Spectral Analysis of Internet Topology Graphs

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

***Abstract*— The discovery of power-laws and spectral properties of the Internet topology indicates a complex underlying network infrastructure. Analysis of spectral properties of the Internet topology has been usually based on the normalized Laplacian matrix of graphs capturing Internet structure on the Autonomous System (AS) level. In this paper, we first extend the previous analysis of the Route Views data to include datasets collected from the RIPE project. Spectral analysis of collected data from the RIPE datasets confirms the previously observed existence of power-laws and similar historical trends in the development of the Internet. Presented spectral analysis of both the adjacency matrix and the normalized Laplacian matrix of the associated graphs also reveals new historical trends in the clustering of AS nodes and their connectivity. The connectivity and clustering properties of the Internet topology are further analyzed by examining element values of the corresponding eigenvectors.**

I. Introduction

Analyzing Internet topologies relies on capturing data information about Internet Autonomous Systems [1] and exploring properties of associated graphs on the AS-level. The Route Views data [2] and RIPE [3] datasets collected from Border Gateway Protocols (BGP) routing tables have been extensively used by the research community [4]–[6]. In this paper, we extend our previous analysis of the Route Views data [8] to include RIPE [3] datasets. These datasets collected from BGP routing tables indicate that Internet topology is characterized by the presence of various power-laws [4], [5]. It has also 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 [7], [8].

It is well known that eigenvalues associated with a network graph are closely related to its topological characteristics. 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 [8]. 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 [7]. In our earlier studies [8], [9] we employed spectral analysis to analyze the Route Views and RIPE datasets in order to find distinct clustering features of the Internet AS nodes. In this paper, we examine eigenvectors of both the adjacency matrix and the normalized Laplacian matrix and illustrate that both matrices may be used to identify clusters of connected AS nodes.

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

The paper is organized as follows: We provide a brief introduction to spectral analysis and power-laws in Section II. Analysis of Internet topology based on the RIPE dataset is presented in Sections III. Connectivity and clusters in Internet topology graphs are examined in Section IV. We conclude with Section V.

II. 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, various power-laws have been associated with Internet graph properties [4]–[6], [8]. 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.

III. Power-Laws and the Internet Topology: RIPE datasets

Most existing Autonomous System (AS) numbers are assigned by regional Internet Assigned Numbers Authority (IANA) registries. 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.

TABLE I
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 various graph properties from Route Views and RIPE datasets collected in 2003 and 2008. The Route Views BGP routing tables were collected from multiple geographically 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 a recent study of the Route Views datasets [8], we observed the presence of power-laws when various properties of AS nodes such as node degree and frequency of node degree were analyzed. Power-laws also appeared in the plots of eigenvalues of the adjacency matrix and the normalized Laplacian matrix vs. the order of the eigenvalues. Extended analysis that included the RIPE datasets also revealed similar properties. For example, the dependencies between the graph eigenvalues and the eigenvalue index shown in Fig. 1 and Fig. 2 are similar to the reported graphs of the Route Views datasets [8]. Plotted on a log-log scale are eigenvalues in decreasing order. Only the 150 largest eigenvalues are plotted.

Confidence intervals for the power-law exponents were calculated using six samples randomly selected from each analyzed dataset. Each dataset is smaller than 30, with unknown standard deviation. Hence, we used the t-distribution to predict the confidence intervals at the 95% confidence level:

$$\bar{X} - t_{x/2}(s/\sqrt{n}) < \mu < \bar{X} + t_{x/2}(s/\sqrt{n}),$$

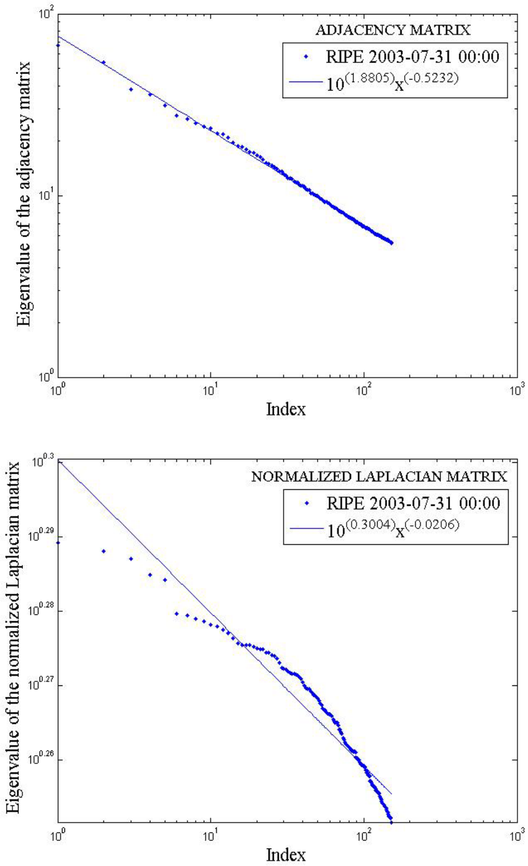

Fig. 1. RIPE 2003 dataset: The eigenvalue power-law exponents ε for the adjacency matrix (top) is –0.5232 with the correlation coefficients –0.9989. The eigenvalue power-law exponents L for the normalized Laplacian matrix (bottom) is –0.0206 with the correlation coefficients –0.9636.

where $\bar{X}$ is the sample mean, n is the number of samples, $t_{x/2}$ is the t-distribution, s is the sample standard deviation, and μ is the population mean. The predicted confidence intervals for the RIPE 2008 datasets are shown in Table II.

TABLE II
Confidence Intervals of Eigenvalue Power-Law Exponents

Dataset	Matrix	Lower level	Higher level
RIPE 2003	Adjacency	-0.5105	-0.5033
	Normalized Laplacian	-0.0206	-0.0188
RIPE 2008	Adjacency	-0.4985	-0.4914
	Normalized Laplacian	-0.0204	-0.0188

IV. Spectral Analysis of the Internet Topology: Route Views and RIPE datasets

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*

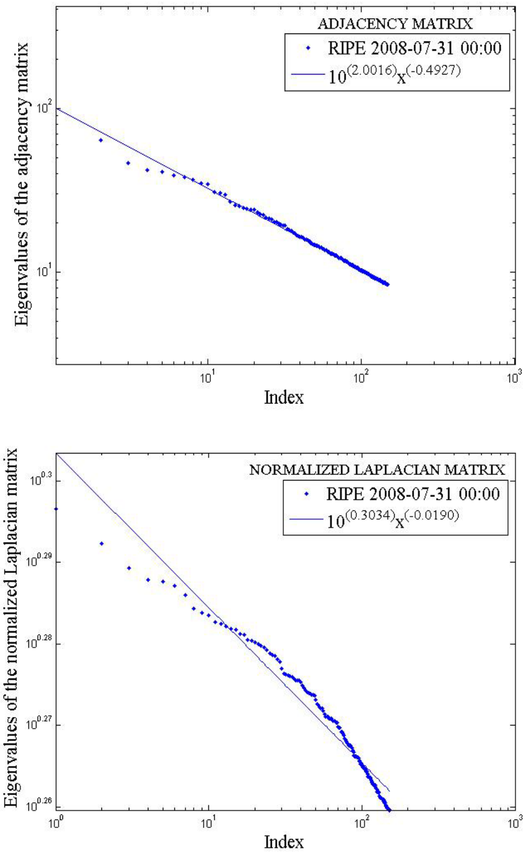

Fig. 2. RIPE 2008 dataset: The eigenvalue power-law exponents ε for the adjacency matrix (top) is –0.4927 with the correlation coefficients –0.9970. The eigenvalue power-law exponents L for the normalized Laplacian matrix (bottom) is –0.0190 with the correlation coefficients –0.9758.

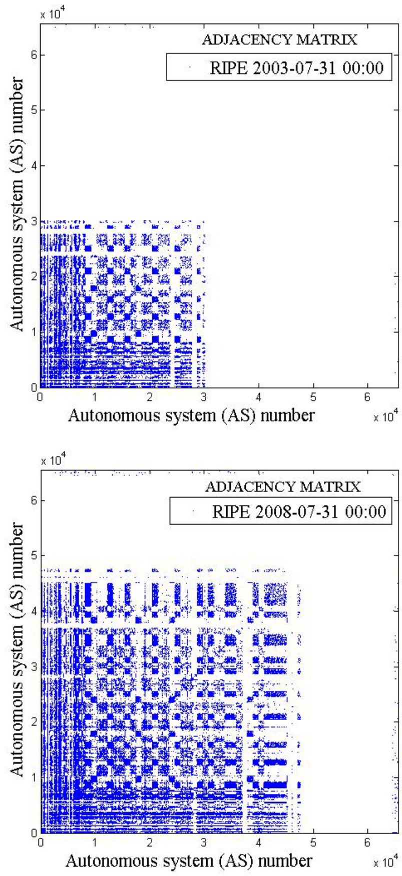

Fig. 3. RIPE 2003 and RIPE 2008 datasets: Patterns of the adjacency matrix. A dot in position (x, y) represents the connection between two AS nodes.

connectivity of a graph [10]. 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 [7].

