POWER-LAWS AND SPECTRAL ANALYSIS OF THE INTERNET TOPOLOGY

by

Laxmi Subedi
M.E.M, St. Cloud State University, 2005
B.E, Tribhuvan University, 2003

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

In the
School of Engineering Science

SIMON FRASER UNIVERSITY

Spring 2010

APPROVAL

Name: **Laxmi Subedi**

Degree: **Master of Applied Science**

Title of Thesis: **Power-laws and spectral analysis of the Internet topology**

Examining Committee:

Chair: **Dr. Mirza Faisal Beg**
Associate Professor of School of Engineering Science

Dr. Ljiljana Trajković
Senior Supervisor
Professor of School of Engineering Science

Dr. Stephen Hardy
Supervisor
Professor of School of Engineering Science

Dr. Qianping Gu
Internal Examiner
Professor of School of Computing Science

Date Defended/Approved:

ABSTRACT

In this thesis, we present the analysis on power-laws and spectral properties of the Internet topology at AS level based on Border Gateway Protocol (BGP) routing datasets collected from two repositories (Route Views and RIPE) over the period of five years.

Analysis of collected datasets revealed that the two datasets have similar historical trends in the development of the Internet. Furthermore, the power-law exponents have not substantially changed over time while spectral analysis revealed notable changes in the clustering of AS nodes and their connectivity. This finding indicates that power-laws do not capture all properties of the Internet graph and are only a measure used to characterize the Internet topology.

The spectral analysis of both the adjacency and the normalized Laplacian matrices of the associated graphs revealed new historical trends in the clustering of AS nodes and their connectivity. Clusters of connected nodes were observed while examining the elements values of the eigenvectors corresponding to the second smallest and the largest eigenvalues.

Keywords: Traffic collection; Internet topology; Autonomous System (AS); Broader Gateway Protocol (BGP); Power-law; Spectral analysis.

ACKNOWLEDGEMENTS

I would like to express gratitude to my senior supervisor, Prof. Ljiljana Trajković, for her constant guidance, suggestions, and direction for this work. Her continued help throughout my graduate studies is invaluable.

My sincere thanks to Prof. Steve Hardy, Prof. Qianping Gu, and Prof. Mirza Faisal Beg for being my committee members and providing valuable comments and suggestions. Many thanks to Prof. Kohshi Okumura for his valuable feedback.

Thanks are due to my colleagues Mohammad Najiminaini, Shaun Nguyen, and Eman Elghoneimy for their help. I would like to thank Reza Qarehbaghi and Sukhchandan Lally for reviewing my thesis and providing valuable suggestions. Many thanks also to the members of Communication Networks Laboratory for their collegiality.

Lastly, thanks go to my husband Sanjay Regmi for his help and motivation and to my son Pranav for being quiet and letting me work on my research in the evenings.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

APNIC	Asia Pacific Network Information Centre
ARIN	American Registry for Internet Numbers
AS	Autonomous System
ASN	Autonomous System Number
BGP	Border Gateway Protocol
BRITE	Boston University Representative Internet Topology Generator
CAIDA	Cooperative Association for Internet Data Analysis
CCDF	Complementary Cumulative Distribution Function
CIDR	Classless Inter-Domain Routing
IANA	Internet Assigned Numbers Authority
IP	Internet Protocol
IRR	Internet Routing Registry
ISP	Internet Service Provider
LANIC	Latin America and Caribbean Internet Address Registry
LIR	Local Internet Registry
MRT	Multi-threaded Routing Toolkit
NLANR	National Laboratory for Applied Network Research
NIR	National Internet Registry
RIPE	Réseaux IP Européens

RIR	Regional Internet Registry
RIS	Routing Information Service
RRC	Remote Route Collector
TCP/IP	Transmission Control Protocol/Internet Protocol

1. INTRODUCTION

1.1 Overview

The Internet network has grown from a few hundred nodes of Autonomous System (AS) to many thousands of AS nodes with millions of users without centralized control. Despite the increase in size and complexity, the performance of the Internet network is remarkable. The Internet network structure, called the Internet topology, has a great impact on the performance of network protocols and applications. Thus, understanding the properties of the Internet topology is important for the development of new protocols, algorithms, and new network infrastructure. These properties are also useful to realistically model the Internet topology for evaluation of various protocols and algorithms and for testing purposes.

Various properties of the Internet topology such as power-laws and clustering nature of ASes have been identified. These properties of the Internet topology have been analyzed by observing the graphs capturing the Internet structure at AS level [18], [23], [29], [33], [48], [52]. Thus, analysis of the Internet topology graphs relies on mining data and capturing information about Autonomous Systems [1].

The routing table of Border Gateway Protocol (BGP) routers contains reachability information of ASes from which the Internet topology at AS level can be inferred. BGP routers use BGP routing protocol, which is an inter-autonomous system routing protocol that exchanges routing reachability information between peers. Every BGP router connected to ASes receives reachability information from its neighbours and this information consists of the list of ASes between the source AS and the destination AS. Such AS information from BGP routing tables is used to infer the logical connection between ASes and to infer the Internet topology at AS level.

Route Views and Réseaux IP Européens (RIPE) projects collect the BGP routing table and provide the research community an access to their database for analysis. Route Views began the collection of BGP routing datasets from participating ASes located in North America in November 1997. Meanwhile, RIPE started the collection process from participating ASes located mostly in Europe in October 1999. Other tools such as *traceroute* provide the route including intermediate routers information over the network between two systems while the Looking Glass project provides servers that give public a remote access to view routing information of the user network. However, these datasets do not provide historical views of the datasets as available from Route Views and RIPE datasets. The research community [18], [23], [29], [33], [48], [52] has extensively used Route Views and RIPE datasets to observe the Internet topological characteristics. We, thus, adopt the datasets from Route Views and RIPE projects for the analysis of power-laws and spectral properties of the Internet topology.

The analysis of power-laws and spectral properties of the Internet topology has been based on the connectivity matrix of a graph called the adjacency matrix. This matrix has element value 1 in the position (i, j) if there is an edge between node i and j. Otherwise, the element value is zero. The eigenvalues, called spectrum, and eigenvectors of such matrix are related to many graph properties such as connected components and the diameter of the network. Thus, the spectral analysis is associated with the analysis of the eigenvalues and the eigenvectors of the matrices.

Power-law is a polynomial expressed in the form of $y \propto x^a$, where y and x are the measures of interest and exponent a is a constant. The existence of various power-laws such as node degree vs. node rank, frequency of node degree vs. node degree, number of nodes within a number of hops vs. number of hops, and eigenvalue of the adjacency matrix vs. its index were observed in 1999 [29]. These power-laws were observed while analyzing the adjacency matrix of AS level Internet graph derived from Route Views datasets. The subsequent revisions regarding the existence of power-laws were also performed [48]. However, the datasets from Route Views reveal heavy tailed node degree distribution that is close to Weibull distribution (a continuous probability distribution). The power-laws are present only in the tail of the distribution. Thus, additional datasets are needed to capture the Internet topology in order to analyze its properties [23].

Another method to study the Internet topology is to employ the spectral analysis, which provides information about structural properties of graphs such as clustering and connectivity of graph nodes. The eigenvalues and the associated eigenvectors of a graph have been used to find the clusters and connectivity of AS nodes in the Internet topology [22], [34]. The eigenvectors corresponding to small eigenvalues tend to capture the local characteristics such as connectivity of nodes. Similarly, the eigenvectors corresponding to the large eigenvalues capture the global characteristics of the graph such as clusters of connected nodes based on geographic regions [34]. The clusters of ASes based on the largest eigenvalue are consistent over time and are considered to be a robust characteristic to represent the Internet topology [34].

Based on the power-laws property of the Internet topology, a number of topology generator tools such as the Boston University Representative Internet Topology Generator (BRITE) [44] were developed. However, a combination of spectral properties and power–laws properties may be needed while developing the Internet topology generators [52].

In this thesis, we present historical trends in the development of the Internet topology by analyzing BGP routing datasets collected from two repositories (Route Views and RIPE) over a period of five years. We analyze the evolution of following power-laws: 1) node degree vs. rank; 2) CCDF of node degree vs. node degree; and 3) eigenvalue vs. index. We also observe the spectral properties based on both the adjacency matrix and the normalized Laplacian matrix in order to find the clustering and connectivity characteristics of AS nodes. Finally, we analyze the connectivity and clustering properties of the Internet topology by examining element values of the eigenvectors corresponding to the second smallest and the largest eigenvalues.

1.2 Motivation

Since the discovery of power-laws in 1999 [29], the Internet topology has increased in size and complexity. The number of Autonomous Systems has increased approximately ten times over the last ten years [3], as shown in Figure 1.1. The constant growth of the Internet has made it difficult to develop its representative model. However, certain characteristics of the Internet topology remain unchanged in spite of its exponential growth. We analyze the Internet topology in search of such invariants at AS level, analyze the existence of power-laws, and perform spectral analysis based on the adjacency and the normalized Laplacian matrices.

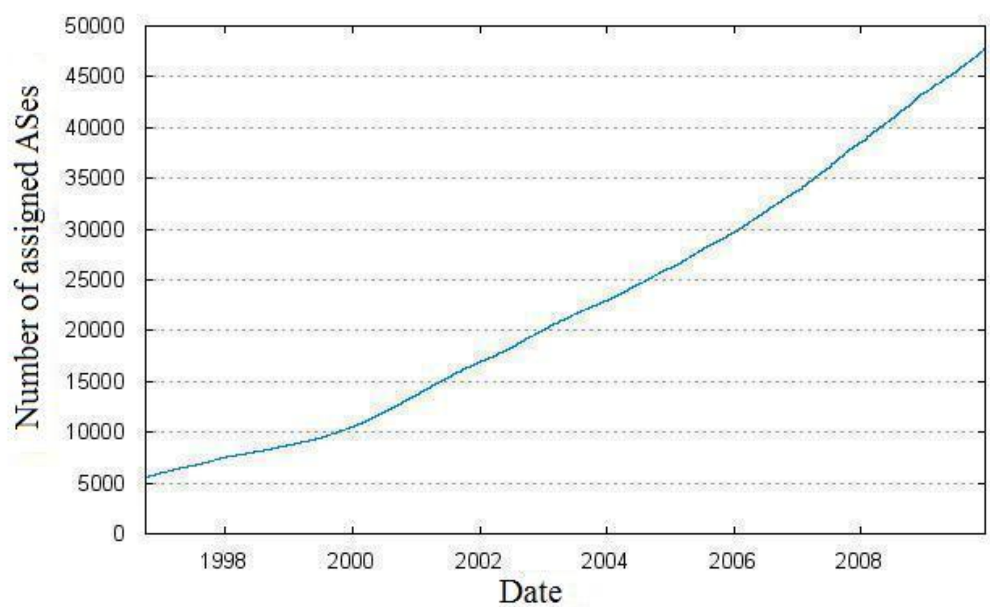


Figure 1.1 The cumulative number of assigned AS numbers over time [3].

In a simple example of a small world network with 20 nodes [55], we observed that the elements of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix indicate clusters of connected nodes. Values of these elements divide nodes into clusters depending on the degree of the nodes. No such clustering was observed based on the elements of the eigenvector corresponding to the second smallest eigenvalue. Section 5.3 presents detailed description of these observations.

In search of such clustering properties of ASes in the Internet topology over the time, we perform spectral analysis based on both the adjacency matrix and the normalized Laplacian matrix. We observe patterns of connected AS nodes in the Internet topology based on the adjacency matrix. We also examine the elements of the eigenvectors corresponding to the second smallest and the largest eigenvalues of both matrices.

1.3 Findings

The analysis of Route Views and RIPE datasets shows similar trends in the development of the Internet topology. Despite the growth of the Internet and increasing number of users, the exponent values of various power-laws such as node degree vs. rank, CCDF of node degree vs. node degree, and the eigenvalue vs. index have not substantially changed over the period of five years. We also observe a new property that the eigenvalues based on the normalized Laplacian matrix also exhibit eigenvalue vs. index power-law property similar to eigenvalue power-law based on the adjacency matrix. They have, as expected, different values for power-law exponents.

By plotting the elements of the adjacency matrix, we observe various patterns of connected AS nodes over the years. While power-laws properties of the Internet topology graphs have not substantially changed over the years, the spectral analysis of the adjacency and the normalized Laplacian matrices of the associated graphs reveals notable changes and new historical trends in the clustering of AS nodes and their connectivity.

1.4 Organization of thesis

The organization of the thesis is as follows. The Internet structure and routing, BGP routing protocols, Autonomous System, and description of Route Views and RIPE datasets are presented in Chapter 2. Chapter 3 presents the definition of various power-laws and a short introduction to spectral analysis. Related work on the analysis of the Internet topology is also included. Chapter 4 contains the analysis of power-laws using Route Views and RIPE datasets. Chapter 5 describes spectral analysis of the Internet topology based on Route Views and RIPE datasets. We conclude with Chapter 6. MATLAB sample code is presented in the Appendix.

2. INTERNET ROUTING AND DESCRIPTION OF BGP DATASETS

In this Chapter, we introduce the Internet structure, its routing procedures, notion of Autonomous System, BGP routing protocol, and give snippets of a routing table. We describe Route Views and RIPE datasets and also provide sample of collected data and an example of the Internet topology graph at AS level.

2.1 Internet structure

The Internet architecture is based on the Transmission Control Protocol/Internet Protocol (TCP/IP) suite. TCP/IP provides a reliable transmission of data between two end hosts in the Internet. The two interconnected nodes can communicate with TCP/IP protocol regardless of their geographical location and this very feature enables the Internet architecture to grow to a global scale.

An individual node accesses the Internet through a local Internet service provider (ISP). The local ISPs are connected to a regional network and the regional networks in turn are connected to a national network thereby culminating into the global network.

Every host on the Internet has a unique numerical address, called an Internet Protocol (IP) address that is used to route packets across the Internet to and from that host. When a host communicates with another host in the Internet, each packet from the source host contains the source and destination host IP addresses and is sent to the nearest router. The IP addresses of the source and destination are stored in the header of every packet that flows across the Internet.

An Internet router typically connects several different networks. For example, university network consists of several interconnected routers where only few routers are connected to the wide area network (WAN). Each packet goes up the hierarchy of the Internet network to reach its destination network where local routers deliver packet to the destination address. The router uses a routing algorithm to route packet from the source to the destination host.

The router employs the Interior Gateway Protocols (IGP) to route Internet packets within a local area network such as a university network. The two main types of IGP protocols are Routing Information protocol (RIP) [10] and Open Shortest Path First (OSPF) protocol [11]. RIP uses the distant vector algorithm, also called the Bellman-Ford algorithm to calculate the routing paths. OSPF protocol uses the Shortest Path First (SPF) algorithm, also called the Dijkstra's algorithm. OSPF routers keep the topology map of the network and send updates of the routing information to other routers in the network. The convergence time of the SPF algorithm is faster than the distant vector algorithm. In addition to the RIP and OSPF protocols, there are various proprietary network protocols such as Interior Gateway Routing Protocol (IGRP) and Enhanced IGRP developed by Cisco Systems. Enhanced IGRP uses the distance vector algorithm and distance information as IGRP. However, the convergence properties and the operating efficiency of Enhanced IGRP is better than IGRP.

The Exterior Gateway Protocol (EGP) such as BGP-4 is used for routing packets between two different routing domains. BGP-4 is an inter-autonomous system routing protocol used to exchange routing reachability information between ASes. Every BGP-4 router that connects ASes receives reachability information from its neighbours. It then chooses routes with the shortest path, updates its routing table, and announces the path to other neighbouring routers according to the routing policy. The network reachability information consists of the list of every AS between the source and the destination ASes. Thus, AS information from BGP routing tables is used to infer logical connection between ASes and to infer the Internet topology at AS level.

2.2 Autonomous Systems (ASes)

The Internet is composed of a collection of routing domains called Autonomous Systems. The AS is a network or a group of networks with a common routing policy. For example, an AS consists of a university network, a business enterprise, or a corporation network. The network within an AS uses a common IGP to route packets. However, two ASes use BGP to share routing information. BGP information at each AS router is kept consistent by receiving BGP update messages from BGP routers of other ASes. Each AS is identified by a unique number known as Autonomous System Number (ASN) assigned by the Internet Assigned Number Authority (IANA).

IANA allocates the Internet Protocol addresses from the pool of unallocated addresses to the Regional Internet Registries (RIR) according to the global policy. An AS consists of a range of IP addresses and the Internet Service Providers (ISPs) assign these IP addresses to its users. ISPs obtain IP addresses from a Local Internet Registry (LIR), National Internet Registry (NIR), or from their appropriate RIR:

- African Regional Internet Registry (AfriNIC), Africa region
- Asia Pacific Network Information Centre (APNIC), Asia Pacific region
- American Registry for Internet Numbers (ARIN), North America region
- Latin American and Caribbean Internet Address Registry (LACNIC), Latin America and Caribbean Islands region
- Réseaux IP Européens Network Coordination Centre (RIPE NCC), Europe, Middle East, and Central Asia region.

The Internet topology can be analyzed at two different granularities: router level and the Autonomous System level, also called the Inter-domain level. At Autonomous System level, each AS domain is represented as a node. Links between two nodes represent the logical connection between two ASes. Thus, the AS graph represents the connections between ASes. Each AS is represented by an ASN. The AS numbers range from 0 to 65,535. The existing ASes are assigned by the regional IANA registries. IANA designates the remaining AS numbers for private use. Certain AS numbers are reserved and do not appear in the Internet graph at AS level.

The assigned AS numbers are listed in Table 2.1. IANA assigned only 33,984 AS numbers in 2003 and 49,150 AS numbers in 2008. In 2003, most unassigned AS numbers were between 34,000 to 64,000. In 2008, the range of unassigned AS numbers was between 49,000 to 64,000.

Table 2.1 AS numbers assigned by IANA.

Number/Date	2003-07-31	2008-07-31
Assigned AS numbers	0-30979 = 30980	0-30979 = 30980
	31810-33791=1981	30980-48127 =17147
	64512-65534 = 1022	64512-65534 = 1022
	65535 = 1	65535 = 1
Total number of assigned ASes	33984	49150

2.3 Border gateway protocol (BGP)

BGP is a robust and scalable routing protocol widely used in TCP/IP networks to exchange routing reachability information with other BGP systems. BGP maintains routing tables, transmits routing updates, and provides routing decisions based on routing metrics such as link bandwidth, network delay, number of hops, path cost, and load. BGP uses the Classless Inter-Domain Routing (CIDR) in order to reduce the size of the Internet routing table. CIDR allows routers to group routes together in order to minimize the number of routing information carried by the core routers.

BGP uses TCP as its transport protocol. That eliminates explicit implementation of retransmission, acknowledgment, and sequencing mechanisms. Two systems exchange messages to open and to confirm the connection through the TCP connection. BGP routers exchange routing information using four types of messages [9]:

- open
- updates
- notification
- keep-alive.

Open

After a TCP connection is established, each BGP peer sends an *open* message to open an initial connection. This is the first message sent between peers after the TCP connection is established. BGP sends *keep-alive* message between peers to confirm the *open* message. The *open* message has additional header that contains protocol version, sender ASN, hold time, BGP identifier, authentication code, and authentication data.

Updates

The *update* message is used to transfer and update routing information between BGP peers. As the routing table changes, incremental updates are sent to peers using update message. The update information allows routers to construct a consistent view of the network topology that describes the relationships between various ASes. The *update* message has additional header that contains total path attributes length, path attributes, and network layer reachability information.

Notification

The *notification* message is sent to any connected peer in response to errors or special conditions. If a connection between the connected routers encounters an error, *notification* message is sent to announce the error and to close the active connection between the routers. The header of *notification* message contains error code, error sub-code, and data.

Keep-alive

The *keep-alive* message is used by BGP router to determine whether the peers are reachable or not. BGP router sends the *keep-alive* messages periodically between the peers in order to ensure the active connection between them. Thus, these messages help avoid active sessions from expiring. The *keep-alive* messages contain only BGP message header format: marker, length, and type of the message.

2.3.1 Cisco router data

Cisco systems implemented customized version of BGP that runs on the proprietary operating system used by its routers [8]. The routing table information from Cisco routers can be collected in different formats using various tools and commands such as *show ip bgp* command that displays entries in the BGP routing table of Cisco routers. Table 2.2 shows snippets of datasets from a BGP routing table [2]. The first row shows symbols of status codes that appear in the first two character positions of each line of a route and indicates the status of the route. The second row shows symbols of origin codes that appear in the last character of each line of a route and indicate BGP origin attribute. The remaining rows shown in Table 2.2 contain the following information:

Table 2.2 Sample data of a BGP routing table.

Status codes: s suppressed, d damped, h history, * valid, > best, i internal					
Origin codes: i IGP, e EGP, ? incomplete					
Network	**Next hop**	**Metric**	**LocPrf**	**Weight**	**Path**
4.25.52.128/26	203.62.252.21		0	78	1221 16779 1 189 ?
*	203.62.252.21			2	1221 16779 1 189 ?
*	198.32.162.18	51804	0	8	4513 1755 1 189 ?
* >	4.0.0.2	18740	0	9	1 189 ?

The *network* column contains BGP prefix for a route. This column includes prefix length or mask unless the network has a pre-CIDR length of 0, 8, 16, or 24 bits corresponding to a default route (0 bits) or a class A, B, or C. The blank field indicates another route for the same prefix that appeared last.

The *next hop* column contains the address to which traffic for the prefix will be forwarded. It shows BGP next hop attribute. The address 0.0.0.0 indicates that the next hop is directly connected.

The *metric* column contains a non-transitive metric value or cost. The lowest metric value is preferred while selecting a route. It has an upper bound of $2^{32}-1$.

The *LocPrf* column contains a local administrative preference attribute of BGP. A higher value is preferred while selecting a route.

The *weight* column contains the local value that is not exchanged between BGP peers. It shows an administrative preference particular to Cisco routers. The highest value is preferred while selecting a route.

The *path* column contains BGP AS path attribute. It shows the Autonomous Systems through which a route has been exchanged before it was received by the router. If the field is empty, the route was generated by the local Autonomous System. When considered in the BGP route selection process, path having a few numbers of ASNs is preferred.

The BGP routers share routing information with each other during withdrawal or announcement of a route. The syntax of routing information for withdrawn route is |*BGP protocol*| *unix time in second*| *Withdraw or Announce*| *Peer IP*| *Peer AS*| *Prefix*|. Example of routing table data for such withdrawn route is |*BGP4MP* | *1052452930*| *W*| *198.58.5.254*| *3727*| *194.127.245.0/24*|.

The syntax of routing information for route announcement is |*BGP protocol*| *unix time in seconds*| *Withdraw or Announce*| *Peer IP*| *Peer AS*| *Prefix*| *AS_PATH*| *Origin*| *Next_Hop*| *Local_Pref*| *MED*| *Community*| *Atomic AGG*| *AGGREGATOR*|. Example of routing table data for such route announcement is |*TABLE_DUMP*| *1050122432*| *B*| *213.140.32.184*| *12956*| *0.0.0.0/0*| *12956*| *IGP*| *213.140.32.184*| *0*| *0*| | *NAG*| |.

2.3.2 Zebra server data

The Zebra software manages the TCP/IP based routing protocols such as BGP4 [8]. The server that uses the Zebra routing software is called Zebra server. This server has built-in mechanism to collect BGP routing table datasets by using the *dump bgp routes-mrt* command. It dumps complete BGP routing data stream of each peer and various state information in Multi-threaded Routing Toolkits (MRT) format. This format is used to export routing protocol messages, state changes, and routing information base contents. The *route_btoa* tool is used to read MRT data and to extract it in ASCII format. There are two forms of data representation: human-readable and machine-readable. The human-readable form displays a paragraph for each MRT record. The machine-readable form presents the same data separated by "|" and occupies a single line. We use machine-readable format to extract routing information.

The syntax of routing information to withdraw routes is |*BGP protocol*| *unix time in seconds*| *Withdraw or Announce*| *Peer IP*| *Peer AS*| *Prefix*|. Example of routing table data for such withdrawn route is |*BGP4MP*| *1052452930*| *W*| *198.58.5.254*| *3727*| *194.127.245.0/*|.

The syntax of routing information to announce routes is |*BGP protocol*| *unix time in seconds*| *Withdraw or Announce*| *Peer IP*| *Peer AS*| *Prefix*| *AS_PATH*| *Origin*| *Next_Hop*| *Local_Pref*| *MED*| *Community*| *Atomic AGG*| *AGGREGATOR*|. Example of routing table data for such announced route is |*BGP4MP*| *1052452919*| *A*| *198.58.5.254*| *3727*| *195.28.224.0/19*| *3727 2914*

6730 8640| *IGP*| *198.58.5.254*| *0*| *0*| *2914:420 2914:2000 2914:3000 3727:380*| *AG*| *195.141.213.58*|.

2.4 Description of Route Views and RIPE datasets

Route Views project collects BGP routing tables from multiple geographically distributed BGP Cisco routers and Zebra servers every two hours. Two Cisco routers and two Zebra servers are located at University of Oregon, USA. The remaining five Zebra servers are located at Equinix-USA, ISC-USA, KIXP-Kenya, LINX-Great Britain, and DIXIE-Japan [2]. Most participating ASes in this project are located in North America.

The Routing Information Service (RIS) is a Réseaux IP Européens Network Coordination Centre (RIPE NCC) [12] project that collects and stores the Internet routing data using Remote Route Collectors (RRCs) at various Internet Exchanges. An RRC is a daemon running to collect default-free BGP routing information. These RRCs are peered with local operators to collect routing datasets. Several RRCs have been deployed in Europe, North America, and Asia. However, most of the participating ASes are located in Europe. The datasets are collected from seventeen different locations: RIPE NCC-Amsterdam, LINX-London, SFINX-Paris, AMS-IX-Amsterdam, CIXP-Geneva, VIX-Vienna, Otemachi-Japan, Stockholm-Sweden, San Jose-USA, Zurich-Switzerland, Milan-Italy, New York-USA, Frankfurt-Germany, Moscow-Russia, Palo Alto-USA, Sao Paulo-Brazil, and Miami-USA. The RRCs in each location collect entire routing tables. The collected routing data are then transferred every eight hours through an incremental file transfer utility called *rsync* to a central storage area at the RIPE center in Amsterdam.

The routing datasets collected from Route Views and RIPE projects contain BGP routing information from the participating ASes located in North America and in Europe. In contrast to the centralized way of collecting routing data in Route Views projects, RIPE applies a distributed approach to the data collection. The regional growth of ASes is larger and more dynamic in Europe than in North America [28]. Route Views and RIPE projects collect datasets from Border Gateway Protocols (BGP) routing tables and have been extensively used by the research community [18], [23], [29], [33], [48], [52]. Route Views began the routing data collection process in November 1997 while RIPE is active since October 1999. Other sources of routing data such as *traceroute* and Looking Glass servers [6] do not provide such historical information. *Traceroute* provides the route including intermediate routers information over the network between two systems. Looking Glass project [6], on the other hand, provides servers that give public remote access to view routing information of the user network. These servers are run by ISPs. Thus, we analyze various graph properties of the Internet topology based on the datasets collected from Route Views and RIPE projects.

In this thesis, we evaluate and compare the Internet graph properties from Route Views and RIPE datasets over the period of five years: 2003-2008. We analyze the datasets collected at 00:00 am on July 31, 2003 and at 00:000 am on July 31, 2008. Datasets from two different locations from Route Views and datasets from ten different locations from RIPE are selected for 2003. Datasets from six different locations from Route Views and datasets from ten different locations from RIPE are selected for 2008. The number of ASes observed in the analyzed datasets from Route Views and RIPE projects are shown in Table 2.3.

Table 2.3 Number of ASes observed in Route Views and RIPE datasets.

Date	2003-07-31 00:00	2008-07-31 00:00
Route Views	15,826	29,166
RIPE	15,777	29,197

2.4.1 Sample datasets

Samples of datasets from Route Views and RIPE projects are:

Sample data from Route Views project:

- TABLE_DUMP| 1050122432| B| 204.42.253.253| 267| 3.0.0.0/8| 267 2914 174 701| IGP| 204.42.253.253| 0| 0| 267:2914 2914:420 2914:2000 2914:3000| NAG| **|**
- TABLE_DUMP| 1050122432| B| 213.140.32.184| 12956| 0.0.0.0/0| 12956| IGP| 213.140.32.184| 0| 0| | NAG| |.

Sample data from RIPE project:

- TABLE_DUMP| 1041811200| B| 212.20.151.234| 13129| 3.0.0.0/8| 13129 6461 7018 80| IGP| 212.20.151.234| 0| 0| 6461:5997 13129:3010| NAG| |

- TABLE_DUMP| 1041811200| B| 193.148.15.85| 3257| 3.0.0.0/8| 3257 1239 7018 80| IGP| 193.148.15.85| 0| 99| 3257:3000 3257:3030 3257:5044| NAG| |.

2.4.2 BGP datasets and Internet topology

The routing table datasets of a BGP router are used to infer the Internet topology graph at AS level. For example, the AS path 267-2914-174-701 of the first sample data of Route Views project in Section 2.4.1 can be located in the sample Internet graph shown in Figure 2.1.

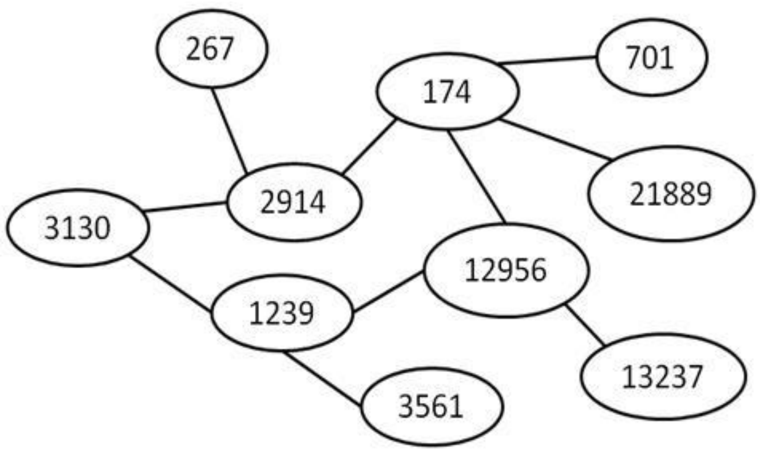


Figure 2.1. Example of the Internet graph derived from BGP routing table datasets.

3. POWER-LAWS AND SPECTRUM OF THE INTERNET GRAPHS

Power-laws and spectral properties of the Internet topology have been analyzed based on the adjacency matrix and the normalized Laplacian matrix of graphs that capture the Internet topology at AS level [34][17], [48], [52]. In this Chapter, we define the Internet topology at AS level in terms of matrices. We describe various power-laws associated with the Internet topology graph and summarize related work. We also define eigenvalues and eigenvectors of a matrix and explain their significance in spectral analysis of the Internet topology. Furthermore, we present work related to the analysis of the Internet graph using spectral analysis.

3.1 Internet topology and graph theory

An Internet AS graph $G(V,E)$ is an undirected, unweighted graph without self-loops and multiple edges from one node to another with V set of vertex and E set of edge. The graph $G(V,E)$ 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 a node. Two nodes i and j are called adjacent if they are connected by a link. The Internet network at AS level can be represented by the adjacency matrix $A(G)$ as:

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

A diagonal matrix $D(G)$ associated with $A(G)$, with row-sums of $A(G)$ as the diagonal elements of $D(G)$, indicates the degree of each node. The Laplacian matrix of a graph $G(V,E)$ 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 $NL(G)$ is defined as:

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

where d_i and d_j are the degrees of nodes i and j, respectively. The eigenvalues of the normalized Laplacian matrix lie in the range between 0 and 2. This property enables the comparison of the distribution of the eigenvalues of two distinct graphs if there is a large difference in their size. Various power-laws may be associated with the graph properties [29], [48].

