

Ljiljana Trajković

Analysis of Internet Topologies

Abstract

The discovery of power-laws and spectral properties of the Internet topology illustrates a complex underlying network infrastructure that carries a variety of the Internet applications. Analysis of spectral properties of the Internet topology is based on matrices of graphs capturing Internet structure on the Autonomous System (AS) level. The analysis of data collected from the Route Views and RIPE projects confirms the existence of power-laws and 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 matrices associated with the Internet graphs reveals notable changes in the clustering of AS nodes and their connectivity. It also reveals new historical trends in the clustering of AS nodes and their connectivity. These properties of the Internet topology may be further analyzed by examining element values of the corresponding eigenvectors.

Digital Object Identifier 10.1109/MCAS.2010.937882

I. Properties of the Internet Topology: An Overview

Analyzing the Internet topology using randomly generated graphs, where routers are represented by vertices and transmission lines by edges, has been widely replaced by mining data that capture information about Internet Autonomous Systems and by exploring properties of associated graphs on the AS-level [1]. The Route Views [2] and RIPE [3] datasets collected from Border Gateway Protocol (BGP) routing tables have been extensively used by the research community [5], [8], [14], [16]. The discovery of power-laws and spectral properties of the Internet topology indicates a complex underlying network infrastructure.

Analysis of the collected datasets indicates 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 [8], [14]. Some of these early conclusions were subsequently revised by considering a more complete AS-level representation of the Internet topology [4], [5]. These extended maps have heavy tailed or highly variable degree distributions and only the distribution tails have the power-law property. It has been observed that the power-law exponents associated with Internet topology have not substantially changed over the years in spite of the Internet exponential growth [11], [12]. Power-laws also appear in the eigenvalues of the adjacency matrix and the normalized Laplacian matrix vs. the order of the eigenvalues. They also show invariance regardless of the exponential growth of the Internet.

While various power-law exponents associated with the Internet topology have remained similar over the years, indicating that the power-laws

Spectral analysis of the Internet graphs reveals historical trends in the clustering of AS nodes and their connectivity.

do not capture every property of a graph and are only one measure used to characterize the Internet, spectral analysis of both the adjacency matrix and the normalized Laplacian matrix of the associated graphs reveals new historical trends in the clustering of AS nodes and their connectivity. The eigenvectors corresponding to the largest eigenvalues of the normalized Laplacian matrix have been used to identify clusters of AS nodes with certain characteristics [11]. Spectral analysis was employed to analyze the Route Views and RIPE datasets in order to find distinct clustering features of the Internet AS nodes [6]. For example, the connectivity graphs of these datasets indicate visible changes in the clustering of AS nodes and the AS connectivity over the period of five years [15], [12]. Clusters of AS nodes can be also identified based on the eigenvectors corresponding to the second smallest and the largest eigenvalues of the adjacency matrix and the normalized Laplacian matrix [15]. The connectivity and clustering properties of the Internet topology can be further analyzed by examining element values of the corresponding eigenvectors.

II. The Internet Route Views and RIPE Datasets

Analyzing Internet topologies relies on capturing data information about Internet Autonomous Systems [1] and exploring properties of associated graphs on the AS-level [10]. These datasets collected from BGP routing tables indicate that Internet topology is characterized by the presence of various power-laws [8], [14]. It has also been observed that the power-law exponents associated with the Internet topology have not substantially changed over the years in spite of the Internet exponential growth [11], [12], [15].

It is well known that eigenvalues associated with a network graph are closely related to its topological characteristics [7]. It is not surprising that power-laws also appear in the plots of eigenvalues of the adjacency matrix and the normalized Laplacian matrix vs. the order of the eigenvalues. These power-laws also exhibit historical invariance [12]. The eigenvectors corresponding to the largest eigenvalues of the normalized Laplacian matrix have also been used to identify clusters of AS nodes with certain characteristics [11]. In reported studies [6], [12], [15] spectral analysis was employed to analyze the Route Views and RIPE datasets in order to find distinct clustering features of the Internet AS nodes. Examination of

eigenvectors of both the adjacency matrix and the normalized Laplacian matrix illustrates that both matrices may be used to identify clusters of connected AS nodes.

