EVOLUTION OF THE INTERNET TOPOLOGY FROM A REGIONAL PERSPECTIVE

by

Jose Carlos Acedo

A Thesis Submitted to the Faculty of the

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2015

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under the rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgement the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: ____________________

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

Loukas Lazos Associate Professor of Electrical and Computer Engineering	Date

ACKNOWLEDGMENTS

First and foremost, I would like to thank my academic advisor Dr. Loukas Lazos. Your passion for your work encouraged me to pursue my own goals, while your attention to details instilled in me a unique perspective on which to approach problems. I am deeply grateful for your taking me on as a student. Our academic work together has been a truly rewarding and enriching experience, while the lessons learned will help guide me throughout the course of my life in whatever avenue I pursue.

I would also like to thank my family and friends for supporting me all this time, and encouraged me to keep going and accomplish anything I set my mind.

TABLE OF CONTENTS

LIST OF FIGURES 9

LIST OF TABLES 11

ABSTRACT 13

1 INTRODUCTION 14
  1.1 Motivation and Scope 14
  1.2 Main Contributions and Thesis Organization 16

2 PRELIMINARIES AND RELATED WORK 18
  2.1 The Internet Architecture 18
  2.2 Internet Topology Data 22
    2.2.1 Data Collection 22
    2.2.2 Data Incompletness 23
  2.3 Topology Analysis 24
    2.3.1 Measurements, Properties, and Modeling 24
    2.3.2 Inferring ASes Relationships 25
    2.3.3 Classifying ASes 27
    2.3.4 Evolution of the Internet Topology 28

3 DATA SOURCES AND PROCESSING 29
  3.1 Data 29
    3.1.1 Data Sources 29
    3.1.2 Data Processing 30
    3.1.3 Data incompleteness 33
  3.2 Topology Attributes 34
    3.2.1 AS Geopolitical Information 34
    3.2.2 AS Classification 36

4 RESULTS 40
- 4.1 Evolution per Region 40
  - 4.1.1 RIPE 40
  - 4.1.2 ARIN 43
  - 4.1.3 APNIC 45
  - 4.1.4 LACNIC 46
  - 4.1.5 AFRINIC 48
- 4.2 Connectivity 51
- 4.3 Topology Structure Differences 55

5 CONCLUSION 59

APPENDIX A: REGION, SUBREGIONS, AND COUNTRIES 61

APPENDIX B: RIPE ADDITIONAL DATA 64

APPENDIX C: ARIN ADDITIONAL DATA 69

APPENDIX D: APNIC ADDITIONAL DATA 71

APPENDIX E: LACNIC ADDITIONAL DATA 76

APPENDIX F: AFRINIC ADDITIONAL DATA 80

APPENDIX G: CONNECTIONS BETWEEN REGIONS 84

REFERENCES 87

LIST OF FIGURES

2.1 Autonomous Systems represent a collection of networks administrated by a single entity. The Internet consists of interconnected ASes. 20
2.2 Routes advertised by AS differ depending on the AS relationship . . 21
2.3 The hierarchical routing structure of the Internet at the AS level. . 22

3.1 Examples of c2p link inference, (a) AS B is provider, (b) AS A is provider, (c) link discarded. 32
3.2 Number of AS nodes and AS links from 01/98 to 01/2015. 34
3.3 AS growth by RIR and by geographic subregion from 01/1998 to 01/2015. 38
3.4 Traditional AS classification. 39
3.5 Evolution of ASes by business role 39

4.1 Percentage of ASes in RIPE, when divided by subregion and type. . 41
4.2 Common link connections in RIPE. 42
4.3 Percentage of ASes by type. 43
4.4 Percentage of links by type. 44
4.5 Percentage of ASes in APNIC when divided by each subregion and each type. 46
4.6 Common link connections in APNIC. 47
4.7 Percentage of ASes in LACNIC when divided by each subregion and each type. 48
4.8 Common links in LACNIC. 49
4.9 Percentage of ASes in AFRINIC when divided by subregion (a), and by type (b). 50
4.10 Percentage of the most common link relationships. 51
4.11 Fig (a) percentage of links that connect to a different region. Fig(b) percentage of links use by top most connected regions, out of all inter-region links 52
4.12 Percentage of the most common links between ARIN-RIPE. 53

4.13 Percentage of inter-region links in the last five years. 55

4.14 Number of ASes and AS births per region. 56

4.15 (a) Average number of intra-links per AS for the different regions, (b) average number of inter-links per AS for the different regions. . 57

B.1 Number of ASes per Type in RIPE. 64

B.2 Proportion of ASes in each region per type 67

B.3 Percentage of links between differnt types of ASes out of the total number of links connecting two AS in RIPE 68

C.1 Number of ASes in ARIN by type 69

C.2 Percentage of links between different types of ASes out of the total number of links connecting two AS in Arin. Notice the scale is differnent in CAHP 70

D.1 Number of ASes per Type in RIPE 71

D.2 Percentage of links between differnt types of ASes out of the total number of links connecting two AS in APNIC 75

E.1 Number of ASes per Type in LACNIC 76

E.2 Percentage of links between differnt types of ASes out of the total number of links connecting two AS in LACNIC 79

F.1 Number of ASes per Type in RIPE 80

F.2 Percentage of links between differnt types of ASes out of the total number of links connecting two AS in AFRINIC 83

G.1 RIR to RIPE subregions 84

G.2 Preference regions to connect inter-region links 85

G.3 Fig (a) percentage of common links that connect ARIN-AFRINIC. Fig(b) percentage of common links that connect RIPE-AFRINIC . 86

G.4 Fig (a) percentage of common links that connect ARIN-LACNIC. Fig(b) percentage of common links that connect RIPE-LACNIC . . 86

LIST OF TABLES

3.1 The AS graph for three sample years 33

A.1 AFRINIC country list 61
A.2 APNIC country list 62
A.3 ARIN country list 62
A.4 LACNIC country list 63
A.5 RIPE country list 63

B.1 RIPE Types per Country 2015 64

C.1 Arin Types per Country 2015 69

D.1 APNIC Types per Country 2015 71

E.1 LACNIC Types per Country 2015 76

F.1 AFRINIC Types per Country 2015 80

ABSTRACT

Over the last few decades, the Internet ecosystem has been continuously evolving to meet the demands of its ever-increasing user base. Drastic changes in the Internet infrastructure have improved its capacity and throughput performance, enabling a wealth of new services. For Internet Service Providers (ISPs), anticipating and accommodating the rapidly shifting traffic demands has been a technological, economical, and political challenge. Thus far, this challenge has been met in an "organic" fashion, for the most part, based on unilateral actions of many different players such as ISPs, content providers, public policy makers, international organizations, and large enterprises. This symbiotic relationship among many and often competing change factors has led to a system of enormous complexity that was not a product of well-founded engineering principles. Despite the continuous efforts of the scientific and enterprise communities to discover and to model the Internet, understanding its structure remains a hard challenge.

In this thesis, we provide a new perspective on the Internet's evolutionary patterns at the Autonomous System (AS) level. While many studies have focused on the mathematical models that express the growth of the AS graph topology as a whole, little research has been performed to correlate this growth with geographic, economic, and political data, as well as related business interests. We divide the Internet to five distinct regions using the well-established Internet registry classification and show that the structural properties and evolutionary patterns differ from region to region. We further analyze the business relationships that dominate each region, as well relationships between regions. Conclusions from our analysis is used to explain global as well as local Internet structure phenomena.

CHAPTER 1

INTRODUCTION

1.1 Motivation and Scope

Over the last few decades, the Internet ecosystem has been continuously evolving to meet the demands of its ever-increasing user base [25]. Drastic changes in the Internet infrastructure have improved its capacity and throughput performance, enabling a wealth of new services. For Internet Service Providers (ISPs), anticipating and accommodating the rapidly shifting traffic demands has been a technological, economical, and political challenge [10, 106]. Thus far, this challenge has been met in an "organic" fashion, for the most part, based on unilateral actions of many different players such as ISPs, content providers, public policy makers, international organizations, and large enterprises. This symbiotic relationship among many and often competing change factors has led to a system of enormous complexity that was not a product of well-founded engineering principles. Despite the continuous efforts of the scientific and enterprise communities to discover and to model the Internet [101, 62, 29, 81, 40, 2, 44, 51, 94], understanding its structure remains a hard challenge.

In particular, a series of critical research questions have been long-standing:

- What does the Internet look like?
- Who are the key players and how do they interrelate?
- How has the Internet evolved over time and in different regions?
- Can we use prior evolutionary patterns to predict the future Internet structure?

- How can we best engineer the Internet infrastructure to meet the growing traffic demands?
- Is there a strong correlation between economics and the Internet evolution?

Investigation of the aforementioned research questions has led to several theories drawing from the fields of network tomography [11, 112], graph theory [76, 75], statistical and inferential modeling [35, 30, 37], and information visualization [15, 97, 68, 8, 9], to name a few. Yet the scientific community is far from achieving a satisfactory level of understanding the Internet's dynamic structure. A significant factor for this difficulty lies in the Internet's scale and complexity. Today, over 3.5B IPv4 and 9B IPv6 addresses have been assigned to 67K Autonomous Systems (ASes) (Oct. 2014 [91]), which are interconnected by billions of physical and logical links. This complex system is primarily shaped by unseen business relationships between the participating parties.

A significant portion of prior research has focused on analyzing the graph properties of the Internet topology at the Autonomous System (AS) level [75, 55, 63, 29, 35], using metrics such as node degree distribution, betweenness, and average hop count, among others. However, macroscopic metrics fail to capture the local properties of the AS relationships. Moreover, most prior models are simple abstractions that do not factor in the different node/link types. Moreover, these metrics do not reveal the regional variations of the Internet evolution based on economic, political, and business criteria; *where in the world is the Internet growing faster? what is the cause of that growth? how connected is the Internet globally? how will it evolve per region?* Finally, the existing techniques offer little practical guidance with respect to resource allocation. For instance, content distribution network (CDN) administrators face the daunting tasks of caching content to different geographic locations, forming new peering relationships, and forecasting their performance/cost tradeoffs [111, 67]. Evaluating the impact of their decisions requires a detailed view of the Internet infrastructure and its dynamics.

1.2 Main Contributions and Thesis Organization

In this thesis, we propose a new perspective for the study of the evolution of the Internet topology. Rather than analyzing the topology as a single interconnected graph, we split the Internet by region according to the RIR classification and additional geographical criteria. We then study the evolution of the Internet topology at each of the subregions and reveal that no universal evolutionary trend exists.

Our study analyzes the evolution of the ASes by type, as these are reflected by the business type of the organization that owns a particular AS. Moreover, we analyze the evolution of links by type to reveal useful information about the hierarchical Internet structure. Specifically, we classify the ASes into four business types, taking into consideration economic aspects and function within the Internet. The four types are as follows: Enterprise Customers (ECs), Small Transit Providers (STPs), Large Transit Providers (LTPs) and Content and Access Host Providers (CAHPs).

Our results show that most of the growth on the Internet overall and within each region in particular over the period of our study is primarily attributed to the increase in the number of ECs. LTPs, while not contributing to the topology growth, have maintained their position in the core of the Internet by having the largest degree and staying globally connected. The relative presence of STPs has decreased, while more peering-oriented ISPs and content providers have emerged. CAHPs are the ASes responsible for the flattening of the Internet topology, and more recently, they have been the driving force for the strong interconnection between regions.

With respect to the evolution of the Internet topology within each region, our findings are as follows. ARIN, the region that covers North America, maintains a highly-hierarchical structure over the 15 years of our study. Most of the intra-links are between providers to customers. The region is dominated by a small number of LTPs, which account for more than 50% of the links in the region. ARIN used to be the region with the largest number of ASes, but it was surpassed by RIPE in 2009. RIPE is the region that covers most of Europe. It is characterized by a large number of CAHPs, and in contrast to ARIN, RIPE has a flat topology due to a large number of peering connections. APNIC, AFRINIC, the regions for Asian and African countries, have a very sparce topology and rely extensively on connections to other regions, and in specific, RIPE. LACNIC, the region that covers Latin American and Caribbean countries has the lowest connection degree with other regions.

There are many differences and similarities between the regions. They have a different topology structure and behave in their own unique ways, but all of them have been continuously growing at varying exponential rates. The regions with the smallest number of ASes are currently the ones growing the fastest. Furthermore, all of them but LACNIC tend to primarily interconnect with other regions. The regions are becoming more interconnected and it is mostly trough peering ASes.

The remainder of the Thesis is organized as follows. In Chapter 2, we present some preliminaries on the Internet organization and present related works on the analysis of the Internet topology. In Chapter 3, we describe the data extraction and analysis methodology adopted in this thesis..The results obtained from the regional analysis of the Internet topology are presented in Chapter 4. We summarize our findings in Chapter 5.

