is consistent with the α-θ relationship in Figure 3. For the router-level topology, the estimated value of θ is close to 1.5. Since its degree variability is lower (the power-law exponent $\alpha \approx 2.1$), the result is also consistent with Figure 3.

4 Routing Bias and Topology Discovery

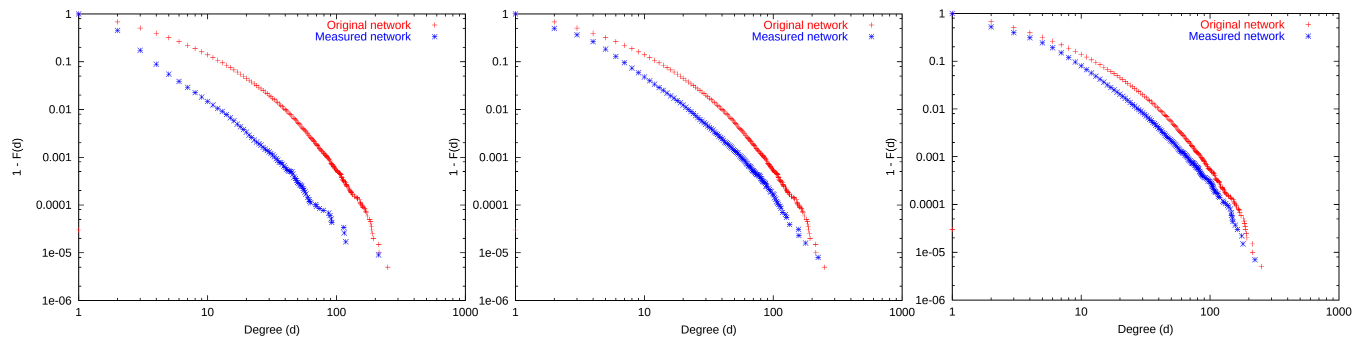
This section illustrates that traceroute-based router-level topology discovery may result in incomplete topology due to routing bias. Note, there are different causes of incompleteness of measured topology: routing bias and limited vantage points. This section distinguishes their effects.

4.1 Traceroute-based Topology Discovery

One challenging problem in modeling the Internet and evaluating various large-scale content delivery techniques is the lack of accurate and complete router-level Internet topology. To obtain an accurate topology, there are many projects. Researchers typically rely on various probing methods and heuristics to infer the internal structure of the router-level topology. By conducting a large amount of probing, it is possible to assemble an overall picture of the topology, though the accuracy and completeness of such a picture are not validated.

One probing method is to use traceroute, a tool that reports the interfaces along the path from a source to a destination. With traceroute, a source host can send a ping packet towards a destination. One of the parameters on a ping packet is *time-to-live* (TTL). This is set to a number. As the packet moves through a router in the path to the destination, this number is decremented by one. If any router sees the TTL value is zero, the it sends the packet back to the source. By setting TTL to different values (starting from one), the source can detect all routers along the path to the destination, assuming that the interfaces respond to the ping packets.

In a typical traceroute-based measurement study, there are often multiple active traceroute sources, or *vantage points*. Each source can assemble the traceroute paths to many destinations and obtain a view of the topology. If one collects and merges views from a large number of sources, then the resulting topology is a more complete reflection of the true topology. Fewer sources certainly leads to less complete topology.



(a) One source

(b) Four sources

(c) Eight sources

Figure 5: Effect of limited vantage points on topology discovery. Small number of sources cause more heavy-tailed degree degree distribution.

This is called the effect of limited vantage points. This effect was studies in [12]. For self-inclusion, the next subsection briefly illustrates this effect.

4.2 Effect of Vantage Points

To study this effect, the following experiment is conducted. First, a synthetic network is generated using the PLRG model. This network has 201772 nodes and its average degree is 3.4. The degrees follow a Weibull distribution (note the PLRG model allows degree distributions other than power-law). Then in this network, some end-hosts are chosen randomly as sources, and the shortest path to all other end-hosts are computed. By assembling the shortest paths, we can obtain a measured topology and its degree distribution. Figure 5 shows three observed degree distributions, corresponding single source, four sources, and eight sources in experiment. For comparison, the figure also shows the degree distribution (Weibull) of the original network.