III. Spectrum of a Graph and Power-Laws

An Internet AS graph G represents a set of AS nodes (vertices) connected via logical links (edge). The number of edges incident to a node is called the degree of the node. Two nodes are called adjacent if they are connected by a link. The graph is defined by the adjacency matrix $A(G)$:

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

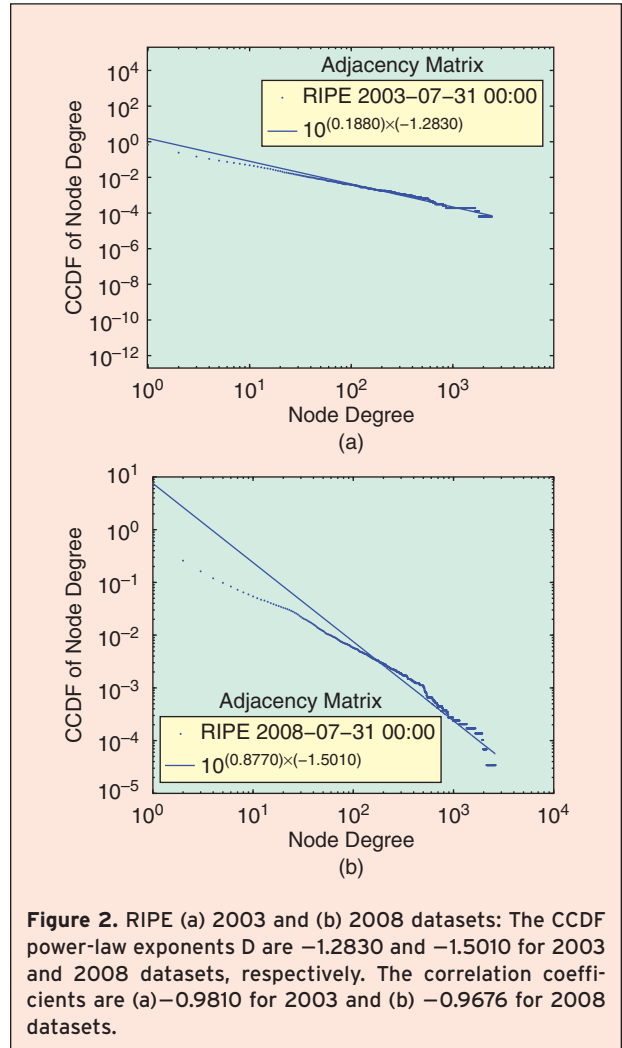
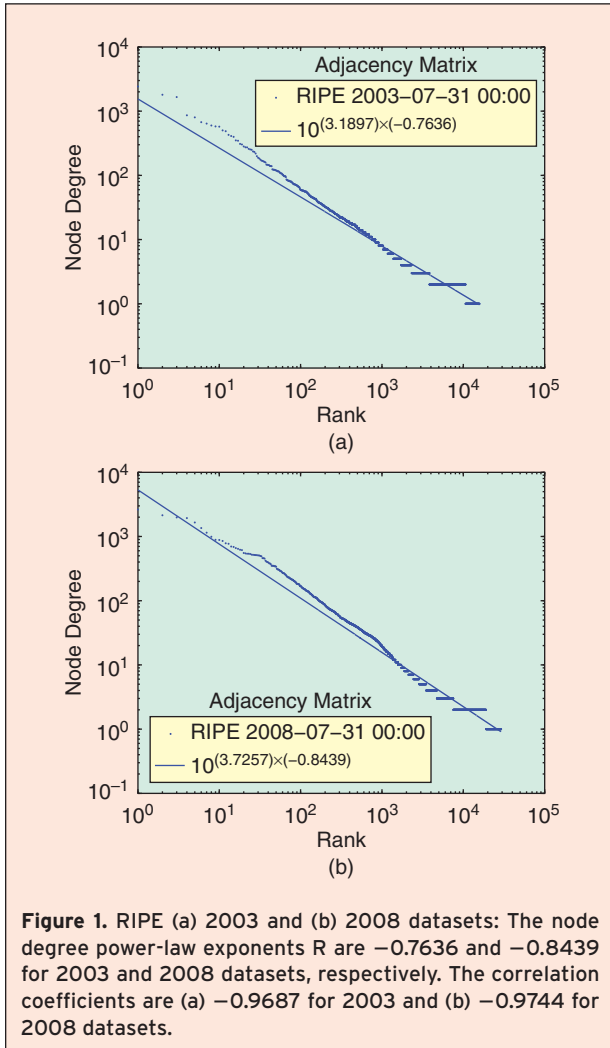
A diagonal matrix $D(G)$ associated with $A(G)$, with row-sums of $A(G)$ as the diagonal elements, indicates the connectivity degree of each node. The Laplacian matrix is defined as $L(G) = D(G) - A(G)$. It is also known as Kirchhoff matrix and a *matrix of admittance*. The normalized Laplacian matrix $NL(G)$ 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. The spectrum of $NL(G)$ is the collection of all its eigenvalues and contains 0 for every connected graph component.

