Analysis of Internet Topologies: A Historical View

Mohamadreza Najiminaini, Laxmi Subedi, and Ljiljana Trajković
Simon Fraser University, Vancouver, Canada
{mna28, lsa38, ljilja}@cs.sfu.ca

***Abstract*— Discovering properties of the Internet topology is important for evaluating performance of various network protocols and applications. The discovery of power-laws and the application of spectral analysis to the Internet topology data indicate a complex behavior of the underlying network infrastructure that carries a variety of the Internet applications. In this paper, we present analysis of datasets collected from the Route Views project. The analysis of collected data shows certain historical trends in the development of the Internet topology. While values of various power-laws exponents have not substantially changed over the recent years, spectral analysis of the normalized Laplacian matrix of the associated graphs reveals notable changes in the clustering of Autonomous System (AS) nodes and their connectivity.**

I. Introduction

Certain characteristics of the Internet topology remain unchanged in spite of its exponential growth. Analyzing the Internet topology and finding properties of associated graphs rely on mining data and capturing information about Autonomous Systems (ASes) [1]. In this paper, we examine datasets from the Route Views project at the University of Oregon [2]. These datasets are collected from Border Gateway Protocols (BGP) routing tables and have been extensively used by the research community [3]–[8].

Analyzing the Internet topology using randomly generated graphs, where routers are represented by vertices and transmission lines by edges, has been widely replaced by exploring properties of the Internet topology on AS-level [4]. Datasets collected from BGP routing tables indicate that the Internet topology is characterized by the presence of various power-laws observed when considering a node degree vs. node rank, a node degree frequency vs. degree, and a number of nodes within a number of hops vs. number of hops [3], [7].

Power-laws also appear in the eigenvalues of the adjacency matrix and the normalized Laplacian matrix vs. the order of the eigenvalues. Eigenvalues associated with a network graph are closely related to its topological characteristics such as the number of edges, spanning trees, and connected components, the diameter of the network, presence of cohesive clusters, long paths and bottlenecks, and the randomness of the network graph. The normalized Laplacian spectrum of the Internet topology on AS level shows invariance regardless of the exponential growth of the Internet [8]. The eigenvectors corresponding to the largest eigenvalues of the Laplacian matrix can also be used to find clusters of ASes with certain characteristics [9]. Datasets from Route Views have been analyzed using spectral analysis in order to find distinct clustering features of the Internet AS nodes [10].

The existence of power-laws in the Internet topology indicates highly skewed distributions of various topology properties measured by power-law exponents [3], [7]. Some of these conclusions were subsequently revised by considering a more complete AS-level representation of the Internet topology [5], [6]. Early research reports showed that power-laws are only consistent with maps of ASes from the Route Views datasets [2] and are inconsistent when analyzing extended maps that have a more complete view of AS connections. These extended maps have heavy tailed or highly variable degree distributions and only the distribution tales have the power-law property.

Graph topologies similar to the Internet topology were generated using various power-laws and distributions of node degrees [11]. Other graph properties causing power-laws, such as preferential connectivity of a new node to existing nodes, incremental growth of the network, distribution of nodes in space, and locality of edge connection, have been exploited to generate synthetic Internet topologies [12]. Preferential connectivity and incremental growth are main parameters in generating Internet-like topologies. Comparing graphs with the power-law degree distribution that are created by topology generators with the Internet AS graph, relies on the interconnection structure of the Internet AS graph [13]. The underlying structure of power-law topologies was also applied in developing graph layout algorithms for visualizing large networks such as the Internet.

The paper is organized as follows: We provide a brief survey of the Internet topology research in Section 1. Spectral analysis and power laws are introduced in Section 2. Analysis of Internet data is presented in Sections 3 and 4. We conclude with Section 5.