3.2 Power-laws and the Internet topology

Various power-laws properties are derived using linear regression line. In this Section, we define power-laws and linear regression based on the least square approximation method and summarize work related to power-laws properties of the Internet topology.

3.2.1 Power-laws

Power-laws have been extensively used to describe and model the Internet at AS level. Power-laws are expressed in the form of $y \propto x^{a}$, where y and x are the measures of interest and exponent a is a constant. The variables y and x has linear relationship when plotted in log-log scale:

$$
\begin{aligned}
&y = 10^{b} \times x^{a} \\
&\log y = \log 10^{b} + \log x^{a} \\
&\log y = b + a \times \log x \\
&y' = b + ax'.
\end{aligned}
$$

The presence of power-laws in node degree vs. node rank, frequency of node degree vs. node degree, complementary cumulative distribution function (CCDF) of node degree vs. node degree, number of nodes within a number of hops vs. number of hops, and eigenvalue of the adjacency matrix vs. its index were observed in [29] and [48].

The node degree power-law is observed while examining the degree of a node. The number of edges incident to a node is called a node degree. When the nodes are sorted in decreasing order of node degree and plotted vs. the rank according to its index in the sequence, the power-law is observed as $d_{v} \propto r_{v}^{R}$, where d_v is the degree of a node v, r_v is the rank of node v, and R is the exponent of the power-law.

The frequency vs. degree power-law implies $f_d \propto d^O$, where f_d is the frequency of degree d and the constant O is the power-law exponent. The frequency f_d of a degree d is the number of nodes with degree d.

The CCDF of node degree vs. node degree power-law implies $D_d \propto d^D$, where D_d is the CCDF of a node degree d and D is the CCDF power-law exponent. The CCDF is defined as $F_c(x) = P(X > x)$, where $P(X > x)$ is the probability that the random variable X has a value greater than x. Thus, CCDF of a node degree d indicates the percentage of nodes that have degree greater than d and shows the distribution of the degree of a node.

Similarly, the eigenvalue vs. its index power-law implies $\lambda_i \propto i^{\varepsilon}$, where λ_i is the eigenvalue of the matrix associated with the increasing sequence of numbers i, the constant ε is the power-law exponent associated with the eigenvalue of the matrix.

3.2.2 Linear regression

Linear regression approach is used to model the power-law relationship between two variables. The exponent of a power-law is calculated by determining the slope of the estimated linear regression line. The technique is based on the least square approximation.

For each set of data points, we derive a straight line $y = b + ax$ through the available set of points x and y such that sum of squares of distances from set of data points to the straight line $y = b + ax$ is minimum, where

$$b = \frac{\Sigma y \Sigma x^2 - \Sigma x \Sigma yx}{n\Sigma x^2 - (\Sigma x)^2}$$ *and*

$$a = \frac{n\Sigma yx - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}.$$

The correlation coefficient, also called the cross-correlation coefficient, provides the quality measure of a least square fitting to the original data. For linear least squares fitting, the coefficient a in $y = b + ax$ is given by

$$a = \frac{n\Sigma yx - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}.$$

The coefficient a' in $x = b' + a'y$ is given by

$$a' = \frac{n\Sigma yx - \Sigma x \Sigma y}{n\Sigma y^2 - (y)^2}.$$

The correlation coefficient denoted by $r = aa'$ is expressed as

$$r = \frac{(n\sum xy - \sum x \sum y)^2}{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}.$$

The correlation coefficient is calculated between linear regression line and plotted data. The correlation coefficients for linear fits to data having different data distributions are shown in Figure 3.1. The correlation coefficient value of 1.0 indicates that the data points are exactly on a line. Thus, a high correlation coefficient indicates the existence of power-law.

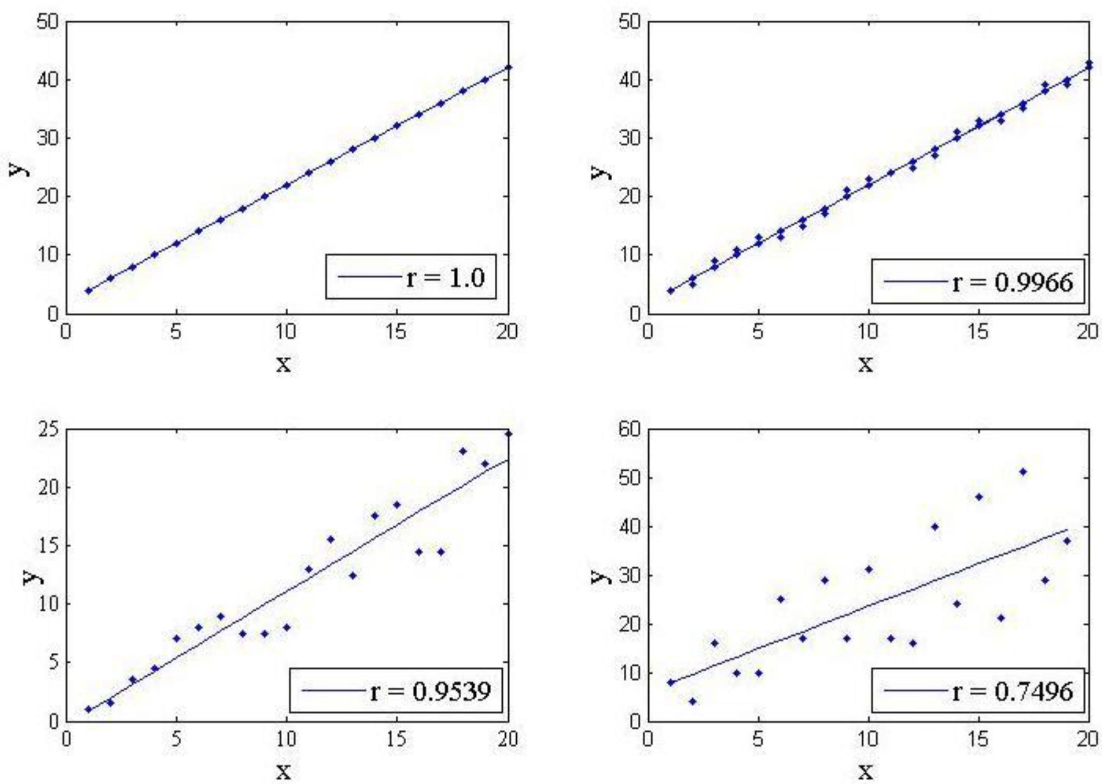


Figure 3.1 Correlation coefficients for linear fits to data having different data distributions.

3.2.3 Power-laws and the Internet topology

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 at AS-level [33]. Datasets collected from BGP routing tables of Route Views project [2] indicate that the Internet topology is characterized by the presence of various power-laws. Four different power-laws were observed when considering a node degree vs. node rank, frequency of node degree vs. node degree, a number of nodes within a number of hops vs. number of hops, and eigenvalue of the adjacency matrix vs. its index [29], [48]. The existence of these power-laws in the Internet topology indicates highly skewed distributions of various topology properties measured by power-law exponents [29], [48]. Some of these conclusions were subsequently revised by considering a more complete AS-level representation of the Internet topology [18], [23].

Q. Chen et al., [23] reported that BGP datasets collected by Route Views project from limited vantage points represent a partial view of the Internet topology. Thus, the degree distribution power-laws observed in [29] may not exist in the Internet topology. These power-laws are only consistent with graphs of ASes from Route Views datasets [2] and are inconsistent when analyzing extended graphs that have a more complete view of AS connections. The comparison of the connectivity of ASes derived from BGP routing tables of forty-one individual ASes and the information from the Looking Glass project [6] to the connectivity of ASes derived from Route View BGP datasets showed that the AS paths derived from BGP routing table do not completely capture the Internet topology. Furthermore, datasets from Route Views project reveal heavy tailed node degree distribution that is close to Weibull distribution. Thus, only the tail exhibits power-laws [23]. Weibull distribution is a continuous probability distribution with the probability density function:

$$f(x) = \begin{cases} (k/\lambda)(x/\lambda)^{k-1} e^{-(x/\lambda)^k} & if\ x \geq 0 \\ 0 \quad if\ x < 0, \end{cases}$$

where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter.

P. Mahadevan et al., [42] compared the power-law exponents calculated using three different datasets, *traceroute* data from the Skitter project of Cooperative Association for Internet Data Analysis (CAIDA) [4], Route Views datasets [2], and RIPE WHOIS database [12]. The Skitter project uses a tool called *skitter* to collect the BGP routing table. RIPE WHOIS database provides routing datasets of RIPE Internet routing registries (IRR). The IRR provide routing information in order to validate BGP messages and to map the AS number at origin to a list of networks. This routing information is stored in WHOIS database of individual registries. The routing data from RIPE WHOIS database is obtained using *whois* command (*whois -h whois.ripe.net AS <ASN>)*. The comparison of the power-law exponents revealed that Route Views and Skitter project datasets obey power-laws characteristics. However, RIPE datasets do not follow power-laws due to excessive number of nodes with minimum node degree [42].

In order to find the relationship between ASes, D. Magoni et al., [40] analyzed various metrics such as connectivity and degree distribution of ASes in the Internet topology graph emanated from BGP routing datasets collected by Route Views project. The result showed that in addition to power-laws reported in [29], the Internet topology exhibits four different power–laws associated with the shortest path between the ASes [40]. The shortest path between any two nodes is the minimum number of nodes that must be traversed from one node to another. The comparison of the power-law exponents for the paths having various length between a pair of ASes confirmed their existence in the shortest path between a pair of ASes. However, the observed variations in power-law exponents suggest that the power-law exponents might not be consistent over time [40].

Rationale of power-laws

In order to analyze the cause of power-laws in the Internet topology, A. Medina et al., [43] considered four different factors: 1) preferential connectivity of a new node to existing nodes, 2) incremental growth of the network, 3) distribution of nodes in space, and 4) locality of edge connection. Preferential connectivity indicates that a new node is more likely to be connected to existing nodes that are highly connected than to nodes that are less connected. Incremental growth implies that networks are formed by continual addition of new nodes. Hence, the size of the network gradually increases. Distribution of nodes in space indicates that nodes are distributed in space according to skewed (heavy-tailed) distribution. Locality of edge connection implies the tendency of a new node to connect to the existing nodes that are closer in distance. A. Medina et al., [43] generated topologies having nodes between 500 and 15,000 with and without incremental growth and preferential connectivity using BRITE [44], analyzed these synthetically generated topologies, and observed the presence of power-laws. The result showed that some of the generated topologies do not obey node degree vs. rank and frequency of node degree vs. node degree power-laws. However, eigenvalue power-law existed in all topologies but with different value of the power-law exponent. The simulated topologies using preferential connectivity and incremental growth have similar power-laws as the real Internet topologies. Thus, the Internet topology having power-laws exhibits the properties of preferential connectivity and incremental growth and the power-law exponents are the metrics used to verify the simulated Internet topology [43].

Power-laws and topology generators

The Internet topology is considered to have a scale free network structure. In order to find the best metrics that generates a large scale graph structure, H. Tangmunarunkit et al., [51] generated the Internet topology graph using various topology metrics such as neighbourhood size (expansion), the size of a cut-set for a balanced bi-partition graph (resilience), and the minimum communication cost spanning tree (distortion). They reported that the network generator based on the degree distribution more accurately captured the large scale structure of the measured topology. Furthermore, the degree-based generator produces a form of hierarchical topology that closely resembles the hierarchical nature of the Internet topology [51]. Considering the Internet topology as a hierarchical structure, T. Bu et al., [15] introduced an algorithm based on the decomposition technique to understand how well the power-law graphs capture the interconnection structure such as hierarchical structure of the Internet graph. The Internet AS graph and the graphs produced by topology generators based on power-laws distribution show similar hierarchical structure [15].

P. Mahadevan et al., [42] analyzed the synthetic topology generated using Power-Laws Random Graphs (PLRG) [13] and compared it with *traceroute* data from the Skitter project of Cooperative Association for Internet Data Analysis (CAIDA) [4], Route Views datasets [2], and RIPE WHOIS database [12]. The PLRG generator uses power-law exponents as an input to create a topology. For a given number of nodes and power-law exponent, PLRG assigns degrees to nodes based on the power-law distribution. It then randomly matches degrees among all nodes generating a topology. The analysis indicated that PLRG model failed to recreate RIPE WHOIS graph since its node distribution does not follow power-laws. However, the comparison of the graphs emanated from Skitter and Route Views datasets to those emanated from PLRG topology generator revealed that the generated topologies do not accurately capture the important properties such as joint degree distribution and clustering properties of Skitter and Route Views graphs [42].

Various characteristics of the Internet topology indicate complex behaviour of the Internet. Based on the observed properties of the Internet, different topology generators such as Waxman, PLRG, Barabási-Albert (BA) model, and the Internet Topology Generator (Inet) were developed. In Waxman topology model, the nodes of the network are uniformly distributed in the plane and edges are added according to probabilities that depend on the distances between nodes. BA model implements the power-laws properties based on the preferential attachment of nodes in order to generate the Internet topology. The Inet model calculates the frequency-degree and rank-degree distributions. It then assigns degrees to each node according to these distributions forming a topology. The comparison of the topology generators reveals that no single topology generator can capture all the characteristics observed in the present state of the Internet [15].

3.3 Spectral analysis of the Internet topology

In this Section, we define eigenvalues and eigenvectors of a matrix and their role in spectral analysis of the Internet topology. We also present work related to the analysis of clustering and connectivity properties of AS nodes in the Internet topology based on spectral analysis.

3.3.1 Eigenvalues and eigenvectors

An AS graph constructed from the BGP routing tables is undirected and does not contain self-loops. Thus, the adjacency matrix of such graph is symmetric and has a complete set of real eigenvalues and an orthogonal eigenvector basis. The definition of eigenvalue follows:

Let x be an n-dimensional real vector such that x is a function of the vertices of graph G, then x is called the eigenvector of A with eigenvalue λ if and only if it satisfies the eigenvalue equation $Ax = \lambda x$. The eigenvalue λ of A corresponding to the eigenvector x is a scalar quantity.

Every $n \times n$ real symmetric matrix A has spectrum on n orthogonal eigenvectors $e_1, e_2, e_3, \ldots, e_n$ with real eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3, \cdots, \geq \lambda_n$. The set of eigenvalues of the adjacency matrix is known as the spectrum of a graph. The analysis of spectrum can be performed for any matrix with real spectrum. The eigenvalues of a matrix are closely related to certain graph invariants. Furthermore, the eigenvalues of a network graph are associated with topological characteristics of the network such as number of edges, spanning trees, connected components, diameter of the network, presence of cohesive clusters, long paths and bottlenecks, and randomness of the network.

The second smallest eigenvalue

The second smallest eigenvalue of the Laplacian matrix of a graph G is called the *algebraic connectivity* of the graph [32]. This eigenvalue is greater than 0 iff G is a connected graph. This indicates that the number of times 0 appears as an eigenvalue of the Laplacian matrix is equal to the number of connected components in a graph. The algebraic connectivity also provides information about an average distance of graph nodes.

The algebraic connectivity is related with the usual vertex and edge connectivity of a graph [31], [32]. If a graph G gets disconnected by removing minimum k vertices together with the adjacent edges, then the graph is said to have k vertex connectivity. Similarly, if a graph G gets disconnected by removing minimum k edges, then the graph is said to have k edge connectivity. The relationships between the algebraic connectivity and the vertex and edge connectivity are described in [31], [32]. For example, $a(G) \leq v(G) \leq e(G)$, where $a(G)$ is the algebraic connectivity, $v(G)$ is the vertex connectivity and $e(G)$ is the edge connectivity of a non-complete graph G.

The magnitude of the second smallest eigenvalue indicates the robustness of a network. Its value depends on the number of vertices and the diameter of a graph. The diameter of a graph is the shortest path between two widest vertices and it gives the maximum number of vertices that need to be passed in order to reach from one vertex to another vertex of a graph. For a connected graph with n vertices and D diameter, the algebraic connectivity lies between 1 and $1/nD$. For a large graph, the algebraic connectivity is small and is closer to $1/nD$. Thus, the algebraic connectivity depends on the number of vertices and in the way vertices are connected. In general, the algebraic connectivity decreases with the number of vertices, and increases with the average degree in case of random graphs.

The large eigenvalues

The spectrum of the adjacency matrix has been extensively used to find the cluster of ASes having similar characteristics such as connectivity pattern [34]. The eigenvectors corresponding to the large eigenvalues contain information relevant to clustering. The large eigenvalues and the corresponding eigenvectors of the adjacency matrix provide information suggestive to the intracluster traffic patterns of the Internet topology [34].

3.3.2 Clusters and connectivity in the Internet topology

The spectrum of the Internet topology graph is considered as the metric for clustering and connectivity analysis. In addition to power-laws properties, properties such as connectivity and clustering properties of the Internet graph were also studied [34], [52]. The eigenvectors corresponding to the small eigenvalues tend to capture the local characteristics that can be determined from the data. Similarly, the eigenvectors corresponding to the large eigenvalues tend to capture the global characteristics of the graph and its semantics such as clusters of connected nodes.

C. Gkantsidis et al., [34] used a spectral filtering method in order to find the clusters of connected nodes in the Internet topology based on the elements values of the eigenvector corresponding to the large eigenvalues of the adjacency matrix. They analyzed graphs derived from various BGP routing table datasets and graphs generated from four different topology generator models: Inet-2.1, Waxman with preferential connectivity, improved Generalized Linear Preference (GLP) heuristics, and PLRG topology generators [34]. The GLP model generates the Internet topology using power-laws such that the probability that node increases its degree is a function of the degree. The results show that synthetically generated topology graphs lack view of the entire Internet topology because the topology generators address only AS or router-level topologies. The analysis of BGP datasets reveals that the clustering varies in the core and at the edges of the network. It also varies at different geographic locations. However, the clustering based on the largest eigenvalue is consistent over time and can be used as a robust characteristic to represent the Internet topology. Thus, good clustering methods are needed in order to identify the clustering properties of the Internet topology [34].

The eigenvalue spectrum of real world graphs provides information about their structural properties. I. Farkas et al., [30] studied the properties of real world graphs such as the Internet in terms of spectral density by analyzing the spectrum of graphs. They developed an algorithm to analyze the nature of spectral density of real world graphs represented as random graph, small world, and scale free models. Plots of the complete spectrum of the graphs show that the spectrum of scale free graph has a triangle-like spectral density in the centre with power-law decay in the tail, while the spectrum of small world graphs shows a complex pattern of spectral density with several random sharp peaks [30].

D. Vukadinovic et al., [52] used the AS topology emanated from BGP routing table collected by NLANR project and from synthetic topology generated using Inet-2.1 topology generator in order to analyze the spectrum of the normalized Laplacian matrix. The normalized Laplacian spectrums of synthetic graphs show variation and significant difference, whereas the spectrums of AS graphs from BGP routing table show invariance over time despite exponential growth of the Internet. Thus, the normalized Laplacian spectrum of a graph provides concise fingerprint of the real Internet topology [52]. Furthermore, analysis of sub-graphs derived from synthetic graphs shows the presence of new power-laws within the connected components at the core of the network topology. Thus, this combination of structural properties and power–laws is needed when developing topology generators [52].

4. POWER-LAWS AND THE INTERNET TOPOLOGY

In this Chapter, we observe four different power-laws: node degree vs. rank, CCDF of node degree vs. node degree, eigenvalue of the adjacency matrix vs. index, and eigenvalue of the normalized Laplacian matrix vs. index over the period of five years using Route Views and RIPE 2003 and 2008 datasets. We use linear regression based on the least square approximation method to model the relationship between y-axis and x-axis parameters. Furthermore, we calculate the confidence intervals of power-law exponents on randomly selected data from Route Views and RIPE 2003 and 2008 datasets.

4.1 Node degree

In this Section, we observe the node degree vs. rank power-law. 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 a log-log scale. Node degrees in decreasing order vs. rank are shown in Figure 4.1 and Figure 4.2. The points represent measured data and the solid line represents the least square approximation. The plots imply $d_v \propto {r_v}^R$, where v is the node number and R is the node degree power-law exponent. The node degree power-law exponent is the slope of the degrees of nodes plotted vs. the rank of the nodes on a log-log scale.

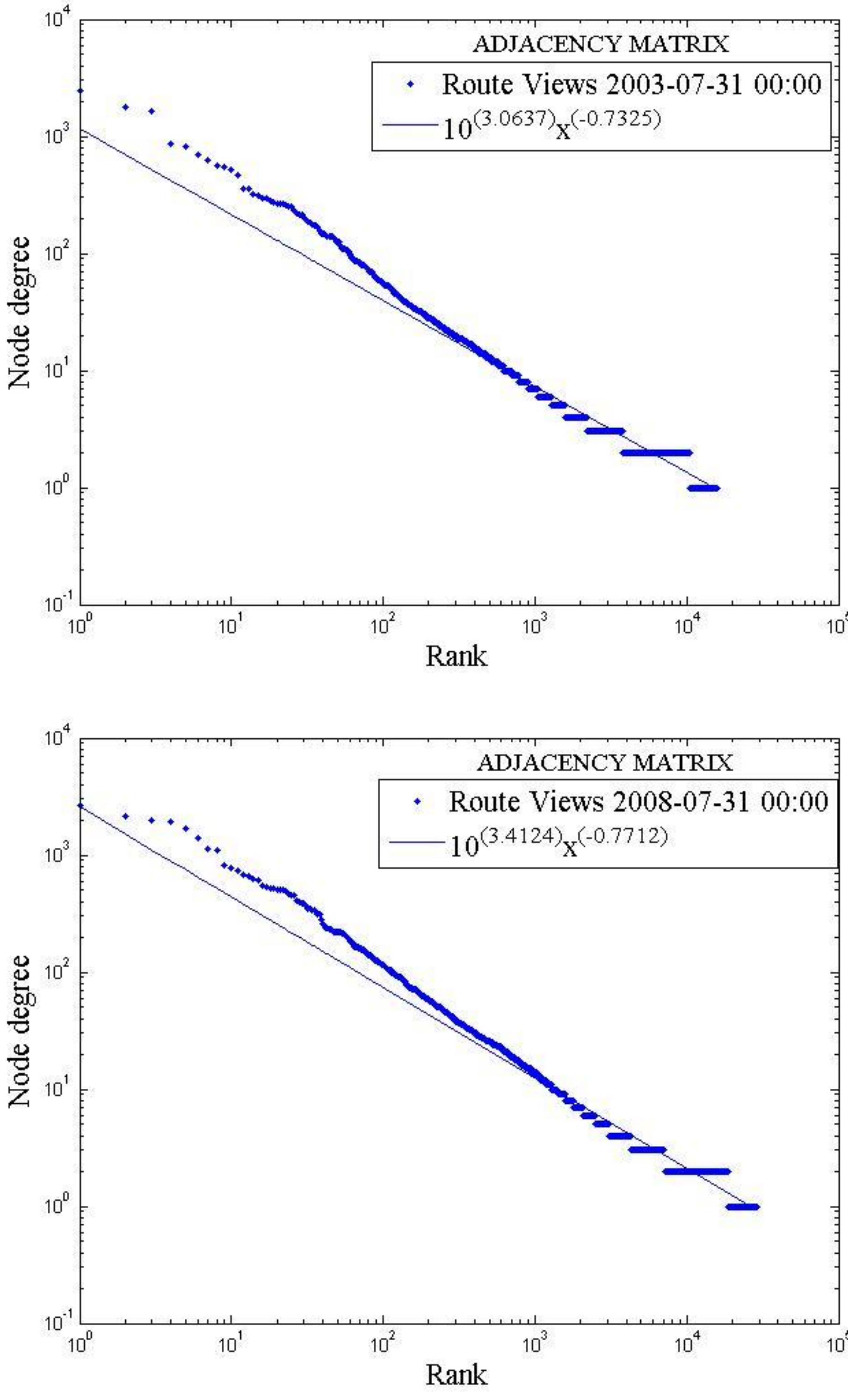


Figure 4.1 Route Views 2003 and 2008 datasets: The node degree power-law exponent R for Route Views 2003 (top) is –0.7325 with correlation coefficient –0.9661. The node degree power-law exponent R for Route Views 2008 (bottom) is –0.7712 with correlation coefficient –0.9721.

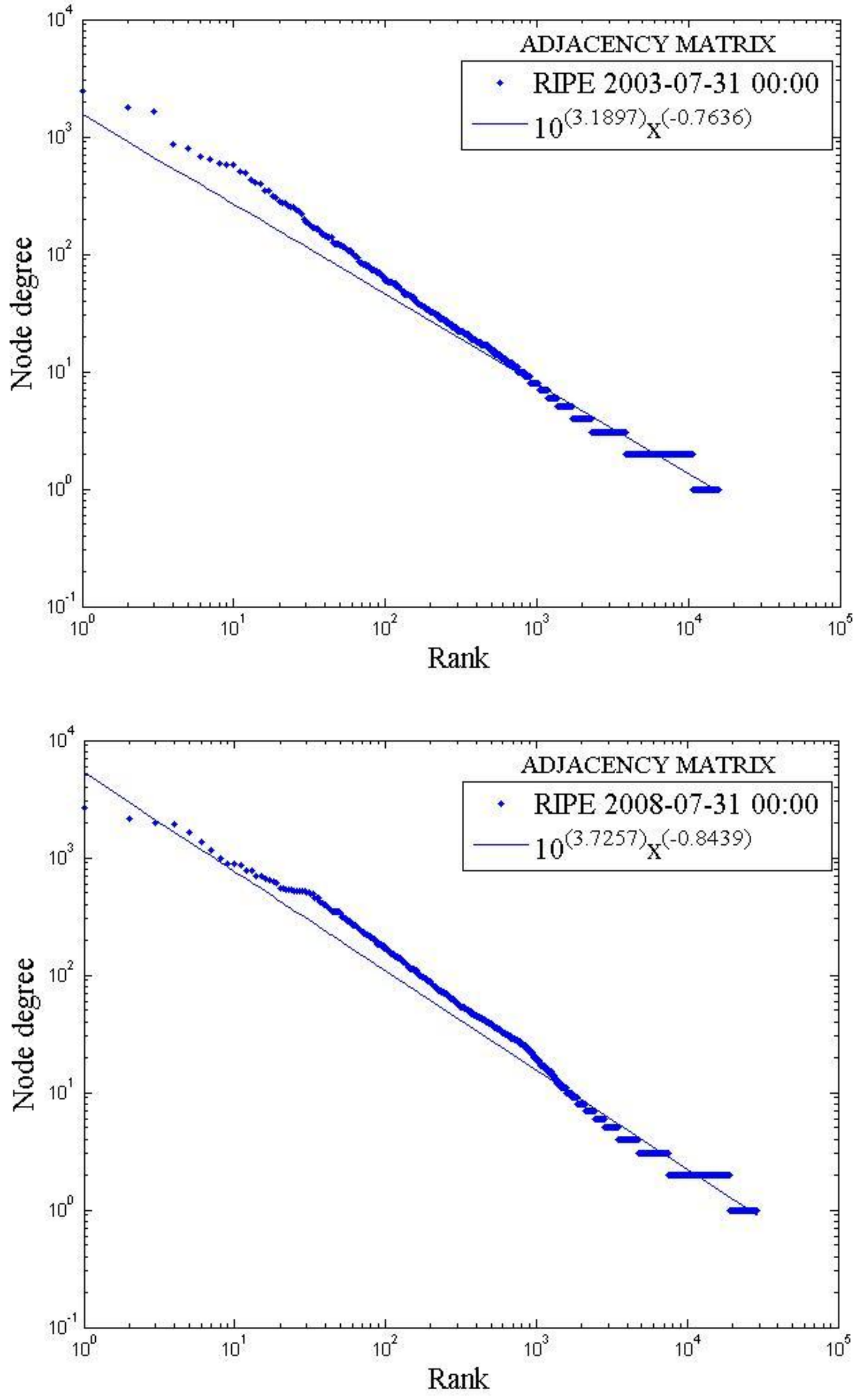


Figure 4.2 RIPE 2003 and 2008 datasets: The node degree power-law exponent R for RIPE 2003 (top) is –0.7636 with correlation coefficient –0.9687. The node degree power-law exponent R for RIPE 2008 (bottom) is –0.8439 with correlation coefficient –0.9744.

The node degree power-law exponents R are –0.7325 and –0.7712 for Route Views 2003 and 2008, respectively. Similarly, the node degree power-law exponents R are –0.7636 and –0.8439 for RIPE 2003 and 2008, respectively. The correlation coefficients are above 96 percent for both 2003 datasets and above 97 percent for both 2008 datasets. The difference of the power-law exponent values is small for 2003 and 2008 datasets.

The first twenty ASes with the highest node degree from all four datasets are identified in order to observe node degrees and the position of highly connected nodes over the years. Table 4.1 lists the first twenty ASes having larger node degree. The node degrees are comparatively larger in 2008 datasets. The node degrees of Route Views and RIPE 2008 datasets are comparable. Similarly, the node degrees of Route Views and RIPE 2008 datasets are comparable. The first 10 ASes with the largest node degrees in Route Views 2003 datasets are similar to those appeared in RIPE 2003 datasets. The first 10 ASes are similar in Route Views and RIPE 2008 datasets. Furthermore, we observe that 80 percent of the ASes in 2003 datasets appear in the first 20 position in 2008 datasets.

Table 4.1 The first 20 ASes with the highest node degree for each dataset.