By analyzing plots of node degree vs. node rank, node degree frequency vs. degree, and eigenvalues vs. the order index, number of nodes within a number of hops vs. number of hops, various power-laws have been associated with Internet graph properties [5], [8], [14], [12], [15]. Linear regression of the analyzed data determines the correlation coefficient between the regression line and the plotted data. A high correlation coefficient indicates the existence of a power-law. 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.

Eigenvalues of matrices associated with Internet topology graphs also exhibit power-law properties. The eigenvalues λ_{ai} of the adjacency matrix and λ_{Li} of the normalized Laplacian matrix are sorted in decreasing order and plotted vs. i , where i represents the order of the eigenvalue. Power-laws for the adjacency matrix and the normalized



Laplacian matrix imply that $\lambda_{ai} \propto i^\varepsilon$ and $\lambda_{Li} \propto i^L$, respectively, where ε and L are their respective eigenvalue power-law exponents.

IV. Power-Laws and the Internet Topology

Most existing Autonomous System (AS) numbers 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. We only consider assigned and designated AS numbers, which range from 0 to 65,535 [1]. The assigned AS numbers are listed in Table I. In 2003, 33,983 AS numbers were assigned by IANA. This number increased to 49,149 in 2008. The remaining AS numbers between 49,000 to 64,000 are mostly left unassigned.

Various graph properties may be observed from the Route Views and RIPE datasets collected over the period of five years, from 2003 to 2008. The Route Views BGP routing tables were collected from multiple geo-

graphically distributed BGP Cisco routers and Zebra servers. Most participating ASs were in North America. In contrast to the centralized way of collecting routing data in Route Views, RIPE applied a distributed approach to the data collection and most participating ASs resided in Europe. The RIPE project Routing Information Service (RIS) collected and stored default-free BGP routing data using Remote Route Collectors (RRCs) at various Internet exchanges deployed in Europe, North America, and Asia. These RRCs peered with local operators to collect the entire routing tables every eight hours. The collected raw data was then transferred via an incremental file transfer utility to a central storage area at the RIPE center in Amsterdam. We used RIPE datasets collected from sixteen distinct locations. Analyzed datasets were collected at 00:00 am on July 31, 2003 and at 00:00 am on July 31, 2008 [2].