CHAPTER 2

PRELIMINARIES AND RELATED WORK

In this chapter, we introduce the reader to the basics of the Internet architecture. In Section 2.1, we outline the basic Internet structure. In Section 2.2, we discuss data collection and data clean up methodologies. In Section 2.3, we present state-of-the-art techniques for analyzing and modeling the Internet topology; measurements and properties in 2.3.1, link classifications in 2.3.2, node classifications in 2.3.3, and finally evolutionary studies in 2.3.4.

2.1 The Internet Architecture

In this section, we provide details of the organizational structure of the Internet. The first concept necessary to understand the Internet architecture is that of a computer network. A computer network consists of a set of computers, also known as hosts, that connect via links to exchange data. These connections are possible by the combination of hardware and software. The hardware constitutes the physical components of the network such as computers, cables, and a combination of specialized network devices such as hubs, repeaters, bridges, switches, and routers. The software refers to the computer programs that administrate the exchange of data through the hardware, and implement the protocols defined the standards that make device-to-device communication possible.

The Internet could be defined as a global system of interconnected heterogeneous computer networks. The network interconnection is made possible by routers, which are responsible for routing data from one computer network to another. Routers connect hosts within the same and different networks, providing the physical structure of the Internet. They forward data, choose the best route for that data, and exchange updates on network status, routing paths, and other vital information essential for the Internet. In order to perform all routing-related tasks between the large number of diverse networks, routers use the Internet protocol suite, which defines a set of protocols for addressing and communicating on the Internet.

The Internet Protocol (IP) is the principal protocol in the Internet protocol suit. It is responsible for addressing and routing. Each host in the Internet requires a unique address known as an IP address. Routers use IP addresses to deliver data from the source host to the destination host. The IP address space is administrated by the Internet Assigned Numbers Authority (AINA), and it is distributed in a hierarchical way, as it is detailed in RFC 7020 [60]. AINA delegates five regional Internet registries (RIRs) to allocate blocks of IP address space to Local Internet registries (LIRs) such as Internet Service Providers (ISPs) and other networks. The LIRs and other networks need to meet certain requirements in order to be assigned IP space. The main requirements are that they have to be administrated by a single entity and they have to be independent of other networks as described in RFC 1930 [58]. Such networks are referred as *autonomous systems* (ASes).

An autonomous system (AS) is a consolidation of many routers and networks operating under a single administrative authority or domain. An example of an AS could be a company, a university, or an ISP. They connect with each other through private relationships. However, due to the increasing complexity of the Internet, routing is performed by two protocols. Intra-domain routing, which is routing within the AS using an Interior Gateway Protocol (IGP), and inter-domain routing, which is routing among ASes using an Exterior Gateway Protocol (EGP).

ASes autonomously determine their own internal communication policies, and IGP. The most widely used IGP protocols are Open Shortest Path First (OSPF), Routing Information Protocol (RIP), and Intermediate System to Intermediate System (IS-IS). On the contrary, ASes must use the same EGP, which currently is the Border Gateway Protocol (BGP) version 4 described in RFC 4271 [87]. The difference between BGP and any other IGP is that BGP has to consider the

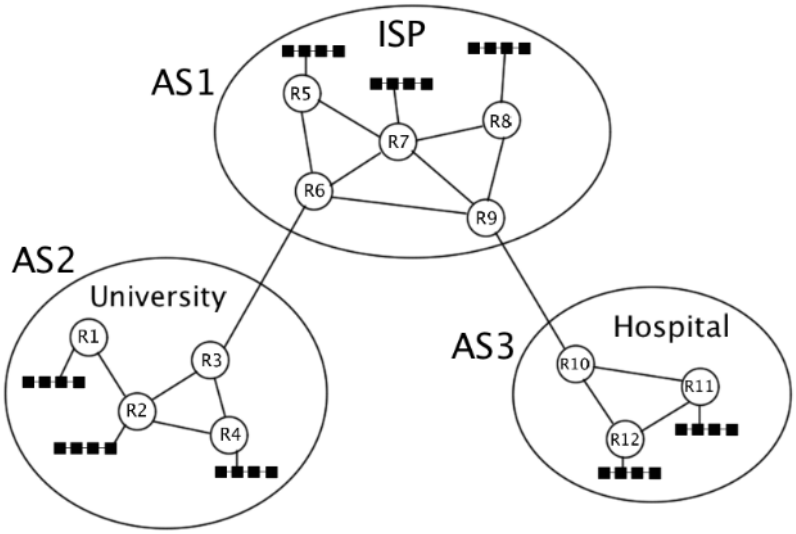
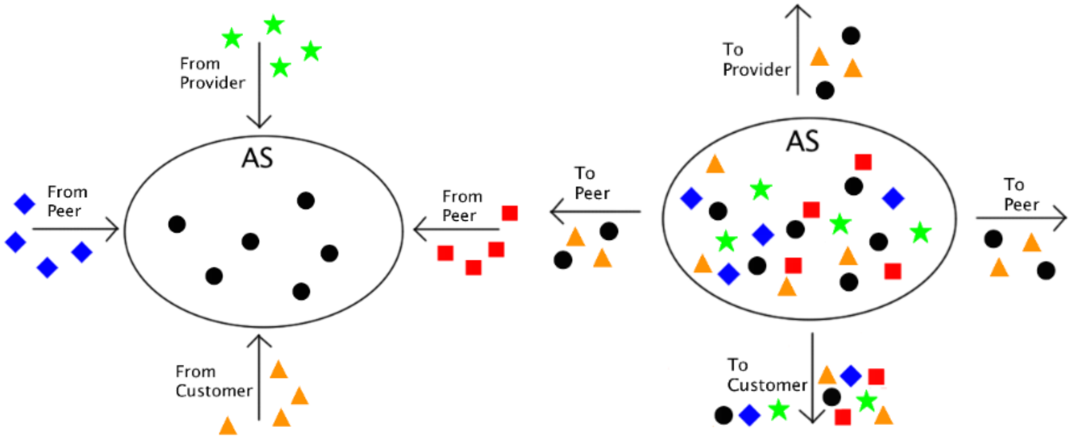


Figure 2.1: Autonomous Systems represent a collection of networks administrated by a single entity. The Internet consists of interconnected ASes.

commercial agreements and policies among ASes instead of only the simple shortest path selection. Figure 2.1 illustrates a simple representation of the ASes and their connectivity. We show three ASes: an ISP, a university and a hospital. Each AS consists of routers (depicted by circles) and hosts (depicted by squares). In this example, routers R3, R6, R9, and R10 will use BGP while the remaining routers will use their own IGP.

The BGP protocol is responsible of connecting the ASes, providing the logical structure of the Internet. BGP routers identify each AS by a unique identification number called autonomous system number (ASN), which is given by RIRs at the moment of registration. Additionally, BGP routers used the classless inter-domain routing notation (CIDR), which is a compact representation of a block of IP address space specified in RFC 4632 [39]. When a BGP router first joins the Internet, it establishes a connection with the directly connected BGP routers and downloads their entire routing tables, which will be permanently maintained in its memory. Each entry in the routing tables contains the IP prefix, the next hop, and the entire AS path to reach that IP prefix. With those tables, it executes an algorithm to



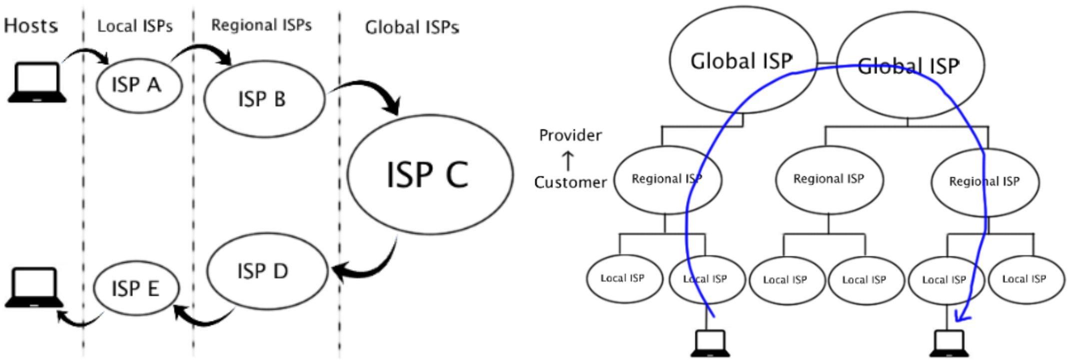
(a) Import Routes (b) Export Routes

Figure 2.2: Routes advertised by AS differ depending on the AS relationship

produce its own routing table taking into consideration the specific policies and cost for each route as well as another table that will be used to send to other BGP routers. After that, the only messages exchanged between ASes are updates for either a withdraw or a new preferred route. In the case those updates contain new information the routing selection algorithm will be executed again.

Deciding what routes to forward or share depends in the economic benefits and the policies between the ASes. The business policies among ASes could be extremely diverse and complex, but it is possible to simplify the role of an AS in the economic relationships in three types. An AS is a provider if it gets paid by another AS to transit data. An AS is a customer if it pays another AS to transit data. Finally, an AS is a peer if it transit data with another AS without charge. Figure 2.2 shows how theses economic relationships interact. Figure 2.2(a) presents an AS with his routes circles, as well as the routes that imports from a peer, a customer, and a provider. Figure 2.2(b) shows the routes that it advertises back. It could be seen that the AS only advertises its customer and its own routes to providers and peers, while it advertises all routes to customers.

The BGP protocol was originally designed to satisfy the economic needs of ISPs



(a) Routing Hierarchy (b) Hierarchical Structure

Figure 2.3: The hierarchical routing structure of the Internet at the AS level.

and to create a hierarchical distributed Internet as shown in Figure 2.3. However, the self-organization of ASes along with the broad differences among the AS types and the constantly evolving AS relationships have altered the hierarchical Internet architecture. The constant evolution of the AS relationships has made it increasingly difficult to capture and characterize the structure of the Internet. For further details on the protocols and devices, and laws that govern the Internet, interested readers are referred to [86, 84].

2.2 Internet Topology Data

2.2.1 Data Collection

Collecting Internet topology data remains one of the biggest challenges in all Internet topology studies. This is because the connections between ASes are formed in an organic way, without centralized control. These connections are based on business relationships and policies that evolve over time.

The first attempts to gather a global overview of the the AS topology were based on BGP monitor routing tables [47, 35]. Such routing tables remain the most popular and reliable sources of information today. Zhang [113] showed that by combining different monitors and taking temporal snapshots, it is possible to create a more complete graph relative to the individual snapshots given by each monitor separately [116, 100]. A complementary method to BGP routing tables is the use of probing and traceroute [48, 74]. This method has been highly debatable; some researchers say that the data is biased and should not be used [1, 115, 114] while others have attempted to consolidate trace route data with BGP data [20]. One of the biggest projects and the main source for traceroute data is the CAIDA ark project [13], previously called skitter, probing from more than 100 monitors around the world.

A third source of AS topology information is the whois databases, which are records controlled by RIRs [66]. This source is the least popular because it is manually maintained by AS operators. In most RIRs, reporting of topology information is only required during the AS registration process. There have been several proposals to improve the data availability [61, 45], but these three sources (BGP tables, probing, and whois databases) are the best information sources to date. A study comparing the advantages and limitation of AS topology data is presented in [75].

2.2.2 Data Incompletness

It is well known to the research community, that the current view of the Internet topology based on the available data sources is far from complete. Chang *et al.* [16] were one of the first to point out that relying exclusively on BGP routing tables results in missing data. They found that while only a small portion of ASes were missing, a significant amount of links were not present. Following works attempting to calculate the amount of missing data and their impact on the AS graph [22, 19]. Most of researchers agree that more than 40% of the links are missing and they are mostly peering relationships [59, 80, 79, 5].

Originally, it was thought that the only reason for missing links was that some peering and backup links were active only for short periods of time, which made them hard to capture by monitors [78]. New studies have showed that the location of the monitors also plays a big role on the amount of captured information. A small set of ASes could be responsible for most of those missing links [52, 54].

2.3 Topology Analysis

While the Internet has been a subject of study since inception, it was not until the mid-nineties with the its commercialization in the United States that an interest about the growth and shape of the Internet increased among researchers. The first study that introduced the notion of an AS level graph was done by Govindan and Reddy [47]. They described a graph where the nodes were representing Internet domains and the links were the route exchanges between corresponding domains. They used snapshots from available BGP monitor data to create the graph and noticed that despite the significant Internet growth over the years, the degree and path distribution remained unchanged. However, the term Internet topology is first introduced in the seminal work of Faloutsos *et al.* [35]. They were the first ones to perform a systematic analysis of the AS topology assuming that the graph constructed with BGP monitor data was accurate and somewhat complete. They discovered that the AS degree distribution obeys simple power laws. This paper is one of the most cited papers when discussing about Internet topology and their data collection techniques, the construction of its graph, and their approaches and analysis led the increased research interest on this area.