Route Views 2003		RIPE 2003		Route Views 2008		RIPE 2008	
ASN	degree	ASN	degree	ASN	degree	ASN	degree
701	2422	701	2420	701	2645	701	2625
1239	1800	1239	1802	7018	2132	7018	2138
7018	1661	7018	1660	174	1983	174	1979
209	862	209	862	3356	1932	3356	1937
3356	815	3356	801	1239	1670	1239	1653
3561	689	3561	674	209	1386	209	1353
3549	628	3549	640	3549	1129	3549	1148
2914	567	2914	599	4323	1094	4323	990
702	553	8220	579	6939	823	13030	887
6461	513	702	574	19151	779	6939	883
4513	469	6461	502	6461	740	6461	852
1	356	4589	487	2828	676	19151	785
4323	355	3303	428	7132	668	9002	771
16631	318	13237	410	2914	629	4589	698
8220	309	6730	398	9002	607	2914	688
7132	295	1	350	1299	542	8928	659
3257	294	4323	346	702	532	2828	642
6347	279	16631	309	8220	520	8220	621
3786	270	3257	303	7575	515	7132	615
4766	267	7132	282	6667	506	3257	553

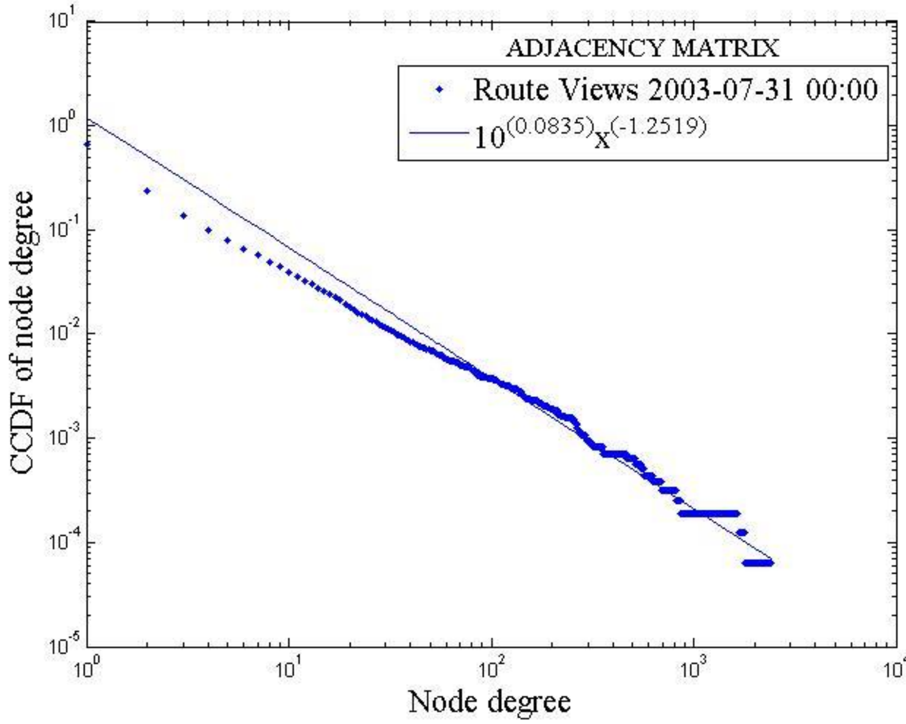
4.2 Complementary Cumulative Distribution Function (CCDF) of node degree

In this Section, we analyze the distribution of node degrees. We use CCDF D_d of a node degree d. It indicates the percentage of nodes that have degree greater than degree d and provides the distribution of the degree of nodes. The CCDF of node degree vs. node degree plotted on a log-log scale are shown in Figure 4.3 and Figure 4.4. It implies $D_d \propto d^D$, where D is the CCDF power-law exponent. The CCDF power-law exponents are –1.2519 and –1.3696 for 2003 and 2008 Route Views datasets, respectively and –1.2830 and –1.5010 for 2003 and 2008 RIPE datasets, respectively. The correlation coefficient is above 98 percent for each 2003 dataset and above 96 percent for each 2008 dataset. The difference of the power-law exponent values for Route Views and RIPE 2003 and 2008 datasets is small.

In a perfect power-law distribution, the node degree power-law exponent is related to CCDF power-law exponent as $R = 1/D$, where R is the node degree power-law exponent [29], [48]. We use the value of R from Section 4.1 to calculate the CCDF power-law exponent for each dataset. Table 4.2 lists the values of CCDF power-law exponents. The relationship $R = 1/D$ holds theoretically in a perfect power-laws distribution. However, we have inferred the CCDF power-law exponents from the empirical datasets. This may have caused the discrepancies in the exponent values as shown in Table 4.2.

Table 4.2 CCDF power-law exponents.

Values	Route Views 2003	Route Views 2008	RIPE 2003	RIPE 2008
D	–1.2519	–1.3696	–1.2830	–1.5010
1/R	–1.3650	–1.2967	–1.3096	–1.1850



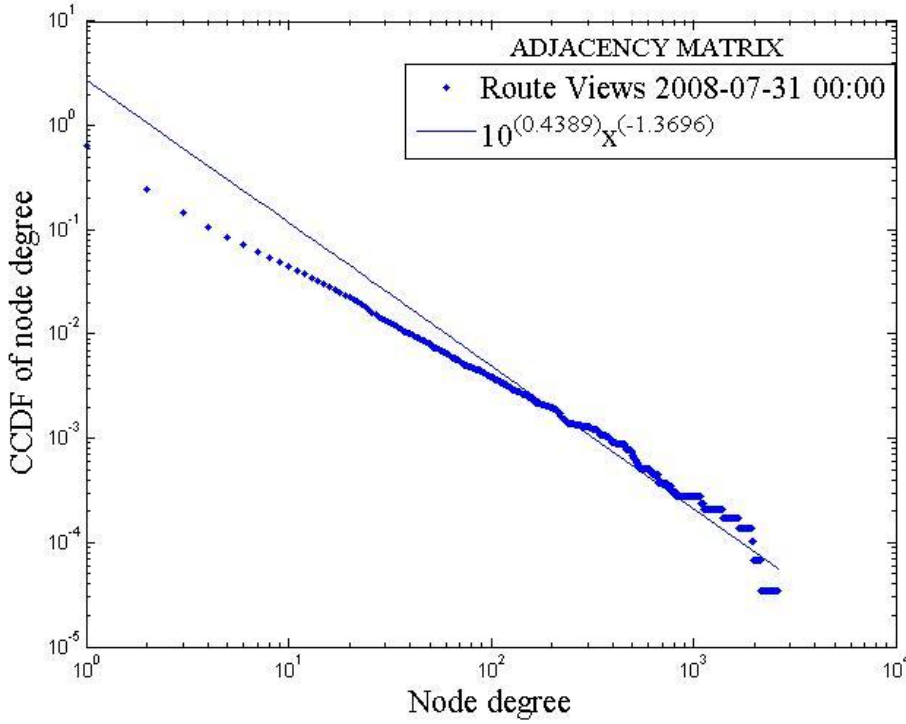
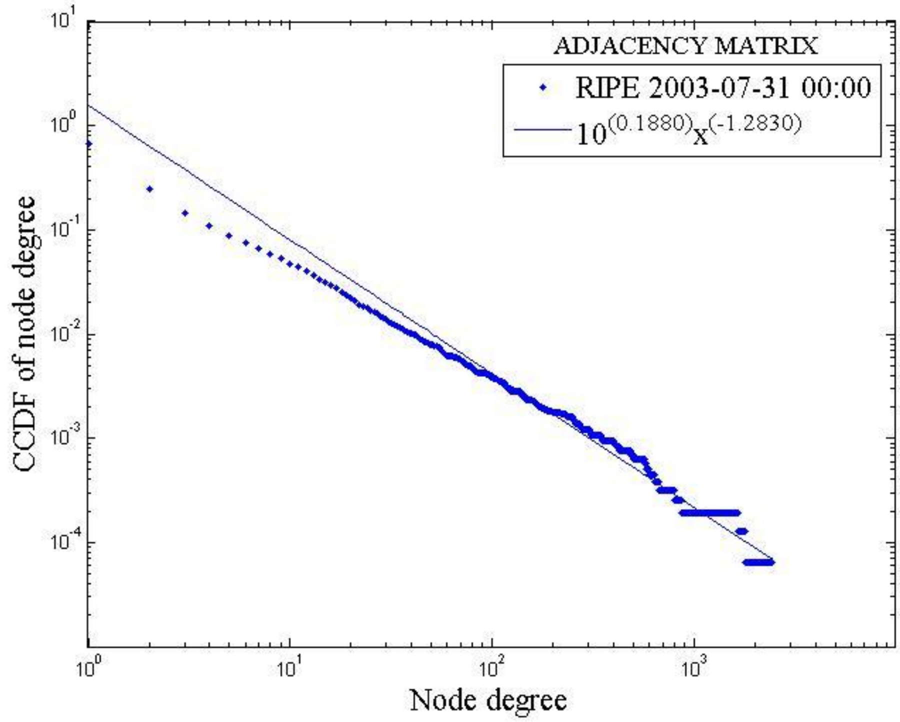


Figure 4.3 Route Views 2003 and 2008 datasets: The CCDF power-law exponent D for Route Views 2003 (top) is –1.2519 with correlation coefficient –0.9810. The CCDF power-law exponent D for Route Views 2008 (bottom) is –1.3696 with correlation coefficient –0.9626.



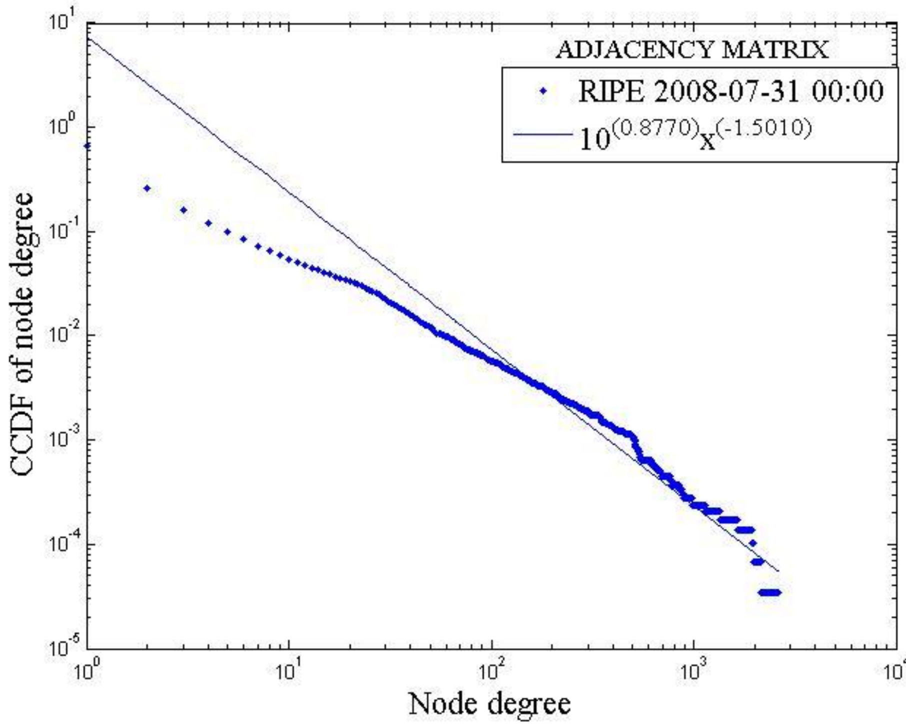


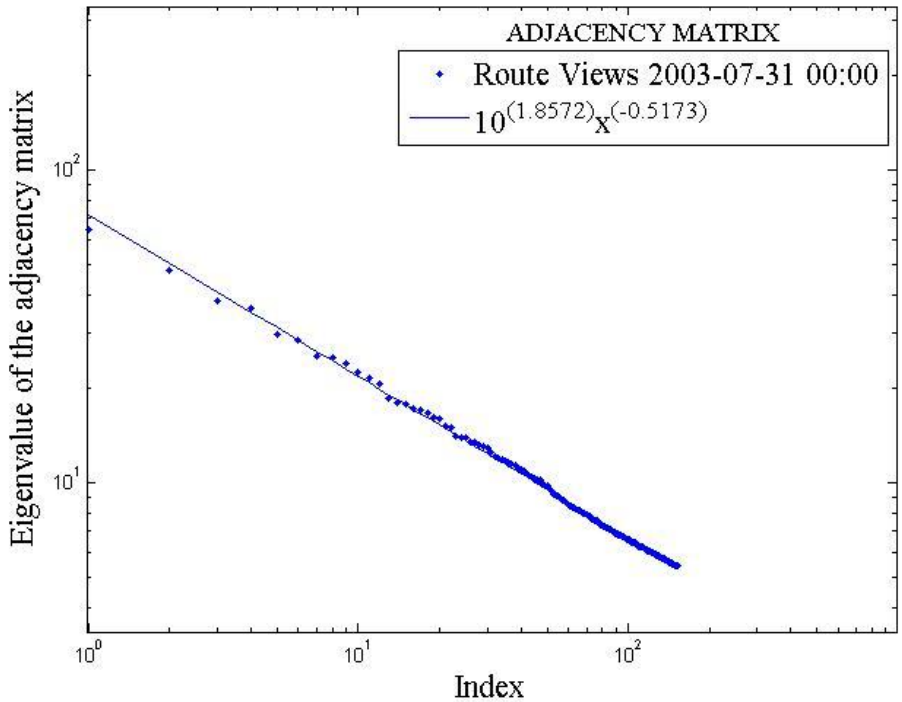
Figure 4.4 RIPE 2003 and 2008 datasets: The CCDF power-law exponent D for RIPE 2003 (top) is –1.2830 with correlation coefficient –0.9812. The CCDF power-law exponent D for RIPE 2008 (bottom) is –1.5010 with correlation coefficient –0.9676.

4.3 Eigenvalues of the adjacency matrix

The eigenvalues of a graph indicate its topological properties such as diameter, number of edges, and number of connected components in a graph. The eigenvalue power-law based on the adjacency matrix implies $\lambda_{ai} \propto i^{\varepsilon}$, where λ_{ai} is an eigenvalue, i is an index, and ε is an eigenvalue power-law exponent of the adjacency matrix.

The eigenvalues λ_{ai} of the adjacency matrix are sorted in decreasing order and plotted vs. the associated increasing sequence of numbers i representing the order of the eigenvalue. The power-law dependence between the graph eigenvalue and the eigenvalue index is shown in Figure 4.5 and Figure 4.6 for Route Views 2003 and 2008 and RIPE 2003 and 2008 datasets, respectively. Plotted on a log-log scale are eigenvalues of a graph matrix in decreasing order. Only the first 150 largest eigenvalues are considered.

The eigenvalue power-law exponents are –0.5173 and –0.4860 for Route Views 2003 and 2008 datasets, respectively and –0.5232 and –0.4927 for RIPE 2003 and 2008 datasets, respectively. The exponent values for Route Views and RIPE 2003 datasets are comparable. Similarly, Route Views and RIPE 2008 datasets have comparable exponent values. The values of the exponent have small difference for Route Views and RIPE 2003 and 2008 datasets. The correlation coefficients are above 96 percent for all four datasets.



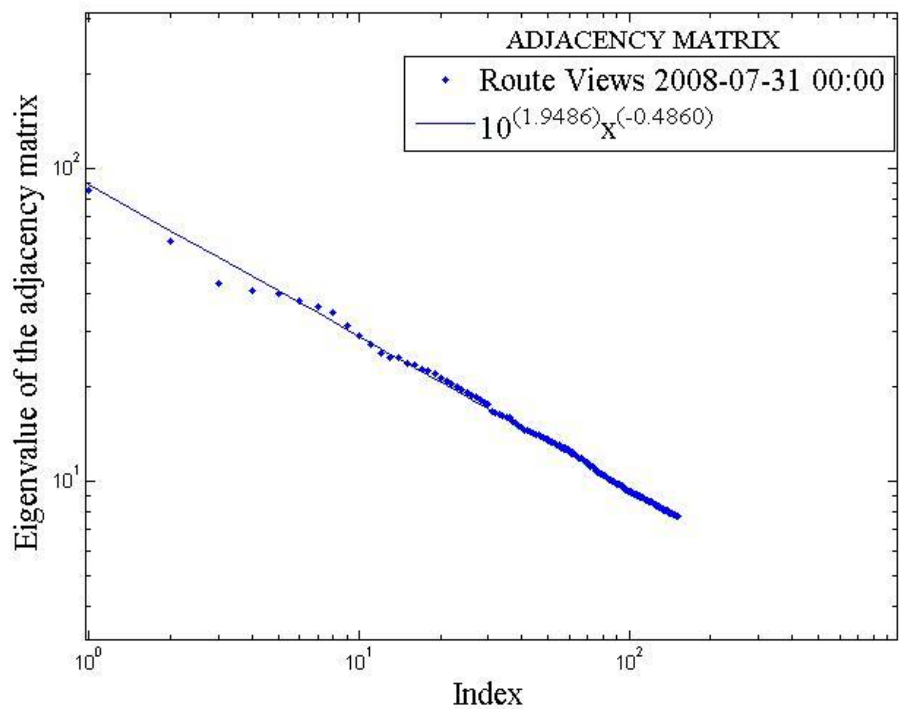
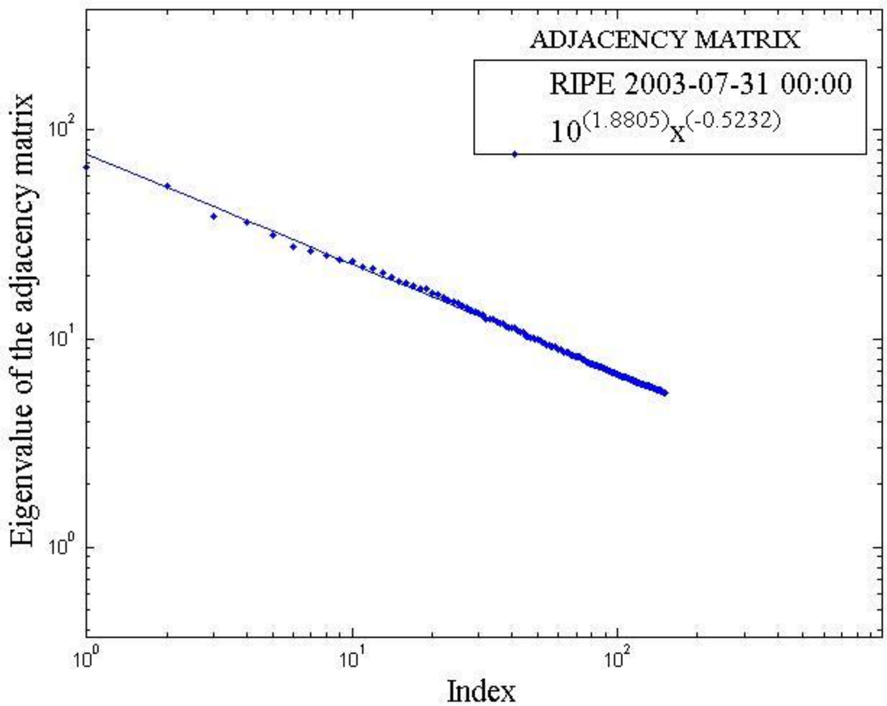


Figure 4.5 Route Views 2003 and 2008 datasets: The eigenvalue power-law exponent $\mathcal{E}$ based on the adjacency matrix for Route Views 2003 (top) is –0.5713 with correlation coefficient –0.9990. The eigenvalue power-law exponent $\mathcal{E}$ based on the adjacency matrix for Route Views 2008 (bottom) is –0.4860 with correlation coefficient –0.9982.



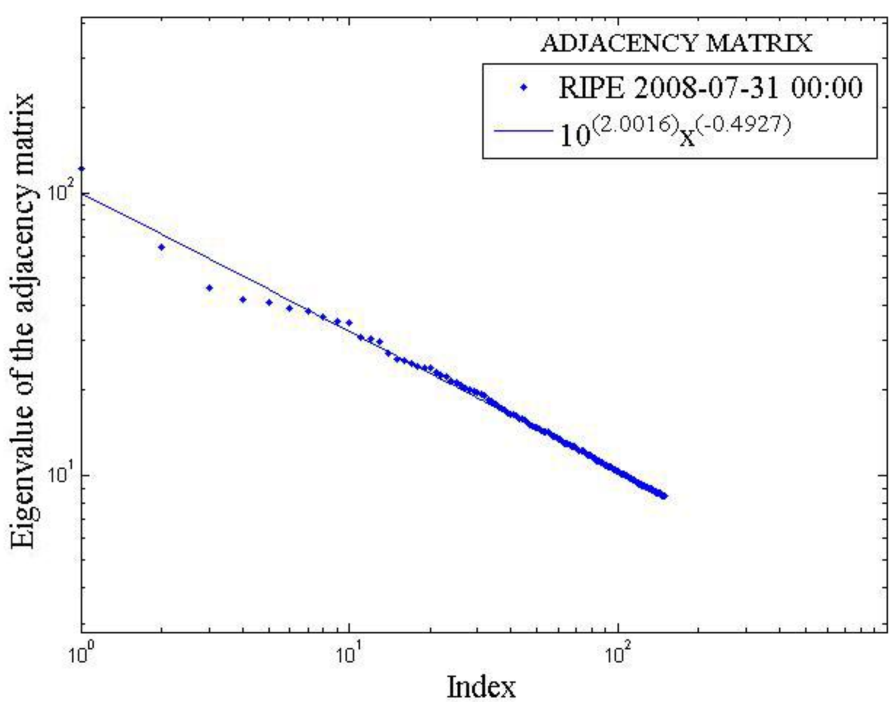


Figure 4.6 RIPE 2003 and 2008 datasets: The eigenvalue power-law exponent $\mathcal{E}$ based on the adjacency matrix for RIPE 2003 (top) is –0.5232 with correlation coefficient –0.9989. The eigenvalue power-law exponent $\mathcal{E}$ based on the adjacency matrix for RIPE 2008 (bottom) is –0.4927 with correlation coefficient –0.9970.

The first 5,000 largest eigenvalues are calculated and plotted vs. the order for all four datasets as shown in Figure 4.7. The MATLAB code was run on Quad core CPU with 2.39 GHz of processing speed and 3.93 GB of RAM and took four days to calculate the first 5,000 largest eigenvalues for each 2003 dataset and six days for each 2008 dataset. The plot that consists of only the first 600 eigenvalues shown in Figure 4.7 indicates that Route Views and RIPE 2008 datasets have larger eigenvalues in comparison to the eigenvalues of Route Views and RIPE 2003 datasets. The eigenvalues of RIPE 2008 datasets have larger value than Route Views 2008 datasets. Furthermore, the eigenvalues of RIPE 2003 datasets have larger value than Route Views 2003 datasets. The first twenty largest eigenvalues of the datasets are listed in Table 4.3.

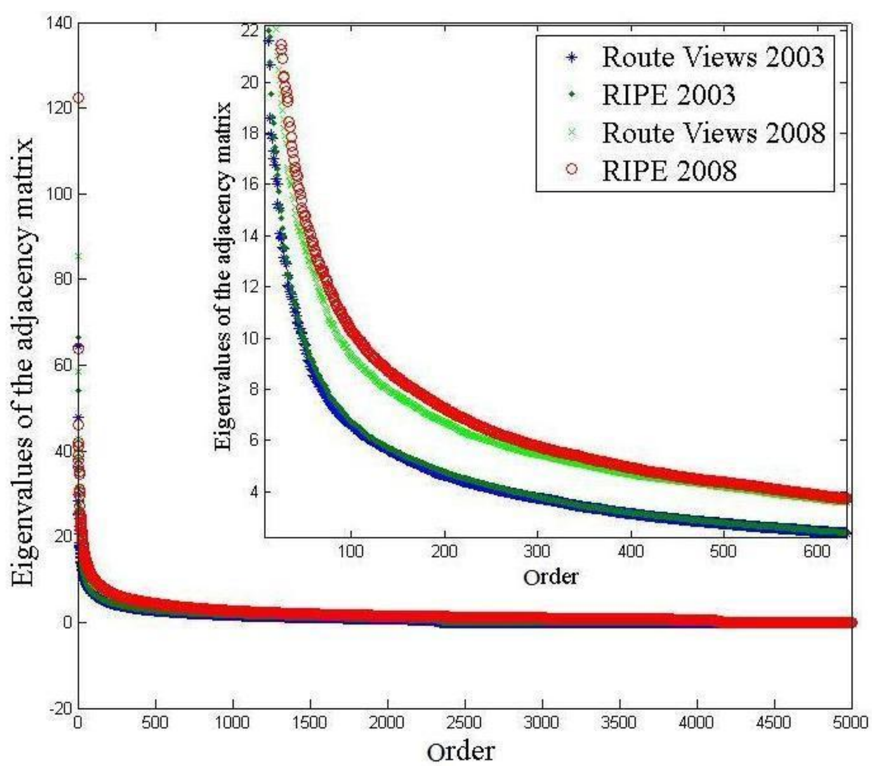


Figure 4.7 Route Views and RIPE 2003 and 2008 datasets: The first 5,000 largest eigenvalues plotted in descending order. Route Views and RIPE 2008 datasets have higher eigenvalues than Route Views and RIPE 2003 datasets.

Table 4.3 The first 20 largest eigenvalues of Route Views and RIPE 2003 and 2008 datasets.

order	Route Views 2003	Route Views 2008	RIPE 2003	RIPE 2008
1	64.30	85.43	66.65	122.28
2	47.75	58.56	54.19	63.94
3	38.15	42.77	38.24	46.14
4	36.23	40.85	36.14	41.98
5	29.88	39.69	31.21	41.08
6	28.50	37.85	27.38	38.93
7	25.47	36.21	26.41	37.94
8	25.06	34.66	25.06	36.47
9	24.13	31.58	23.86	35.08
10	22.51	29.34	23.32	34.47
11	21.61	27.40	22.02	30.97
12	20.69	25.69	21.77	30.54
13	18.58	25.00	20.75	29.68
14	17.94	24.82	19.55	27.03
15	17.78	23.89	18.67	25.74
16	17.31	23.69	18.42	25.35
17	16.99	22.81	17.85	24.83
18	16.75	22.46	17.44	24.30
19	16.22	22.04	17.24	24.06
20	16.01	21.36	16.63	24.00

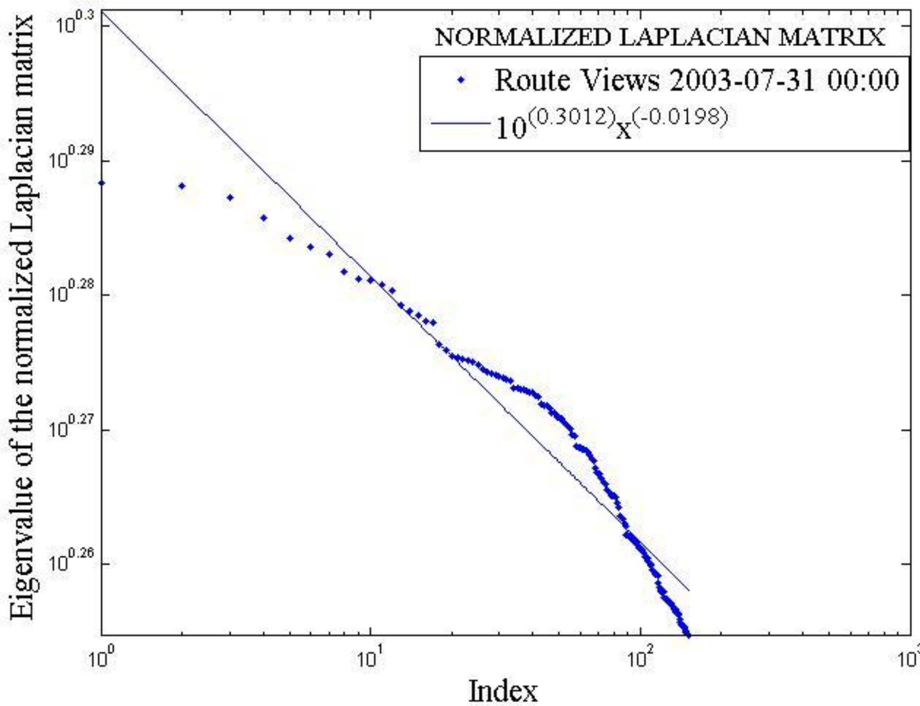
The power-law exponent and correlation coefficient of the eigenvalue power-law using the first 5,000 largest eigenvalues for Route Views 2008 datasets are –24.73 and –0.7232, respectively. The comparable values of power-law exponent and correlation coefficient are also observed for the remaining three datasets. This indicates that the eigenvalue power-law exists only in the tail of the distribution.

4.4 Eigenvalues of the normalized Laplacian matrix

The presence of eigenvalue power-law reported in [29] is based on the eigenvalues of the adjacency matrix. We also analyze the eigenvalue power-law based on the eigenvalues of the normalized Laplacian matrix. The newly observed power-law dependence between the eigenvalue of the normalized Laplacian matrix and the eigenvalue index is shown in Figure 4.8 and Figure 4.9 for Route Views 2003 and 2008 datasets and RIPE 2003 and 2008 datasets, respectively.

The eigenvalue power-law based on the normalized Laplacian matrix vs. its index implies $\lambda_{Li} \propto i^{L}$, where λ_{Li} is the eigenvalue, i is the index, and L is the eigenvalue power-law exponent of the normalized Laplacian matrix. We use similar procedure as in Section 4.3 to calculate the eigenvalue power-law exponent of the normalized Laplacian matrix. The eigenvalue power-law exponents are –0.0198 and –0.0177 for Route Views 2003 and 2008 datasets, respectively and –0.0206 and –0.0190 for RIPE 2003 and 2008 datasets, respectively. The difference of the power-law exponent values for Route Views and RIPE 2003 and 2008 datasets is small. The correlation coefficients are above 95 percent for all four datasets.

In order to observe the first 5,000 largest eigenvalues of the normalized Laplacian matrix, we ran the MATLAB code in Quad core CPU with 2.39 GHz processing speed and 3.93 GB of RAM for ten days without a success.



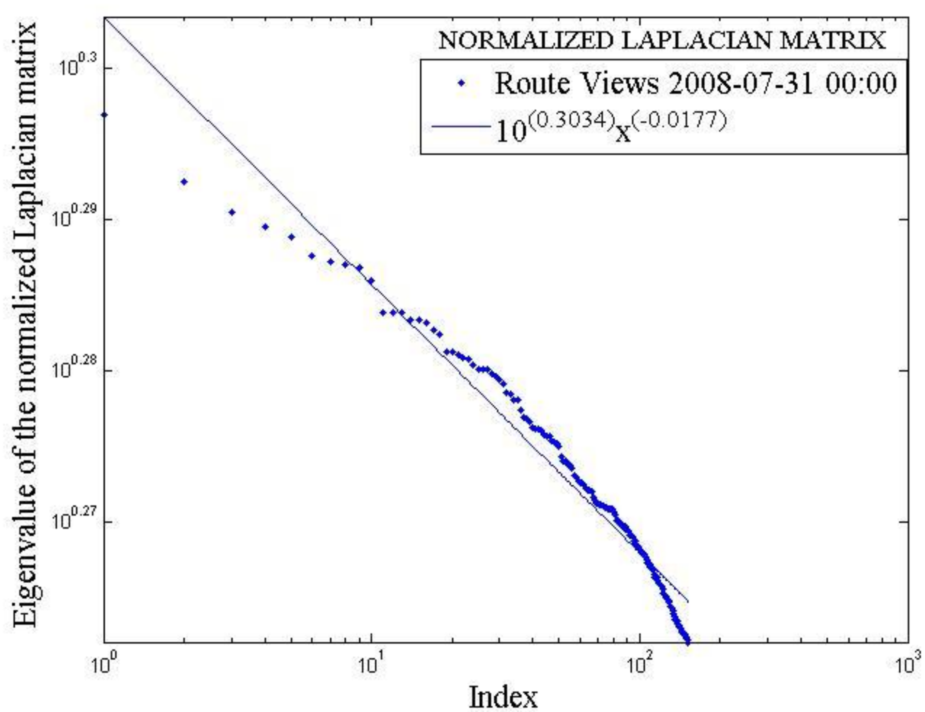
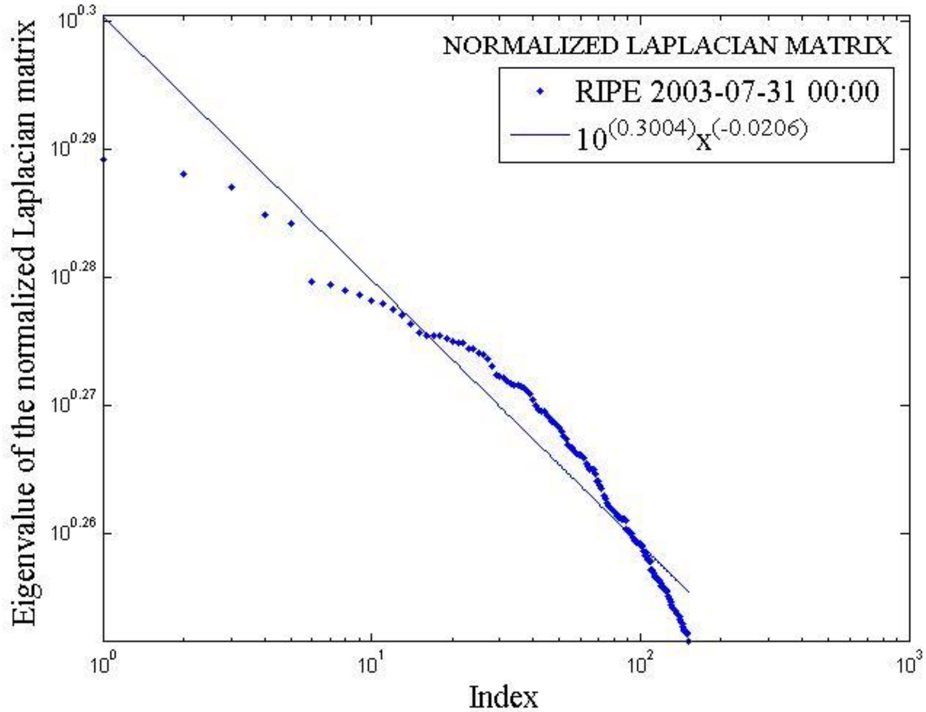


Figure 4.8 Route Views 2003 and 2008 datasets: The eigenvalue power-law exponent L based on the normalized Laplacian matrix for Route Views 2003 (top) is –0.0198 with correlation coefficient –0.9564. The eigenvalue power-law exponent L based on the normalized Laplacian matrix for Route Views 2008 (bottom) is –0.0177 with correlation coefficient –0.9782.