In recent studies of the Route Views and RIPE datasets [12], [15] the presence of power-laws was observed when various properties of AS nodes such as node degree

In spite of the Internet growth, increasing number of users, and the deployment of new network elements, power-law exponents have not changed substantially.

and frequency of node degree were analyzed. 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 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. Power-laws appeared in the plots of eigenvalues of the adjacency matrix and the normalized Laplacian matrix vs. the order of the eigenvalues. For example, the dependencies between the graph eigenvalues and the eigenvalue index shown in Fig. 3 and Fig. 4 are similar to the reported graphs of the Route Views datasets [12]. Plotted on a log-log scale are eigenvalues in decreasing order. Only the 150 largest eigenvalues are plotted. 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.

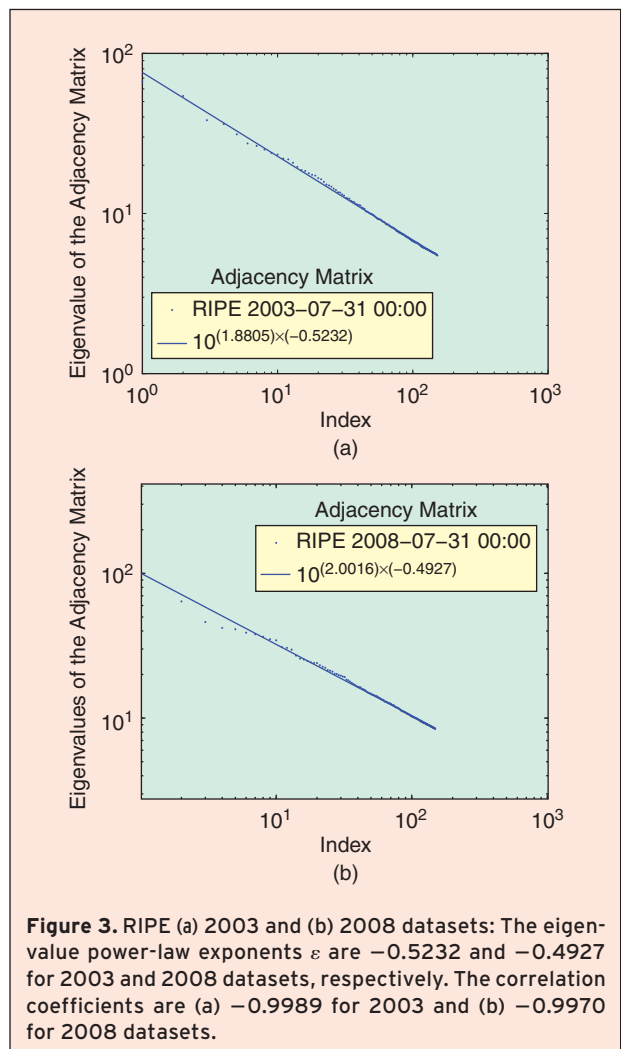
V. Spectral Analysis of the Internet Topology