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. On both the Route Views datasets [8], 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 [8].

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. 3. (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 [11], elements of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix indicated clusters of connected nodes. 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.

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. 4 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

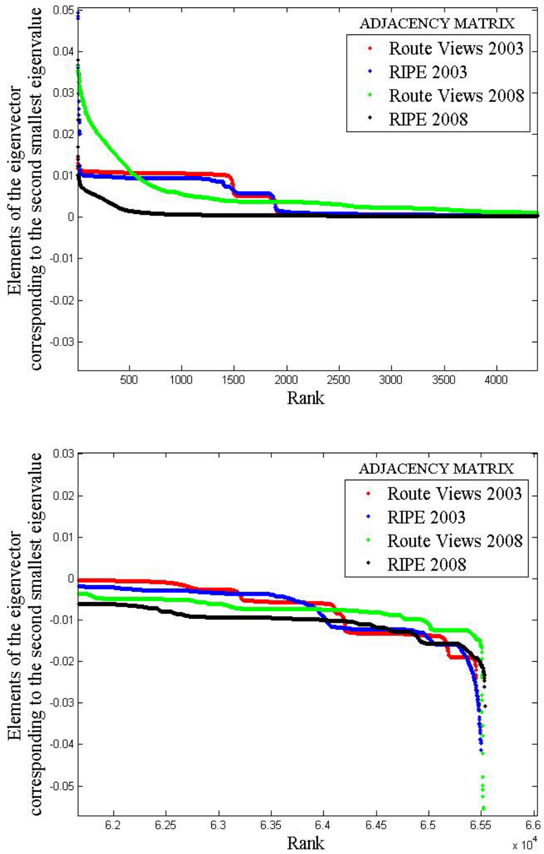

Fig. 4. Route Views and RIPE 2003 and 2008 datasets: Elements of the eigenvectors corresponding to the second smallest eigenvalue of the adjacency matrix. Shown are nodes at the lowest (top) and the highest (bottom) ends of the rank spectrum.

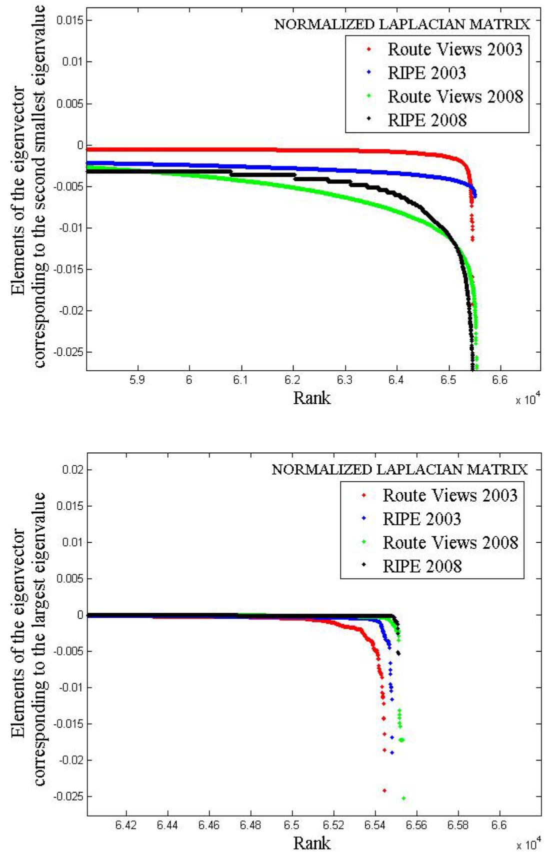

Fig. 5. Route Views and RIPE 2003 and 2008 datasets: Elements of the eigenvector corresponding to the second smallest (top) and the largest (bottom) eigenvalue of the normalized Laplacian matrix. Shown are nodes at the highest end of the rank spectrum.

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. 5 (top) for the second smallest eigenvalue and in Fig. 5 (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.

V. Conclusions

We have analyzed datasets from the Route Views and RIPE projects and 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. We identified clusters of AS nodes 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.

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). Rseaux IP Europens [Online]. Available: http://www.ripe.net/ris.

[4] 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.

[5] 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.

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

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

[8] M. Najiminaini, L. Subedi, and Lj. Trajkovic, "Analysis of Internet topologies: a historical view," *IEEE Int. Symp. Circuits Syst.*, Taipei, Taiwan, May 2009, pp. 1697–1700.

[9] 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.

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

[11] D. J. Watts and S. H. Strogatz, "Collective dynamics of small world networks," *Nature*, vol. 393, pp. 441-442, June 1998.