II. Spectrum of a Graph and Power-Laws

An Internet AS graph G represents a set of ASes connected via logical links. The number of edges incident to a node in an undirected graph is called the degree of the node. Two nodes are called adjacent if they are connected by a link. The network can be represented by the adjacency matrix A(G):

$$A(i,j) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are connected} \\ 0 & \text{otherwise} \end{cases}$$

This research was supported by the NSERC Discovery Grant 216844-07, the NSERC RTI Grant 330480-06, and the BCNET.

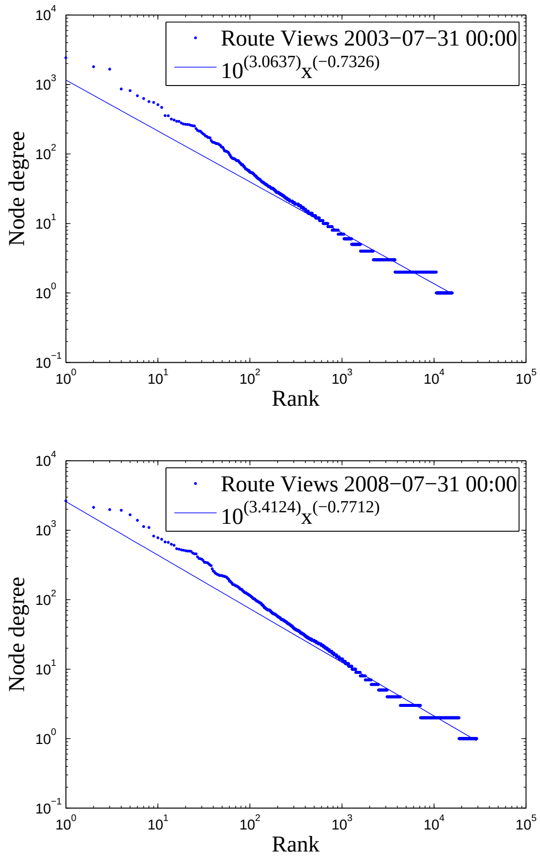

Fig. 1. Route Views 2003 (top) and 2008 (bottom) datasets: The node degree power-law exponents R are –0.7326 and –0.7712 for 2003 and 2008 datasets, respectively. The correlation coefficients are 0.9661 for 2003 (top) and 0.9686 for 2008 (bottom) datasets.

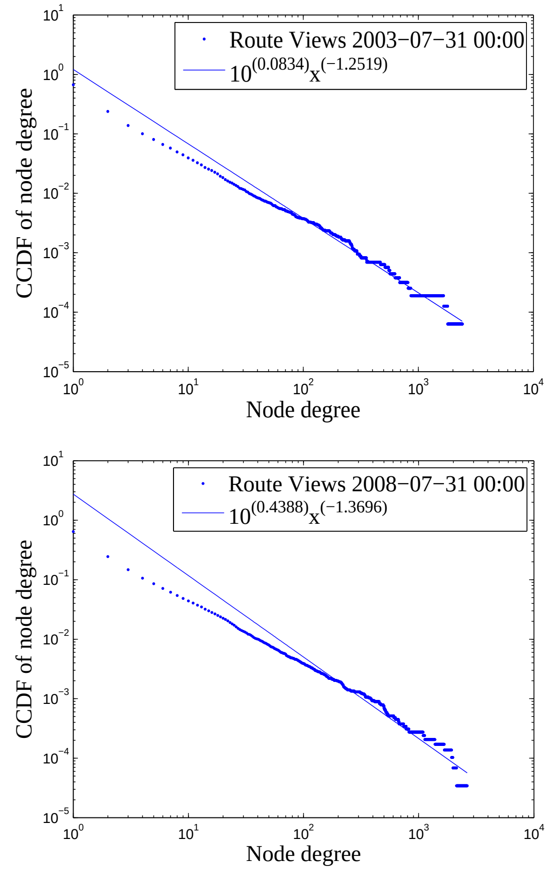

Fig. 2. Route Views 2003 (top) and 2008 (bottom) datasets: The CCDF power-law exponents D are –1.2519 and –1.3696 for 2003 and 2008 datasets, respectively. The correlation coefficients are 0.9810 for 2003 (top) and 0.9626 for 2008 (bottom) datasets.