2.3.1 Measurements, Properties, and Modeling

Since the discovery of the power laws of the Internet by Faloutsos *et al.* [35], the Internet topology has been analyzed under various metrics and methods. Some of the studies were dedicated to explain the power laws [21, 12] as a tradeoff optimization such as deployment vs. operational cost and overhead vs. transmission cost [34]. Others studied the graph properties of the AS graph and reported metics such as node degree, degree distribution, path length, etc. [103, 73, 101]. They found that the topology was growing exponentially with respect to the number of nodes and links. They also identified the "rich get richer" phenomenon, in which nodes obtain new links with a rate proportional to their degree.

A similar phenomenon known as rich-club connectivity was reported by Zhou *et al.* [120]. This phenomenon refers to the extend to which nodes of high degree are also connected to each other through paths of length less than two. More complex metrics were introduced in later years such as k-core decomposition, k-dense distribution, and weighted spectral distribution [4, 36, 82]. However, all prior analyses are limited to global AS graph views without taking into account the AS types, the link types, and the location of the ASes. The generalization to a global graph topology normalizes the regional variations in the evolution of the Internet.

Other researchers have focused on constructing models and graph generators to accurately represent the Internet topology and predict its evolution. Most of the works create models for node and link addition which maintain graph properties that have been shown to be less volatile such as node degree distribution, clustering coefficient, rich-club connectivity and betweenness [6, 3, 110, 119]. The proposed models have been continuously revised to capture new phenomena that have appeared with the evolution of the Internet [17, 118, 18, 107, 99]. This line of research has led to many conflicting models that fail to satisfy the Internet's evolutionary trends. Some studies have attempted to explain and bridge the models' inconsistencies [69, 83, 38]. Finally, the need for incorporating economic, geographic, technological, and social factors in the analysis of the Internet topology has been noted in [56, 57, 98, 95, 53].

2.3.2 Inferring ASes Relationships

Routing policies at the AS level are governed by the commercial agreements between the organizations that operate the ASes. These agreements are typically unpublished and could change over time. Detailed knowledge of the AS relationships is a critical factor for the understanding of the Internet architecture, performance, resilience, and security. Inferring the AS relationship types is used to determine how traffic flows through the Internet, alternative paths, the cost of various traffic flows, etc. Moreover, it can be used to determine strategies for improving the Internet infrastructure in a systematic fashion.

One of the first AS link classification efforts was done by Huston [64] who classified the business relationships between ISPs in customer-provider, peering, mutual-transit, and mutual-backup agreements. Gao developed an algorithm to infer AS relationships from topology data [40]. She classified the relationships in three types: customer to provider (c2p), peer-to-peer (p2p), and sibling. Her solution relied on the assumption that BGP paths are hierarchical, or valley-free. That is, after traversing a provider-to-customer or peer-to-peer link, the AS path cannot traverse another customer-to-provider or peer-to-peer link. That assumption came from the idea that since AS relationships are commercial agreements, ASes won't transit traffic for free collecting the cost. Her algorithm also relies on node degree, assuming that the ASes with the highest degree lie on the top of the AS hierarchy, while nodes with similar degree are likely to be peers. She classified 90.5% of the links that existed during the year of the study as c2p, 8% as p2p and 1.5% as siblings with an achieved accuracy of 96.1%, 89.22%, and more than 50%, respectively, according to the 6.3% of the data they could validate through AT&T and WHOIS lookup services.

Subramanian *et al.* [102] formalized the problem of inferring AS relationships and concluded that it is a NP-problem if the valley-free assumption did not hold. The latter was proved by Di Battista *et al.* [26]. Dimitropoulos *et al.* [27] proposed a solution based on a MAX-2-SAT problem formulation. However, the proposed implementation does not complete in a practical length of time for the current size of the topology. Subsequent algorithms included the assumption that all Tier 1 networks should form a clique, but most of them still seek to maximize the number of valley-free paths [108, 50, 113, 79]. The latest algorithms [72, 43] rely in three assumption: an AS enters into a provider relationship to become globally reachable, there exist a tier 1 clique, and there is no cycle of p2c links. They only looked for two types of relationships c2p and p2p, and their results have an accuracy of 99.6% for c2p and 98.7% for p2p, but they validated 34.6% of their results, which is the largest validation data to date.

2.3.3 Classifying ASes

An AS network could be an enterprise, a school, a hospital, or it could be a small or global ISP. Knowing the AS type enables the analysis of trends and behaviors in the AS topology Unfortunately, AS type information is not readily available from AS information sources. Several techniques have been developed that infer the AS type from the AS graph. Govindan and Reddy [47] presented one of the earliest AS type classifications, when the Internet topology was predominately hierarchical. Their proposed method exploited the node degree to classify ASes in four hierarchical levels.

Subramanian *et al.* [102] claimed that node degree alone is not sufficient for an accurate AS classification. He exploited the inferred link relationships to classify ASes to five levels: dense core (level 0), transit core (level 1), outer core (level 2), small regional ISPs (level 3), and customers (level 4). Dhamdhere *et al.* [23] combined node degree with link relationships over a period of ten years to classify the ASes in four business types: enterprise customers, small transit providers, large transit providers, and content/access/hosting providers. Dimitropoulos *et al.* [28] used the registered name of the ASes in RIRs and machine learning to classify ASes to eight categories: ISPs, Internet Exchange Points (IXP), network information centers, universities/schools, military networks, governmental networks, hospitals, and companies. The same method was used by Baumann *et al.* [7], but they enriched the classification to 18 different industry types. However, both studies were only able to classify 50 to 60 percent of the total number of ASes.

2.3.4 Evolution of the Internet Topology

Several studies have analyzed the evolution of the AS graph over time. One of the first works that looked into the Internet evolution was by Norton *et al.* [77]. They described the Internet in three tiers and reasoned that economic and competitive interests are the primary driver of AS peering. They discovered that historically ASes of the same tier tend to peer. Economides *et al.* [31] studied the evolution of the Internet backbone. Their study concluded that the backbone remains robust and diverse throughput the years, without the tendency of collapsing to a few nodes. Oliveira *et al.* [81] performed one of the most complete analyses of the AS graph evolution. They found that most of the births and deaths of ASes occur at the edge of the topology.

Edwards *et al.* [32] examined how the eight most commonly-used topological measures change over eight years. They concluded that the distributions of most of the measures remain unchanged, except for the average path length and clustering coefficient. Leskovec *et al.* [71] showed that the topology is becoming more dense with time. Gill *et al.* [42] attributed the increasing topology density to content providers that tend to bypass ISPs by deploying their own wide area networks. They described the change from the strict hierarchical Internet structure observed in the early years to a highly meshed p2p structure as the *flattening of the Internet*.

Further works confirmed that the Internet was flattening [46, 24, 55, 70, 117]. Dhamdhere *et al.* [23, 25] showed that content, access, and host providers are the most active ASes with respect to adding/deleting links in the AS topology. The Internet flattening was attributed to those providers. They also investigated the geographic distribution of content, access and host providers and found that there are increasing in number at a larger rate in Europe compared to America. Their findings suggested that the topology could be different in different regions of the world, which is one of the ideas that we explored in this thesis.

CHAPTER 3

DATA SOURCES AND PROCESSING

In this chapter, we discuss the data sources used in our study. We further present the methods applied to combine data from multiple sources and filter erroneous data.

3.1 Data

3.1.1 Data Sources

To study the Internet evolution, we created AS graphs monthly from January 1998 to January 2015. An AS graph consisted of AS nodes and AS links with the following attributes.

- AS node
    - AS number
    - AS type (EC, CAHP, STP, LTP)
    - Geolocation (region, sub-region, country)
- AS link
    - Incident AS nodes
    - Link type (c2p, p2p)

The Internet topology graphs were constructed using two public data sources: the UCLA Internet AS-level topology archive repository [104], and the CAIDA AS relationships dataset [14]. The two repositories were selected because they incorporate an extensive network of BGP monitors and because they are the only public sources that maintain historic information from as early as 1998. These sources have also been used in most prior Internet topology studies.

The UCLA Internet AS-level Topology Archive was our primary source for inferring the AS nodes and AS links. This repository processes raw BGP data records daily from several BGP data collectors including RouteViews [96], RIPE RIS [88], PCH [85], and Internet2 [65]. The collectors peer with BGP routers and record every BGP path advertisement sent or received by the routers. The UCLA gets the routing tables of each one of the collectors, as of now 133 different collectors, and with the paths from the tables they create two topologies. One topology using only the entries to IPv4 addresses while the other uses only the entries to IPv6 addresses. We only used IPv4 topologies since they would reflect the historical changes on the Internet.

We further acquired the CAIDA AS relationship dataset to assign a type to each AS link [14]. In this dataset, the AS links are classified into two types: c2p links and p2p links. The link type is inferred from raw BGP paths advertisements using the algorithm described in [72]. We decided to use the CAIDA dataset because the link inference algorithm was shown to have an an accuracy of 99.6% for c2p links and 98.7% for p2p links [72].

3.1.2 Data Processing

Inferring the basic AS graph. The basic AS graph, which consists only of AS nodes and AS links, was derived from the UCLA data repository. As the graph topologies are reported daily, the first step was to group the daily data on a monthly basis. One problem with the UCLA dataset is that it can contain falsely advertised links due to configuration mistakes in routing tables, path poisoning or router failures. This transient events last only a few hours. Therefore, in order to remove possible false paths, we eliminated AS links that appeared only once over the span of a month.

Inferring the AS relationship type: The next step in processing the topology data was to incorporate the CAIDA data for inferring the AS relationship type. We did not filter the CAIDA dataset because filtering of erroneous information is part of the inference algorithm in [72]. However, we had to infer the link type for about 10% of the links of the basic AS graph because the CAIDA dataset did not contain any information on those links.

For links of unknown type, we applied the following inference methods[1]. For each link, we first observed the number of peers of the incident ASes and the node degree. If the incident ASes had two or more peers in common and a similar node degree[2], the link connecting them was likely to be a peering link, so it was classified as p2p. This classification was also used in Gao's algorithm [40], and it was validated by Zhou, when he found that the Internet topology shows a rich-club connectivity[120].

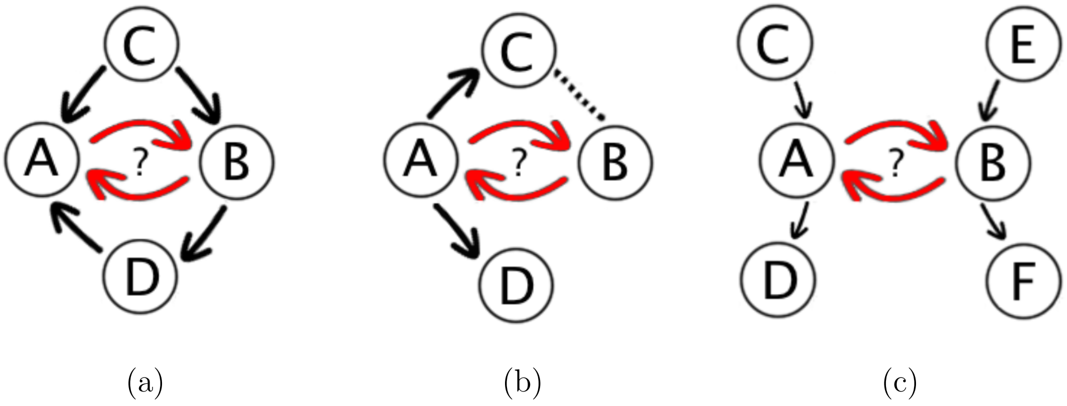
For the rest of the links, we applied the following method under the assumptions that the topology is valley-free and that there are no c2p cycles [40, 72]. First, we assigned the AS with the higher node degree AS A to be the provider and the AS with the lower node degree AS B as the customer. We then tested if this assignment satisfied the assumptions by making sure that neither A nor any of its peers and providers were customers or peers of B or any of B's customers[3]. Then we switched the provider and customer roles and tested AS B under the same conditions. Three results were possible. AS A and AS B both failed the tests. This means that the link between A and B could be a peering link, but it could also mean that the data is unreliable, so we discarded such links to decrease the number of errors. Another possible outcome is for AS A and AS B to both passed the tests. This means that there is not enough information to infer the link type, so the link was discarded, with one exception. If AS B has only one neighbor and it is AS A, then AS A was

---

[1] Note that the algorithm in [72] could not be applied without access to raw BGP path advertisements.

[2] We assume similarity if the $min/max \geq .8$.

[3] We compare ASes no more than two hops away from AS A and AS B because considering ASes at further distance did not yield any significant improvement



(a) (b) (c)

Figure 3.1: Examples of c2p link inference, (a) AS B is provider, (b) AS A is provider, (c) link discarded.

assigned to be provider of AS B in a c2p link. The same applied for the reverse condition. Finally, the third possible outcome is that only one AS passed the test. In this case, the AS that passed the test was assigned as the provider and the otherAS was assigned to be the customer in a c2p link.