The figure shows when only one source is used, the measured degree distribution is very close to a power-law, which is sharply different from the original Weibull distribution. When the number of sources increases, the measured degree distribution becomes closer to the original one, and it appears less heavy-tailed. However, even when there are eight sources and there are measured paths from each source to all other end-hosts, the observed topology is still far from being complete. These results indicate that with a small number of vantage points, it is dangerous to claim the observed degree distribution reflects the true underlying network.

One drawback of above experiments is that the incompleteness of the measured topology is due to not only the limited number of vantage points, but also the routing bias described in the previous sections. Remind that with routing bias, there is a super-linear preference for well-connected nodes in the shortest paths. A question is, how important is routing bias on topology discovery? The next subsection distinguishes the effect of routing bias from that of limited vantage points.

4.3 Effect of Routing Bias

The principle how traceroute works suggests that the effectiveness of traceroute-based topology discovery largely depends on the routing algorithm (each router has a routing table generated using a shortest-path-first algorithm, and this routing table determines the next hop in the path to the destination). Routing bias may affect the accuracy of traceroute-base measurement, in addition to the choice of vantage points.

In order to study the pure effect of routing bias, we need first to eliminate the effect of limited number of vantage points. Therefore, in the experiments in this subsection, measurements are taken from a very large

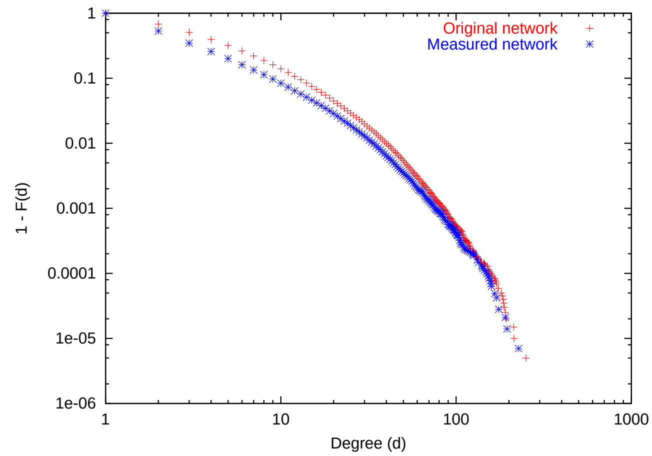


Figure 6: Effect of routing bias on topology discovery, with a synthetic network as the underlying network. Routing bias causes incompleteness of the measured network.

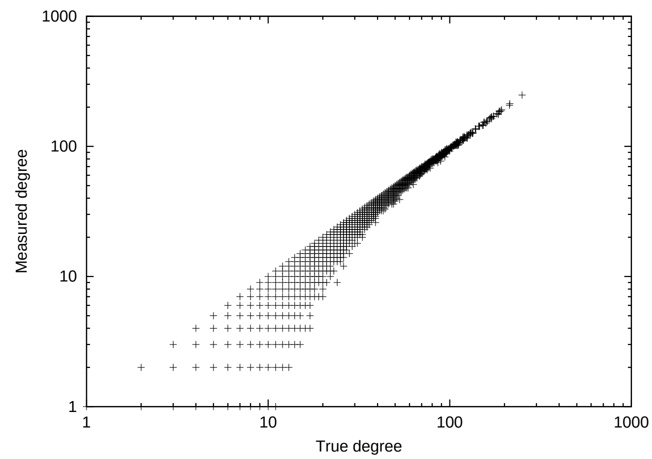


Figure 7: In the synthetic network, routing bias causes links associated to less well-connected nodes are less likely to be discovered.

number of sources. Two networks are used in the experiments. The first is the synthetic network used in the last subsection. The other network is the Skitter router-level topology described in Table 1.

In the synthetic network, 1000 random sources are picked, and for each source 500 random destinations are picked. The measured network is assembled from a half million paths (there are over three million links in total). Figure 6 shows the degree distribution of the measured network. For comparison, the figure also shows the the degree distribution of the original network. From this figure we can see that the degree distribution of the measured network is close to that of the original network. The two curves fit less well when degree is small. It implies that routing bias (the found super-linear preference for high-degree nodes) indeed changes the shape of the degree distribution.