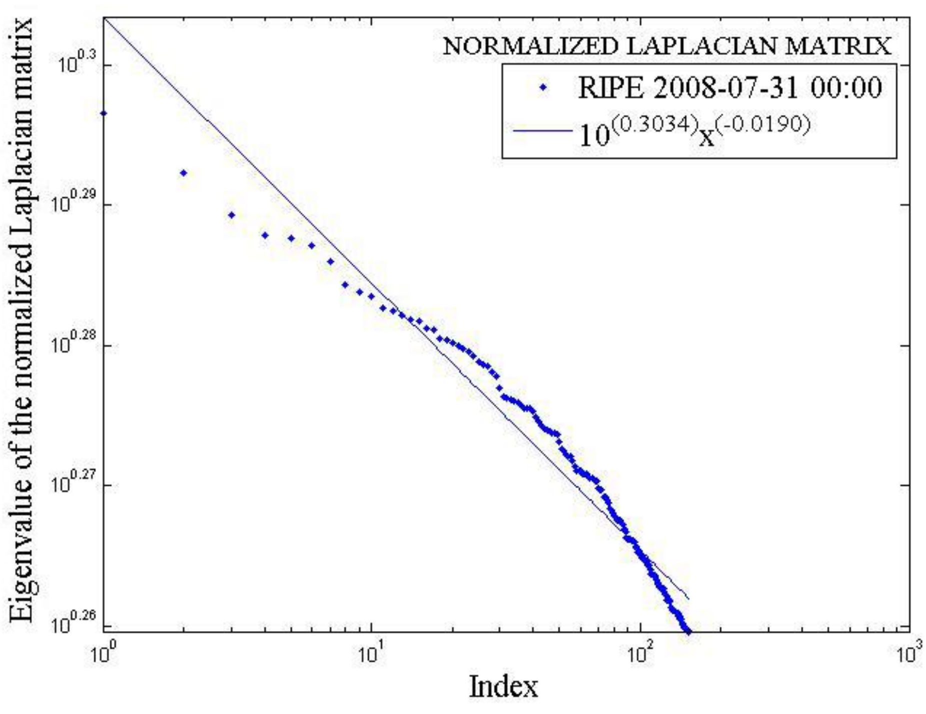


Figure 4.9 RIPE 2003 and 2008 datasets: The eigenvalue power-law exponent L based on the normalized Laplacian matrix for RIPE 2003 (top) is –0.0206 with correlation coefficient –0.9636. The eigenvalue power-law exponent L based on the normalized Laplacian matrix for RIPE 2008 (bottom) is –0.0190 with correlation coefficient –0.9758.

4.5 Confidence intervals

The four different datasets collected on July 31st 2003 and 2008 from Route Views and RIPE projects indicate the presence of power-laws. In order to further verify the result and to estimate the mean value of the power-law exponents, we collect twenty-four random sample datasets and calculate the power-law exponents for each sample. The sample datasets and the calculated power-law exponents for Route Views and RIPE 2003 and 2008 datasets are shown in Table 8.1 – Table 8.3 in Appendix.

The number of samples of each dataset is smaller than 30, with unknown standard deviation. Hence, we use the t-distribution to compute the confidence intervals of power-law exponents at 95 percent confidence level:

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

where $\overline{X}$ is the sample mean, $t_{x/2}$ is the t-distribution, s is the sample standard deviation, n is the number of samples, and μ is the population mean. We estimate the confidence intervals using six random samples selected from each 2003 and 2008 Route Views and RIPE datasets. The estimated confidence intervals of the power-law exponents are listed in Table 4.4.

The plot of the confidence intervals of the node degree power-law exponents and CCDF power-law exponents of four datasets is shown in Figure 4.10. The width of the confidence intervals is similar for Route Views 2008, RIPE 2003, and RIPE 2008 datasets. The confidence interval for Route Views 2003 datasets is comparatively wider. The node degree and CCDF of node degree power-law exponents have comparable values for both RIPE and Route Views 2003 and 2008 datasets. The analysis of node degree and CCDF of node degree power-law exponents in Sections 4.1 and 4.2, respectively also reveals insignificant increase of node degree and CCDF of node degree power-law exponents for 2008 datasets. The correlation coefficients of the node degree power–law exponent are above 96 percent for all random datasets. Meanwhile, the correlation coefficients of the CCDF power-law exponent are above 90 percent for all random datasets.

Table 4.4 Confidence intervals of power-law exponents at 95 percent confidence level.

Exponent	Range	Route Views 2003	RIPE 2003	Route Views 2008	RIPE 2008
R	higher	–0.7328	–0.7467	–0.7776	–0.8443
	lower	–0.6909	–0.7307	–0.7712	–0.8281
D	higher	–1.2692	–1.2714	–1.4086	–1.5115
	lower	–1.1794	–1.2307	–1.3622	–1.4411
ε	higher	–0.5118	–0.5105	–0.4886	–0.4985
	lower	–0.5046	–0.5033	–0.4840	–0.4914
L	higher	–0.0193	–0.0206	–0.0184	–0.0204
	lower	–0.0187	–0.0188	–0.0173	–0.0188

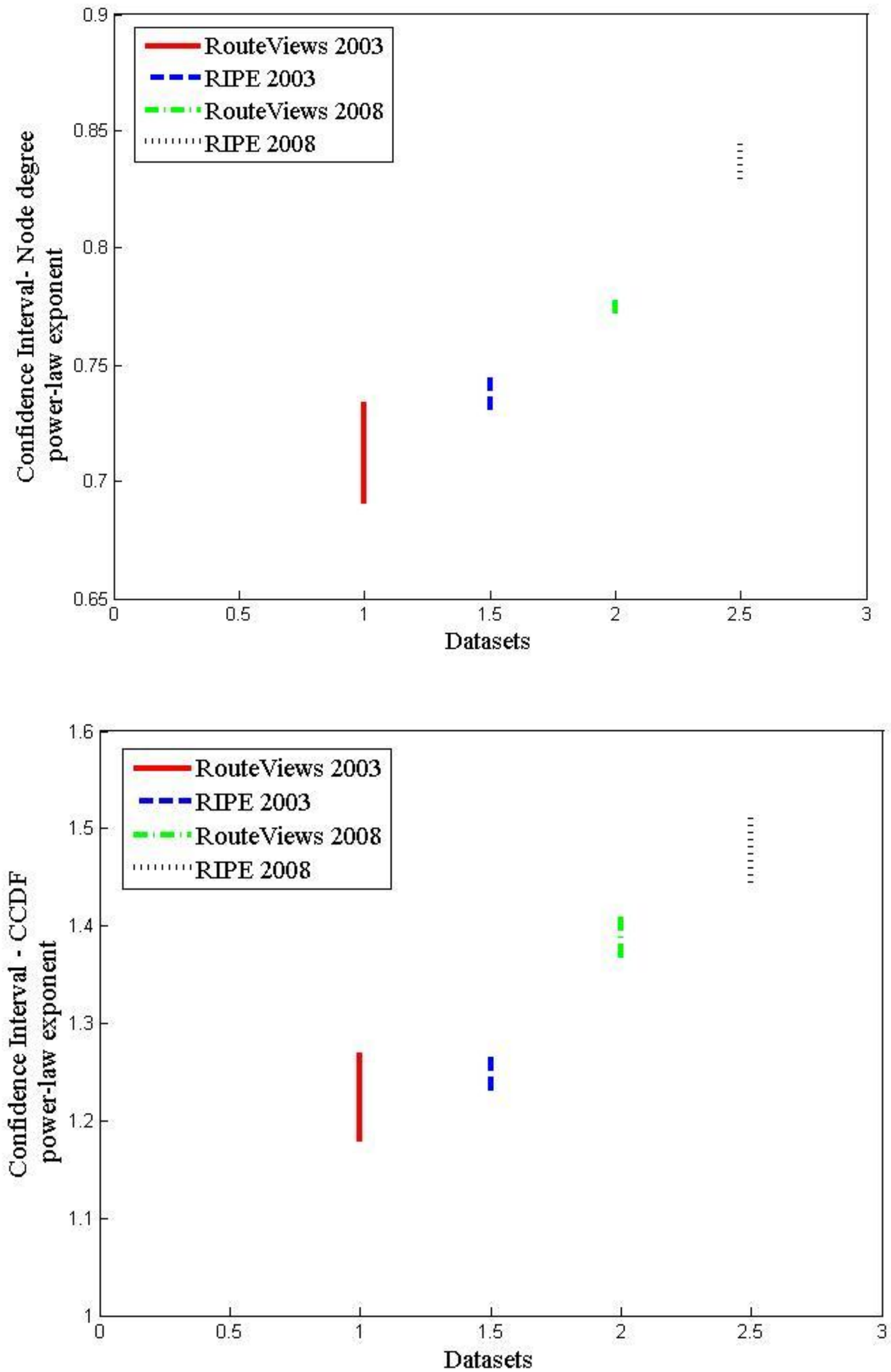
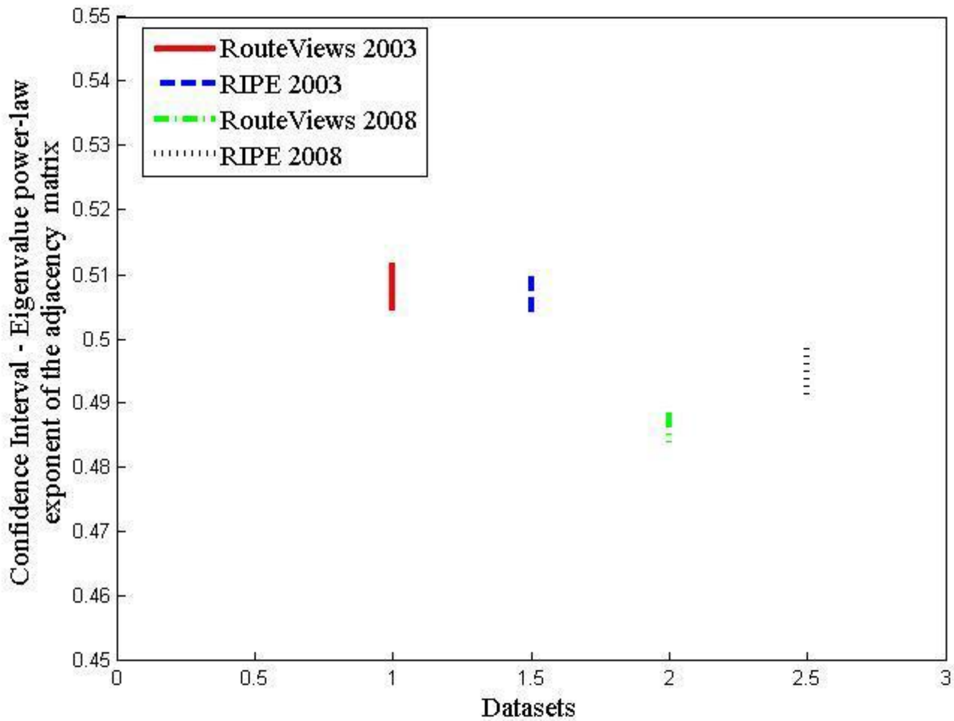


Figure 4.10 Confidence intervals: Node degree power-law exponent (top) and CCDF power-law exponent (bottom).

The plots of the confidence intervals of the eigenvalue power-law exponent based on the adjacency and the normalized Laplacian matrices for Route Views and RIPE 2003 and 2008 datasets are shown in Figure 4.11. The confidence intervals of power-law exponents for all four datasets have comparable width. The values of the eigenvalue power-law exponent based on the adjacency matrix have small difference between Route Views and RIPE 2003 and 2008 datasets. Figure 4.11 (top) reveals that the eigenvalue power-law exponents based on the adjacency matrix have not significantly changed over the last five years. Figure 4.11 (bottom) reveals that the eigenvalue power-law exponents based on the normalized Laplacian matrix are also comparable over the last five years.

The correlation coefficients of eigenvalue power-law exponents based on the adjacency matrix and the normalized Laplacian matrix for all datasets are above 99 percent and 95 percent, respectively.

In all four power-laws, confidence intervals of power-law exponents have small width and are comparable. The values of the correlation coefficients are also larger for all power-laws. This indicates the presence of four power-laws in the Internet topology over the period of five years. The small shift in the values of power-law exponents indicates that power-law exponents have not significantly changed over the last five years.



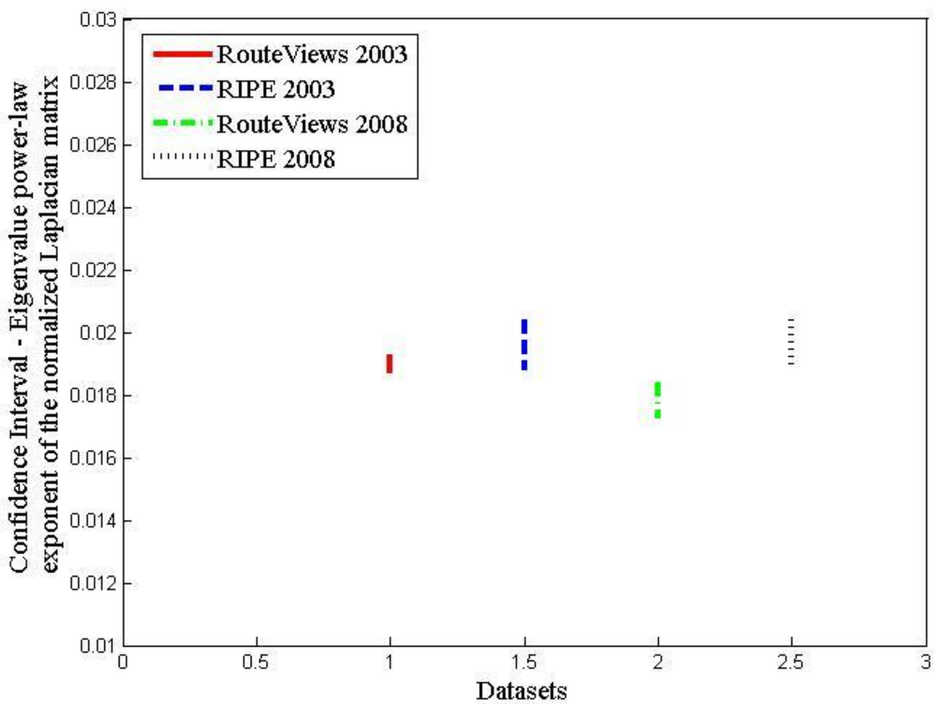


Figure 4.11 Confidence intervals: Eigenvalue power-law exponent based on the adjacency matrix (top) and based on the normalized Laplacian matrix (bottom).

5. SPECTRAL ANALYSIS AND THE INTERNET TOPOLOGY

In this Chapter, we analyze the spectrum of the adjacency matrix and the normalized Laplacian matrix of the Internet graph at AS level derived from Route Views and RIPE 2003 and 2008 datasets. We plot the patterns of connected AS nodes over the years based on the adjacency matrix. We compare the connectivity status based on the second smallest and the largest eigenvalues in order to observe the clusters of connected nodes. Finally, we analyze the elements values of the eigenvectors based on the second smallest and the largest eigenvalues.

5.1 Clusters of ASes based on the adjacency matrix

The element value of the adjacency matrix A_{xy} is 1 if nodes x and y are connected and 0 if nodes x and y are not connected. The patterns of connected AS nodes in Route Views and RIPE datasets for 2003 and 2008 are shown in Figure 5.1 and Figure 5.2, respectively. A dot in the position (x, y) in the plot of the adjacency matrix represents the connection patterns between AS nodes. No connectivity is shown between the unassigned AS nodes. The clusters are wider in case of Route Views 2008 and RIPE 2008 since the number of ASes in 2008 datasets is larger than in 2003 datasets.

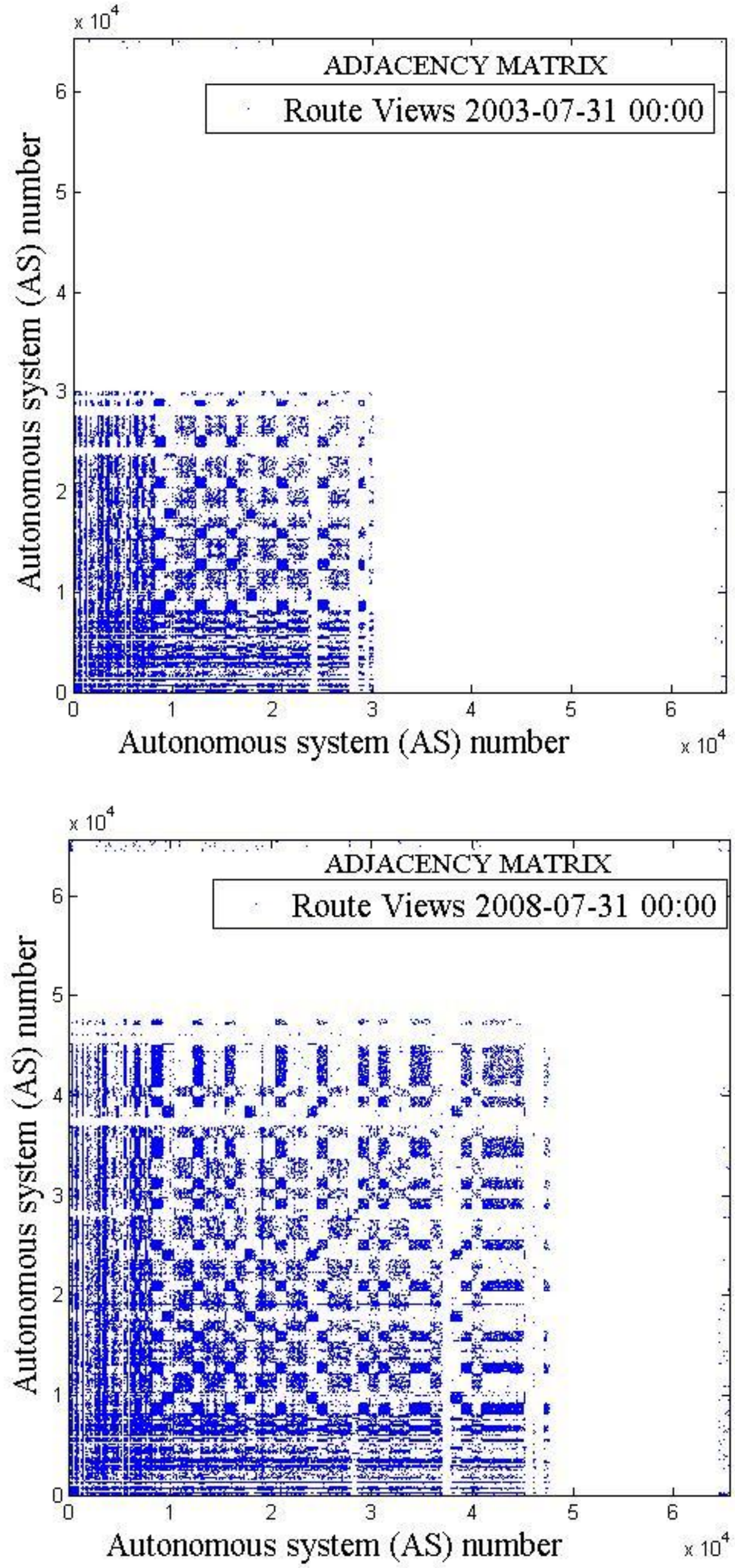
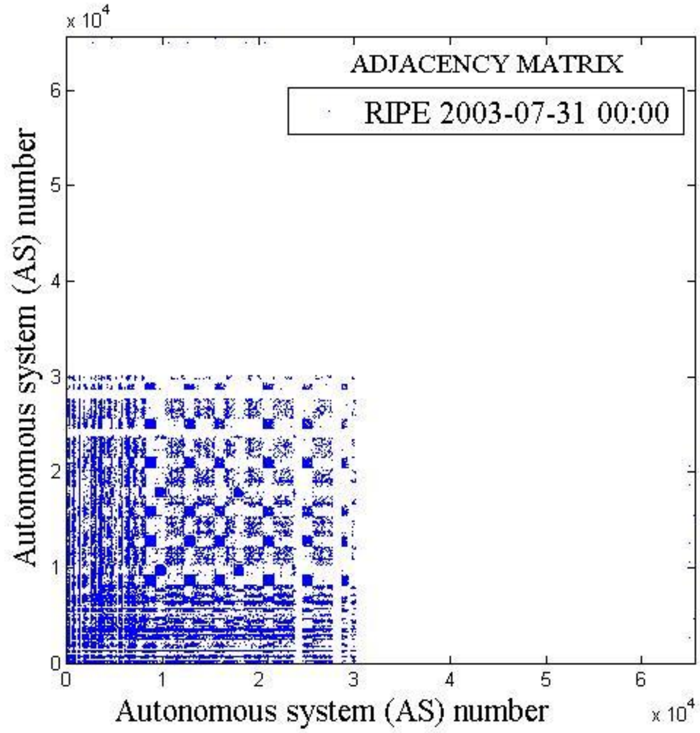


Figure 5.1 Route Views 2003 and 2008 datasets: Patterns of the adjacency matrix for Route Views 2003 (top) and 2008 (bottom) datasets. A dot in position (x, y) represents the connection between two AS nodes.



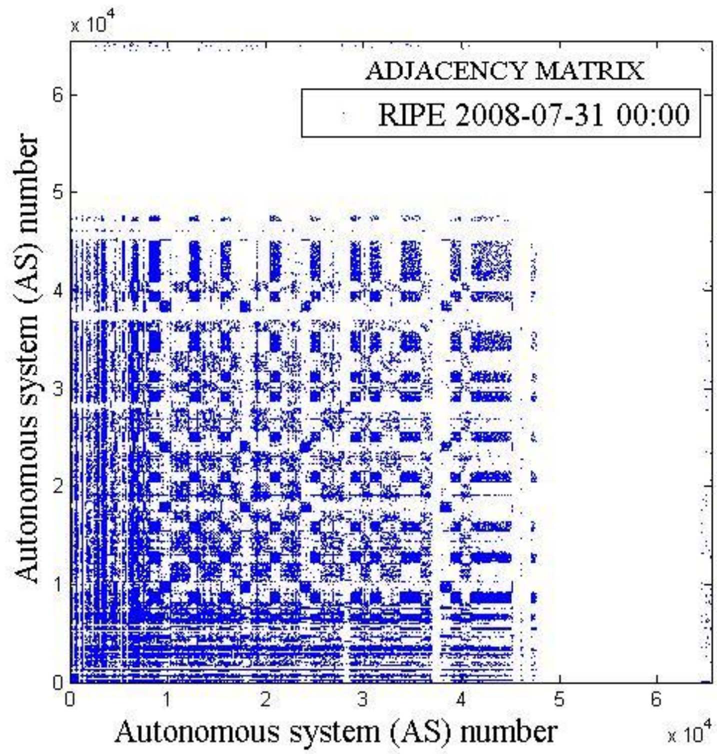


Figure 5.2 RIPE 2003 and 2008 datasets: Patterns of the adjacency matrix for RIPE 2003 (top) and 2008 (bottom) datasets. A dot in position (x, y) represents the connection between two AS nodes.

The plots depict interesting clusters associated not only with the ASes having higher node degree but also with the ASes with medium and lower node degrees. The existence of higher connectivity inside a particular cluster and relatively lower connectivity between clusters is also visible as shown in Figure 5.3. Similar patterns of clusters are observed when comparing Route Views and RIPE 2003 and 2008 datasets.

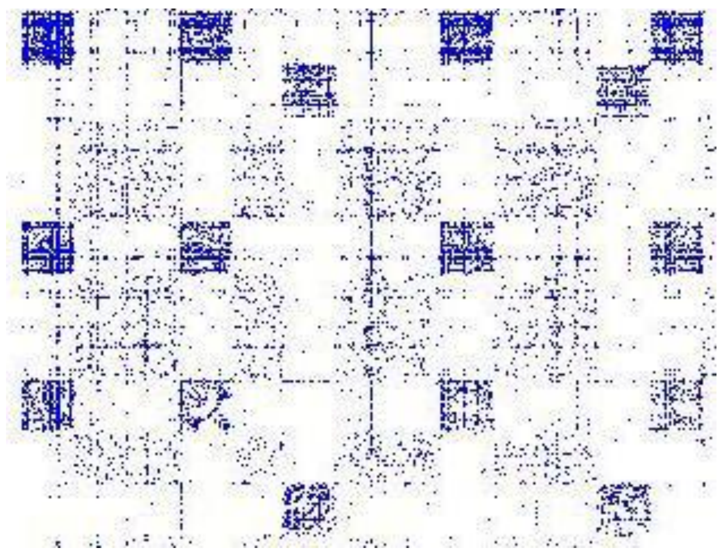


Figure 5.3 Route Views 2008: Zoomed view of the patterns of ASes based on the adjacency matrix.

5.2 Connectivity status based on the elements of the eigenvectors

In this Section, we analyze the elements of the eigenvectors corresponding to the second smallest and the largest eigenvalues of the adjacency matrix and the normalized Laplacian matrix in order to observe the clusters of ASes having similar connectivity. The second smallest eigenvalue, called *algebraic connectivity* [31], [32] of a normalized Laplacian matrix is related to the connectivity characteristic of the graph. The elements of the eigenvector corresponding to the large eigenvalue also contain information relevant to clustering [34].

In order to determine clusters of connected AS nodes in the Internet graphs, we consider the elements of the eigenvectors corresponding to the second smallest and the largest eigenvalues of the adjacency and the normalized Laplacian matrices. Each element of the eigenvector is associated with the AS node having same index in the Internet graph. All AS nodes are sorted in ascending order based on the corresponding elements values of the eigenvector. The sorted AS vector is then indexed and the connectivity status is equal to 1 if an AS is connected to another AS or zero if an AS is isolated or is not present in the datasets. We have used only assigned ASes. This sorting separates the connected nodes from the disconnected nodes and generates the clusters of connected AS nodes.

In order to create a connected graph from a matrix, let us consider a graph with six nodes [N1, N2, N3, N4, N5, N6]. Let us assume that the elements of the eigenvector corresponding to an eigenvalue of a matrix of the graph are [0.35, –0.35, –0.35, 0.41, 0.50, 0.61]. Let us assume that the node N4 is not connected and all other nodes are connected. The elements of the eigenvector are assigned to the nodes: [N1(0.35), N2(–0.35), N3(–0.35), N4(0.41), N5(0.50), N6(0.61)] and arranged in the ascending order [–0.35, –0.35, 0.35, 0.41, 0.50, 0.61]. The nodes are then sorted based on the index of the corresponding element of the eigenvector. The resulting order of nodes is [N2, N3, N1, N4, N5, N6]. The value of the connectivity status is 1 if a node is connected or 0 if a node is isolated. Thus, the value of connectivity status is 0 for node N4 and 1 for the remaining nodes. The assigned connectivity status values vs. the order of the nodes are plotted as shown in Figure 5.4. This sorting makes the ASes having similar element values stay closer generating the clusters of connected nodes.

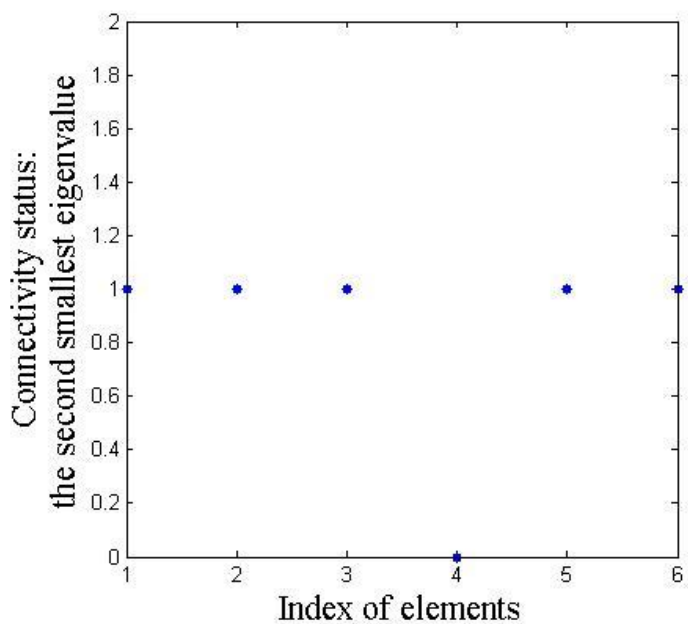


Figure 5.4 Example of connectivity status based on the eigenvector of a matrix.

5.2.1 Analysis based on the adjacency matrix

In this Section, we observe the connectivity status based on the elements of the eigenvectors corresponding to the second smallest and the largest eigenvalues of the adjacency matrix. We follow similar procedure as in Section 5.2 to plot the clusters of connected AS nodes.

The connectivity status based on the second smallest eigenvalue of the adjacency matrix is shown in Figure 5.5 and Figure 5.6 for Route Views 2003 and 2008 and RIPE 2003 and 2008 datasets, respectively. The clusters of AS nodes are similar for Route Views 2003 and RIPE 2003 datasets. Route Views 2008 datasets also reveal similar clusters of nodes to RIPE 2008 datasets. However, Figure 5.5 and Figure 5.6 indicate visible changes in the connectivity status of AS nodes while comparing the connectivity status of Route Views and RIPE 2003 datasets with Route Views and RIPE 2008 datasets.

We also calculate the elements of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix for each dataset. The connectivity status of each dataset is shown in Figure 5.7 and Figure 5.8. Similar to the connectivity status based on the second smallest eigenvalue, the connectivity status based on the largest eigenvalues for Route Views 2003 and RIPE 2003 datasets is similar. Route Views 2008 datasets also have similar connectivity status to RIPE 2008 datasets. The comparison of the connectivity status of 2003 datasets and 2008 datasets shows visible changes over the period of five years.