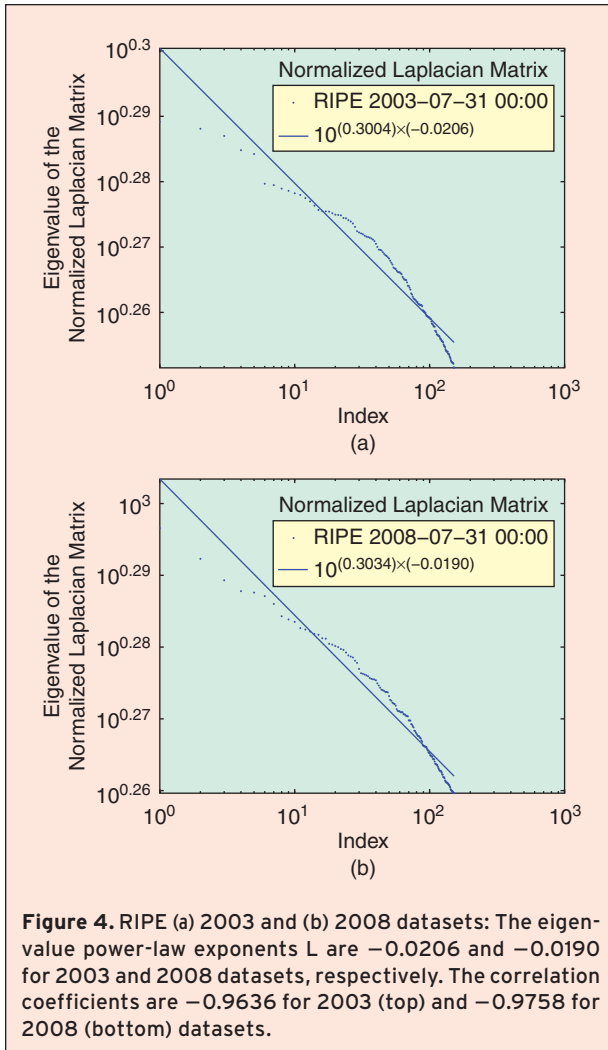
The second smallest eigenvalue of a normalized Laplacian matrix is related to the connectivity characteristic of the graph. Connectivity measures the robustness of a graph and can be designated as vertex or edge connectivity. Vertex (edges) connectivity of a graph is the minimal number of vertices (edges) whose removal would result in a disconnected graph. The second smallest eigenvalue of a graph reflects the vertex and edge connectivities and it is called *the algebraic connectivity* of a graph [9]. Its value is zero if and only if the graph is not connected. It has also been observed that 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 nodes with similar connectivity patterns constituting clusters [11].

Table 1.
Autonomous System (AS) numbers.

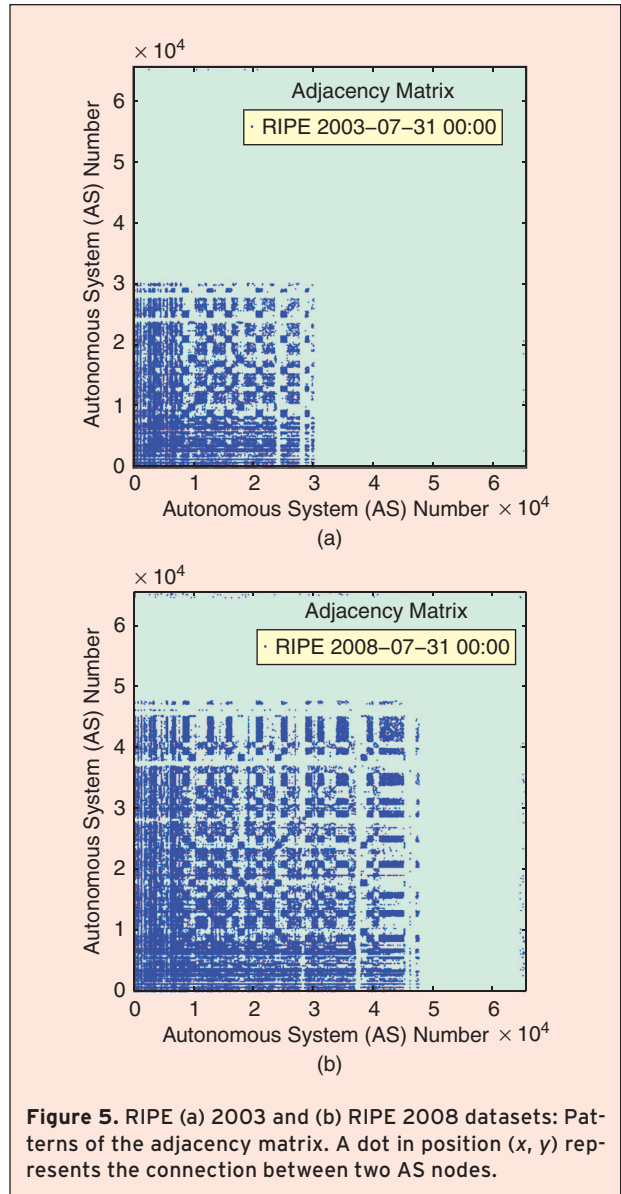
Date	2003-07-31	2008-07-31
Assigned AS numbers	1-30979 (30979)	1-30979 (30979)
(number of AS nodes)	31810-33791 (1981)	30980-48127 (17147)
	64512-65534 (1022)	64512-65534 (1022)
	65535 (1)	65535 (1)