Figure 3.1 shows three simple cases for the unknown link AS A–AS B. The arrow is pointing from the provider to the customer, while the dotted line indicates peering, and it assumes that the ASes have multiple stub customers not shown in the figure. In Figure 3.1(a), by selecting AS A as the provider we will produce a cycle A-B-D-A. This c2p cycle is produced by AS D, which is the provider of AS A, while being a customer of AS B, which makes the test fail. A c2p cycle should not exist for routing to converge [41]. On the other hand, AS B passes the test, so it is assigned as the provider while AS A is assigned as the customer on a c2p link. In the example of Figure 3.1(b), AS A passes the test, but AS B fails as a provider. This happen because AS C is a peer of AS B and has AS A as the provider, which is assumed to be the customer. This is not a c2p cycle, but it makes the topology non valley-free. If AS B were the provider of AS A then AS C could send data uplink by using its peer instead of its provider producing a valley [40]. Therefore, in this case AS A is assigned as the provider and AS B is assigned as the customer

Table 3.1: The AS graph for three sample years

Year	UCLA		CAIDA		DELETED		FINAL	
	Nodes	Links	Nodes	Links	%Nodes	%Links	Nodes	Links
1999	5859	9670	5676	12190	%0.64	%1.74	5872	12871
2006	23856	69088	23296	79849	%0.74	%3.12	23681	87792
2014	48587	204704	46063	253207	%0.78	%3.90	48223	297574

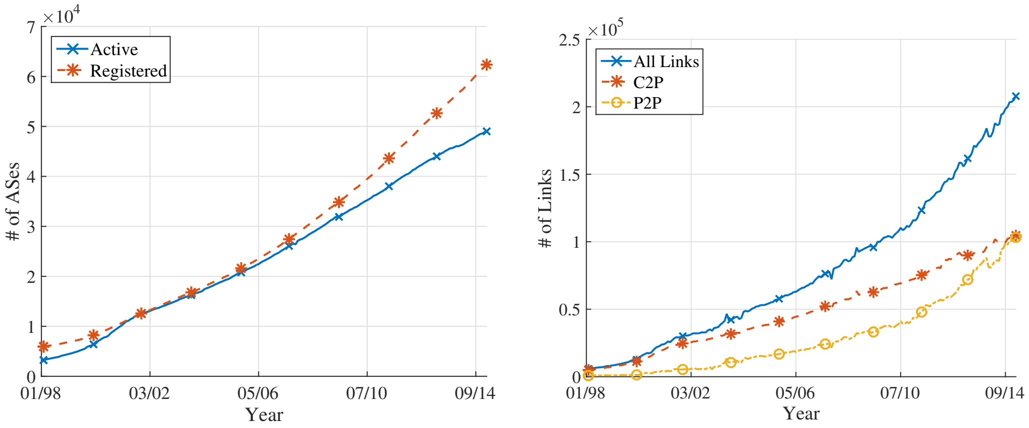
in a c2p link. Figure 3.1(c) shows an example where the information is not enough to infer the customer and provider, so the link is discarded.

Table 3.1 shows the percentage of AS nodes and AS links that were deleted from the AS graph topology during the AS graph inference process, for a sample of three years. We observe that a relatively small amount of AS nodes and AS links could not be properly classified by merging the UCLA and CAIDA repositories and the application of the aforementioned inference algorithms.

In Figure 3.2(a), we show the total number of ASes that were reported in the UCLA repository over the last 17 years. These are labeled as *active* ASes because they are actively participating on the Internet by sending BGP path advertisements. For comparison, we also report the number of ASes registered at IANA, including the ones that may no longer be in use. Only active ASes are considered in the evolution analysis of the AS graph. Figure 3.2(b) shows the total number of links in our topology graph over the period of 17 years, based on the UCLA repository data. The graph shows only the number of links (c2p and p2p) after the AS link classification algorithm has been applied.

3.1.3 Data incompleteness

Several works have demonstrated that the Internet topology data available in public repositories is incomplete [16, 113, 22]. Although almost all active ASes are reported (see Figure 3.2(a)), a significant portion of p2p and backup links at



(a) Number of AS nodes

(b) Number of AS links

Figure 3.2: Number of AS nodes and AS links from 01/98 to 01/2015.

the edge of the Internet is missing [19, 22]. This is because ASes do not advertise all their links, but restrict BGP path advertisements to preferable paths. Despite this limitation, the publicly available datasets capture the dynamics of primary links that remain active most of the time. The often-missing transient links and backup links at the Internet's periphery play a marginal role in the overall Internet structure from an evolutionary study point of view. For this reason and due to the lack of any alternative reliable source of historical data, we focused our study on the links reported by the UCLA and CAIDA repositories.

3.2 Topology Attributes

3.2.1 AS Geopolitical Information

One of the key goals of our study is to understand how the Internet has evolved at different regions of the world. This is in contrast with most prior works that analyze the AS topology as a single highly-connected graph. A natural selection for the division of the world into regions is to adopt the IANA specification of the five RIR: AFRINIC [89], APNIC [90], ARIN [91], LACNIC [92], RIPE NCC [93]. This is because each RIR constitutes a single administrative domain for the registration of ASes and closely resembles the geopolitical division of the world into continents.

As a first step, we associated each AS node with the RIR that it was first registered. Figure 3.3(a) shows the percentage of ASes per RIR, over a period of 16 years. We can observe that ARIN and RIPE dominate the Internet by consistently owning more than 80% of the ASes over the years. However, the difference between those two has been srinking. In 1998, ARIN accounted for more than 50% of the ASes, while RIPE accounted for approximately 25% of all ASes. However, by 2009, RIPE surpassed ARIN in the number of ASes. Nowadays, RIPE accounts for 46% of all ASes while ARIN accounts for 35% of the ASes. The rest of the RIRs saw small increases in the percentage of ASes. AFRINIC for example, has the smallest percentage of ASes by 2014 equaling to 1% of the total.

We further divided the RIRs into 13 subregions. The subregions closely followed the United Nations macro-geographical regions and subregions division [105], which is primarily based on geographic position, population, and economic criteria. The UN geo-scheme was slightly modified to account for subregions with very little Internet presence. In particular, we merged data from Western Africa and North Africa, as well as data from Middle and Eastern Africa with Southern Africa. Furthermore, Oceania was left undivided. Figures 3.3(b)-3.3(f) show the AS growth in each subregion, categorized by RIR.

In addition to region and subregion, each AS was tagged with the country in which it was registered. This information was obtained by examining the RIR delegation files. The country information was used to explore further details in some of interesting cases. The list of countries that belong to each subregion are listed in Appendix A.

3.2.2 AS Classification

Several ways have been proposed for categorizing ASes to different types. The most widely used convention is to group ASes into three categories: stub ASes, multi-homed ASes, and transit ASes. In this classification, stub and multi-homed ASes are customers that do not carry any transit traffic. All ASes used for routing traffic are classified as transit.

Other classifications further extend the AS types by function, business type, and the services they provide [102, 28, 23]. These are meant to provide a more descriptive characterization of the nature and function of each AS. In our analysis, we adopted the classification in [25].

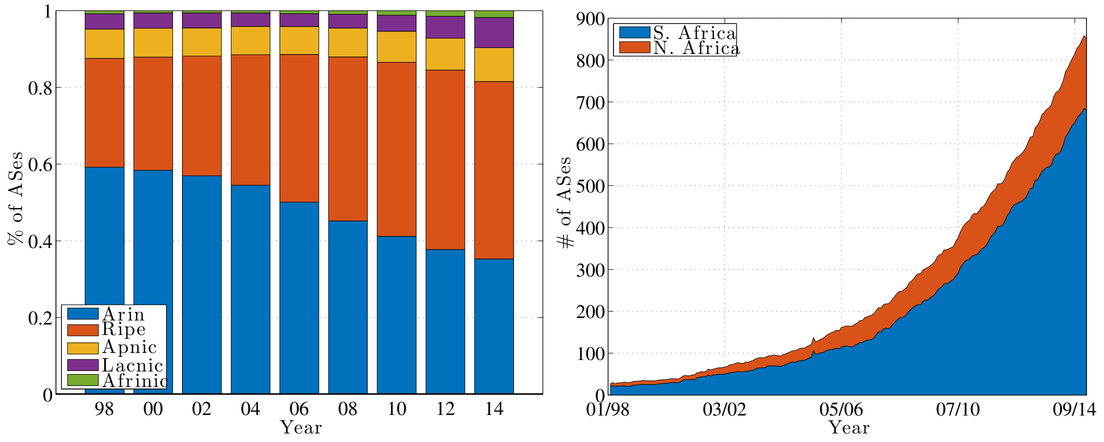
Figure 3.4(a) shows the number of ASes per type, using the conventional classification. We observe that stub and multi-homed ASes have grown at similar rates. Transit ASes are the least represented type, growing at the lowest rate. Figure 3.4(b) shows the number of ASes born each month by type. An AS is considered to be born if it has not appeared in any BGP advertisement before. It can be seen that new ASes are more likely to be born as stubs. That is, they become customers of a single transit AS. However, with the progress of time, stub ASes connect to more than one transit ASes (for connectivity, performance and reliability reasons) and thus, become multi-homed. One problem with the traditional classification is that it offers little to no information about peering relationships, which is essential to understand the dynamics of the Internet topology.

To tackle the limitations of the traditional AS classification, we adopted the business classification approach by Dhamdhere [25]. This classification groups ASes according to their business role to the following four types: Large Transit Provider (LTP), Small Transit Provider (STP), Content Access and Host Provider (CAHP) and Enterprise Customer (EC). LTPs are international ISPs with large number of AS customers and wide geographical presence. STPs are regional and national ISPs. CAHPs are small and local ISPs, or business that do not offer transit services but host content (e.g, CDN nodes). Finally, ECs are end-user organizations such as universities, and companies, government agencies, etc. The classification is performed by applying the following empirical rules:

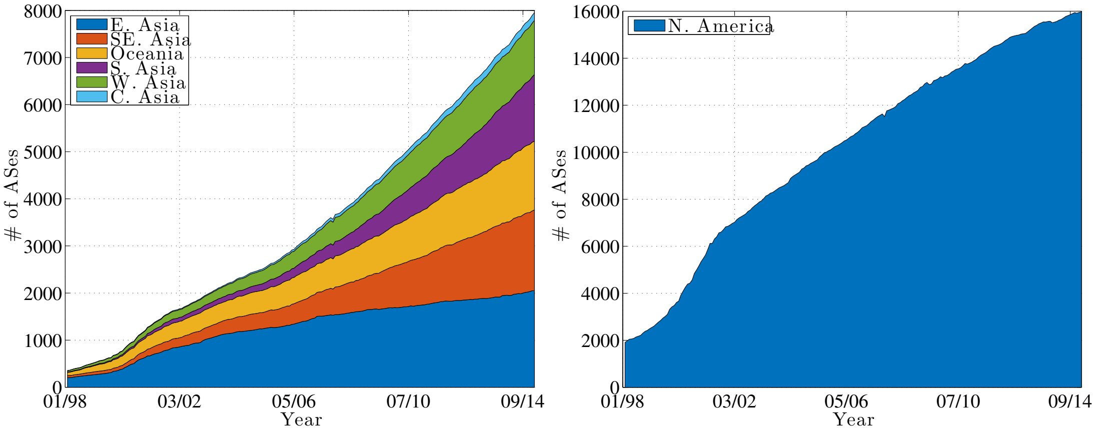
$$\begin{aligned}
EC: &\quad C < 2.1, R \leq 1 \\
STP: &\quad 2.1 \leq C < 180, \quad R < 4 \quad \text{and} \quad 48 \leq C < 180, \quad R \geq 4 \\
LTP: &\quad C \geq 180 \\
CAHP: &\quad C \leq 2.1, \quad R > 1 \quad \text{and} \quad 2.1 \leq C < 48, \quad R \geq 4
\end{aligned}$$

In the rules above, C denotes the average customer degree and R denotes the average peer degree. This classification rule was developed by manually selecting a training set of 50 ASes from each AS type and applying machine-learning to optimize the boundaries between different AS types. The algorithm resulted in a classification accuracy between 76% to 82%, depending on the AS type.