A diagonal matrix D(G) associated with A(G), with row-sums of A(G) as the diagonal elements, indicates the connectivity degree of each node. The Laplacian matrix is defined as L(G) = D(G) – A(G). The eigenvalues of L(G) are closely related to certain graph invariants. For example, the spectrum of L(G) is the collection of all eigenvalues and contains 0 for every connected graph component.

The normalized Laplacian matrix NL(G) of a graph is defined as:

$$NL(i,j) = \begin{cases} 1 & \text{if } i = j \text{ and } d_i \neq 0 \\ -1/\sqrt{d_i d_j} & \text{if } i \text{ and } j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

where d_i and d_j are degrees of nodes i and j, respectively.

Various power-laws may be associated with graph properties [3], [5]. Node degree vs. rank, frequency of node degree vs. node degree, and eigenvalues vs. index are calculated and plotted on a log-log scale. Linear regression of the analyzed data is used to determine the correlation coefficient between the regression line and the plotted data. A high correlation coefficient between the regression line and the plotted data indicates the existence of a power-law, which implies that node degree, frequency of node degree, and eigenvalues follow a power-law dependency on the rank, node degree, and index, respectively. The power-law exponents are calculated from the linear regression lines $10^{(a)}x^{(b)}$, with segment a and slope b when plotted on a log-log scale.

III. Power-Laws and the Internet Topology

We have compared various graph properties from the Route Views datasets collected over the period of five years, from 2003 to 2008. The BGP routing tables are collected from multiple geographically distributed BGP Cisco routers and Zebra servers. Analyzed datasets were collected at 00:00 am on July 31, 2003 and at 00:00 am on July 31, 2008 [2].

We observe the dependence between the node degree and the rank of each node. The graph nodes v are sorted in descending order based on their degrees d_v and are indexed with a sequence of numbers indicating their ranks r_v. The (r_v, d_v) pairs are plotted on the log-log scale. The power-law implies $d_v \propto r_v^R$, where v is the node number and R is the node degree power-law exponent. Node degrees in decreasing order vs. the rank, plotted on a log-log scale, are shown in Fig. 1. The complementary cumulative distribution function (CCDF) D_d of a node degree d is equal to the number of nodes having degree less than or equal to d, divided by the number of nodes. The power-law implies that $D_d \propto d^D$, where D is

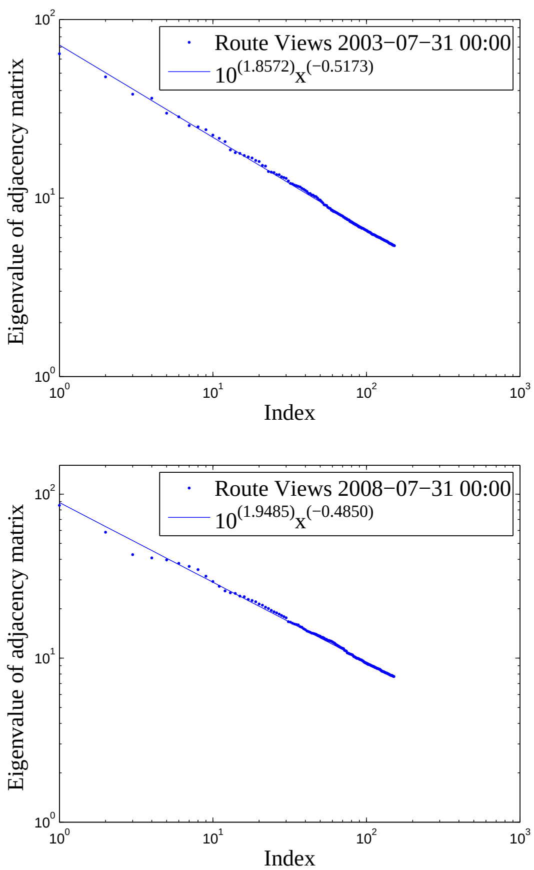

Fig. 3. Route Views 2003 (top) and 2008 (bottom) datasets: The eigenvalue power-law exponents ε are –0.5173 and –0.4850 for 2003 and 2008 datasets, respectively. The correlation coefficients are 0.9990 for 2003 (top) and 0.9882 for 2008 (bottom) datasets.