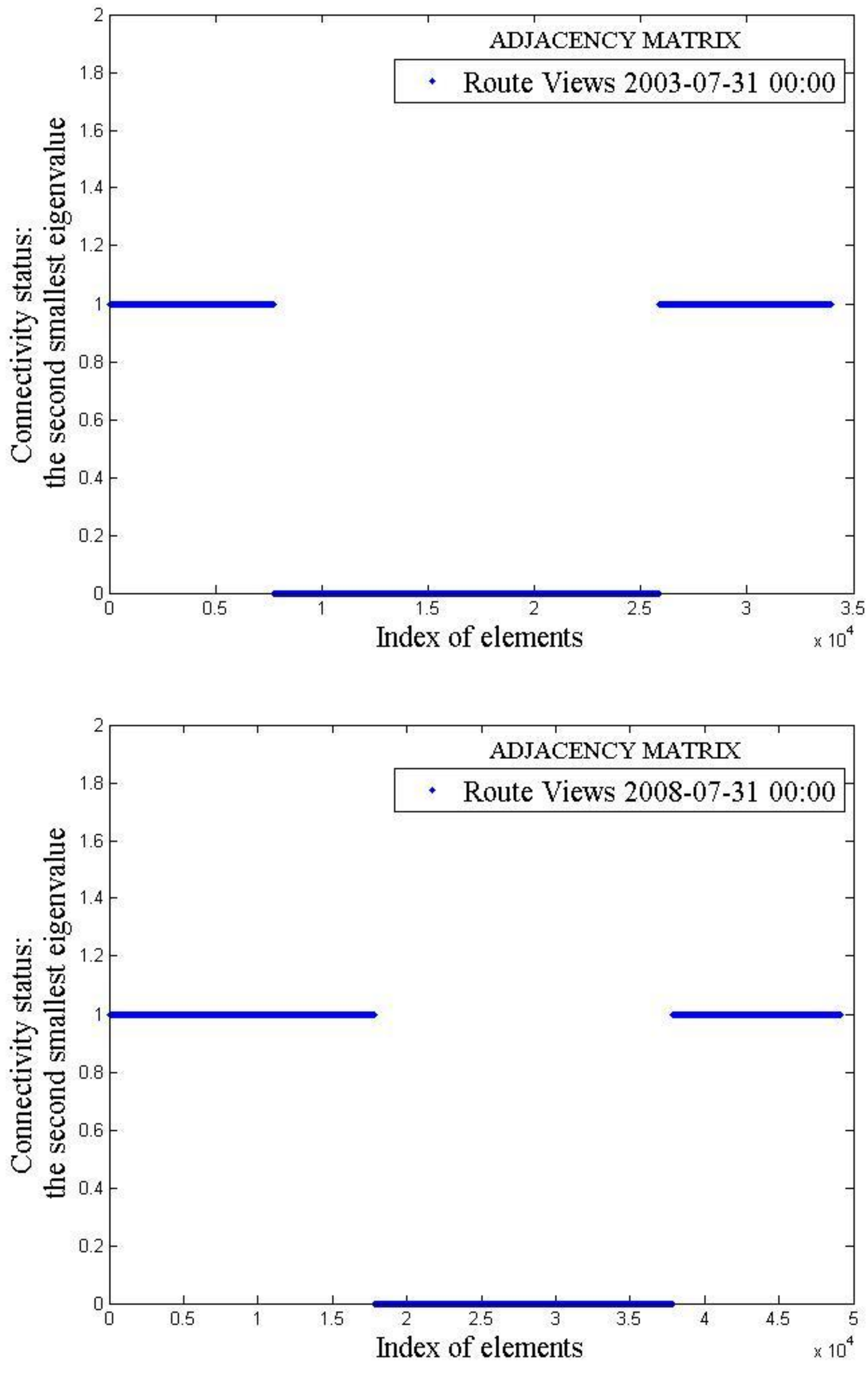


Figure 5.5 Route Views 2003 and 2008 datasets: Spectral views of the AS connectivity based on the second smallest eigenvalue of the adjacency matrix for Route Views 2003 (top) and 2008 (bottom) datasets.

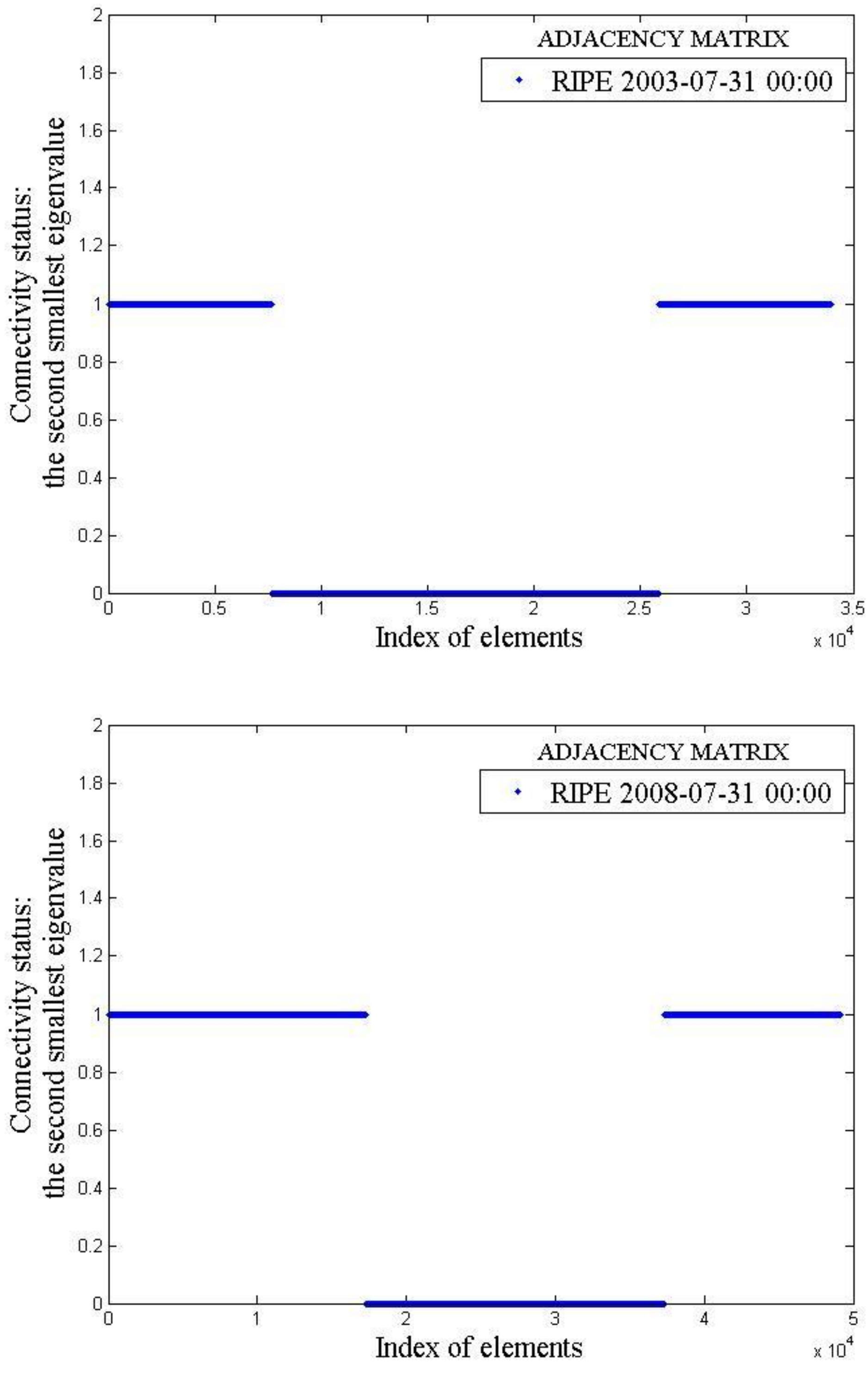
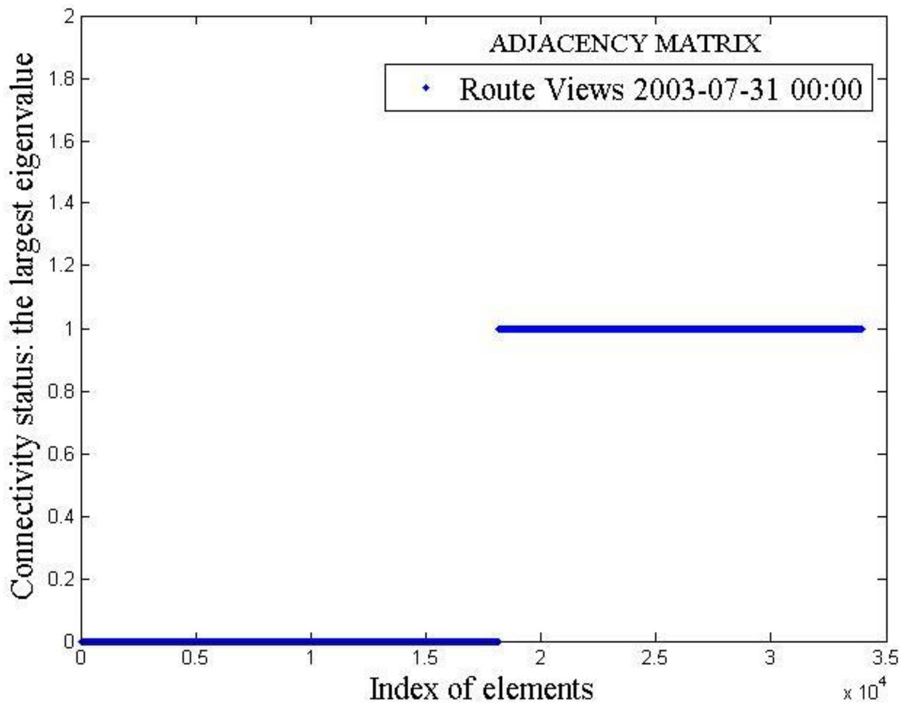


Figure 5.6 RIPE 2003 and 2008 datasets: Spectral views of the AS connectivity based on the second smallest eigenvalue of the adjacency matrix for RIPE 2003 (top) and 2008 (bottom) datasets.



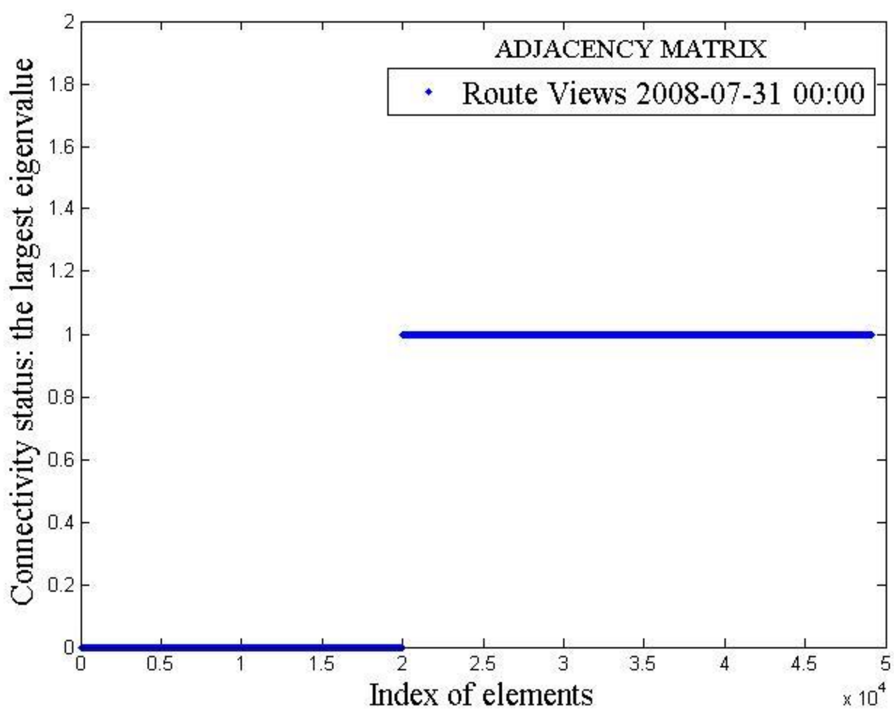
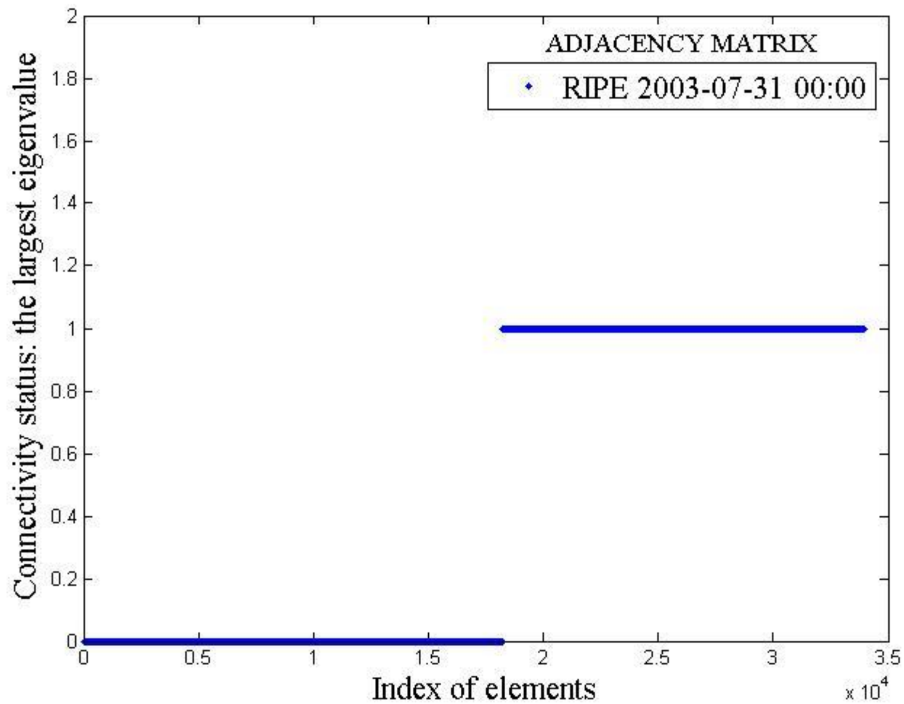


Figure 5.7 Route Views 2003 and 2008 datasets: Spectral views of the AS connectivity based on the largest eigenvalue of the adjacency matrix for Route Views 2003 (top) and 2008 (bottom) datasets.



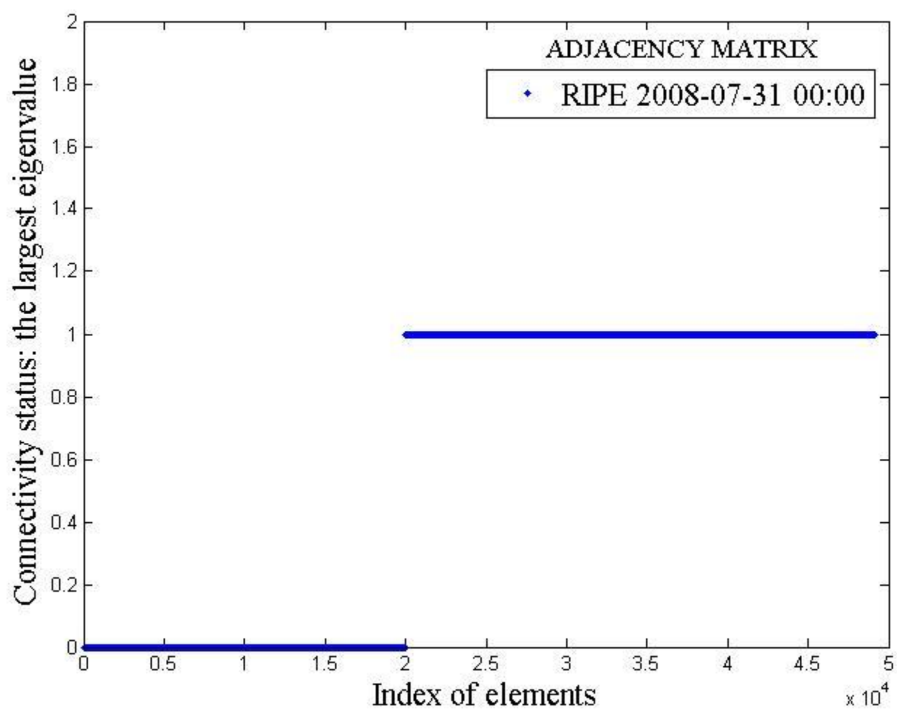


Figure 5.8 RIPE 2003 and 2008 datasets: Spectral views of the AS connectivity based on the largest eigenvalue of the adjacency matrix for RIPE 2003 (top) and 2008 (bottom) datasets.

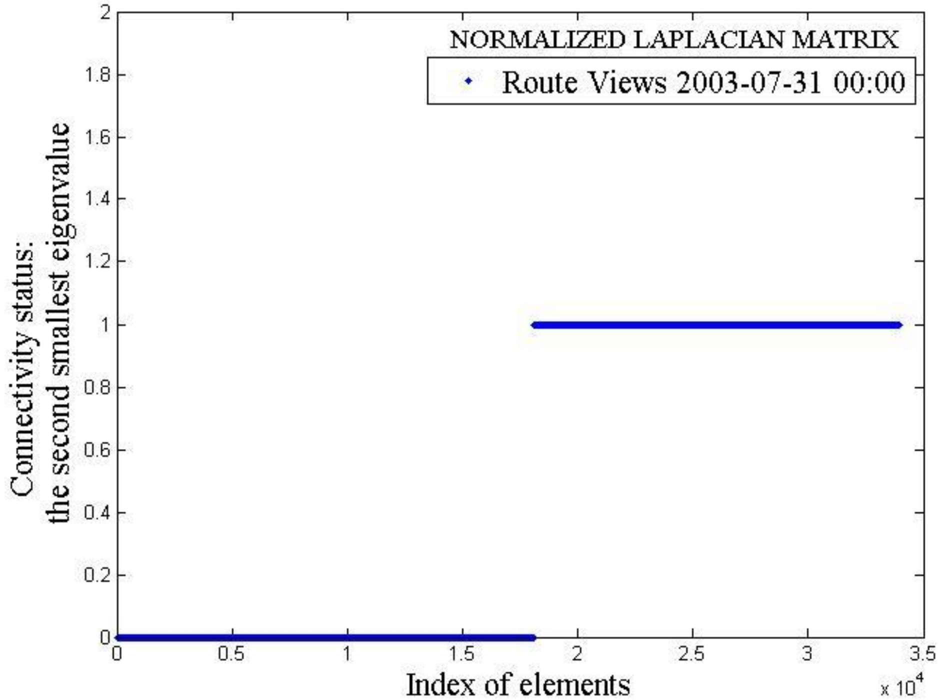
5.2.2 Analysis based on the normalized Laplacian matrix

In this Section, we observe the connectivity status based on the elements of the eigenvectors corresponding to the second smallest and the largest eigenvalues of the normalized Laplacian matrix. We follow similar procedure as in Section 5.2 to plot the clusters of connected AS nodes.

The connectivity status of AS nodes based on the second smallest eigenvalue is shown in Figure 5.9 and Figure 5.10 for Route Views 2003 and 2008 and RIPE 2003 and 2008 datasets, respectively. We observe that the connectivity status of Route Views 2003 datasets is similar to RIPE 2003 datasets. Furthermore, Route Views 2008 datasets reveal similar connectivity patterns as RIPE 2008 datasets.

The connectivity status of AS nodes based on the largest eigenvalue is shown in Figure 5.11 and Figure 5.12. Route Views 2003 datasets shows similar connectivity trends to RIPE 2003 datasets. The connectivity status of Route Views 2008 datasets is also similar to RIPE 2008 datasets. The comparison of the connectivity status of 2003 datasets to 2008 datasets shows visible changes over the last five years.

We note that the connectivity status based on the second smallest eigenvalue of the adjacency matrix is similar to the connectivity graph based on the largest eigenvalue of the normalized Laplacian matrix, and vice versa. This interesting property has its basis in the spectral properties of the two matrices since $L = D - A$, where L is the Laplacian matrix, D is the degree matrix having node degree in the diagonal, and A is the adjacency matrix.



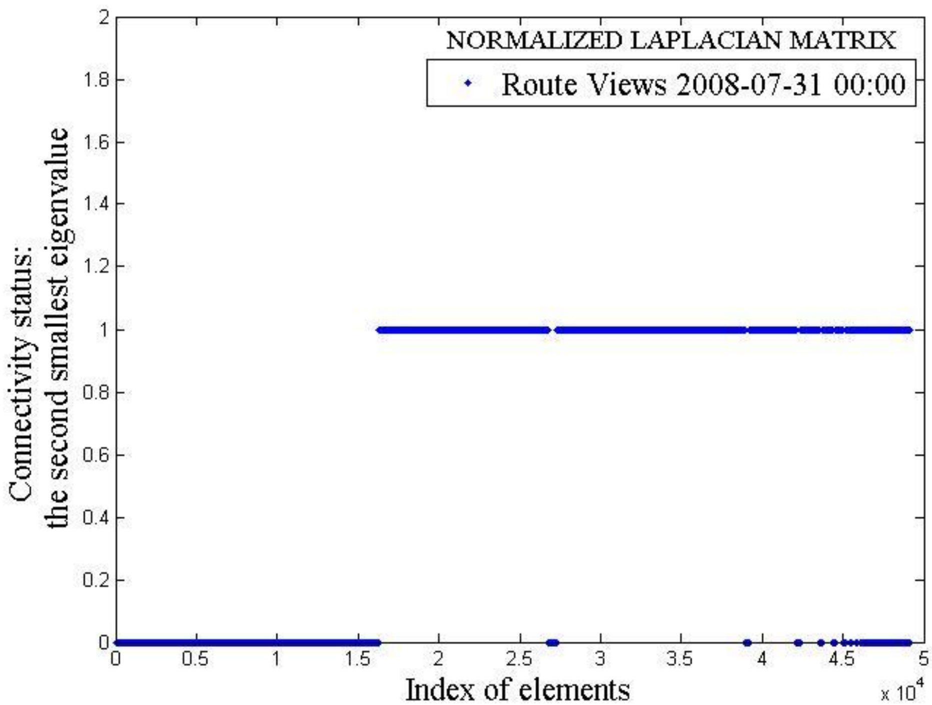


Figure 5.9 Route Views 2003 and 2008 datasets: Spectral views of the AS connectivity based on the second smallest eigenvalue of the normalized Laplacian matrix for Route Views 2003 (top) and 2008 (bottom) datasets.

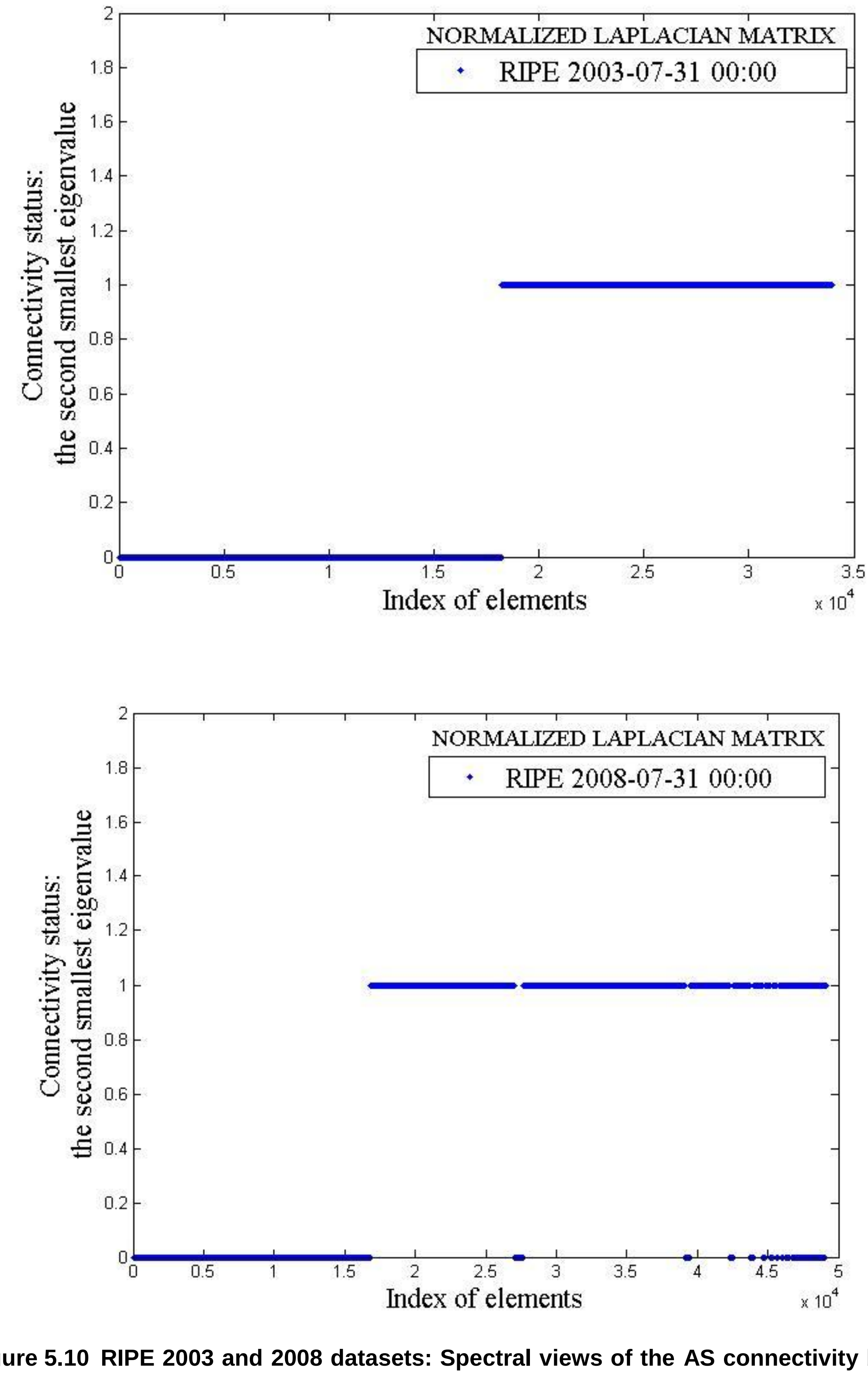
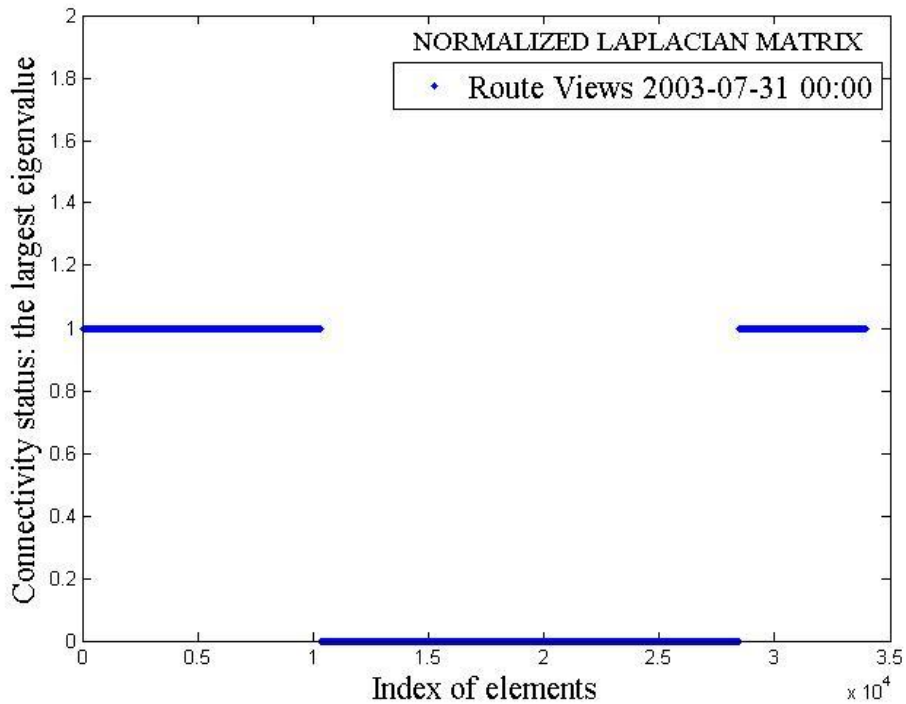


Figure 5.10 RIPE 2003 and 2008 datasets: Spectral views of the AS connectivity based on the second smallest eigenvalue of the normalized Laplacian matrix for RIPE 2003 (top) and 2008 (bottom) datasets.



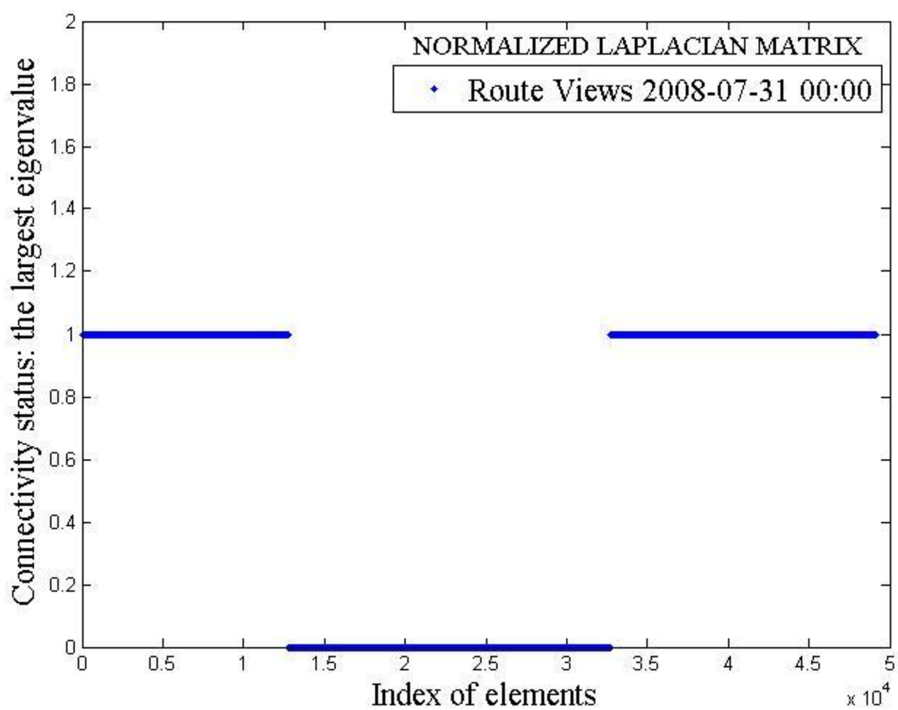


Figure 5.11 Route Views 2003 and 2008 datasets: Spectral views of the AS connectivity based on the largest eigenvalue of the normalized Laplacian matrix for Route Views 2003 (top) and 2008 (bottom) datasets.