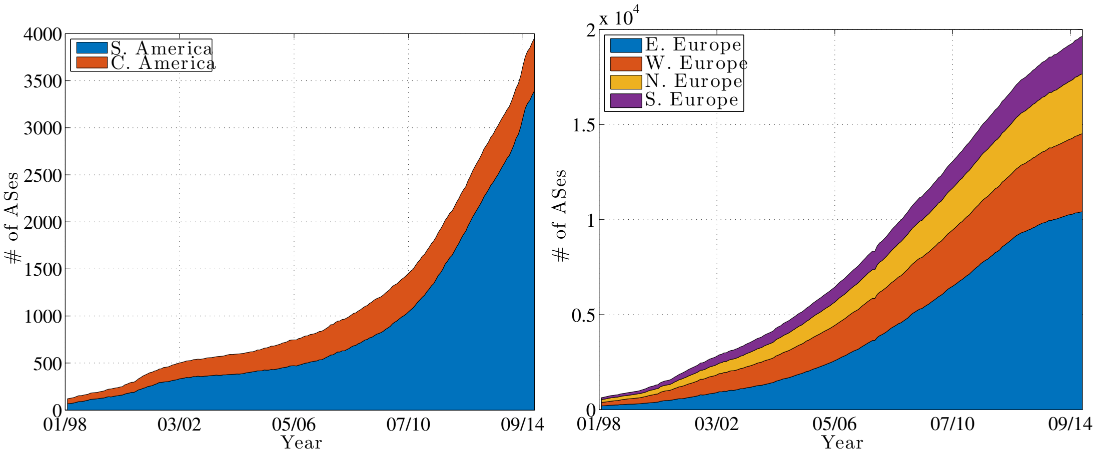
Figure 3.5 shows the evolution of ASes by business role over the course of our study period. Similar to the results reported in [25], we observe that the number of ECs continues to grow at a larger rate than the rest of the AS types. Moreover, the CAHPs growth outpaces the growth of STPs. In our regional analysis presented in Section 4.1, we show that the global growing trends are not universal, but the rate of growth of AS types varies within each region.



(a) All regions (b) AFRINIC

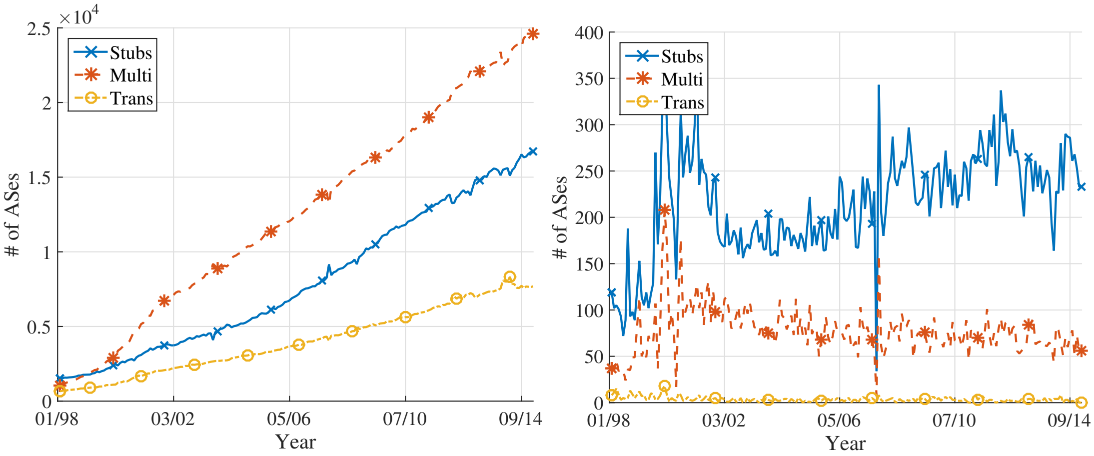


(c) APNIC (d) ARIN



(e) LACNIC (f) RIPE

Figure 3.3: AS growth by RIR and by geographic subregion from 01/1998 to 01/2015.



(a) Num. of ASes per type

(b) Born ASes per type

Figure 3.4: Traditional AS classification.

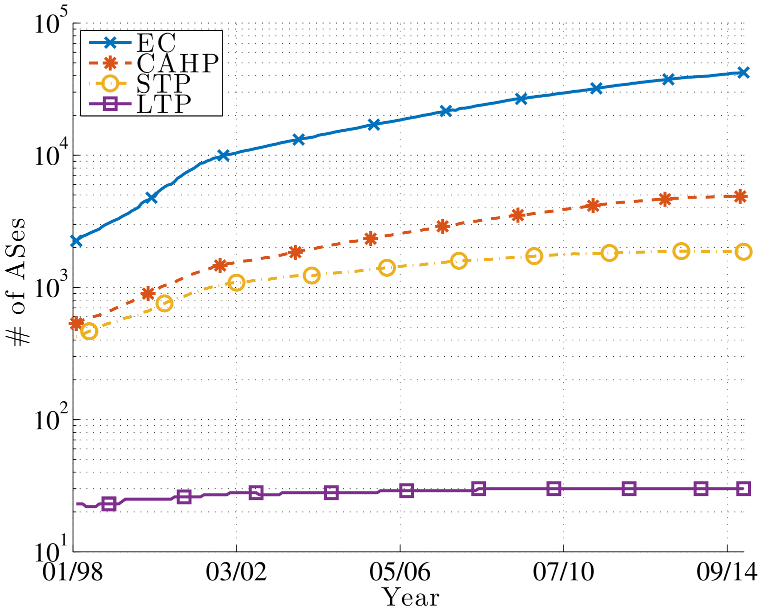


Figure 3.5: Evolution of ASes by business role

CHAPTER 4

RESULTS

In this chapter, we discuss the results of the regional Internet evolution analysis. In Section 4.1, we present the evolution of the AS topology for each world region. We analyze the interconnections between different regions in Section 4.2. Finally, we compare the region differences and similarities in Section 4.3.

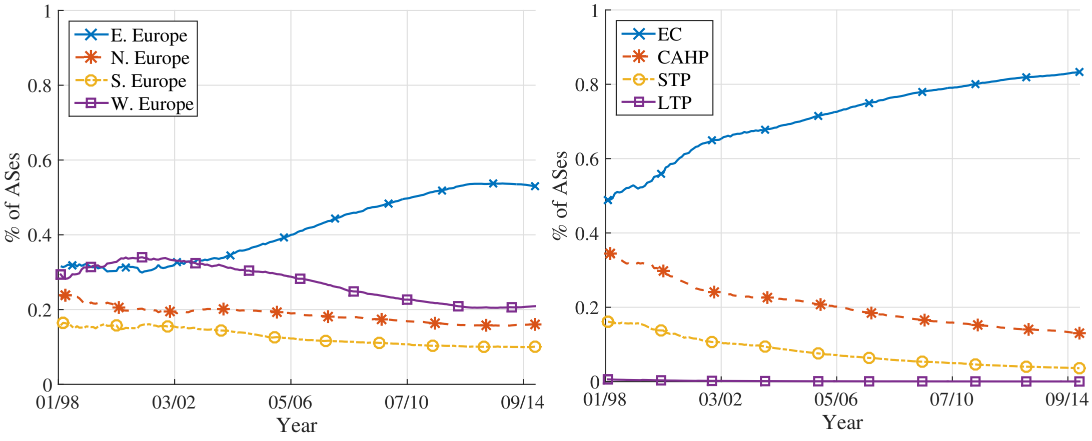
4.1 Evolution per Region

In this section, we analyze the AS topology evolution per region. We present our results by order of importance on the overall Internet topology.

4.1.1 RIPE

The Reseaux IP Europeens (RIPE) Network Coordination Centre represents the ASes present in Europe. It consists of four subregions, namely Easter Europe (E.E.), Northern Europe (N.E.), Southern Europe (S.E.), and Western Europe (W.E.). In 2009 became the region with the largest number of ASes. The majority of them are located in Eastern Europe, followed by Western, North, and Southern Europe, as shown in Figure 4.1(a). The AS topology in RIPE is highly interconnected and structurally flat, as we show in the analysis that follows.

ECs are the most common type of AS in RIPE. As shown in Figure 4.1(b), ECs make up more than 80% of the ASes today, while only 50% of ASes are classified as ECs in 01/1998. Eastern Europe hosts the majority of ECs, with most of them being in Russia, Ukraine, and Poland (see Appendix B). Despite being the most common type of ASes, ECs are only responsible for 32% of the links in the region, and most of those links are connections to STPs. Moreover, in most of the



(a) RIPE by subregion

(b) RIPE by AS type

Figure 4.1: Percentage of ASes in RIPE, when divided by subregion and type.

countries, ECs usually connect to another ASes in the same country For example in January 2015 that was the case for 74% of the EC links in RIPE. An explanation for that could be that is cheaper to connect to a local provider, or because country policies facilitate connections within countries. Finally, we fitted the EC growth to an exponential model and found that ECs have been growing in RIPE according to the following model

$$y = 1782e^{(0.01195x)}. \tag{4.1}$$

CAHPs are the next most common AS type in RIPE and the main contributor to the flattening of the AS topology. Even thought in 2015 they accounted for less than 20% of the ASes in the region (see Figure 4.1(b)), more than 55% of the existing links are incident to at least one CAHP. Out of those links, 88% are peering links (see Figure 4.2), mostly from one CAHPs to another CAHP or an STP. This increase in the number of peering links started in 2002, and it is possibly a result of the creation of the EU Telecoms Framework, which forms a set of common standards and regulations for telecommunication industries of all members of the European Union [33].

One difference between CAHPs and ECs in RIPE is that there is almost the same amount of CAHPs in Western and Easter Europe, despite the larger number of ASes present in Easter Europe. However, the biggest difference between CAHPs and ECs is that links between the former are not constrained to country borders. CAHPs connect without taking in consideration countries; in 2015, more than 70% of the CAHP links were to an AS in a different country, indicating the evolution of peering cross-country peering relationships. Several researchers have attributed this to the dominant presence of IXPs in Europe that encouraged the establishment of settlement-free peering connection among ISPs [5, 49, 109].

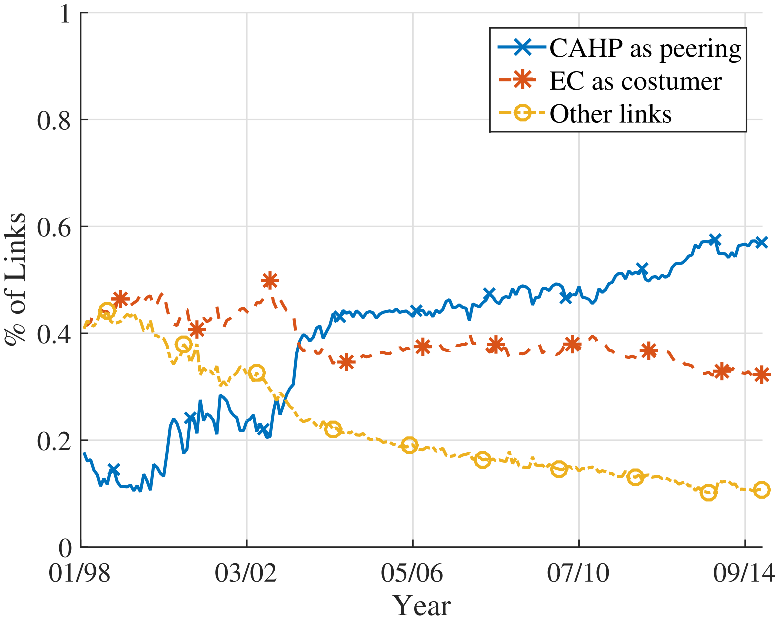


Figure 4.2: Common link connections in RIPE.

STPs serve as a middleman between ECs and the rest of the region. They connect mostly between ECs and CAHPs, and like CAHPs they connect freely between countries. In 2015, 75% of them connected to an AS in a different country. Along with CAHPs, but to a smaller degree, STPs create the highly-interconnected structure that characterizes RIPE.

As expected, the least represented type in RIPE are LTPs. Specifically, there are nine LTPs, including Retn, TransTelekom, VimpelCom, Interoute, and Telia-

Sonera. They do not represent a big percentage of the links in RIPE, but they are very well connected within the region and internationally. This lack of LTPs and the highly connected number of CAHPS, is the reason why we can say that RIPE today has a flat topology.

4.1.2 ARIN

The ASes located in North America are registered to the American Registry of Internet Numbers (ARIN). More than 90% of the ASes in this region are hosted by the United States. It could be said that ARIN is historically the most important region, since it has the oldest ASes. Moreover ARIN hosted the largest number of ASes until 2009, when it was displaced from that position by RIPE. ARIN still plays a significant role on the Internet. Yet by analyzing its topological structure, we can see that it has largely maintained its hierarchical organization to different tiers of transit providers.

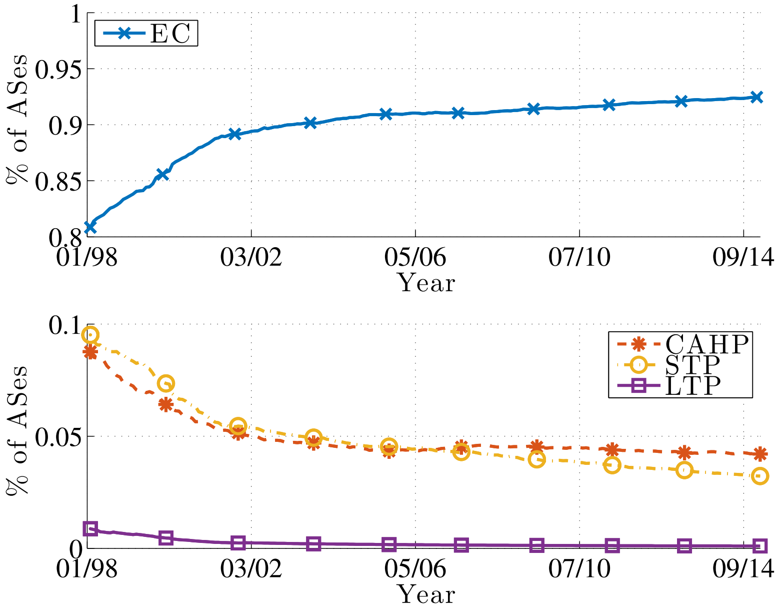


Figure 4.3: Percentage of ASes by type.