We examined the second smallest and the largest eigenvalues and their associated eigenvectors of both the adjacency and the normalized Laplacian matrix for Route Views and RIPE datasets. Each element of an eigenvector was first associated with the AS having the same index. Each AS was then sorted in the ascending order based on that eigenvector values and the sorted AS vector was then indexed. The connectivity status was defined to be equal to 1 if the AS was connected to another AS or zero if the AS was isolated or was absent from the routing table. The connectivity graphs for Route Views and RIPE datasets indicated visible changes in the clustering of AS nodes and the AS connectivity over the period of five years [12], [15].

It is interesting to observe that the connectivity status based on the second smallest eigenvalue and the largest eigenvalue of the adjacency matrix and based on the normalized Laplacian matrix are asymmetric. The connectivity graph based on the second smallest



eigenvalue of the normalized Laplacian matrix is similar to the connectivity graph based on the largest eigenvalue of the adjacency matrix, and vice versa.

In order to observe clusters of connected AS nodes in the RIPE 2003 and RIPE 2008 datasets, we plotted patterns of the adjacency matrix shown in Fig. 5. (No connectivity is shown between the unassigned AS nodes.) The Route Views 2003 and RIPE 2003 datasets showed similarity in clustering patterns. The same observation held for the Route Views 2008 and RIPE 2008 datasets.

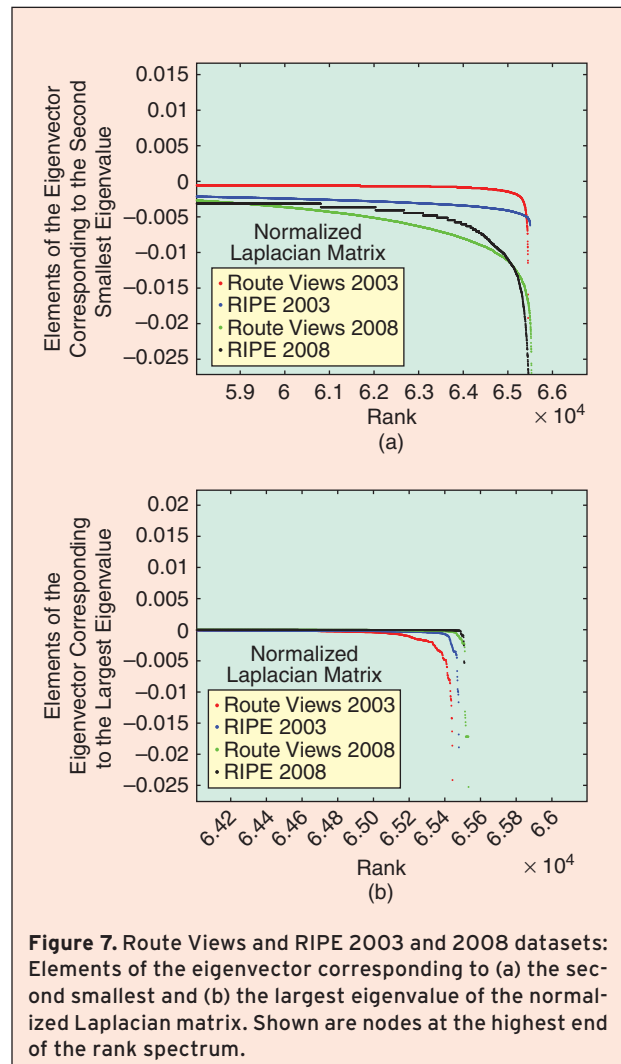
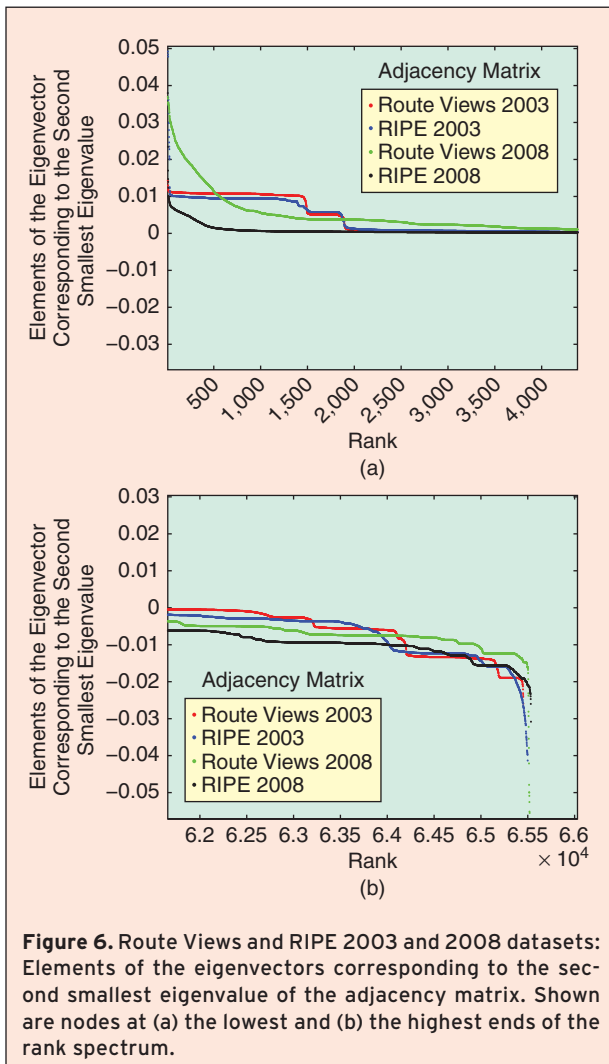
In a simple example of a small world network with 20 nodes [17], elements of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix indicated clusters of connected nodes. Values of the elements