To further understand the effect of routing bias, let us compare the degrees of each node in the original network and in the measured network. Figure 7 shows the result where each point represents a node. There are many points below the diagonal of the plot, indicating many nodes have undiscovered links.[2] This is

[2]The diagonal of the plot means a fully discovered network. If there is an unbiased measurement (which may not exist at all), then the network will be almost fully discovered and all points will be close to the diagonal, given the large total number of links

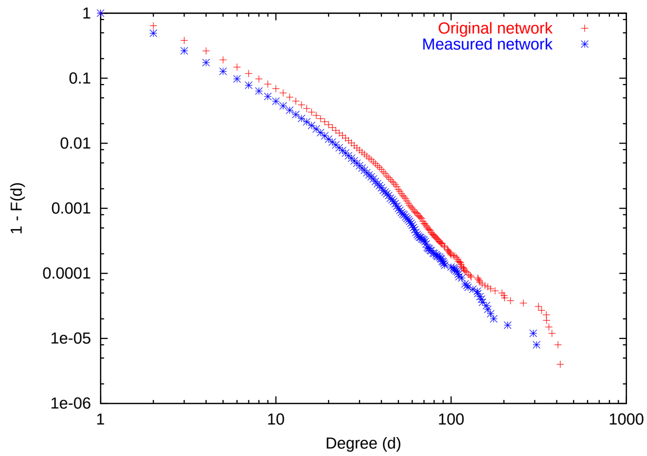


Figure 8: Effect of routing bias on topology discovery, with the Skitter router-level topology as the underlying network. Routing bias causes incompleteness of the measured network.

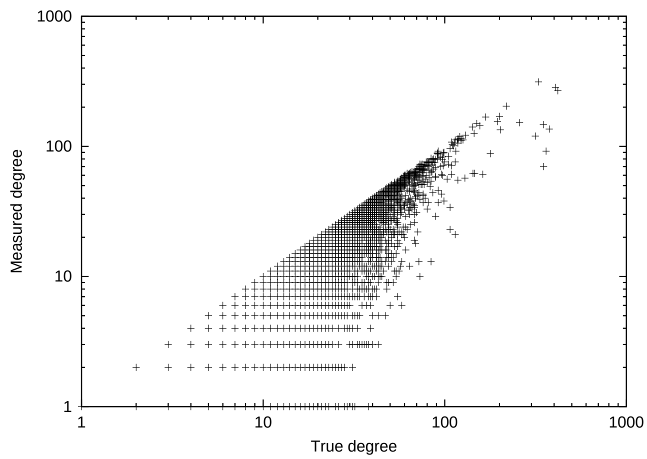


Figure 9: In the Skitter topology, generally routing bias causes links associated to less well-connected nodes are less likely to be discovered.

particular true for the nodes with low degree. For very well-connected nodes, almost all links are discovered; but for less well-connected nodes, it is possible most of its links are not discovered. Clearly, routing bias, the property that links associated to low-degree nodes are less likely to be traversed, causes a more heavy-tailed distribution of the measured network.

With the Skitter router-level topology, similar experiment is conducted.[3] First, 1000 random sources are picked, and for each source 1000 random destinations are picked. A measured network is obtained by assembling one million paths. The degree distribution is shown in Figure 8. It shows that this distribution is still much different from the true distribution in the original network.

Again, let us compare the degrees of each node in the original network and in the measured network. Figure 9 shows the result. We can see there are many points below the diagonal of the plot. A clear observation is for less well-connected nodes, it is more likely many of their associated links are not discovered.

sampled.

[3]This topology may not be the complete router-level topology. However, it is still valuable to find if routing bias affects topology discovery in this network.

This observation explains why the measured degree distribution does not fi t the original one very well when the degree value is small. In addition, for well-connected nodes, it is also possible many links are not discovered. This is certainly not because of the bias due to variable-degree nature of the network. Likely it is because of other structural properties in the real router-level topology. Such structural properties may explain why the curves in Figure 8 do not fi t well when the degree value is large.

5 Related Work

In the past few years, considerable attention has been paid to Internet topology measurement and characterization. One topological feature is the skewed degree distributions. With such distributions, nodes have a non-uniform probability of being connected to others, with some nodes having extremely large numbers of neighbors. The degree distributions of both the AS-level and router-level Internet topologies have often been observed to follow approximately a power-law [2].