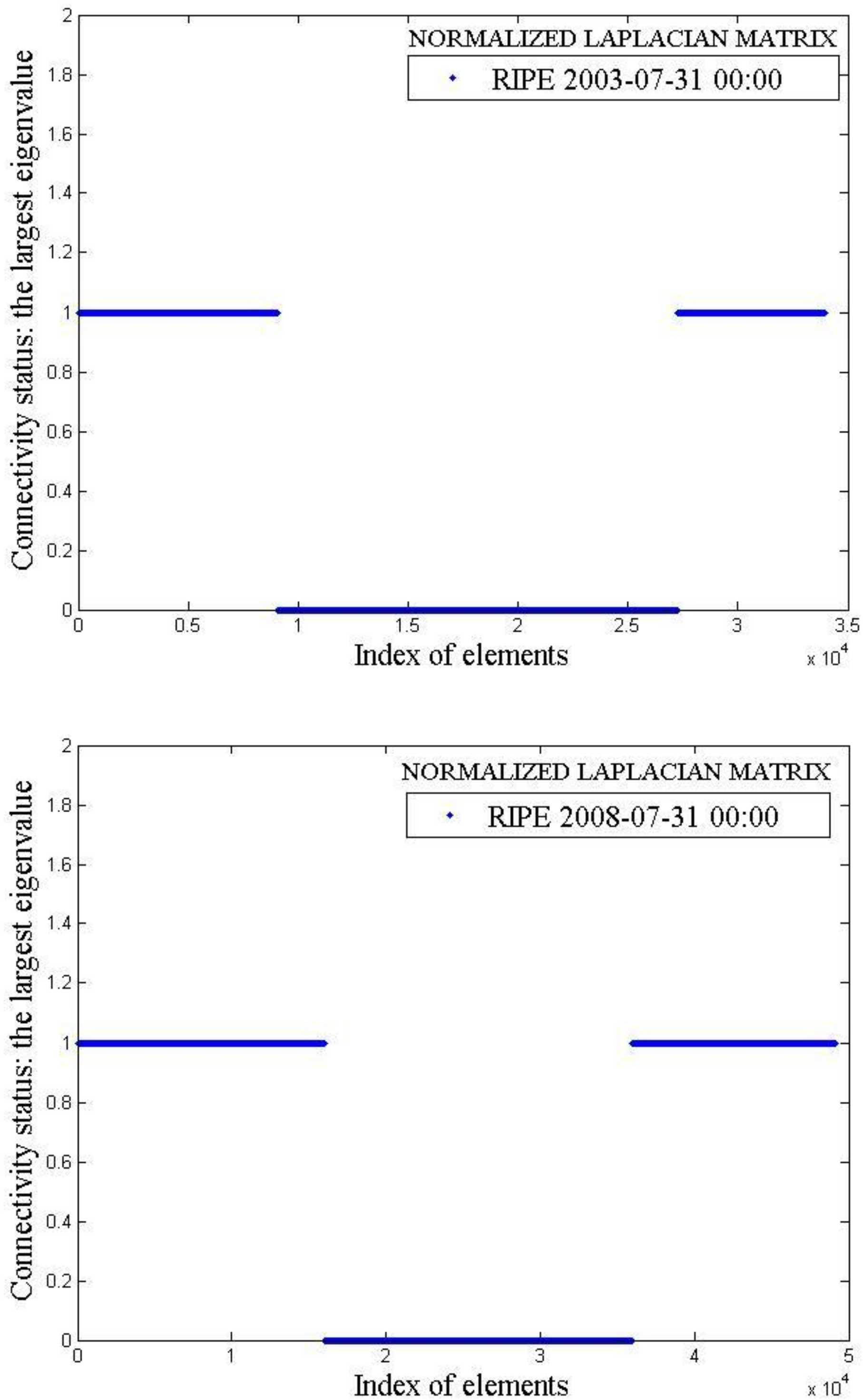


Figure 5.12 RIPE 2003 and 2008 datasets: Spectral views of the AS connectivity based on the largest eigenvalue of the normalized Laplacian matrix for RIPE 2003 (top) and 2008 (bottom) datasets.

5.3 Clusters of ASes based on the elements of eigenvectors

When we examine the elements of the eigenvectors corresponding to the second smallest and the largest eigenvalues of the adjacency matrix of a small world network with 20 nodes [55], we observe that the nodes having similar degrees are grouped together based on the element values of the eigenvector corresponding to the largest eigenvalue. We first calculate the elements of the eigenvectors corresponding to the second smallest and the largest eigenvalues. We then sort the elements in descending order and plot them vs. order as shown in Figure 5.13. We also calculate the index of node based on the index of the corresponding element of the eigenvector. We then plot the node degree of a node vs. the index of the node as shown in Figure 5.14.

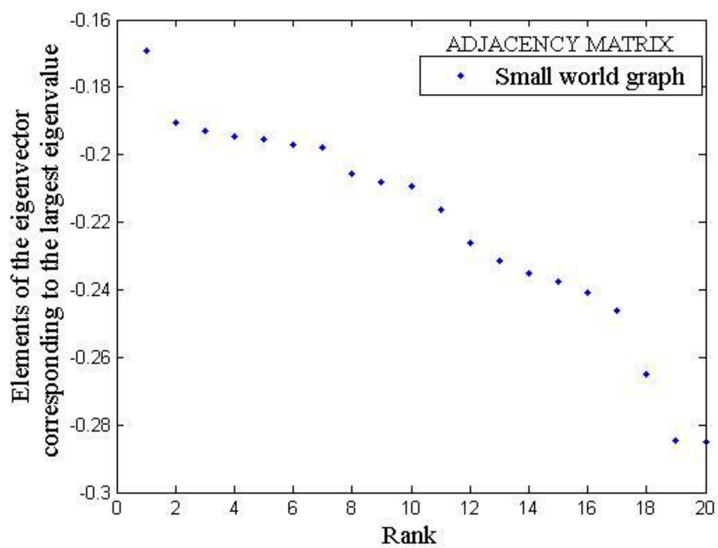


Figure 5.13 Small world graph: Elements of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix sorted in decreasing order.

The values of the elements of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix group nodes having similar node degrees.

The group of nodes having small node degree appear first, followed by the nodes having larger node degrees, as shown in Figure 5.14. No such grouping is observed based on the elements values of the eigenvector corresponding to the second smallest eigenvalue.

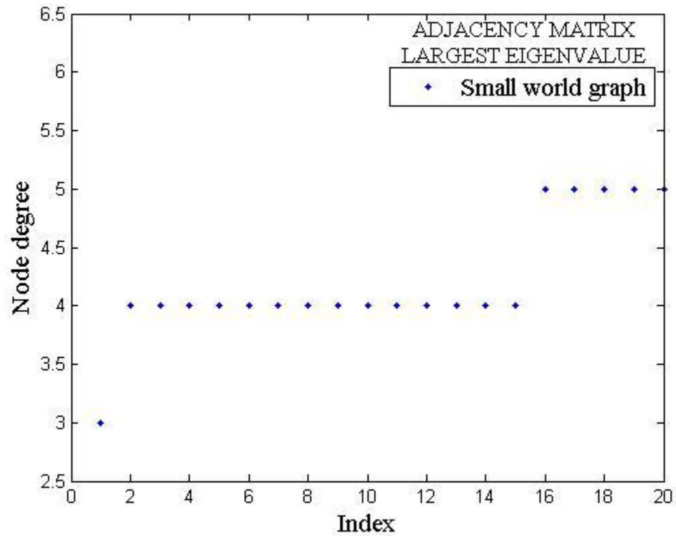


Figure 5.14 Small world graph: Groups of connected nodes based on the elements values of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix.

In search of such clusters of connected AS nodes in the Internet graphs, we examine the elements of the eigenvectors corresponding to the second smallest and the largest eigenvalues of both the adjacency and the normalized Laplacian matrices.

5.3.1 Analysis based on the adjacency matrix

In this Section, we observe the cluster of ASes based on the elements of the eigenvectors corresponding to the second smallest and the largest eigenvalues of the adjacency matrix. The elements of the eigenvectors are sorted in descending order based on the weight of each element and plotted vs. the order. AS nodes are also sorted based on the corresponding elements values of the eigenvector to determine the index of the AS node. The node degree of AS node vs. the index of the AS are then plotted to observe the clusters of nodes.

The elements of eigenvector corresponding to the second smallest eigenvalue of the adjacency matrix are plotted vs. the order as shown in Figure 5.15. Only nodes on the lowest and the highest ends of the rank spectrum are shown. Majority of the nodes ranked in between belong to a cluster that corresponds to similar element value of the eigenvector. Figure 5.15 indicates that very few nodes have large element values of the eigenvector corresponding to the second smallest eigenvalue and comparatively large numbers of nodes have small element values. Route Views datasets reveals that the number of nodes having larger element values of the eigenvector is higher in 2008 in comparison to Route Views and RIPE 2003 datasets. However, the elements values of the eigenvector for RIPE 2008 are comparatively smaller to that for Route Views 2008 datasets. The lowest end indicates that all four datasets have similar element values.

The node degree of each AS node plotted vs. the index of the AS based on the adjacency matrix is shown in Figure 5.16 and Figure 5.17 for Route Views 2003 and 2008 and RIPE 2003 and 2008 datasets, respectively. The element values of the eigenvector corresponding to the second smallest eigenvalue of the adjacency matrix divide nodes into two separate clusters of connected nodes.

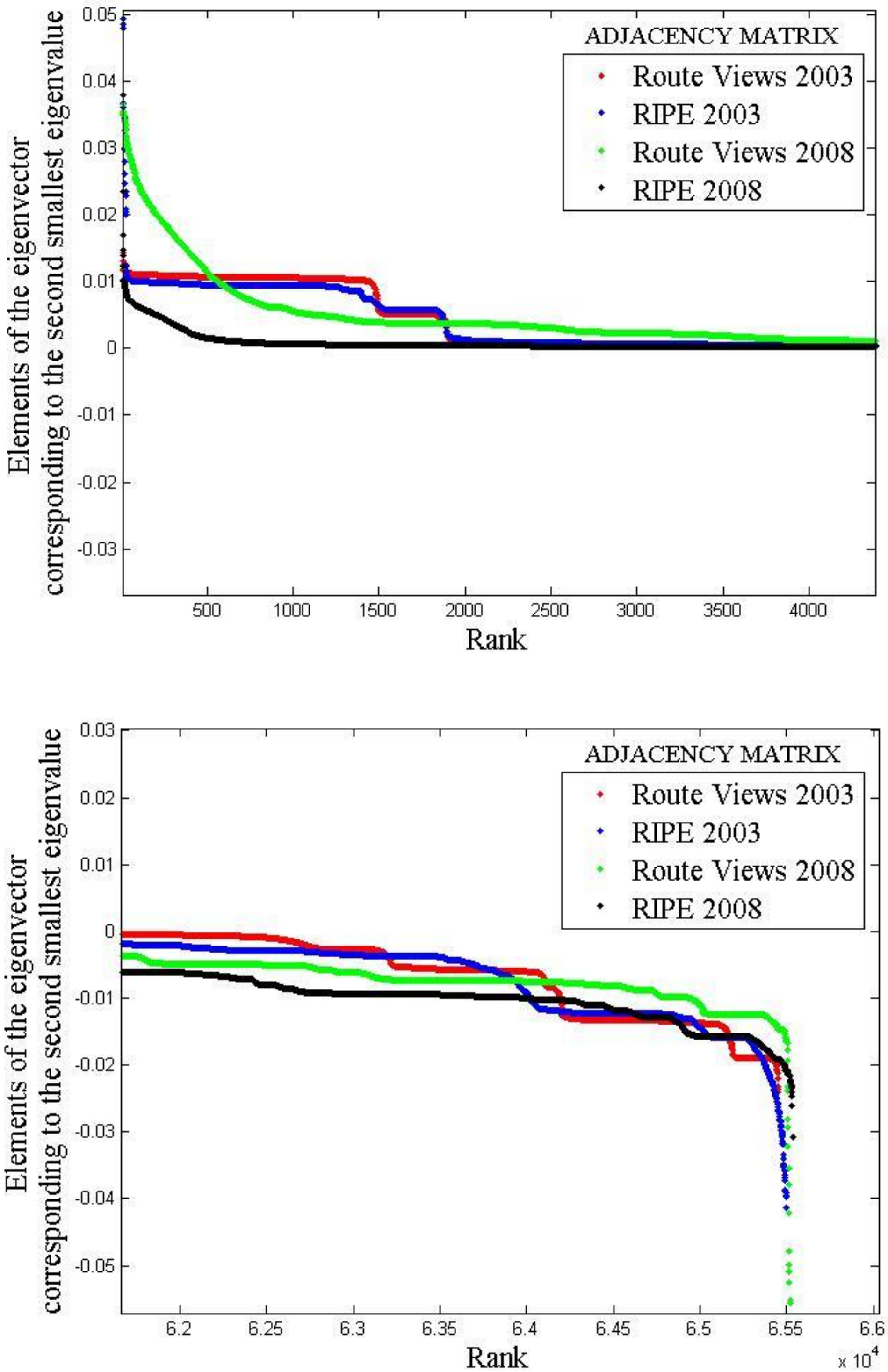
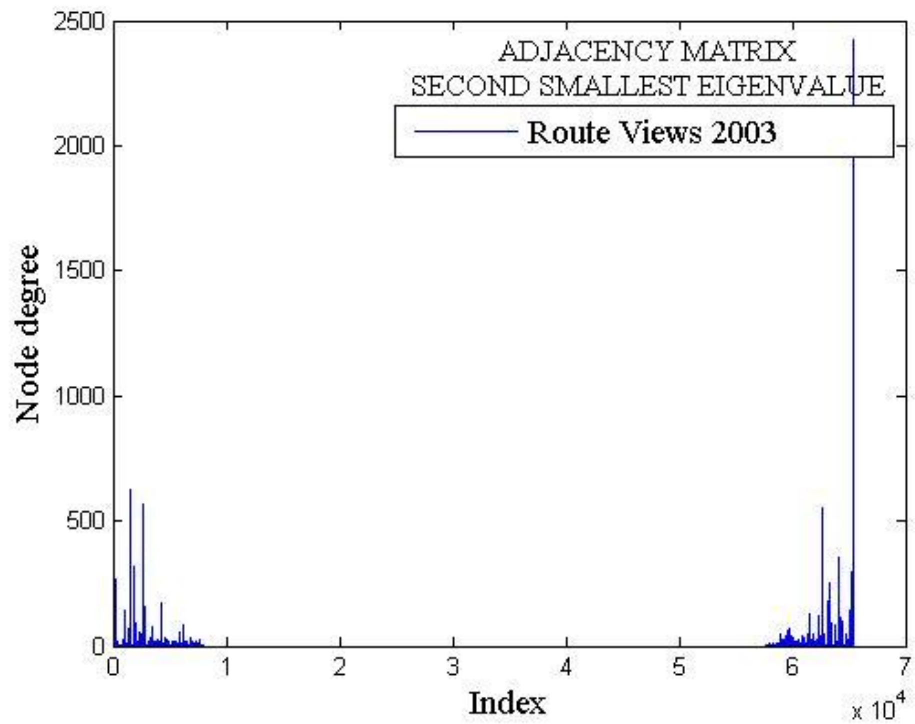


Figure 5.15 Route Views and RIPE 2003 and 2008 datasets: Elements of eigenvectors corresponding to the second smallest eigenvalue of the adjacency matrix. Shown are the nodes at the smallest (top) and the largest (bottom) ends of the rank spectrum.



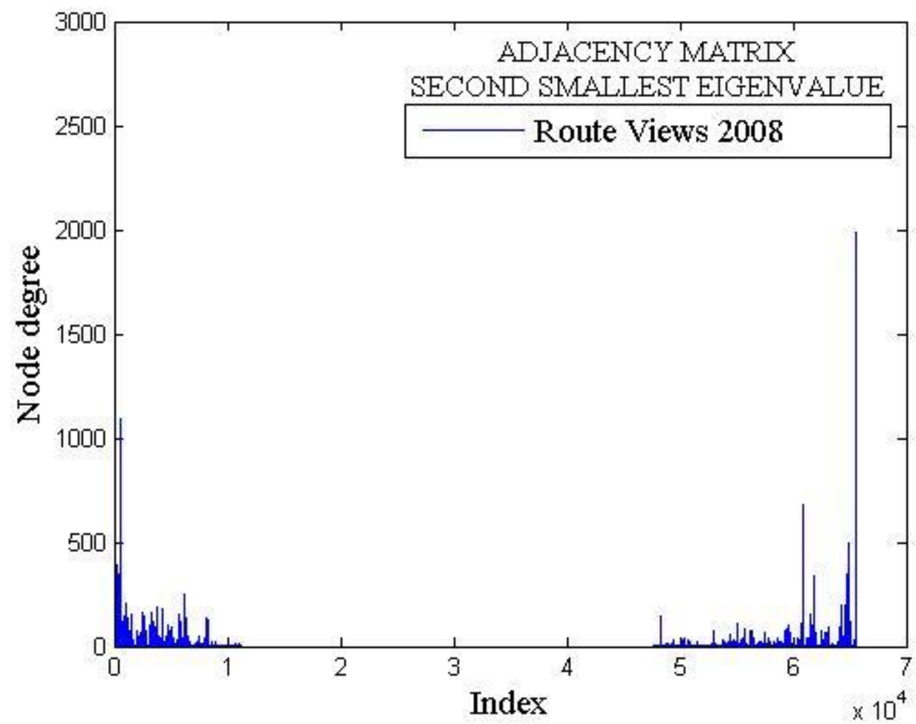


Figure 5.16 Route Views 2003 and 2008 datasets: Clusters of connected nodes based on the elements values of the eigenvector corresponding to the second smallest eigenvalue of the adjacency matrix for Route Views 2003 (top) and 2008 (bottom) datasets.

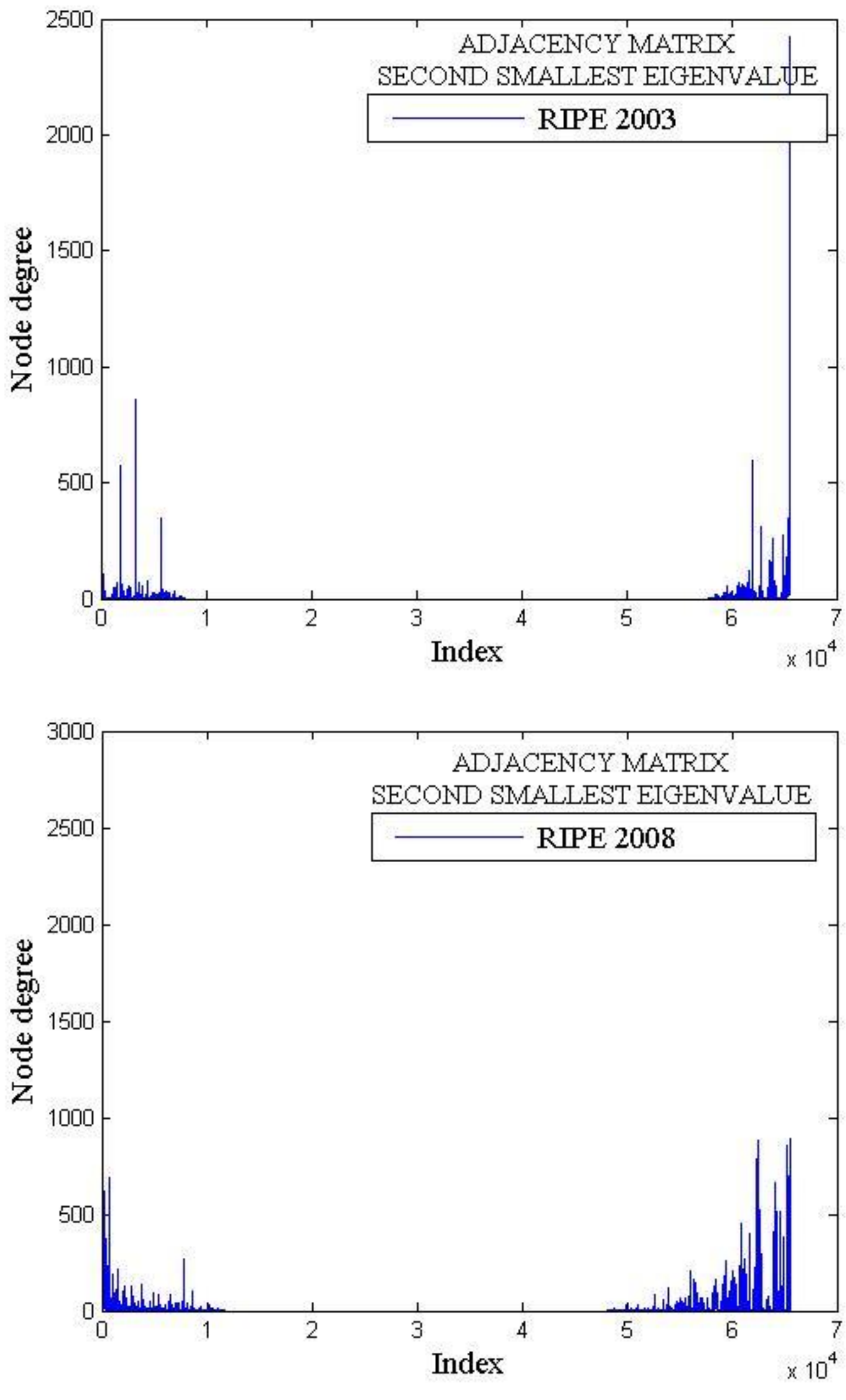


Figure 5.17 RIPE 2003 and 2008 datasets: Clusters of connected nodes based on the elements values of the eigenvector corresponding to the second smallest eigenvalue of the adjacency matrix for RIPE 2003 (top) and 2008 (bottom) datasets.

The elements of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix are arranged in decreasing order and are plotted vs. the order as shown in Figure 5.18 (top). The plots reveal that for large number of nodes the elements of the eigenvector have very small values in the range of 10^{-19}. Only few nodes at the highest end of the rank spectrum have comparatively large negative elements values. The values of the elements at the highest ends of the rank spectrum are visible in Figure 5.18 (bottom).

In order to observe the clusters of AS nodes based on the element values, the index of AS nodes are identified that corresponds to the index of the element values of the eigenvector arranged in descending order. The node degree of AS node vs. the index of the AS based on the adjacency matrix are plotted as shown in Figure 5.19 and Figure 5.20 for Route Views 2003 and 2008 and RIPE 2003 and 2008 datasets, respectively. The elements values of the eigenvector corresponding to the largest eigenvalue also separate nodes into a cluster of connected nodes. The clusters are observed at the highest end of the rank spectrum for each dataset. The length of the cluster is comparable for Route Views and RIPE 2003 datasets. Route Views and RIPE 2008 datasets also have similar length of clusters. However, the length of the cluster is smaller for 2003 datasets. This is due to the smaller number of assigned ASes in 2003 than in 2008.

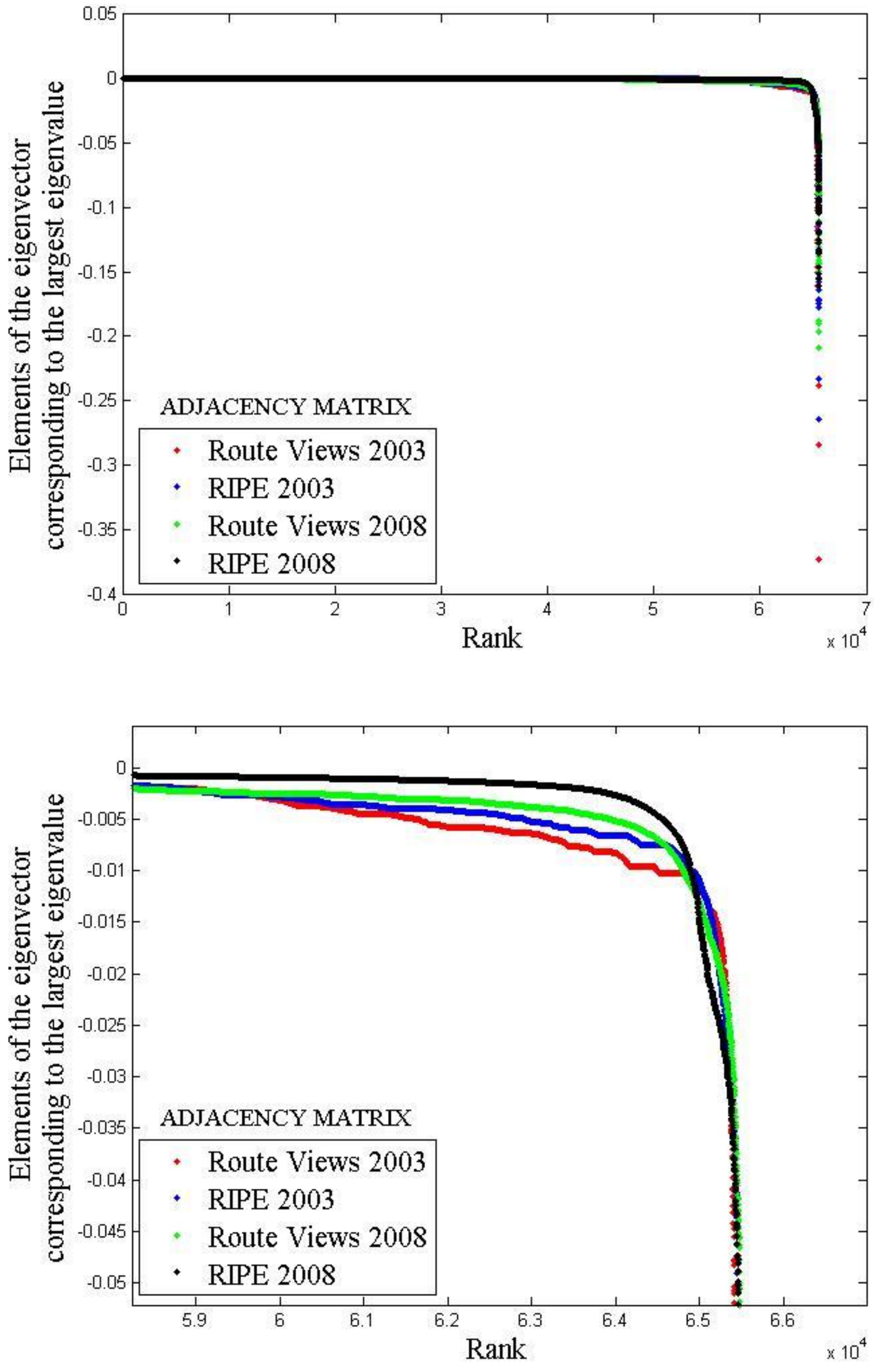


Figure 5.18 Route Views and RIPE 2003 and 2008 datasets: Elements of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix. Shown are the nodes at all (top) and the highest (bottom) ends of the rank spectrum.

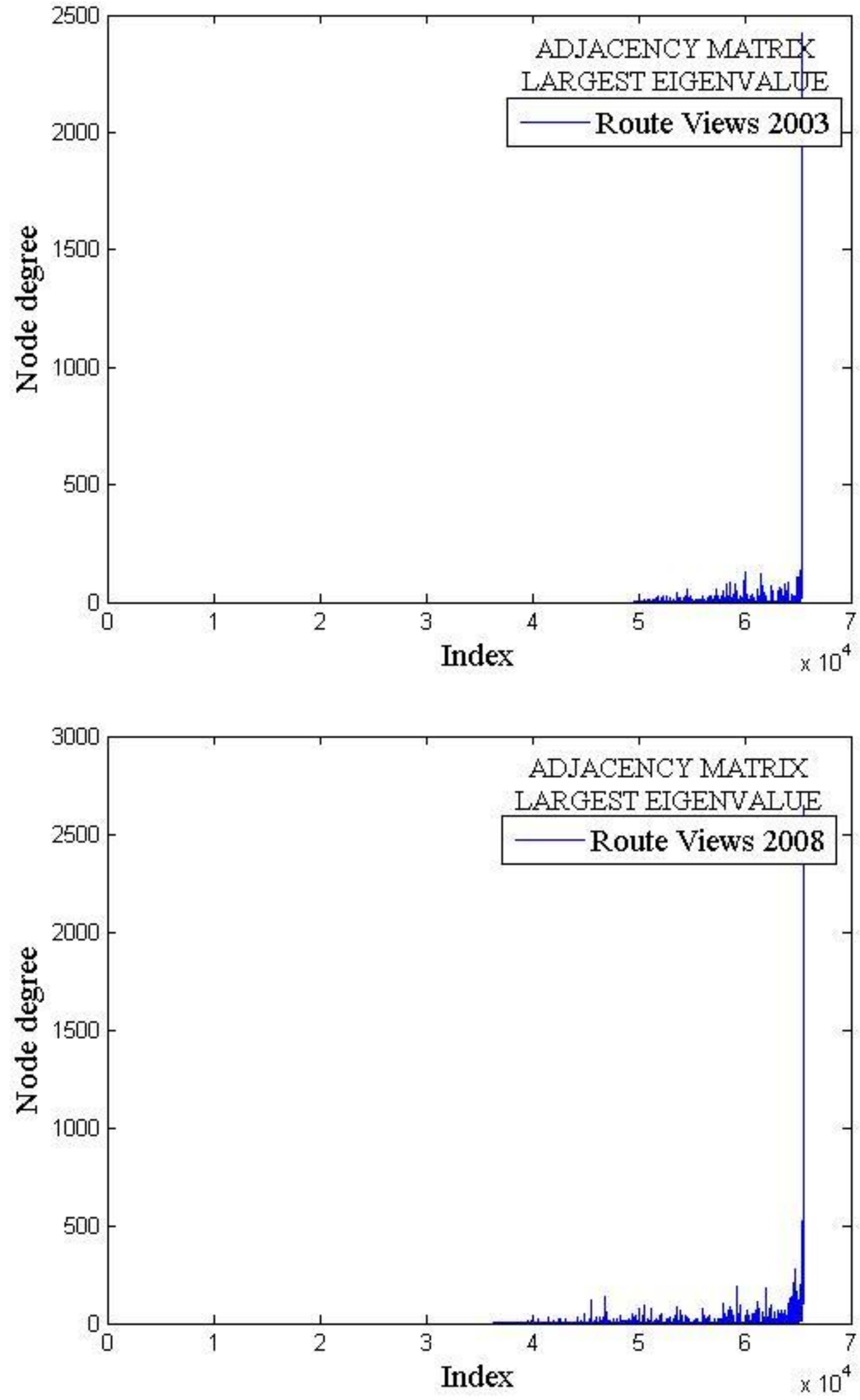


Figure 5.19 Route Views 2003 and 2008 datasets: Clusters of connected nodes based on the elements values of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix for Route Views 2003 (top) and 2008 (bottom) datasets.