Values of the elements of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix divided nodes into clusters depending on the node degrees.

of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix divided nodes into clusters depending on the node degrees.

In search of identifying clusters of nodes in Internet topology, we examined values of the elements of eigenvectors corresponding to the second smallest and the largest eigenvalue of the adjacency matrix and the normalized Laplacian matrix. Sample plots shown in Fig. 6 indicate that values of the elements of eigenvectors separate graph nodes into clusters. Only those nodes on the lowest and the highest ends of the rank spectrum are shown. The majority of the nodes that are ranked

in-between belong to a cluster having almost identical values of the eigenvector. Note that the adjacency matrix provided clustering information similar to the normalized Laplacian matrix. However, the normalized Laplacian matrix revealed additional details regarding spectral properties of graphs. Sample plots of nodes at the higher end of the rank spectrum are shown in Fig. 7 (top) for the second smallest eigenvalue and in Fig. 7 (bottom) for the largest eigenvalue of the normalized Laplacian matrix. They indicate a more prominent separation between the datasets when compared to plots that correspond to the adjacency matrix.



VI. Conclusions

Analysis of the Route Views and RIPE datasets have confirmed the presence of similar power-laws in graphs capturing the AS-level Internet topology in both datasets. Spectral analysis based on both the normalized Laplacian matrix and the more intuitive adjacency matrix emanating from these graphs was used to examine connectivity of Internet graphs. Clusters of AS nodes can be identified based on the eigenvectors corresponding to the second smallest and the largest eigenvalue of these matrices. While many properties of Internet topology graphs have not substantially changed over the years, spectral analysis revealed notable changes in the connectivity and clustering of AS nodes.

Acknowledgments

The author would like to thank Laxmi Subedi for generating MATLAB figures used in this paper. This research was supported by the NSERC Discovery Grant 216844-07, the NSERC RTI Grant 330480-06, and the BCNET.