Most of the ASes in ARIN are ECs, and since the early 2000s they represented more than 90% of the ASes in the region as seen in Figure 4.3. Furthermore, most of the links in ARIN, around 78% in 2015, connect ECs to transit providers, especially to an LTP. Figure C.2 in Appendix C shows the percentage of links between the different AS types.

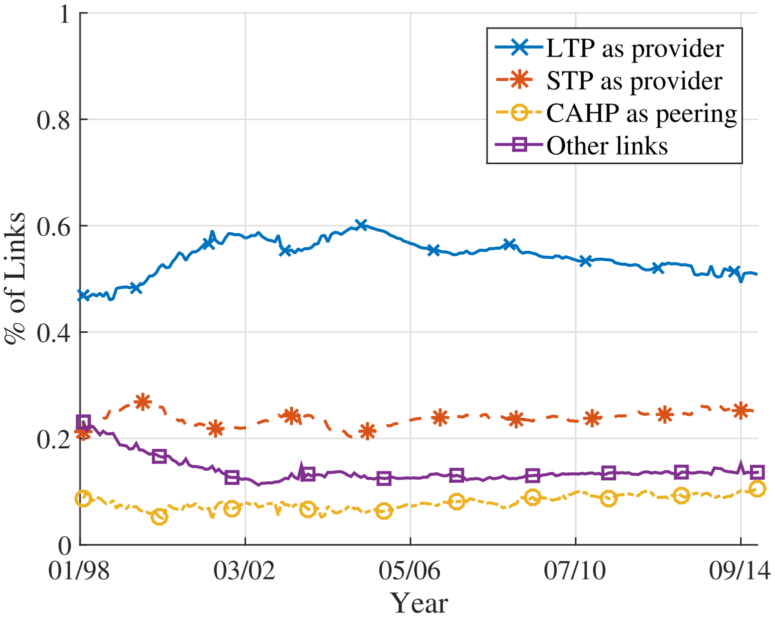


Figure 4.4: Percentage of links by type.

On the other hand, less than 5% of the ASes are classified as CAHPs. This number is very close to the number of STPs. An CAHP is distinguished from an STP due to the large number of peering links. However, in ARIN, less than 10% of the links are p2p links at any point in time. Additionally, CAHPs preferred connections are directly to ECs, or other CAHPs rather than LTPs or STPs.

As noted in Figure 4.3, less than 5% of the ASes are STPs, but they manage to influence the topology structure considerably. They provide access to a large number of ECs by accounting for about 25% of the links to ECs, (see Figure C.2). They are also connected as customers to LTPs and as providers or peers to CAHPs, functioning as an intermediary between the ECs and the rest of the Internet.

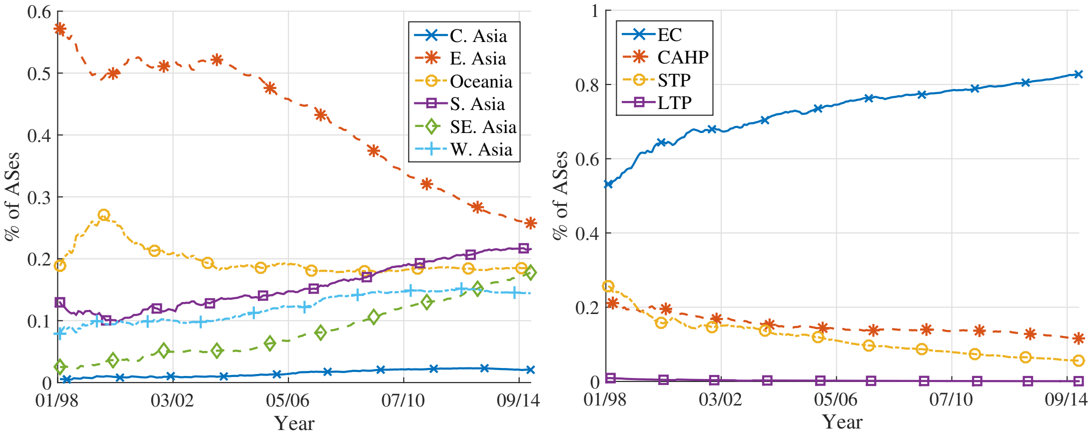
The presence of LTPs in ARIN creates a unique AS topology that is drastically different than the structure we observed in RIPE. The 18 ASes that are classified as LTPs by 01/2015 account for 52% of the links present in 01/2015, and nearly 60% in past years, (see Figure 4.4). In fact, 73% of the links created by LTPs originate from five companies: Level 3, Cogent, AT&T, Qwest, and Verizon. The biggest LTP used to be Verizon until 2009 when AT&T took its place, but that only lasted three years. At present, Level 3 is the biggest LTP and has a degree of 4170. The LTPs provide access to STPs, CAHPs, and large ECs, then STPs provide access to the rest of ECs, forming a hierarchical structure defining the ARIN topology.

4.1.3 APNIC

The ASes located in Asia and Asia Pacific are registered to the Asia-Pacific Network Information Center (APNIC). APNIC has the largest number of subregions, namely Central Asia (C.A.), East Asia (E.A.), Southern Asia (S.A.), South East Asia (SE.A.), Oceania, and Western Asia (W.A.). Most of the subregions host a similar number of ASes (see Figure 4.5). The leading subregions are E.A with countries like South Korea, Japan, and China, and SE.A with Indonesia, Thailand, and Singapore. However, the two countries with the largest number of ASes are Australia and India, which are in Oceania and S.A (see Appendix D). The APNIC region primarily hosts small and regional ISPs, which provide transit services providing to ECs.

Similar to other regions, the number of ECs have grown from around 60% in 2000 to more than 80% in 2015. They are commonly customers of either STPs or CAHPs. Contrary to RIPE and ARIN, most of the ECs in APNIC establish business relationships with other ASes across country boundaries. We believe that geography (several countries are islands, large sparsely populated lands) and economy (some countries in the area are far more technologically advanced than others) play a pivotal role for motivating new ECs connect to providers across borders).

The role of the major providers in this region is assumed by STPs and CAHPs. STPs are the preferred providers, but CAHPs have gradually changed the situation,



(a) APNIC by subregion (b) APNIC by AS type

Figure 4.5: Percentage of ASes in APNIC when divided by each subregion and each type.

as peering with them became popular around 2008. Until 2008, the AS topology APNIC followed the typical hierarchical model. A slow but significant increase in peering relationships is observed since then, as it attested by the percent increase in CAHP peering relationships (see Figure 4.6). However, the links between STPs, CAHPs, and EC have not been enough to fully connect the region, resulting in more inter-connecting links than intra-connecting ones

There are only three ASes that classify as LTP in APNIC, which are owned by Pacnet, Korea Telocom, and LG Uplus, two of them in South Korea, and one in Hong Kong. Despite the smaller number of interconnections of this LTPs relative to those present in ARIN, the LTPs of APNIC have a presence in 60 out of the 70 countries in this region.

4.1.4 LACNIC

The ASes located in Latin American and Caribbean countries are registered to the Latin America and Caribbean Network Information Center (LACNIC). LACNIC consists of two subregions, namely Central America and Caribbean (C.A.),

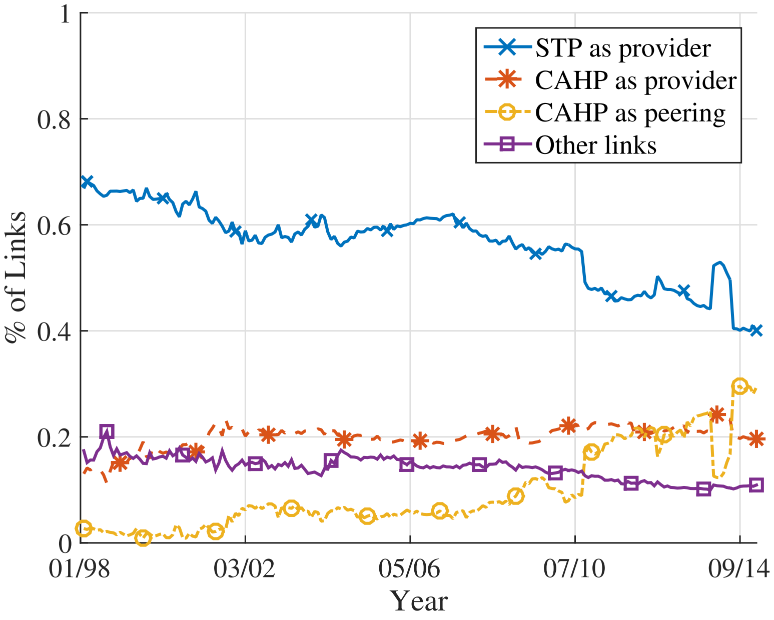
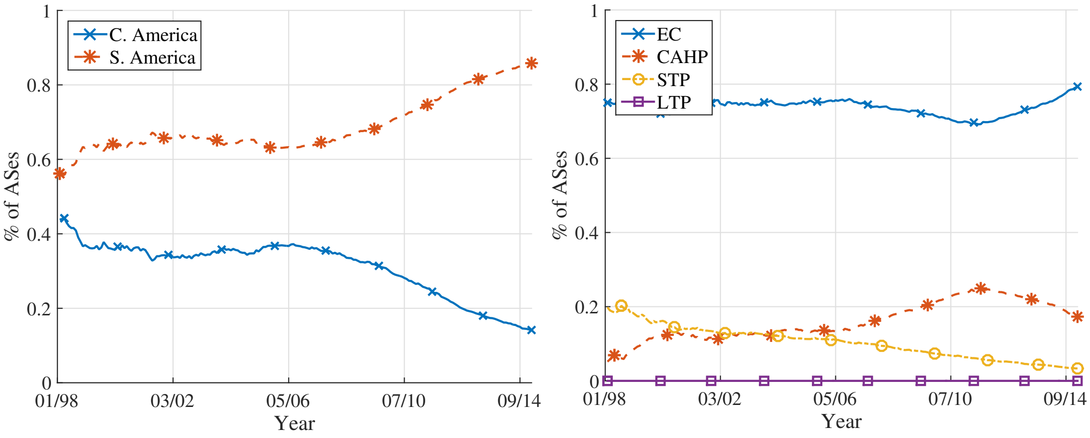


Figure 4.6: Common link connections in APNIC.

and South America (S.A.). In 2015, LACNIC accounted for less than 10% of the ASes on the Internet. In Figure 4.7(a), we show the percentage of ASes in each subregion as a function of time. We observe that the gap between the Central and South America widens continuously. Today more than 80% of the ASes in this region belong to South America, mostly due to the Internet development in Brazil.

As seen in Figure 4.7(b), ECs are the most prevalent type in LACNIC. In 2015, they accounted for about 80% of the ASes in the region and have maintained close to the percentage from the beginning of the period in our data analysis. An interesting phenomenon is observed with respect to the STPs and CAHPs. In 1998, the number of STPs was significantly larger than the CAHPs. However in 2006, the number of CAHPs has surpassed the STPs, with the difference continuously growing until 2011. In conjunction with the explosion of the peering relationships, as seen in Figure 4.8, one can conclude that the Internet in LACNIC has been flattening.

From a country point of view, most of the countries in the region have a small



(a) LACNIC by subregion

(b) LACNIC by AS type

Figure 4.7: Percentage of ASes in LACNIC when divided by each subregion and each type.

number of ASes, which are connect to ASes in different countries (see Appendix E). Brazil accounts for 60% of the ASes in the region and over 95% of the CAHPs. Between 2006 and 2008, there was a big change in the way ASes did business in Brazil causing a huge increase in peering links. We are unaware of the reasons for that change, but the effects of it have made the LACNIC AS topology highly-peered and non-hierarchical. Nowadays, more than 60% of the links in the region peer with an CAHP and 20% of the c2p links list a CAHP AS as the provider. This indicates that the traffic that flows through 80% of the links is transited by a CAHP.