Various models have been proposed to generate variable-degree network, as opposed to the classical Erdős-Rényi model [13] and other structural topology models [23, 1]. One family of variable-degree topology generators [20, 7] are based on the scale-free power-law model [19]. This model uses a preferential connectivity mechanism (or the so-called ”rich-get-richer” phenomenon) to drive the network to power-law distribution. Another model is Inet [21]. It assigns degrees to the nodes following a power-law distribution, and then uses a linear preferential model to realize the assigned node degrees. The random graph model in [14], as described earlier, also uses a random matching to generate power-law degree distributions. A recent study [9] divides topology models into two categories: structural and degree-based models. The main difference is structural models explicitly generate hierarchy while degree-based models generate variable-degree networks without specifying hierarchy. However, [9] found that networks generated using degree-based models automatically exhibits hierarchical properties.

There are debates on whether power-law is a good fi t to Internet degree distribution, and why there are highly variable degrees. Several studies [6, 24] indicated that the scale-free network model [19] may not explain them well. Indeed Internet object sizes may be better captured by other distributions such as Weibull distribution [5]. A recent study [12] revealed that even the observed approximate power-law degree distributions can be the result of sampling biases. Sampling biases mean edges are sampled in a highly biased manner. Therefore, as they have shown, even if the underlying network is created using the Erdős-Rényi model, the observed network still exhibits power-law degree distribution. In this sense, [12] is closely related to this paper. However, [12] as well as [25] mainly consider the effect of vantage points, *i.e.*, when there are few sources of measurement, the observed topology has different degree distribution. This paper distinguishes routing bias from the effect of limited vantage points and shows routing bias due to degree variability causes incompleteness of measured topology.

Understanding Internet topology is certainly critical, but more importantly, we need to understand how the topological properties, such as high variability of node degree, impact the design and evaluation of various network algorithms and protocols. In the past, signifi cant work has been done in that regard. Two example topics of recent interests are evaluation of multicast effi cacy, and susceptibility of the Internet to fault and attack. Several recent studies have addressed the impact of topology on IP multicast using shortest path trees [26, 27, 28, 29]. These studies indicated that multicast tree size depends on the underlying network model assumption, and high variability of degree does matter [22]. Other studies have examined fault-tolerance of the Internet and power-law networks [30, 31, 32, 33, 34]. They found power-law networks are more robust to random faults but susceptible to malicious and targeted attacks. This paper has identifi ed another important aspect—routing bias in variable-degree networks, which may affects various network practices, ranging from traffi c provisioning to routing policy design to network measurement.

6 Conclusions

This paper has focused on routing bias in variable-degree networks. In such networks, degree variability may cause severely skewed routing preference for the nodes. Our empirical study has demonstrated that there is a strong preference for well-connected nodes in variable-degree network, and the probability of routing through a node is a super-linear function of the degree of the node. In addition, the level of routing bias depends on the degree variability of the underlying network. The paper has also demonstrated the relevance of routing bias to the accuracy of traceroute-based router-level topology discovery. Routing bias implies that links associated to less well-connected nodes are much less likely to be discovered. Hence, routing bias can become an important contributor to the incompleteness of measured router-level topology, even when measurements are taken from a sufficiently large number of vantage points.

Routing bias is important not only to network measurement activities that heavily reply on the Internet routing mechanisms, it is also important to other network practices. We are currently investigating the impact of routing bias on new routing algorithm design, and studying routing metrics design to minimize the impact of such bias.

Acknowledgments

Part of the data used in this research was provided by CAIDA, http://www.caida.org. The authors would like to thank them for making Skitter trace data available for this study. The authors would also like to thank the topology project team at University of Michigan, http://topology.eecs.umich.edu, for making their AS graph data sets available, The authors would like to thank John Byers for his valuable suggestions for conducting this research.

References

[1] E. W. Zegura, K. L. Calvert, and M. J. Donahoo, "A quantitative comparison of graph-based models for Internet topology," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 770–783, 1997.

[2] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," in *Proceedings of ACM SIGCOMM*, 1999.

[3] R. Govindan and H. Tangmunarunkit, "Heuristics for Internet map discovery," in *Proceedings of IEEE INFOCOM*, 2000.

[4] W. Willinger, R. Govindan, S. Jamin, V. Paxson, and S. Shenker, "Scaling phenomena in the Internet: Critically examining criticality," in *Proceedings of the National Academy of Sciences*, 2001.

[5] A. Broido and kc claffy, "Internet topology: Connectivity of IP graphs," in *Proceedings of SPIE ITCOM*, 2001.

[6] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "The origin of power laws in Internet topologies revisited," in *Proceedings of IEEE INFOCOM*, 2002.