Ljiljana Trajković received the Dipl. Ing. degree from University of Pristina, Yugoslavia, in 1974, the M.Sc. degrees in electrical engineering and computer engineering from Syracuse University, Syracuse, NY, in 1979 and 1981, respectively, and the Ph.D. degree in electrical engineering from University of California at Los Angeles, in 1986.

She is currently a Professor in the School of Engineering Science at Simon Fraser University, Burnaby, British Columbia, Canada. From 1995 to 1997, she was a National Science Foundation (NSF) Visiting Professor in the Electrical Engineering and Computer Sciences Department, University of California, Berkeley. She was a Research Scientist at Bell Communications Research, Morristown, NJ, from 1990 to 1997, and a Member of the Technical Staff at AT&T Bell Laboratories, Murray Hill, NJ, from 1988 to 1990. Her research interests include high-performance communication networks, control of communication systems, computer-aided circuit analysis and design, and theory of nonlinear circuits and dynamical systems.

Dr. Trajković served as 2007 President of the IEEE Circuits and Systems Society. She was a member of the Board of Governors of the IEEE Circuits and Systems Society (2001–2003 and 2004–2005). She serves as Vice President Publications of the IEEE Systems, Man, and Cybernetics Society (2010–2011) and served as Vice

President Long-Range Planning and Finance (2008–2009) and as a Member at Large of its Board of Governors (2004–2006). She is Chair of the IEEE Circuits and Systems Society joint Chapter of the Vancouver/Victoria Sections. She was Chair of the IEEE Technical Committee on Nonlinear Circuits and Systems (1998). She was Technical Program Co-Chair of ISCAS 2005 and served as Technical Program Chair and Vice General Co-Chair of ISCAS 2004. She served as an Associate Editor of the *IEEE Transactions on Circuits and Systems (Part I)* (2004–2005 and 1993–1995), the *IEEE Transactions on Circuits and Systems (Part II)* (1999–2001 and 2002–2003), and the *IEEE Circuits and Systems Magazine* (2001–2003). She is a Fellow of the IEEE.

References

- [1] (2010, February 10). *Autonomous system numbers* [Online]. Available: <http://www.iana.org/assignments/as-numbers>.
- [2] (2010, February 10). *BGP capture datasets* [Online]. Available: <http://archive.routeviews.org>.
- [3] (2010, February 10). *Réseaux IP Européens* [Online]. Available: <http://www.ripe.net/ris>.
- [4] 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.
- [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] J. Chen and Lj. Trajković, "Analysis of Internet topology data," in *Proc. IEEE Int. Symp. Circuits Syst.*, Vancouver, BC, Canada, May 2004, vol. IV, pp. 629–632.
- [7] F. R. K. Chung, *Spectral Graph Theory*. Providence, Rhode Island: Conference Board of the Mathematical Sciences, 1997.
- [8] 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.
- [9] M. Fiedler, "Algebraic connectivity of graphs," *Czech. Math. J.*, vol. 23, no. 2, pp. 298–305, 1973.
- [10] L. Gao, "On inferring autonomous system relationships in the Internet," *IEEE/ACM Trans. Networking*, vol. 9, no. 6, pp. 733–745, Dec. 2001.
- [11] 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.
- [12] M. Najiminaini, L. Subedi, and L. Trajković, "Analysis of Internet topologies: A historical view," *Proc. IEEE Int. Symp. Circuits and Systems*, Taipei, Taiwan, May 2009, pp. 1697–1700.
- [13] A. Pothén, 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.
- [14] 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.
- [15] L. Subedi and Lj. Trajković, "Spectral analysis of Internet topology graphs," *Proc. IEEE Int. Symp. Circuits and Systems*, Paris, France, June 2010, pp. 1803–1806.
- [16] 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.
- [17] D. J. Watts and S. H. Strogatz, "Collective dynamics of small world networks," *Nature*, vol. 393, pp. 440–442, June 1998.