4.1.5 AFRINIC

Countries within the African continent are registered under the Africa Network Information Center (AFRINIC). This is the smallest region to this day, having less than 2% of the Internet ASes, which is slightly more than 800 ASes in the beginning of 2105. Africa is divided to two subregions, namely Northern Africa, and Southern Africa, with countries in each area hosting on average 15 ASes. Only

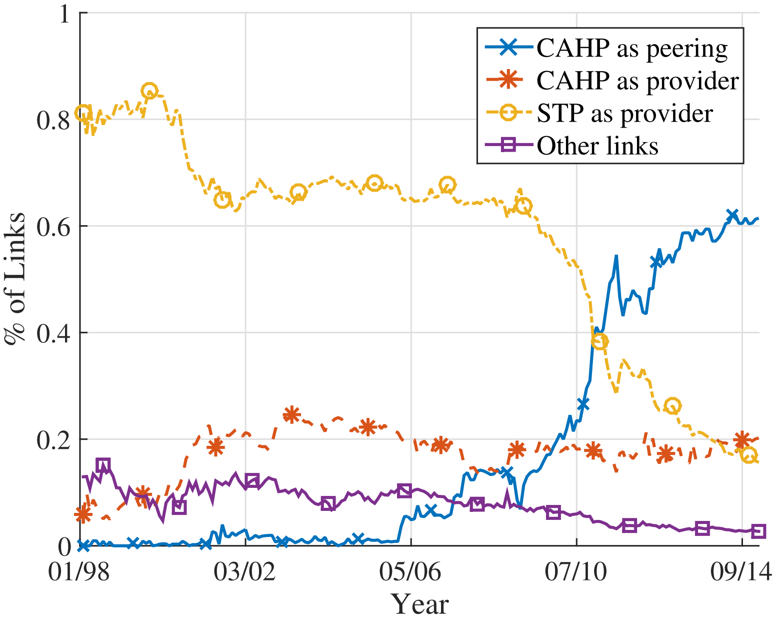
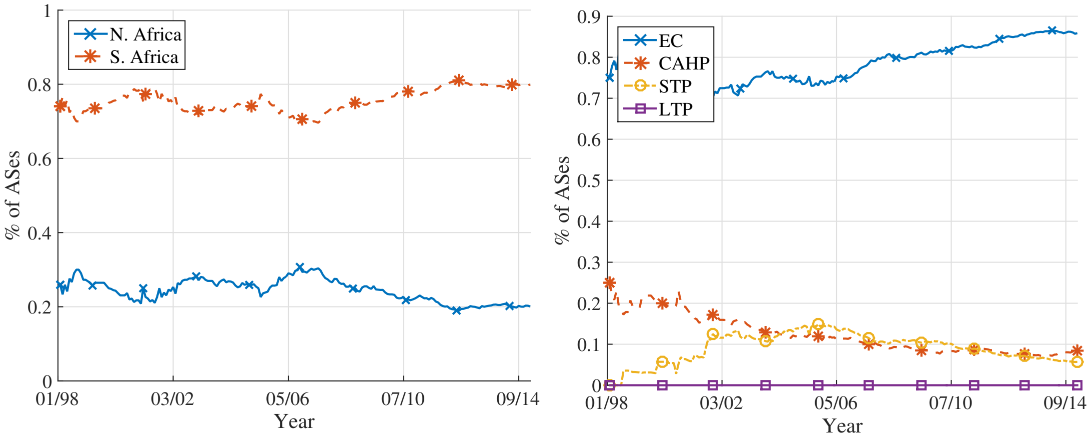


Figure 4.8: Common links in LACNIC.

four countries host more than 50 ASes and the one with the largest number is South Africa with 201 (see Appendix F). We cannot really define the topology structure of AFRINIC, due to its relatively small size. Most of the ASes connect outside the region than within, which is unique compared to the structure of the rest of the regions.

Figure 4.9(a) shows the percentage of ASes within each subregion. We see the distribution of the ASes has largely remained unchanged, with Southern Africa hosting approximately 80% of the ASes. In Figure 4.9(b), we show the AS distribution in the region per AS type. This is the only region where the number of CAHPs was larger than the number STPs in January 1998. For all the other regions, STPs surpassed the CAHPs in the early Internet days. However, the number of ASes in this region is also the smallest and the difference between CAHPs and STPs in early days was less than 5 ASes, which is not significant. It is also interesting to note that the trend in AFRINIC is reversed compared to other RIRs. The number of CAHP has shown a continuous drop while the number of STPs has increased and then stabilized around 10%. This can be justified by the fact that for the majority



(a) AFRINIC subregions (b) AFRINIC AS types

Figure 4.9: Percentage of ASes in AFRINIC when divided by subregion (a), and by type (b).

of the period analyzed, AFRINIC lacked the basic Internet infrastructure. That is, it has not reached a critical mass of infrastructure in terms of transit providers in order to start the process of building peering relationships among the existing ASes.

This is also evident in Figure 4.10, which shows the link distribution. A reverse trend compared to all other RIRs is observed. Links involving CAHPs as providers were far more prevalent until 2008. At that time, c2p links involving an STP as a provider surpassed those involving an CAHP. This trend was again reversed in 2014. The reason for that is that the small and regional ISPs in AFRINIC behave different than the ones in RIPE and ARIN. The lack of infrastructure led the early ISPs in AFRINIC to cooperate and peer between them and with other regions in order to achieved global connectivity. Due to the small number of ASes in AFRINIC this behavior had no impact on the creation of the AS classification rules. Therefore, these ISPs with high number of peers are classified as CAHPs instead of STPs, which is the business type they should be. These peering relationships are mostly with ASes in countries outside the region, and they are essential to connect the AFRINIC with the world Internet. We describe more details about

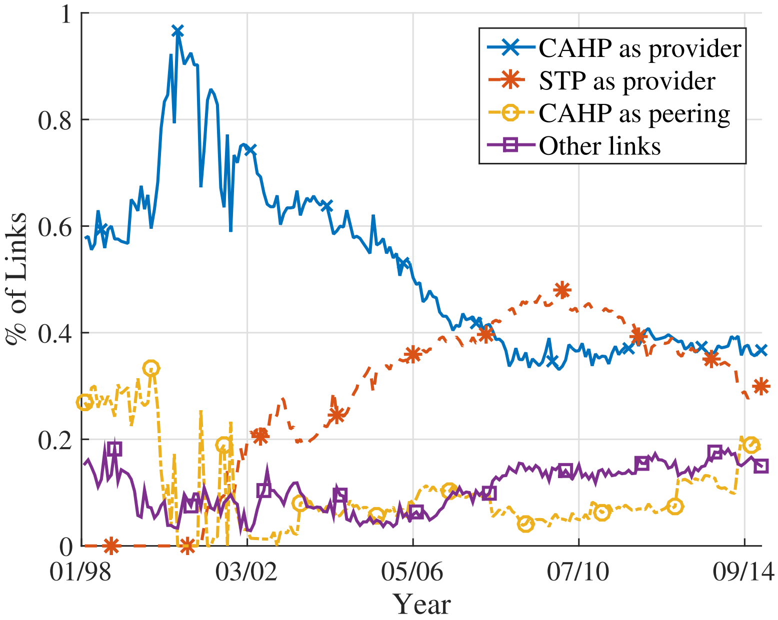


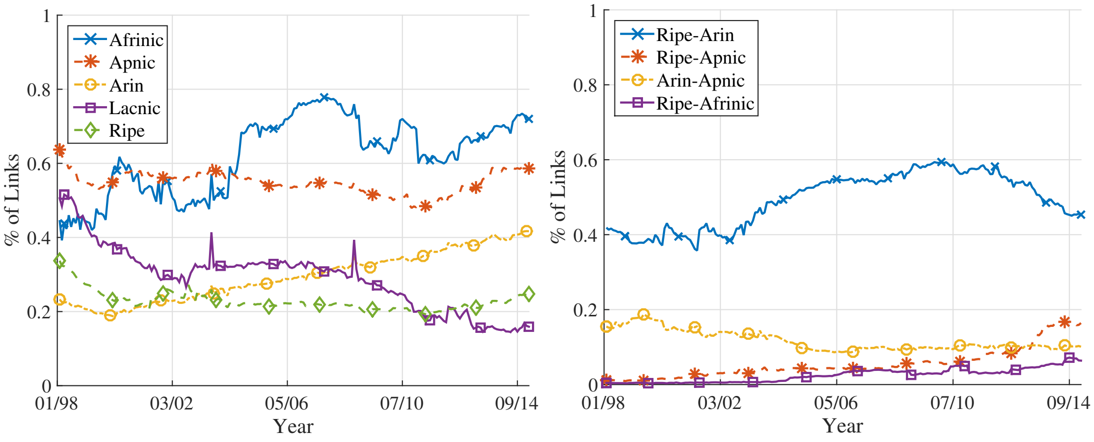
Figure 4.10: Percentage of the most common link relationships.

this inter-region connections in Section 4.2.

4.2 Connectivity

In this section, we analyze how the different regions interconnect. Moreover, we analyze the type of ASes involved in the inter-region connections. In Figure 4.11(a), we show the percentage of links with one of the incident ASes within the region of interest and the second incident AS in another region (i.e., excluding all links that connect two ASes in the same region). Figure 4.11(b), shows the top four regions in terms of the number of inter-region links. We use a percentage to indicate the fraction of the overall links that connect the different regions.

The analysis of the inter-region links leads to several interesting findings. Referring to Figure 4.11(a), we observe two reverse trends between ARIN and RIPE, the two most dominant regions (they host more than 75% of the ASes in the world). In RIPE, the percentage of inter-region links has been slowly degreasing until 2011,



(a) Percentage of inter-region links

(b) Top connected regions

Figure 4.11: Fig (a) percentage of links that connect to a different region. Fig(b) percentage of links use by top most connected regions, out of all inter-region links

indicating that the majority of links are added to connect ASes within the region. This is primarily attributed to the fact that a lot of these links are peering links between ASes that belong to different European countries. On the other hand, the percentage of links connected to other regions has been steadily increasing for ARIN. This is because, ARIN consists of only two countries, and therefore more connections aim at providing enhance connectivity with other countries around the world.

As expected, most inter-region links are between RIPE and ARIN. Of these links, the majority is between the subregions of North America with Northern Europe and Western Europe (see Appendix G.1), which can be attributed to the combination of submarine communications cables and the large number of IXPs in those subregions.

There are three types of relationships that stand out, as it is shown in Figure 4.12. The first one is RIPE–CAHPs peering with ARIN CAHPs. Out of the links connecting the two regions, this type of relationship has increased at a faster pace than any other. It grew from 15% in 2000 to 25% in 2010. Today, it is pre-

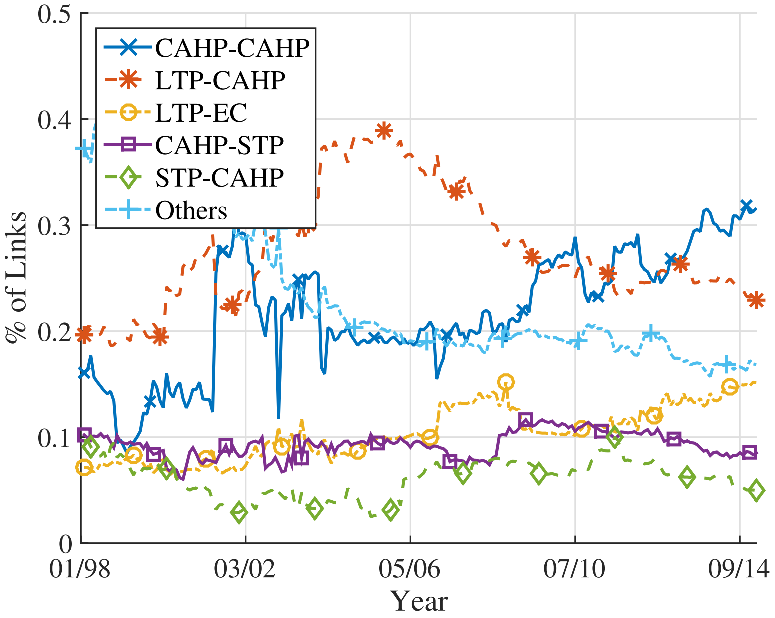


Figure 4.12: Percentage of the most common links between ARIN-RIPE.

dominant relationship with more 30% of the links. The second relationship that stands out is RIPE–CAHPs linking with ARIN LTPs. These connections used to account for more than 38% of the total links connecting the two regions back in 2005. Today, they account for only 23%, but they are still very influential.

In January 2015, we found that there are 4,106 links from RIPE CAHPs to ARIN LTPs, and at the same time there are only 2,794 CAHPs in RIPE, which means that, on average, every CAHP in RIPE could be a customer or a peer with one of the 18 ARIN LTPs. The third type of links that stand out are ARIN LTPs to RIPE ECs. They only account for less than 15% of the links, yet this is quite significant. In 2015, these were around 2,555 links, meaning that ARIN LTPs served as providers to almost 15% of RIPE ECs. These two regions are still and by far the most interconnected regions today, but if we see figure Figure 4.11(b), it shows that inter-region connections reached a peak in 2009. However, the reason for that has nothing to do with RIPE and ARIN. It is because the number of links between other regions has increased; the links between ARIN and RIPE has not decreased and the number continues to grow.