[7] T. Bu and D. Towsley, "On distinguishing between Internet power law topology generators," in *Proceedings of IEEE INFOCOM*, 2002.

[8] L. Subramanian, A. Agarwal, J. Rexford, and R. H. Katz, "Characterizing the Internet hierarchy from multiple vantage points," in *Proceedings of IEEE INFOCOM*, June 2002.

[9] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "Network topology generators: Degree based vs. structural," in *Proceedings of ACM SIGCOMM*, 2002.

[10] N. Spring, R. Mahajan, and D. Wetherall, "Measuring ISP topologies with Rocketfuel," in *Proceedings of ACM SIGCOMM*, August 2002.

[11] R. Mahajan, N. Spring, D. Wetherall, and T. Anderson, "Inferring link weights using end-to-end measurements," in *Proceedings of ACM SIGCOMM Internet Measurement Workshop (IMW)*, 2002.

[12] Anukool Lakhina, John Byers, Mark Crovella, and Peng Xie, "Sampling biases in IP topology measurements," in *Proceedings of IEEE INFOCOM*, 2003.

[13] P. Erdős and A. Rényi, "The evolution of random graphs," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, pp. 17–61, 1960.

[14] W. Aiello, F. R. K. Chung, and L. Lu, "A random graph model for massive graphs," in *ACM Symposium on Theory of Computing (STOC)*, 2000, pp. 171–180.

[15] Rocketfuel: An ISP Topology Mapping Engine, "http://www.cs.washington.edu/research/networking/rocketfuel/," .

[16] National Lab for Applied Network Research, "Global ISP interconnectivity by AS number. http://moat.nlanr.net/AS/," .

[17] Topology Project at University of Michigan, "AS graph data sets. http://topology.eecs.umich.edu/data.html," .

[18] USC Information Sciences Institute, "Internet maps. http://www.isi.edu/div7/scan/mercator/maps.html," .

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

[20] A. Medina, I. Matta, and J. Byers, "On the origin of power laws in Internet topologies," *ACM SIGCOMM Computer Communication Review*, April 2000.

[21] C. Jin, Q. Chen, and S. Jamin, "Inet: Internet topology generator," Tech. Rep. CSE-TR-433-00, EECS Department, University of Michigan, 2000.

[22] S. Jin and A. Bestavros, "Small-world Internet topologies: Possible causes and implications on scalability of end-system multicast," Tech. Rep. BUCS-2002-004, Boston University, 2002.

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

[24] H. Tangmunarunkit, J. Doyle, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "Does AS size determine degree in AS topology?," *ACM SIGCOMM Computer Communication Review*, October 2001.

[25] Paul Barford, Azer Bestavros, John Byers, and Mark Crovella, "On the marginal utility of network topology measurement," in *Proceedings of ACM SIGCOMM Internet Measurement Workshop (IMW)*, November 2001.

[26] J. Chuang and M. Sirbu, "Pricing multicast communications: A cost based approach," in *Proceedings of Internet Society INET*, 1998.

[27] Graham Phillips, Scott Shenker, and Hongsuda Tangmunarunkit, "Scaling of multicast trees: Comments on the Chuang-Sirbu scaling law," in *Proceedings of ACM SIGCOMM*, 1999.

[28] R. Chalmers and K. Almeroth, "Modeling the branching characteristics and efficiency gains in global multicast trees," in *Proceedings of IEEE INFOCOM*, 2001.

[29] C. Adjih, L. Georgiadis, P. Jacquet, and W. Szpankowski, "Multicast tree structure and the power law," in *Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2002.

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

[31] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, "Resilience of the Internet to random breakdowns," *Physical Review Letters*, vol. 4626, pp. 85–89, 2000.

[32] Kihong Park and Heejo Lee, "On the effectiveness of route-based packet filtering for distributed dos attack prevention in power-law internets," in *Proceedings of ACM SIGCOMM*, 2001.

[33] C. R. Palmer, G. Siganos, M. Faloutsos, C. Faloutsos, and P. Gibbons, "The connectivity and fault-tolerance of the internet topology," in *Workshop on Network-Related Data Management (NRDM)*, 2001.

[34] Seung-Taek Park and Alexy Khrabrov, "Static and dynamic analysis of the internet's susceptibility to faults and attacks," in *Proceedings of IEEE INFOCOM*, April 2003.