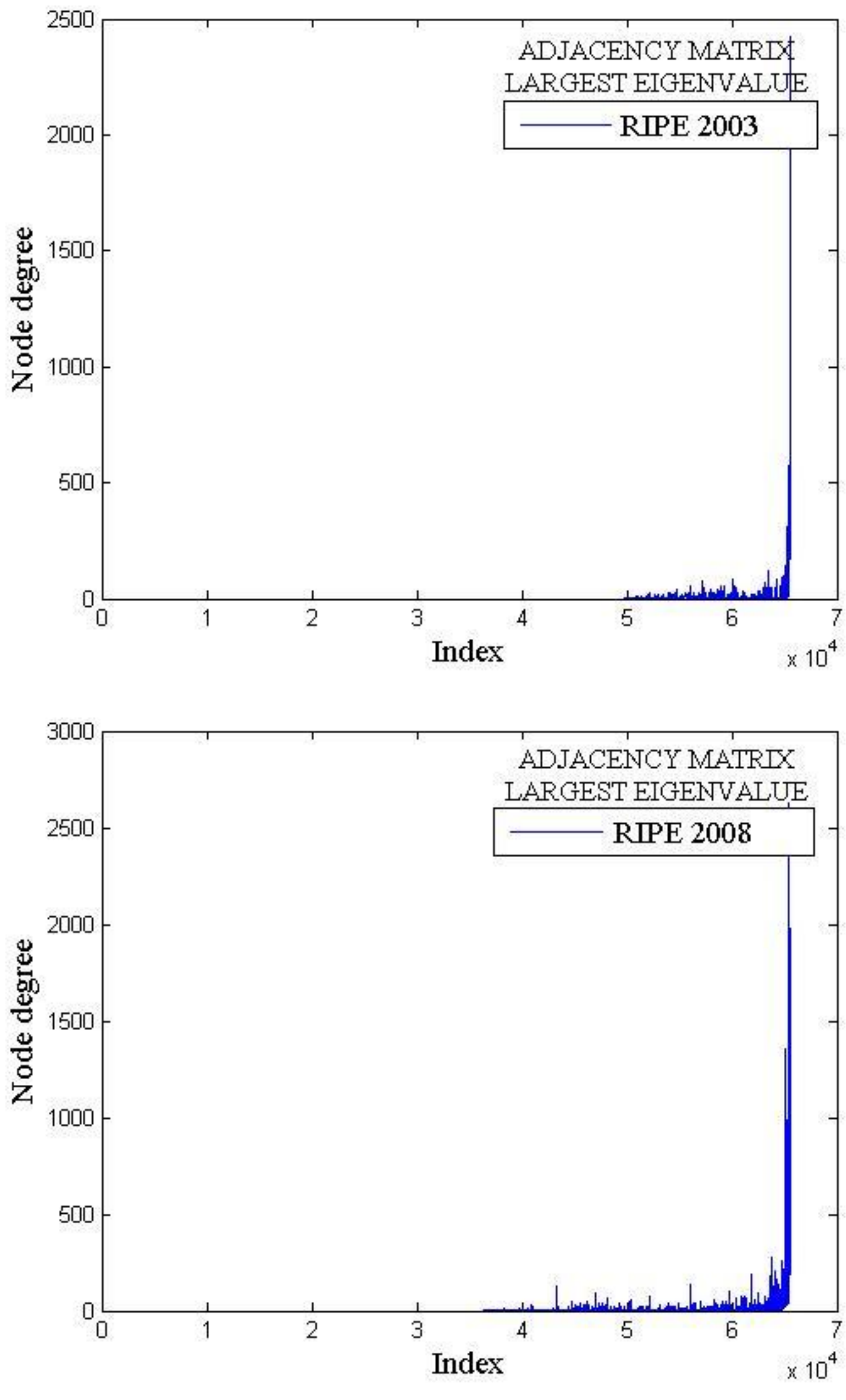


Figure 5.20 RIPE 2003 and 2008 datasets: Clusters of connected nodes based on the elements values of the eigenvector corresponding to the largest eigenvalue of the adjacency matrix for RIPE 2008 (top) and 2008 (bottom) datasets.

5.3.2 Analysis based on the normalized Laplacian matrix

We also examine the elements of the eigenvectors corresponding to the second smallest and the largest eigenvalues of the normalized Laplacian matrix. We adopt similar sorting method as in Section 5.3.1.

The elements of the eigenvector corresponding to the second smallest eigenvalue of the normalized Laplacian matrix are arranged in descending order and plotted vs. the index as shown in Figure 5.21. The plot reveals the clustering behaviour of the elements values similar to the elements values corresponding to the adjacency matrix. Majority of the nodes ranked in between have similar element values. Furthermore, few nodes have large element values.

The node degree of AS node plotted vs. the index of the AS based on the element value of the eigenvector corresponding to second smallest eigenvalue of the normalized Laplacian for Route Views 2003 and 2008 and RIPE 2003 and 2008 datasets are shown in Figure 5.22 and Figure 5.23, respectively. The cluster of connected nodes is present towards the highest end of the rank spectrum. The small elements values of the eigenvector corresponding to the second smallest eigenvalue of the normalized Laplacian matrix correspond to the connected nodes forming a cluster. The nodes having similar node degrees are grouped together within the cluster. Furthermore, the nodes having small node degrees appear in the highest end of the rank spectrum followed by the nodes having higher node degrees as shown in Figure 5.24.

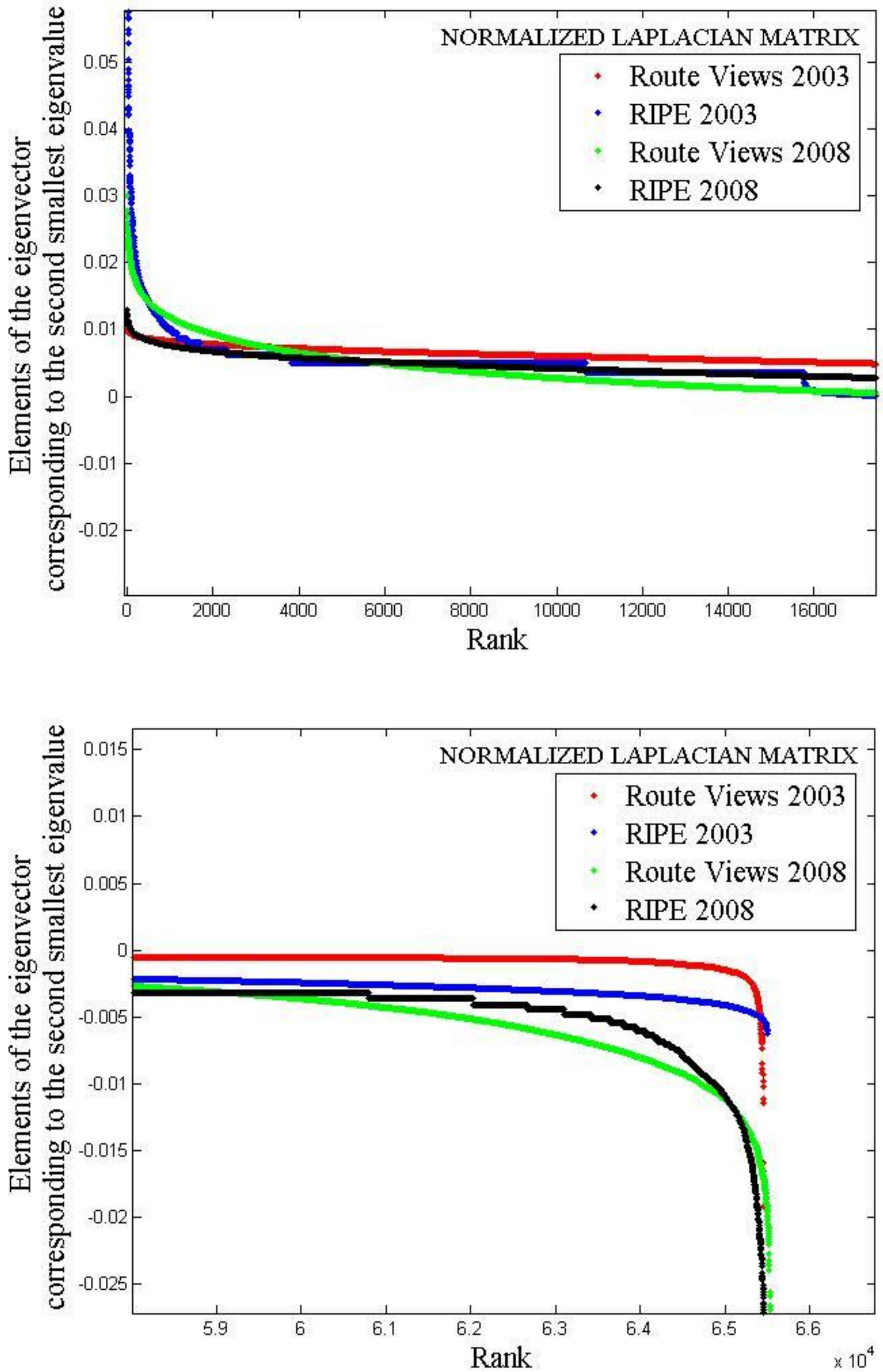


Figure 5.21 Route Views and RIPE 2003 and 2008 datasets: Elements of the eigenvector corresponding to the second smallest eigenvalue of the normalized Laplacian matrix. Shown are the nodes at the lowest (top) and the highest (bottom) ends of the rank spectrum.

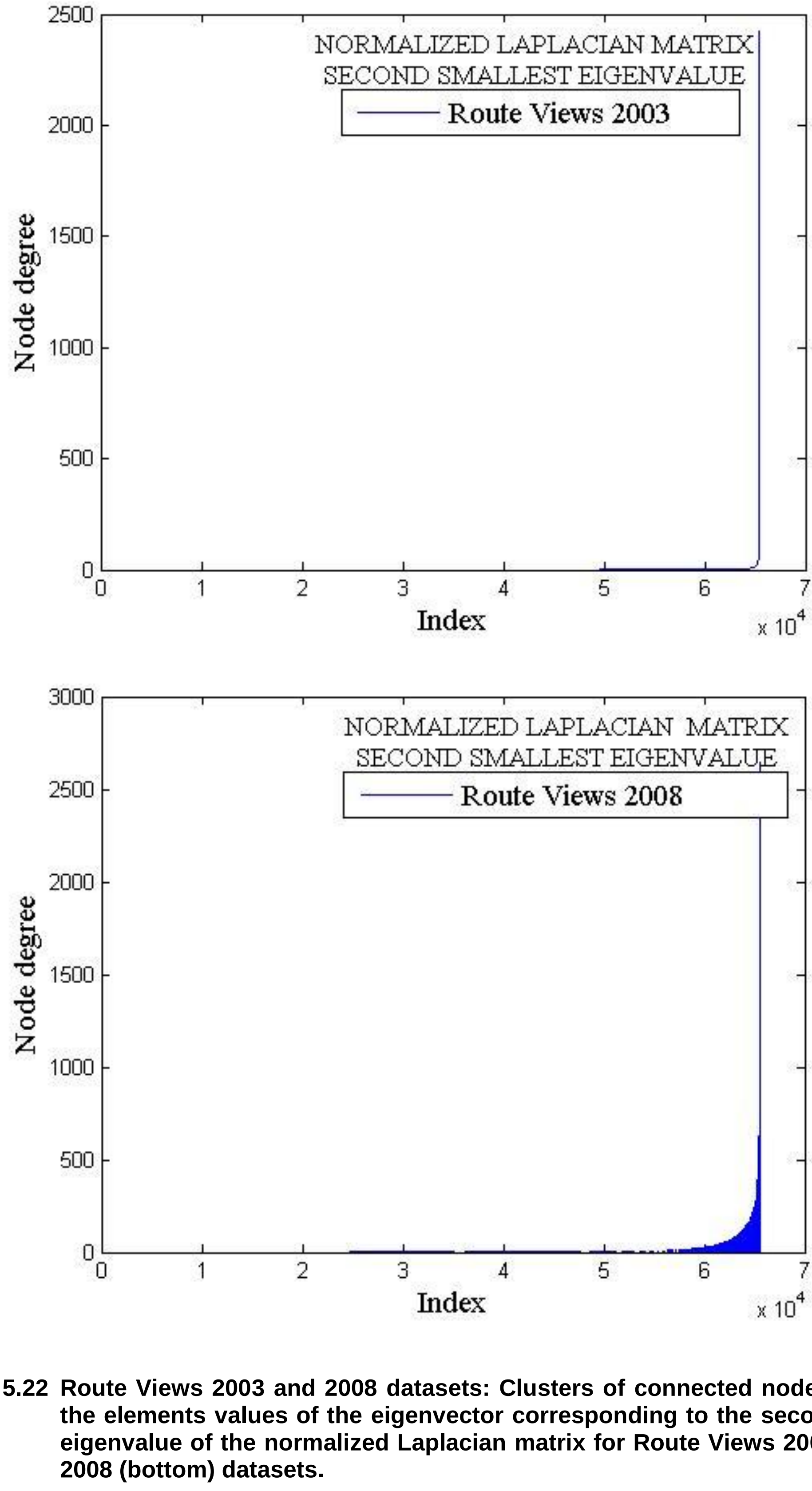


Figure 5.22 Route Views 2003 and 2008 datasets: Clusters of connected nodes based on the elements values of the eigenvector corresponding to the second smallest eigenvalue of the normalized Laplacian matrix for Route Views 2003 (top) and 2008 (bottom) datasets.

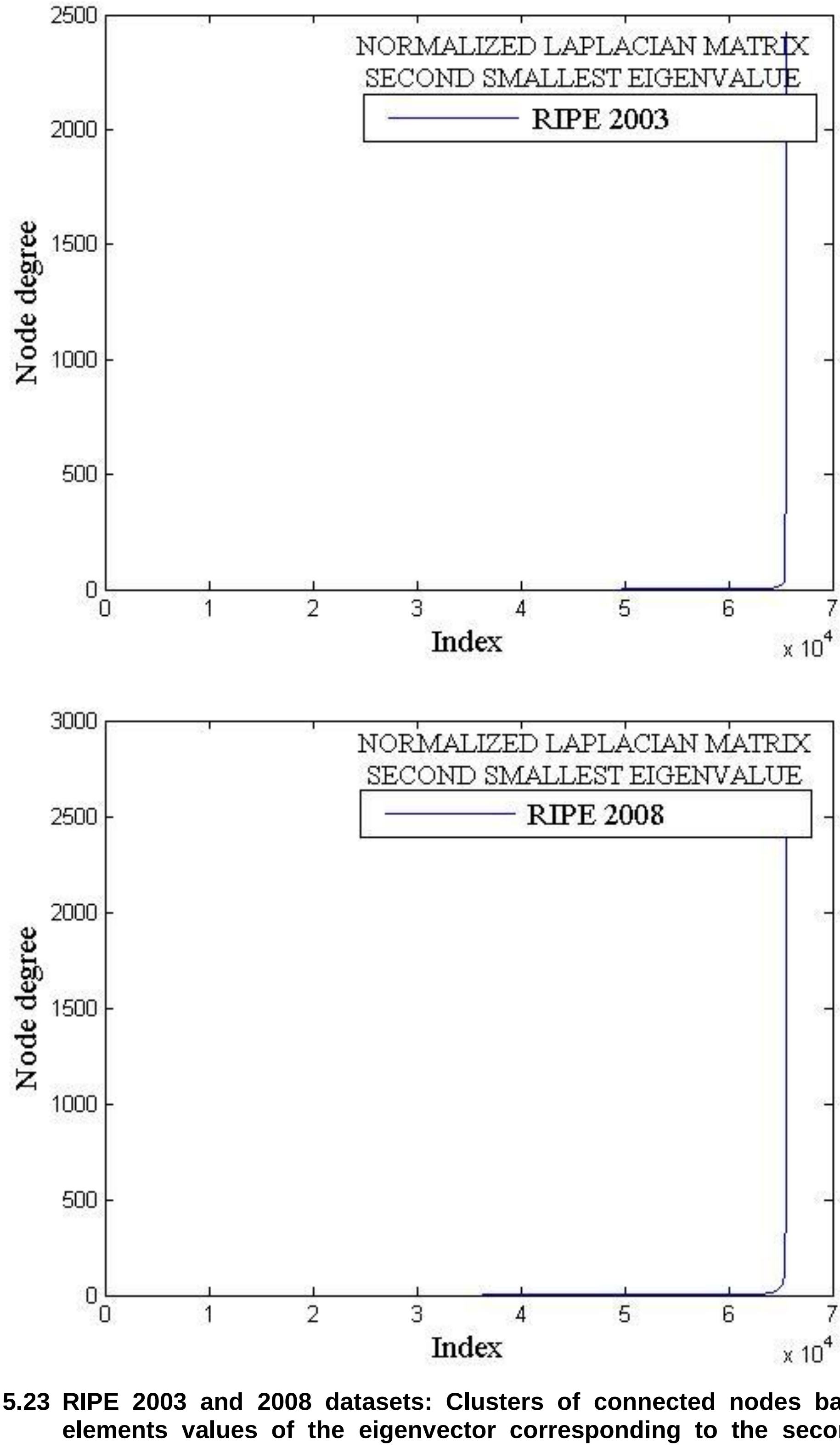


Figure 5.23 RIPE 2003 and 2008 datasets: Clusters of connected nodes based on the elements values of the eigenvector corresponding to the second smallest eigenvalue of the normalized Laplacian matrix for RIPE 2003 (top) and 2008 (bottom) datasets.

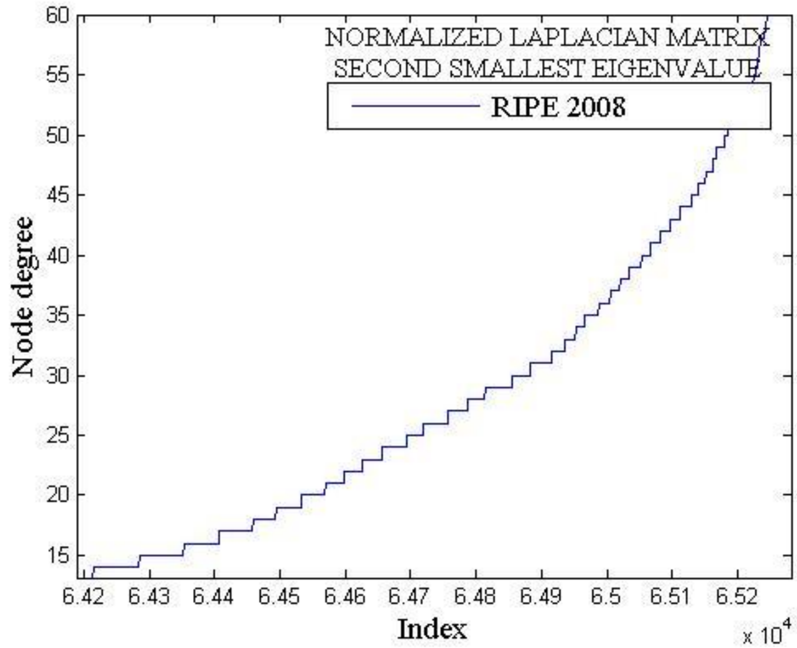


Figure 5.24 RIPE 2008: Zoomed view of node degree vs. rank.

The sorted elements of the eigenvector corresponding to the largest eigenvalue of the normalized Laplacian matrix are plotted vs. the order as shown in Figure 5.25. Most of the elements in the middle of the rank spectrum have very small value (10^{-19}) for all four datasets. The elements of the eigenvector for few nodes at the lowest end of the rank spectrum have positive values while few nodes at the highest end of the rank spectrum have negative value. The pattern is similar for Route Views and RIPE 2003 datasets and for Route Views and RIPE 2008 datasets.

The node degrees distribution based on the decreasing order of the elements values of the corresponding AS nodes is shown in Figure 5.26 and Figure 5.27. Two clusters of connected nodes are visible at the lowest and the highest ends of the rank spectrum for all four datasets.

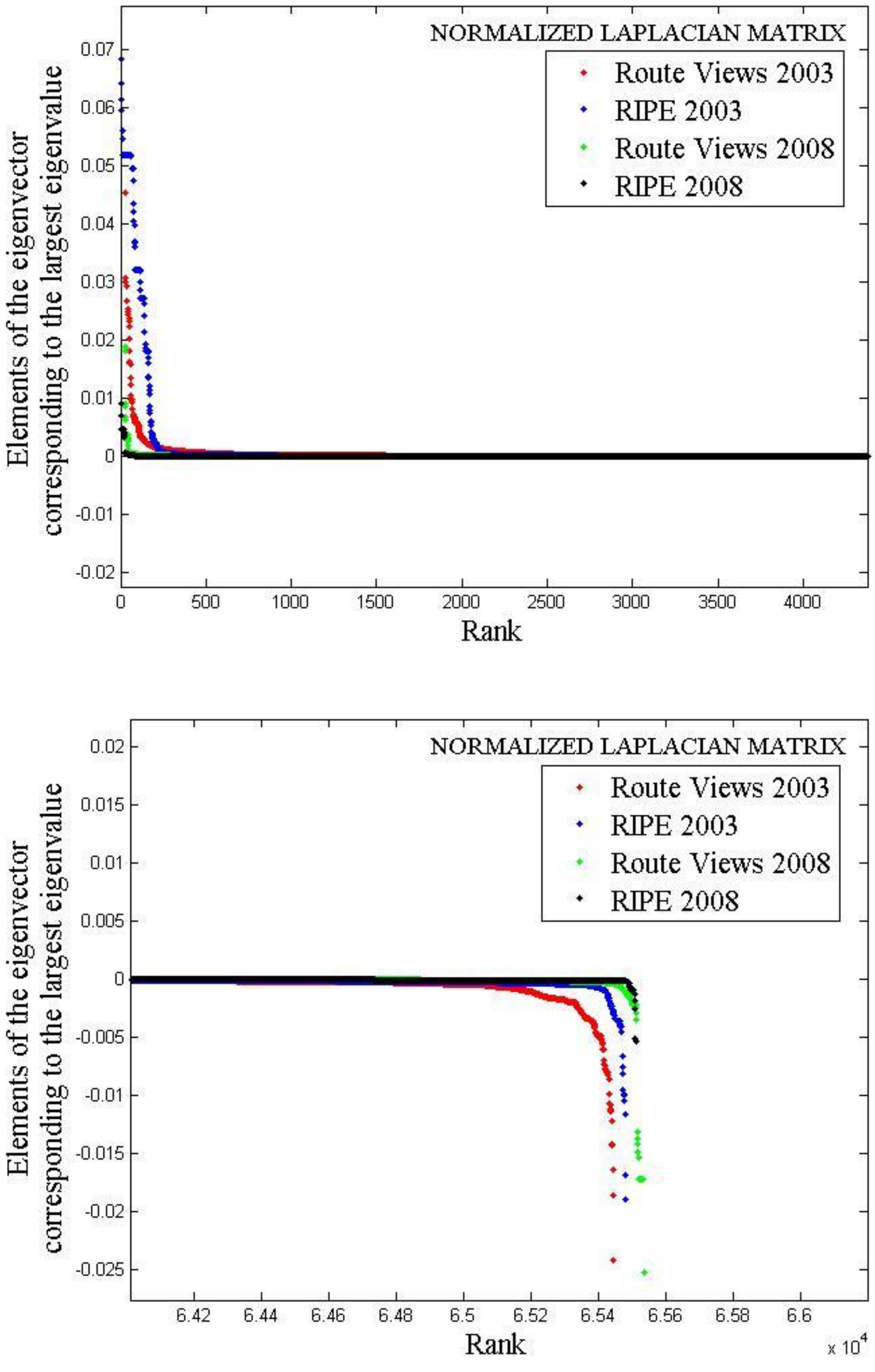


Figure 5.25 **Route Views and RIPE 2003 and 2008 datasets: Elements of the eigenvector corresponding to the largest eigenvalue of the normalized Laplacian matrix. Shown are the nodes at the lowest (top) and the highest (bottom) ends of the rank spectrum.**

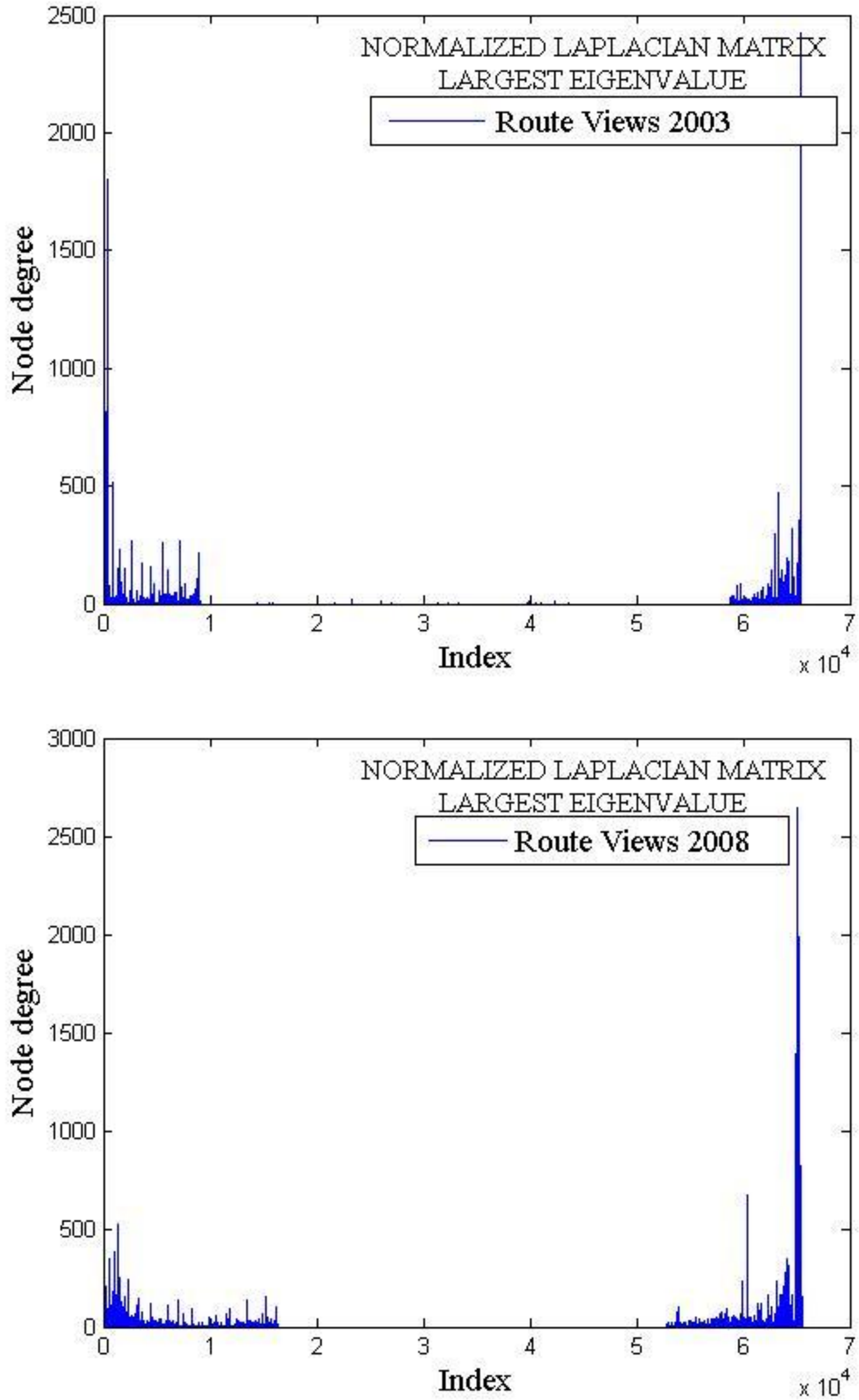


Figure 5.26 Route Views 2003 and 2008 datasets: Clusters of connected nodes based on the elements values of the eigenvector corresponding to the largest eigenvalue of the normalized Laplacian matrix for Route Views 2003 (top) and 2008 (bottom) datasets.

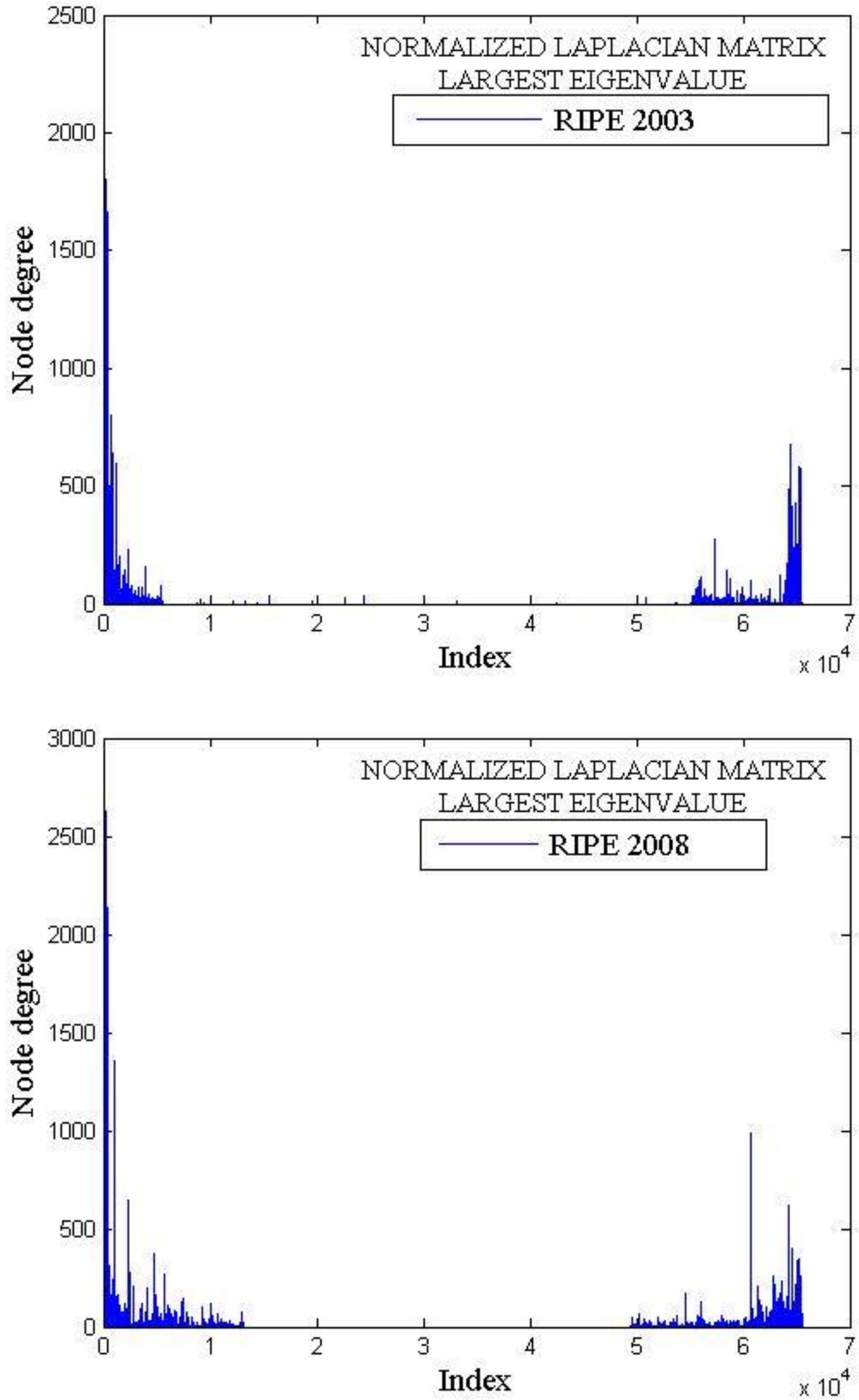


Figure 5.27 RIPE 2003 and 2008 datasets: Clusters of connected nodes based on the elements values of the eigenvector corresponding to the largest eigenvalue of the normalized Laplacian matrix for RIPE 2003 (top) and 2008 (bottom) datasets.