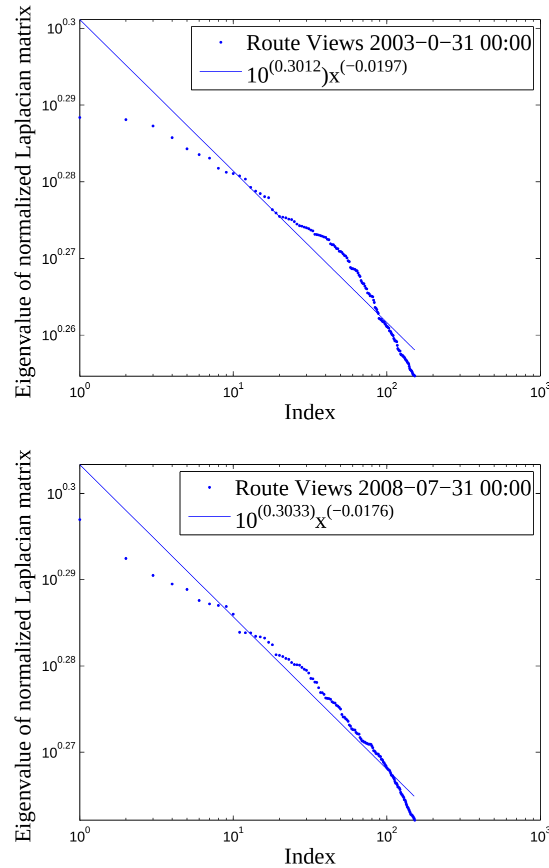

Fig. 4. Route Views 2003 (top) and 2008 (bottom) datasets: The eigenvalue power-law exponents L are –0.0197 and –0.0176 for 2003 and 2008 datasets, respectively. The correlation coefficients are 0.9564 for 2003 (top) and 0.9783 for 2008 (bottom) datasets.

the CCDF power-law exponent. The CCDFs of node degrees, plotted on log-log scale, are shown in Fig. 2.

The eigenvalues λ_{ai} and λ_{Li} of the adjacency matrix and the normalized Laplacian matrix are sorted in decreasing order and plotted vs. the associated increasing sequence of numbers i representing the order of the eigenvalue. Power-laws for the adjacency matrix and the normalized Laplacian matrix imply $\lambda_{ai} \propto i^{\varepsilon}$ and $\lambda_{Li} \propto i^{L}$, respectively, where ε and L are their respective eigenvalue power-law exponents. The dependence between the graph eigenvalues and the eigenvalue index is shown in Fig. 3. Plotted on a log-log scale are eigenvalues of a graph matrix in decreasing order. Only the 150 largest eigenvalues are plotted. The newly observed dependence between the eigenvalues of the normalized Laplacian matrix and the eigenvalue index is shown in Fig. 4. The analysis indicates that in spite of the Internet growth, increasing number of users, and the deployment of new network elements, power-law exponents have not changed substantially.

IV. Spectral Analysis of the Internet Graph

AS numbers range from 0 to 65,535 [1]. Most existing ASes are assigned by regional Internet Assigned Numbers Authority (IANA) registries. The remaining AS numbers are designated by IANA for private use. Certain AS numbers are reserved and do not appear in the Internet graph. In the spectral analysis, we only include the assigned and designated ASes.