An interesting property is observed when comparing the clusters of AS nodes based on the second smallest and the largest eigenvalues of the adjacency and the normalized Laplacian matrices. The cluster of nodes based on the elements values of the eigenvector corresponding to the second smallest eigenvalue of the adjacency matrix is similar to the cluster based on the largest eigenvalue of the normalized Laplacian matrix and vice versa. This property has its basis in the spectral properties of the two matrices since $L = D - A$, where L is the Laplacian matrix, D is the degree matrix, which is the diagonal matrix formed from the node degrees, and A is the adjacency matrix. However, how this relationship affects theoretically in the spectral properties of the adjacency and the Laplacian matrices could be an interesting research. It would also be interesting to investigate the spectral properties of matrices such as Laplacian and signless Laplacian matrix, which is defined as $Q = D + A$.

6. CONCLUSIONS

We have analyzed the Internet topology datasets collected from Route Views and RIPE projects and have confirmed the presence of power-laws in graphs capturing the AS-level Internet topology in both datasets over the past five years. We have evaluated four power-laws: node degrees vs. rank, CCDF of node degree vs. node degree, eigenvalue of the adjacency matrix vs. index, and eigenvalue of the normalized Laplacian matrix vs. index. We found that eigenvalues based on the normalized Laplacian matrix also exhibit power-law similar to eigenvalue power-law based on the adjacency matrix. They have, as expected, different values for power-law exponents. The results showed that the power-law exponents associated with the Internet topology have not significantly changed over the years indicating that the power-laws do not capture every property of graph and are only a measure used to characterize the Internet topology.

Spectral analysis based on the more intuitive adjacency matrix and the normalized Laplacian matrix derived from the Route Views and RIPE datasets was used to examine the clustering of ASes and their connectivity in the Internet graphs. By plotting the elements of the adjacency matrix, we observed similarity in the pattern of clusters of connected AS nodes in the Internet topology over the years. The clusters indicate higher connectivity inside a particular cluster and relatively lower connectivity between clusters. While power-laws properties of the

Internet topology graphs have not substantially changed over the years, spectral analysis revealed notable changes in the connectivity and clustering of AS nodes. The connectivity status based on the adjacency and the normalized Laplacian matrices indicated visible changes in the clustering of connected AS nodes over the past five years.

We also identified clusters of AS nodes based on the eigenvectors corresponding to the second smallest and the largest eigenvalues of the adjacency and the normalized Laplacian matrices. Presented spectral analysis of both the matrices of the associated graphs also revealed new historical trends in the clustering of AS nodes. We observed that clustering based on the second smallest eigenvalue of the adjacency matrix is similar to clustering based on the largest eigenvalue of the normalized Laplacian matrix, and vice versa. The cluster based on the second smallest eigenvalue of the normalized Laplacian matrix consists of the groups of nodes having similar node degree. Furthermore, group of nodes having larger node degree follows group of nodes having smaller node degree within a cluster. It would be interesting to investigate whether the observed clusters have any significant effect in the modeling of the Internet topology and the performance of the network protocols and new algorithms.

7. REFERENCES

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

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

[3] BGP Routing table Analysis [Online]. Available: http://www.potaroo.net/tools/asns/.

[4] Cooperative Association for Internet Data Analysis [Online]. Available: http://www.caida.org/.

[5] Correlation coefficient [Online]. Available: http://mathworld.wolfram.com/CorrelationCoefficient.html.

[6] Looking Glass project [Online]. Available: http://www.traceroute.org/.

[7] MATLAB [Online]. Available: http://www.mathworks.com/.

[8] RFC 1266 [Online]. Available: http://www.rfc-editor.org/rfc/rfc1266.txt.

[9] RFC 1771 [Online]. Available: http://www.rfc-editor.org/rfc/rfc1771.txt.

[10] RFC 2328 [Online]. Available: http://www.rfc-editor.org/rfc/rfc2328.txt.

[11] RFC 2453 [Online]. Available: http://www.rfc-editor.org/rfc/rfc2453.txt.

[12] Reseaux IP Europeens [Online]. Available: http://www.ripe.net/ris.

[13] W. Aiello, F. Chung, and L. Lu, "A random graph model for massive graphs," in *Proceedings of the 32nd Annual Symposium on the Theory of Computing*, Portland, USA, May 2000, pp.171–180.

[14] D. Andersen, N. Feamster, S. Bauer, and H. Balakrishnan, "Topology inference from BGP routing dynamics," in *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, Marseille, France, Nov. 2002, pp. 243–248.

[15] T. Bu and D. Towsley, "On distinguishing between Internet power law topology generators," in *Proceedings of IEEE INFOCOM,* New York, NY, USA, June 2002, pp. 638–647.

[16] S. Butler, *Spectral Graph Theory*. Lectures notes, Nankai University, Tianjin, China, Sept. 2006.

[17] B. E. Carpenter, "Observed relationships between size measures of the Internet," *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 2, pp. 5–12, Apr. 2009.

[18] H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "Towards capturing representative AS-level Internet topologies," *The International Journal of Computer and Telecommunications Networking*, vol. 44, no. 6, pp. 735–755, Apr. 2004.

[19] H. Chang, S. Jamin, and W. Willinger, "Internet connectivity at the AS level: an optimization-driven modeling approach," in *Proceedings of ACM SIGCOMM Workshop on MoMeTools*, Karlsruhe, Germany, Aug. 2003, pp. 33–46.

[20] H. Chang, S. Jamin, and W. Willinger, "To peer or not to peer: modeling the evolution of the Internet's AS-level topology," in *Proceedings of IEEE INFOCOM,* Barcelona, Spain, Apr. 2006, pp.1–12.

[21] H. Chang, S. Jamin, and W. Willinger, "What causal forces shape Internet connectivity at the AS-level ?," *Technical report*, EECS Dept., University of Michigan, 2003.

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

[23] 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 INFOCOM*, New York, NY, USA, Apr. 2002, pp. 608–617.

[24] F. R. K. Chung, *Spectral Graph Theory.* Providence, Rhode Island: Conference Board of the Mathematical Sciences, June 1997, pp. 2–6.

[25] R. Cohen and D. Raz, "The Internet dark matter - on the missing links in the AS connectivity map," in *Proceedings of IEEE INFOCOM*, Barcelona, Spain, Apr. 2006, pp. 1–12.

[26] J. L. Devore, *Probability and Statistics for Engineering and the Sciences: Enhanced*. Duxbury Press, 2008, pp. 254–263/446–499.

[27] J. Doyle and J. M. Carlson, "Power laws, highly optimized tolerance and generalized source coding," *Physical Review Letter,* vol. 84, no. 24, pp. 5656–5659, Mar. 2000.

[28] A. Dhamdhere and C. Dovrolis, "Ten years in the evolution of the Internet ecosystem*,"* in *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, Vouliagmeni, Greece, Oct. 2008, pp.183–196.

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

[30] I. Farkas, I. Dernyi, A. Barabasi, and T. Vicsek, "Spectra of "real-world" graphs: beyond the semicircle law", *Physical Review E*, vol. 64, no. 2, 026704, pp. 1–12, July 2001.

[31] M. Fiedler, "A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory," *Czech Math Journal*, vol. 25, no. 1, pp. 619–633, Sept. 1975.

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

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

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

[35] H. Haddadi, D. Fay, A. Jamakovic, O. Maennel, A. W. Moore, R. Mortier, M. Rio, and S. Uhlig, "Beyond node degree: evaluating AS topology models," *Technical report*, UCAM-CL-TR-725, Computer laboratory, University of Cambridge, July 2008.

[36] H. Haddadi, G. Iannaccone, A. Moore, R. Mortier, and M. Rio, "Network topologies: inference, modelling and generation," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 2, pp. 48–67, July 2008.

[37] G. Huston, "Interconnection, peering and settlements-Part II," *Internet Protocol Journal*, Mar. 1999 [Online]. Available: http://www.cisco.com/warp/public/759/ipj 2-1/ipj 2-1.html.

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

[39] Y. Li, J. Cui, D. Maggiroini, and M. Faloutsos, "Characterizing and modelling clustering features in AS-level Internet topology," in *Proceedings of IEEE INFOCOM 2008,* Phoenix, Arizona, USA, Apr. 2008, pp. 271–275.

[40] D. Magoni and J. J. Pansiot, "Analysis of the autonomous system network topology," *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 3, July 2001, pp. 26–37.

[41] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, K. claffy, and A. Vahdat, “The Internet AS-level topology: three data sources and one definitive metric,” *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 1, Jan. 2006, pp. 17–26.

[42] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, and K. Claffy, “Lessons from three views of the Internet topology,” *CAIDA*, Technical Report, 2005.

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

[44] A. Medina, A. Lakhina, I. Matta, and J. Byers, “Brite: an approach to universal topology generation,” in *Proceedings of the International Workshop on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems*, Cincinnati, OH, Aug. 2001, pp. 346–356.

[45] M. Najiminaini, L. Subedi, and Lj. Trajkovic, “Analysis of Internet topologies: A historical view,” in *Proceedings of IEEE Int. Symp Circuit and Systems*, Taipei, Taiwan, May 2009, pp. 1697–1700.

[46] R. Oliveira, B. Zhang, and L. Zhang, "Observing the evolution of Internet AS topology," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 4, pp. 313–324, Jan. 2007.

[47] 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.

[48] 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.

[49] L. Subedi and Lj. Trajkovic, “Spectral analysis of the Internet topologies,” in *Proceedings of IEEE Int. Symp Circuit and Systems*, Paris, France, May 2010.

[50] H. Tangmunarunkit, J. Doyle, R. Govindan, S. Jamin, and S. Shenker, “Does AS size determine degree in AS topology ?,” *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 5, pp. 7–10, Oct. 2001.

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

[52] D. Vukadinovic, P. Huang, and T. Erlebach, “On the spectrum and structure of Internet topology graphs,” in *Proceedings of Innovative Internet Computing Systems*, Kühlungsborn, Germany, June 2002, vol. 2346, pp. 83–96.

[53] F. Wang and L. Gao, “On inferring and characterizing Internet routing policies,” in *Proceedings of ACM SIGCOMM Internet Measurement Conference*, Miami, FL, Oct. 2003, pp. 15–26.

[54] X. Wang and D. Loguinov, “Wealth-based evolution model for the Internet AS-level topology,” in *Proceedings of IEEE INFOCOM* , Barcelona, Spain, Apr. 2006, pp. 1–11.

[55] D. J. Watts and S. H. Strogatz, “Collective dynamics of small world networks,” *Nature*, vol. 393, pp. 440–442, June 1998.

[56] R. Winter, “Modeling the Internet routing topology - in less than 24h,” in *Proceedings of ACM/IEEE/SCS 23rd Workshop on Principles of Advanced and Distributed Simulation*, Lake Placid, USA, June 2009, pp. 72–79.

8. APPENDIX

8.1 Confidence intervals

The calculated power-law exponents and the correlation coefficients for four power-laws from the sample datasets collected from Route Views and RIPE 2003 and 2008 datasets are shown in Table 8.1 - Table 8.4. The following symbols are used for the power-law exponents and correlation coefficients:

- R – exponent of node degree vs. rank power-law
- D – exponent of CCDF of node degree vs. node degree power-law
- ε – exponent of eigenvalue of the adjacency matrix vs. index power-law
- L – exponent of eigenvalue of the normalized Laplacian matrix vs. index power-law
- r – correlation coefficient.

Table 8.1 Power-law exponents for Route Views 2003 datasets. Datasets were randomly selected.

Dataset number	R		D		ε		L	
	value	r	value	r	value	r	value	r
1	–0.7342	–0.9671	–1.2673	–0.9188	–0.5088	–0.9991	–0.0191	–0.9670
2	–0.7062	–0.9645	–1.2088	–0.9077	–0.5101	–0.9990	–0.0188	–0.9537
3	–0.6910	–0.9614	–1.1809	–0.9163	–0.5060	–0.9988	–0.0189	–0.9712
4	–0.7368	–0.9668	–1.2844	–0.9130	–0.5134	–0.9992	–0.0186	–0.9541
5	–0.7113	–0.9642	–1.2191	–0.9149	–0.5034	–0.9990	–0.0191	–0.9524
6	–0.6917	–0.9612	–1.1853	–0.9046	–0.5074	–0.9989	–0.0194	–0.9582

Table 8.2 Power-law exponents for Route Views 2008 datasets. Datasets were randomly selected.

Dataset number	R		D		ε		L	
	value	r	value	r	value	r	value	r
1	–0.7712	–0.9721	–1.3696	–0.9626	–0.4850	–0.9982	–0.0177	–0.9782
2	–0.7779	–0.9732	–1.3572	–0.9094	–0.4874	–0.9952	–0.0187	–0.9870
3	–0.7715	–0.9721	–1.3848	–0.9182	–0.4843	–0.9980	–0.0174	–0.9749
4	–0.7781	–0.9727	–1.4158	–0.9223	–0.4887	–0.9985	–0.0179	–0.9807
5	–0.7730	–0.9724	–1.3790	–0.9158	–0.4833	–0.9983	–0.0182	–0.9705
6	–0.7749	–0.9722	–1.4062	–0.9212	–0.4882	–0.9985	–0.0173	–0.9662

Table 8.3 Power-law exponents for RIPE 2003 datasets. Datasets were randomly selected.

Dataset number	R		D		ε		L	
	value	r	value	r	value	r	value	r
1	–0.7340	–0.9662	–1.2341	–0.9154	–0.5026	–0.9994	–0.0208	–0.9633
2	–0.7384	–0.9665	–1.2497	–0.9171	–0.5073	–0.9994	–0.0189	–0.9565
3	–0.7319	–0.9654	–1.2304	–0.9155	–0.5035	–0.9994	–0.0199	–0.9615
4	–0.7319	–0.9654	–1.2448	–0.9146	–0.5113	–0.9992	–0.0206	–0.9625
5	–0.7495	–0.9680	–1.2810	–0.9188	–0.5100	–0.9991	–0.0187	–0.9640
6	–0.7465	–0.9674	–1.2661	–0.9122	–0.5069	–0.9994	–0.0195	–0.9526

Table 8.4 Power-law exponents for RIPE 2008 datasets. Datasets were randomly selected.

Dataset number	R		D		ε		L	
	value	r	value	r	value	r	value	r
1	–0.8323	–0.9742	–1.4743	–0.9248	–0.4958	–0.9968	–0.0202	–0.9756
2	–0.8352	–0.9741	–1.4341	–0.9152	–0.4943	–0.9968	–0.0194	–0.9689
3	–0.8240	–0.9745	–1.4616	–0.9271	–0.4899	–0.9973	–0.0183	–0.9621
4	–0.8451	–0.9759	–1.4565	–0.9210	–0.5003	–0.9977	–0.0205	–0.9800
5	–0.8434	–0.9748	1.5233	–0.9301	–0.4937	–0.9969	–0.0195	–0.9769
6	–0.8371	–0.9741	–1.5081	–0.9293	–0.4959	–0.9971	–0.0196	–0.9722

8.2 MATLAB code

This Section contains the sample MATLAB code used to analyze the datasets collected from Route Views and RIPE projects.

The binary files with BGP routing information are downloaded from the databases of Route Views and RIPE projects [2], [12]. The binary files are then converted into the text file using “bgpdump” Linux command. We use MATLAB tool to extract the AS numbers from the text file and to create the adjacency matrix. Further analysis was also performed using MATLAB tool. The sample code follows.

8.2.1 The adjacency matrix

```
% Sample script to change the binary file into text file.%
./bgpdump -m /local-scratch/lsa38/RIPE/2003/rc00/bview.20030715.0000.gz
> /local-scratch/lsa38/RIPE/2003/july15/rc00.txt
./bgpdump -m /local-scratch/lsa38/RIPE/2003/rc00/bview.20030716.0000.gz
> /local-scratch/lsa38/RIPE/2003/july16/rc00.txt
./bgpdump -m /local-scratch/lsa38/RIPE/2003/rc00/bview.20030717.0000.gz
> /local-scratch/lsa38/RIPE/2003/july17/rc00.txt
./bgpdump -m /local-scratch/lsa38/RIPE/2003/rc00/bview.20030718.0000.gz
> /local-scratch/lsa38/RIPE/2003/july18/rc00.txt
```

```
% Sample code to create the adjacency matrix from the text file
containing BGP routing information. %

close all;
clear all;
Adjacency_matrix=sparse(65535,65535);
for i=1:39
fileName=['../RouteViews2008/july31/rc',num2str(i),'.txt'];
fileId=fopen(fileName);
    while 1
      eachLine = fgetl(fileId);
      if isempty(eachLine); else
            count=0;
            for ii=1:length(eachLine)
                  if (eachLine(ii)=='|')
                        count=count+1;
                        if (count==6)
                              startPoint=ii;
                        end
```

```
                        if (count==7)
                              stopPoint=ii;
                        end
                  end
            end
            singleLine=[];
             for ii=startPoint+1:stopPoint-1
                  singleLine=[singleLine,  eachLine(ii)];
            end
             if isempty(singleLine);
            singleLine=str2num(singleLine);
                  for ii=1:length(singleLine)-1
                      Arow=singleLine(ii);
                      Acolumn=singleLine(ii+1);
                      Adjacency_matrix(Arow,Acolumn)=1;
                      Adjacency_matrix(Acolumn,Arow)=1;
                  end
            end
            end
        if (feof(fid))
            break;
        end
     end fclose(fid);
end;
fclose('all');
save ../RV2008Results/july31/AdjacencyMatrix.mat ADJM;
clear all;
load AdjacencyMatrix.mat;
for i=1:length(Adjacency_matrix)
      node_degree(i)=sum(Adjacency_matrix(i,:));
end
save ('NodeDegree', 'node_degree');
```

8.2.2 The normalized Laplacian matrix

```
% Sample code to create the normalized Laplacian matrix. %
clear
load AdjacencyMatrix.mat;
load NodeDegree.mat;
n= length(Adjacency_matrix)
Norm_laplacian_matrix= sparse(1:n,1:n,0);
for i=1:n
      for j=1:n
            if i==j & node_degree(i)~=0
                  Norm_laplacian_matrix(i,j)=1;
                  elseif (Adjacency_matrix(i,j)~=0)
                        Norm_laplacian_matrix(i,j)=
                              1/(sqrt(node_degree(i)*node_degree(j)));
                  end
            end
      end
save('NormLaplacianMatrix','Norm_laplacian_matrix')
```

8.2.3 Power-law: node degree vs. rank

```
%  Sample code to calculate the node degrees from the adjacency matrix
and to plot them in log-log scale as shown in Figure 4.1 and Figure
4.2. %

load AdjacencyMatrix.mat
for i = 1:length(Adjacency_matrix)
      node_degree(i) = sum(Adjacency_matrix(:,i));
end
node_degree=full(node_degree);
node_degree=sort(node_degree,'descend');
loglog(node_degree,'.');
s = nonzeros(node_degree);
n=length(s);
xxx=(1:n);
x=log10(xxx);
y=node_degree(1:n);
y=log10(y);
m=(n*sum(x.*y)-sum(x)*sum(y))/(n*sum(x.^2)-(sum(x))^2)
b=(sum(y)-m*sum(x))/n
r=(n* sum(x.*y)-sum(x)*sum(y))/ sqrt((n*sum(x.^2)-
sum(x)^2)*(n*sum(y.^2)-sum(y)^2))
hold on figure1
loglog(xxx,10^b.*xxx.^m);
ylabel('Node degree','fontsize',16,'fontname','times')
xlabel('Rank','fontsize',16,'fontname','times')
```

8.2.4 Power-law: CCDF of node degree vs. node degree

```
% Sample code to calculate the CCDFs of node degrees and to plot them
in log-log scale as shown in Figure 4.3 and Figure 4.4. %

load NodeDegree.mat
maxDegree=max(node_degree);
n=length(node_degree);
x=zeros(1,maxDegree);
for i=1:maxDegree
    m=0;
    for j=1:n
        if node_degree(j)==i
            m=m+1;
        end
    end
x(i)=m;
end

ccdf=zeros(1,maxDegree);
ccdf(1)=x(1)/sum(x);
for i=2:length(x)
      ccdf(i)=x(i)/sum(x)+ccdf(i-1);
end
one= ones(1,max1);
ccdf=one-ccdf;
```

```
for i = 1:length(ccdf)
      ccdf(i) = sum(ccdf(:,i));
end

ccdf=full(ccdf);
ccdf=sort(ccdf,'descend');
loglog(ccdf,'.');
s = nonzeros(ccdf);
n=length(s)-1;
xxx=(1:n);
x=log10(xxx);
y=ccdf(1:n);
y=log10(y);
m=(n*sum(x.*y)-sum(x)*sum(y))/(n*sum(x.^2)-(sum(x))^2)
b=(sum(y)-m*sum(x))/n
r=(n* sum(x.*y)-sum(x)*sum(y))/sqrt((n*sum(x.^2)-sum(x)^2)*(n*sum(y.^2)
      -sum(y)^2))
hold on figure1
loglog(xxx,10^b.*xxx.^m);
ylabel('CCDF of node degree','fontsize',16,'fontname','times')
xlabel('Node degree','fontsize',16,'fontname','times')
```

8.2.5 Power-law: eigenvalue of the adjacency matrix vs. index

```
% Sample code to calculate the first 150 largest eigenvalues of the
adjacency matrix and to plot them vs. index in log-log scale as shown
in Figure 4.5 and Figure 4.6. %

clear;
load AdjacencyMatrix.mat;
eigen_values= eigs(Adjacency_matrix, 300);
save ('EigenvaluesAdjMatrix.mat','eigen_values');
eigen_values= sort(eigen_values,'descend');
for i=1:1:150
      toplot(i)=eigen_values(i);
end
loglog(toplot,'.');
b=toplot;
n=150;
xxx=(1:n);
x=log10(xxx);
y=b(1:n);
y=log10(y);
m=(n*sum(x.*y)-sum(x)*sum(y))/(n*sum(x.^2)-(sum(x))^2)
b=(sum(y)-m*sum(x))/n
r=(n* sum(x.*y)-sum(x)*sum(y))/ sqrt((n*sum(x.^2)-
sum(x)^2)*(n*sum(y.^2)-sum(y)^2))
hold on figure1
loglog(xxx,10^b.*xxx.^m);
ylabel('Eigenvalue of adjacency matrix',fontsize',16,‘fontname',
'times');
xlabel('Index','fontsize',16,'fontname','times')
```

8.2.6 Power-law: eigenvalues of the normalized Laplacian matrix vs. index

```
% Sample code to calculate the first 150 largest eigenvalues of the
normalized Laplacian matrix and to plot them vs. index in log-log scale
as shown in Figure 4.8 and Figure 4.9. %

clear;
load NormlaplacianMatrix.mat;
eigen_values= eigs(Norm_laplacian_matrix, 150);
save ('EigenvaluesLapMatrix','eigen_values');
eigen_values= sort(eigen_values,'descend');
for i=1:1:150
      toplot(i)=eigen_values(i);
end
loglog(toplot,'.');
b=toplot;
n=150;
xxx=(1:n);
x=log10(xxx);
y=b(1:n);
y=log10(y);
m=(n*sum(x.*y)-sum(x)*sum(y))/(n*sum(x.^2)-(sum(x))^2)
b=(sum(y)-m*sum(x))/n
r=(n* sum(x.*y)-sum(x)*sum(y))/ sqrt((n*sum(x.^2)-
sum(x)^2)*(n*sum(y.^2)-sum(y)^2))
hold on figure1
loglog(xxx,10^b.*xxx.^m);
ylabel('Eigenvalue of normalized Laplacian matrix',
      'fontsize',16,'fontname','times');
xlabel('Index','fontsize',16,'fontname','times');
```

8.2.7 Pattern of connected AS nodes

```
% Sample code to plot the patterns of connected ASes as shown in Figure
5.1 and Figure 5.2. %

clear
load AdjacencyMatrix.mat
spy(Adjacency_matrix)
ylabel('Autonomous system (AS)
number','fontsize',16,'fontname','times')
xlabel('Autonomous System Number
(ASN)','fontsize',16,'fontname','times')
h= legend('RIPE 2003-07-31 00:00');
set(h, 'fontsize',16,'fontname','times');
```

8.2.8 Connectivity status: the second smallest eigenvalue

```
% Sample code to calculate the elements of the eigenvector
corresponding to the second smallest eigenvalue of the normalized
Laplacian matrix and to plot the connectivity status as shown in Figure
5.9 and Figure 5.10. Similar code is used to calculate the elements of
```

```
the eigenvector corresponding to the second smallest eigenvalue of the
adjacency matrix as shown in Figure 5.5 and Figure 5.6. %

clear
load NodeDegree.mat
degree=node_degree;
load AdjacencyMatrix.mat
rr = Adjacency_matrix;
n=49149;
ss=sparse(n,n,0);
count=zeros(1,n);
k=0;
for i=1:48127
    k=k+1;
    count(k)=degree(i);
    ss(k,1:48127)=rr(i,1:48127);
    ss(k,48128:49151)=rr(i,64512:65535);
end
for i=64512:65535
    k=k+1;
    count(k)=degree(i);
    ss(k,1:48127)=rr(i,1:48127);
    ss(k,48128:49151)=rr(i,64512:65535);
end
Manipulated_adjacency_matrix =sparse(n,n,0);
Manipulated_adjacency_matrix = ss
save('ManipulatedAdjacencyMatrix', Manipulated_adjacency_matrix')
save('ManipulatedNodeDegree', 'count')
clear
load NodeDegree.mat
degree= node_degree;
load Norm_laplacian_matrix.mat
rr= Norm_laplacian_matrix;
ss=sparse(49149,49149,0);
count=zeros(1,49149);
k=0;
for i=1:length(rr)
      if (sum(rr(i,:))~=1)
            k=k+1;
            ss(k,1:48127)=rr(i,1:48127);
            count(k)=degree(i);
            ss(k,48128:49151)=rr(i,64512:65535);
      end
end
for i=k+1:49149
    ss(i,i)=1;
end
Manipulated_Norm_Lap =sparse(49149,49149,0);
Manipulated_Norm_Lap =ss;
save('ManipulatedNormLapMatrix','Manipulated_Norm_Lap')
clear
load Manipulated_degree.mat
load ManipulatedNormLapMatrix.mat
[V D] = eigs(Manipulated_Norm_Lap,2,'SA');
x=V(:,2);
x=x';
index=1:1:length(x);
```

```
for i=1:length(x)
    for j=i:length(x)
        if x(j)<x(i)
            tmp=x(j);
            x(j)=x(i);
            x(i)=tmp;
            tmp=index(i);
            index(i)=index(j);
            index(j)=tmp;
        end
    end
end
for i=1:length(Manipulated_Degree)
    if Manipulated_Degree(index(i))==0
        connection(i)=0;
    else
        connection(i)=1;
    end
end
plot(length(connection):-1:1,connection, '.');
ylabel('Connectivity status: smallest
      eigenvalue','fontsize',16,'fontname','times')
xlabel('Index of element','fontsize',16,'fontname','times')
ylim([0 2]);
```

8.2.9 Connectivity status: the largest eigenvalue

```
% Sample code to calculate the elements of the eigenvector
corresponding to the largest eigenvalue of the normalized Laplacian
matrix and to plot the connectivity status as shown in Figure 5.11 and
Figure 5.12. Similar code is used to calculate the elements of the
eigenvector corresponding to the largest eigenvalue of the adjacency
matrix as shown in Figure 5.7 and Figure 5.8. %

clear
load Manipulated_degree.mat
load ManipulatedNormLapMatrix.mat
[V D] = eigs(Manipulated_Norm_Lap,2,'LA');
 x=V(:,1);
x=x';
index=1:1:length(x);
for i=1:length(x)
    for j=i:length(x)
        if x(j)<x(i)
            tmp=x(j);
            x(j)=x(i);
            x(i)=tmp;
            tmp=index(i);
            index(i)=index(j);
            index(j)=tmp;
        end
    end
end
for i=1:length(Manipulated_Degree)
```

```
    if Manipulated_Degree(index(i))==0
        connection(i)=0;
    else
        connection(i)=1;
    end
end
plot(length(connection):-1:1,connection, '.');
ylabel('Connectivity status: largest eigenvalue','fontsize'
      ,16,'fontname','times')
xlabel('Index of lements','fontsize',16,'fontname','times')
ylim([0 2]);
```

8.2.10 Elements of eigenvector: the second smallest eigenvalue

```
% Sample code to calculate the elements of the eigenvector corresponding to the second smallest eigenvalue of the adjacency matrix and to plot them vs. order as shown in Figure 5.15. Similar code is used to plot the elements of the eigenvector corresponding to the second smallest eigenvalue based on the normalized Laplacian matrix as shown in Figure 5.21. %

clear
cd('Routeview_2003')
load AdjacencyMatrix.mat
[V D] = eigs(Adjacency_matrix,2,'SA');
x=V(:,2);
x=sort(x,'descend')
p=plot(x,'.')
set(p,'Color','red')
 cd ../.
cd('RIPE 2003')
clear
hold all
load AdjacencyMatrix.mat
[V D] = eigs(Adjacency_matrix,2,'SA');
x=V(:,2);
x=sort(x,'descend')
p=plot(x,'.')
set(p,'Color','blue')
 cd ../.
cd('Routeview_2008')
hold all
clear
load AdjacencyMatrix.mat
[V D] = eigs(Adjacency_matrix,2,'SA');
x=V(:,2);
x=sort(x,'descend')
p=plot(x,'.')
set(p,'Color','green')
cd ../.
cd('RIPE 2008')
clear
hold all
load AdjacencyMatrix.mat
[V D] = eigs(Adjacency_matrix,2,'SA');
```

```
x=V(:,2);
x=sort(x,'descend')
p=plot(x,'.')
set(p,'Color','black')
ylabel('Elements of eigenvector of second smallest
eigenvalues','fontsize',16,'fontname','times')
xlabel('Rank','fontsize',16,'fontname','times')
h=legend('RouteViews 2003','RIPE 2003','RouteViews 2008','RIPE 2008');
set(h, 'fontsize',16,'fontname','times');
```

8.2.11 Elements of eigenvector: the largest eigenvalue

```
% Sample code to calculate the elements of the eigenvector
corresponding to the largest eigenvalue of the adjacency matrix and to
plot them vs. order as shown in Figure 5.18. Similar code is used to
plot the elements of the eigenvector corresponding to the largest
eigenvalue based on the normalized Laplacian matrix as shown in Figure
5.25. %

clear
cd('Routeview_2003')
load AdjacencyMatrix.mat
[V D] = eigs(Adjacency_matrix,2,'LA');
x=V(:,1);
x=sort(x,'descend')
p=plot(x,'.')
set(p,'Color','red')
cd ../.
cd('RIPE 2003')
clear
hold all
load AdjacencyMatrix.mat
[V D] = eigs(Adjacency_matrix,2,'LA');
x=V(:,1);
x=sort(x,'descend')
p=plot(x,'.')
set(p,'Color','blue')
cd ../.
cd('Routeview_2008')
hold all
clear
load AdjacencyMatrix.mat
[V D] = eigs(Adjacency_matrix,2,'LA');
x=V(:,1);
x=sort(x,'descend')
p=plot(x,'.')
set(p,'Color','green')
cd ../.
cd('RIPE 2008')
clear
hold all
load AdjacencyMatrix.mat
[V D] = eigs(Adjacency_matrix,2,'LA');
x=V(:,1);
x=sort(x,'descend')
```

```
p=plot(x,'.')
set(p,'Color','black')
ylabel('Elements of eigenvector of largest
eigenvalue','fontsize',16,'fontname','times')
xlabel('Rank','fontsize',16,'fontname','times')
h=legend('RouteViews 2003','RIPE 2003','RoutViews 2008','Blue: RIPE
2008');
set(h, 'fontsize',16,'fontname','times');
```