We calculate the second smallest and the largest eigenvalues and associated eigenvectors of normalized Laplacian matrix for each Route Views dataset. Each element of an eigenvector is associated with the AS having the same index. ASes are sorted in the ascending order based on the eigenvector values and the sorted AS vector is then indexed. The connectivity status is equal to 1 if the AS is connected to another AS or zero if the AS is isolated or is absent from the routing table. The second smallest eigenvalue, called “algebraic connectivity” [14], [15] of a normalized Laplacian matrix, is related to the connectivity characteristic of the graph. Elements of the eigenvector corresponding to the largest eigenvalue of the normalized Laplacian matrix tend to be positioned close to each other if they correspond to AS nodes with similar connectivity patterns constituting clusters [9]. The connectivity status based on the second smallest and the largest eigenvalues of the normalized Laplacian matrix is shown in Fig. 5 and Fig. 6, respectively. The spectral graphs indicate visible changes in the clustering of AS nodes and the AS connectivity over the period of five years.

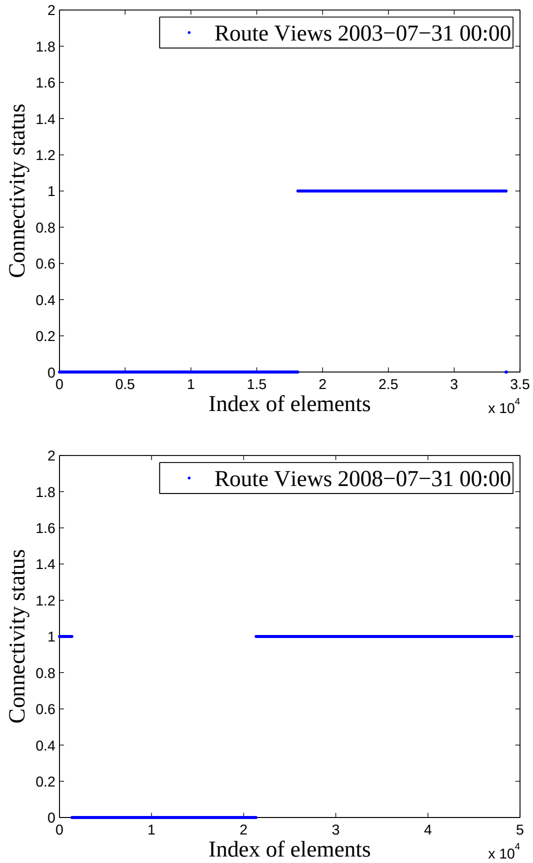

Fig. 5. Route Views 2003 (top) and 2008 (bottom) datasets: Spectral views of the AS connectivity based on the second smallest eigenvalue.

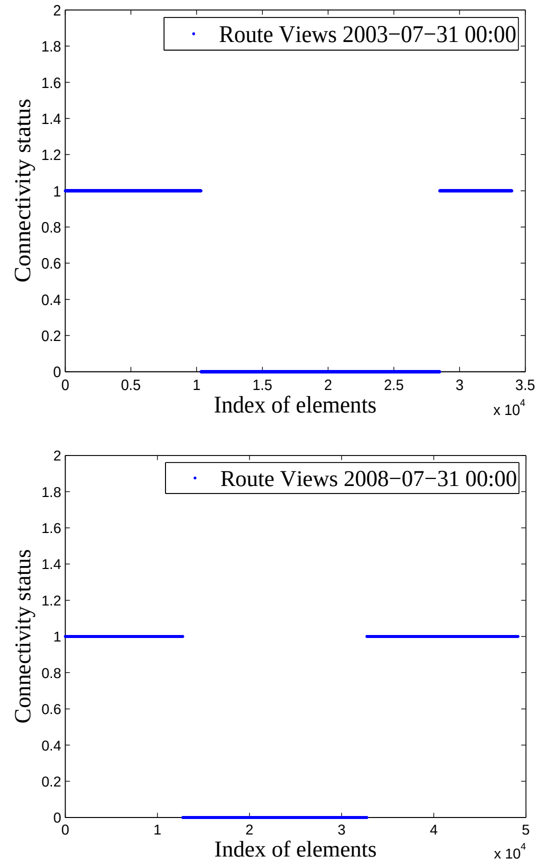

Fig. 6. Route Views 2003 (top) and 2008 (bottom) datasets: Spectral views of the AS connectivity based on the largest eigenvalue.

V. Conclusions

We have evaluated collected data from the Route Views project and have confirmed the presence of power-laws in graphs capturing the AS-level Internet topology. The analysis captured historical trends in the development of the Internet topology over the past five years. While various power-law exponents associated with the Internet topology have remained similar, indicating that the power-laws do not capture every property of a graph and are only one measure used to characterize the Internet, spectral analysis based on the normalized Laplacian matrix indicated visible changes in the clustering of AS nodes and the AS connectivity.

References

[1] Autonomous System Numbers [Online]. Available: http://www.iana.org/assignments/as-numbers.

[2] BGP capture datasets [Online]. Available: http://archive.routeviews.org.

[3] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," in *Proc. ACM SIGCOMM, Computer Communication Review*, vol. 29, no. 4, pp. 251–262, Sept. 1999.

[4] L. Gao, "On inferring autonomous system relationships in the Internet," *IEEE/ACM Trans. Networking*, vol. 9, no. 6, pp. 733–745, Dec. 2001.

[5] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "The origin of power laws in Internet topologies revisited," in *Proc. INFOCOM*, New York, NY, USA, Apr. 2002, pp. 608–617.

[6] H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "Towards capturing representative AS-level Internet topologies," in *Proc. ACM SIGMETRICS Performance Evaluation Review*, New York, NY, June 2002, vol. 30, no. 1, pp. 280–281.

[7] G. Siganos, M. Faloutsos, P. Faloutsos, and C. Faloutsos "Power-laws and the AS-level Internet topology," *IEEE/ACM Trans. Networking*, vol. 11, no. 4, pp. 514–524, Aug. 2003.

[8] D. Vukadinovic, P. Huang, and T. Erlebach, "On the spectrum and structure of Internet topology graphs," in H. Unger et al., Eds., *Innovative Internet Computing Systems*, LNCS 2346. Berlin: Springer-Verlag, 2002, pp. 83–96.

[9] C. Gkantsidis, M. Mihail, and E. Zegura, "Spectral analysis of Internet topologies," in *Proc. INFOCOM*, San Francisco, CA, Mar. 2003, vol. 1, pp. 364–374.

[10] J. Chen and Lj. Trajković, "Analysis of Internet topology data," in *Proc. IEEE International Symposium on Circuits and Systems*, Vancouver, BC, Canada, May 2004, vol. IV, pp. 629–632.

[11] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "Network topology generators: degree-based vs. structural," in *Proc. ACM SIGCOMM, Computer Communication Review*, vol. 32, no. 4, pp. 147–159, Oct. 2002.

[12] A. Medina, I. Matta, and J. Byers, "On the origin of power laws in Internet topologies," in *Proc. ACM SIGCOMM 2000, Computer Communication Review*, vol. 30, no. 2, pp. 18–28, Apr. 2000.

[13] S. Jaiswal, A. Rosenberg, and D. Towsley, "Comparing the structure of power-law graphs and the Internet AS graph," in *Proc. 12th IEEE International Conference on Network Protocols*, Washington DC, Aug. 2004, pp. 294–303.

[14] M. Fiedler, "Algebraic connectivity of graphs," *Czech. Math. J.*, vol. 23, no. 2, pp. 298–305, 1973.

[15] A. Pothen, H. Simon, and K.-P. Liou, "Partitioning sparse matrices with eigenvalues of graphs," *SIAM Journal of Matrix Analysis*, vol. 11, no. 3, pp. 430–452, July 1990.