Of the rest of the regions, the percentage of inter-region links has remained relatively constant for APNIC. Since 1998, more than 50% of the links of ASes in APNIC connect to a different region. ARIN used to be the preferred region to connect, but in 2013 RIPE became the dominant choice (see Appendix G.2). This change did not happen abruptly, RIPE slowly gained popularity due to accessibility, distance, financial reasons, and simply because it slowly became the region with the largest number of ASes. Most of these connections are peering relationships with ASes in Northern Europe and Western Europe, with one exception; the APNIC region of Oceania still has more links to North America than Europe, especially to LTPs.

Two opposite trends are observed for AFRINIC and LACNIC. For the former, the number of inter-region links has been steadily growing. This can be attributed to the effort of the African ISPs to connect to other regions where most of the content is hosted (very few ASes remain within AFRINIC). Originally most of the connections were to ARIN, but it changed in 2003, possibly as a result of the introduction of SAT-3, a submarine communications cable linking Europe to South Africa in 2002. Currently around 80% of those links connect to RIPE. They are predominantly peering relationships with CAHPs, which is one of the reason many ISPs in AFRINIC tend to be considered CAHPs. Nonetheless, there is still a smaller percentage of ASes connecting to ARIN, but these connections are mostly from an ARIN LTP to an AFRINIC EC, CAHP or STP, (see Appendix G.3).

For LACNIC, the number of inter-region links has been steadily decreasing since 1998, reaching to the lowest value of 20% in 2015. LACNIC is the most isolated region and its ASes have the lowest interaction with other regions. In 2015, only 15% of the new links connected to a different region, while in 2010 this percentage was equal to 25%. For the most part, the inter-region connections are between LACNIC ASes as customers of ARIN LTPs, but that could change in the near future as peering connection seem to be gaining popularity, especially to RIPE CAHPs, (see Appendix G.4).

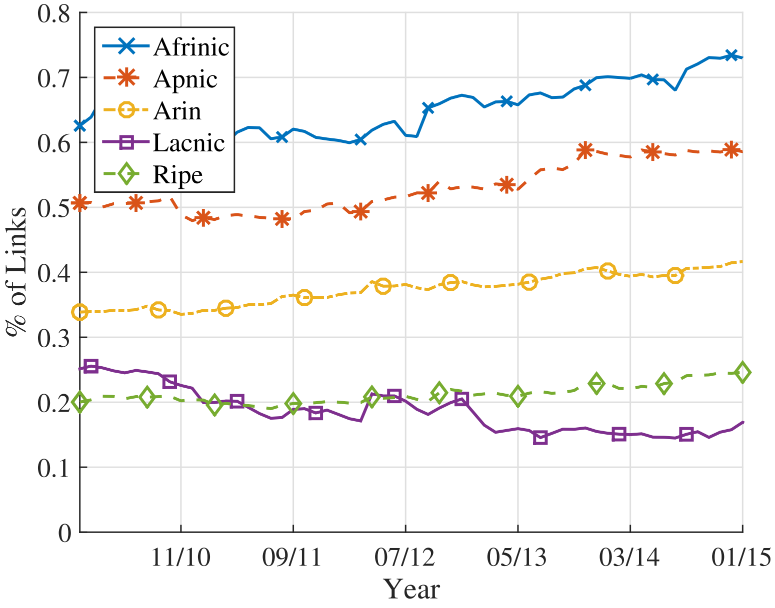


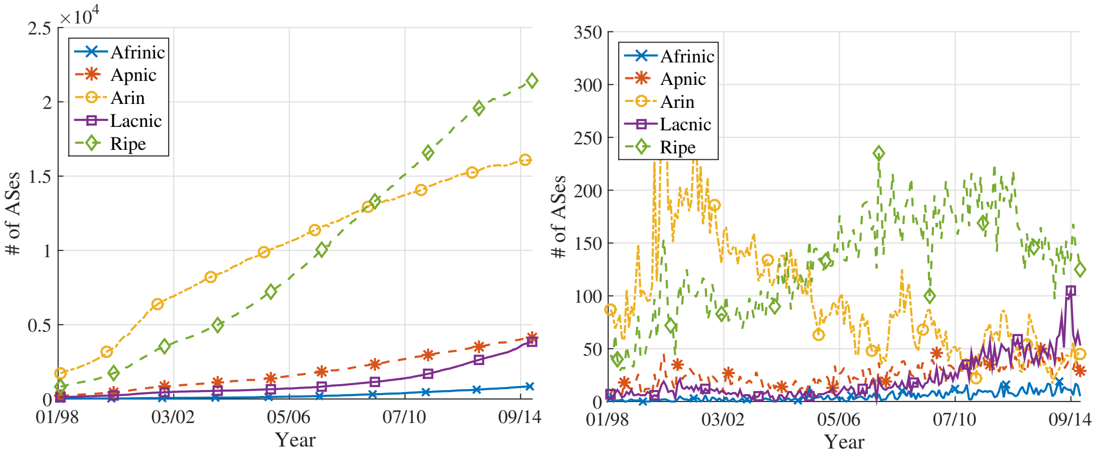
Figure 4.13: Percentage of inter-region links in the last five years.

In Figure 4.13, we show the percentage of inter-region links for each region (same as Figure 4.11(a), but for the last five years). With the exception of LACNIC, all regions have increased their inter-region connectivity. This global trend shows that ASes favor connections with other regions, and relying less in LTPs for global connectivity. These results support the recent theories that the Internet is becoming denser and more flat.

4.3 Topology Structure Differences

In this section, we compare various of topological aspects of the various regions and highlight major differences and similarities between them. Our analysis is centered around AS and link growth.

The most important difference in the AS topology of different regions is the way they grow. In Figure 4.14(a), we show the number of ASes per region. We can see that all of the regions have seen growth at different rates. Also, it is possible to see that the two largest regions have two distinct growing phases. ARIN was growing



(a) Number of ASes per region (b) Number of Born ASes per region

Figure 4.14: Number of ASes and AS births per region.

at a faster pace from 1998 to 2001, then it slowed down. A similar phenomenon is observed with RIPE, but its turning point was 2012. On the other hand, AFRINIC, APNIC, and LACNIC, have continuously increased their growing rate. The same can be observed in Figure 4.14(b), which shows the AS births over the course of our study. The number of ASes born from ARIN and RIPE has decreased while in other regions increased.

In previous works [81, 25], it has been demonstrated that an exponential growth models yield to better regression fits, so we found the exponential fitting for each region and we show them in the next set of equations, ordered by fastest growth exponent.

$$LACNIC \quad : \quad y = (139.9 \pm 7.3) * e^{(0.01602 \pm 0.0003)x}$$

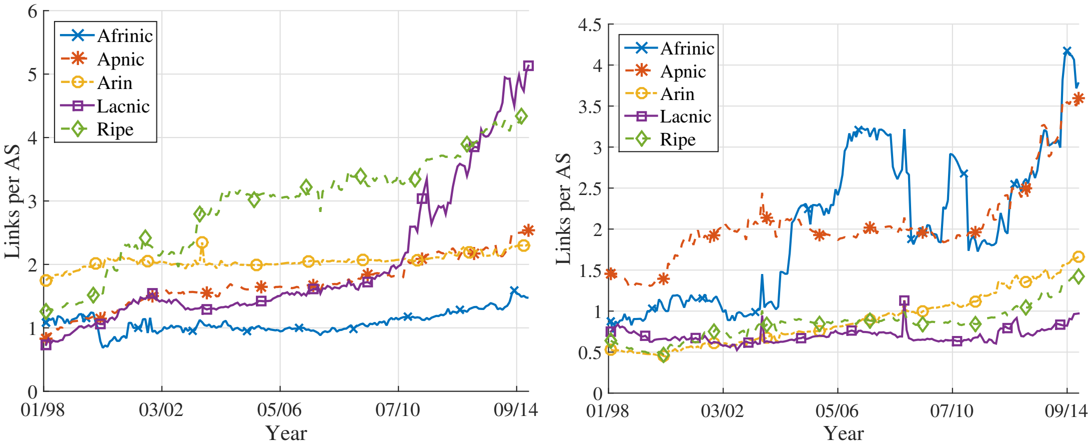
$$AFRINIC \quad : \quad y = (34.66 \pm 1.33) * e^{(0.01591 \pm 0.0002)x}$$

$$APNIC \quad : \quad y = (539.1 \pm 19.9) * e^{(0.01032 \pm 0.0002)x}$$

$$RIPE \quad : \quad y = (2665 \pm 165) * e^{(0.01088 \pm 0.0003)x}$$

$$ARIN \quad : \quad y = (4921 \pm 262) * e^{(0.006455 \pm 0.0003)x}$$

These equations agree with our previous observations; the three regions with the smaller number of ASes are the ones growing faster. This could mean that after a fast exponential growth, the regions get to a point of in which the AS growth stabilizes, becoming slow exponential or even linear.



(a) Intra-links per AS

(b) Inter-links per AS

Figure 4.15: (a) Average number of intra-links per AS for the different regions, (b) average number of inter-links per AS for the different regions.

Figure 4.15 shows the average number of intra-links (within a region) and inter-links (with other regions) per AS, for any AS that has both link types. From Figure 4.15(a), we can see that AFRINIC remains very poorly connected internally, with 1.5 intra-links per AS. ARIN and APNIC have a little bit more than two links per AS. RIPE on the other hand has seen significant growth in the average number of intra-links, especially since 2010, with more than 4 links per AS. This is attributed to the large number of countries within RIPE (relative to ARIN) and the growth in peering relationships among the ASes. The region with the most intra-links is LACNIC, with more than five intra-links per AS. This is expected as the percentage of inter-links within this region has been continuously shrinking, indicating that most of the added links are within the region.

In Figure 4.15(b), we show the average number of inter-links per AS for each of the regions. We observe that AFRINIC and APNIC host ASes with the largest inter-link connections per AS. These also grow at a faster pace than all other regions. LACNIC remains the least connected region, with a relatively constant number of inter-links per AS. RIPE and ARIN have also shown growth specially since 2011.

CHAPTER 5

CONCLUSION

Studying the Internet topology is essential for the development and improvement of the Internet. Previous works have focused on the topology as a simple graph, ignoring economic and geographical aspects of the AS ecosystem. In this thesis, we studied the evolution of the Internet considering those aspects, and detailed the evolution trends for each Internet region, giving a new perspective on the overall Internet architecture.

In particular, we analyzed the evolution of the ASes by type, as these are reflected by the business type of the organization that owns a particular AS. Moreover, we analyzed the evolution of links by type to reveal useful information about the hierarchical Internet structure. Using a business classification of ASes into four business types, we found that most of the growth of the Internet overall and within each region is primarily attributed to the increase in the number of ECs. LTPs, while not contributing to the topology growth, have maintained their position in the core of the Internet by having the largest degree and staying globally connected. The relative presence of STPs has decreased, while more peering-oriented ISPs and content providers have emerged. CAHPs are the ASes responsible for the flattening of the Internet topology, and more recently, they have been the driving force for the strong interconnection between regions.

With respect to the evolution of the Internet topology within each region, our findings were as follows. ARIN, the region that covers North America, maintains a highly-hierarchical structure over the 15 years of our study. Most of the intra-links are between providers to customers. The region is dominated by a small number of LTPs, which account for more than 50% of the links in the region. ARIN used to be the region with the largest number of ASes, but it was surpassed by RIPE in 2009. RIPE is the region that covers most of Europe. It is characterized by a large number of CAHPs, and in contrast to ARIN, RIPE has a flat topology due to a large number of peering connections. APNIC, AFRINIC, the regions for Asian and African countries, have a very sparce topology and rely extensively on connections to other regions, and in specific, RIPE. LACNIC, the region that covers Latin American and Caribbean countries has the lowest connection degree with other regions.

There are many differences and similarities between the regions. They have a different topology structure and behave in their own unique ways, but all of them have been continuously growing at varying exponential rates. The regions with the smallest number of ASes are currently the ones growing the fastest. Furthermore, all of them but LACNIC tend to primarily interconnect with other regions. The regions are becoming more interconnected and it is mostly trough peering ASes.

APPENDIX A

REGION, SUBREGIONS, AND COUNTRIES

In this appendix we provide the table for each the region and subregion of the countries that composed them.

Table A.1: AFRINIC country list

Subregion	Countries
Southern Africa	Botswana, Burundi, Benin, Burkina Faso,Cabo Verde, Comoros, Djibouti, Eritrea, Ethiopia, Kenya, Ivory Coast, Gambia, Ghana, Guinea, Guinea-Bissau, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mayotte, Mozambique, Namibia, Niger, Nigeria, Reunion, Rwanda, Saint Helena, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, South Sudan, Swaziland, Togo, Uganda, United Republic of Tanzania, Zambia, Zimbabwe
Northern Africa	Algeria, Angola, Cameroon, Central African Republic, Chad, Congo, Democratic Republic of the Congo, Egypt, Equatorial Guinea, Gabon, Libya, Morocco, Sao Tome and Principe, Sudan, Tunisia, Western Sahara

Table A.2: APNIC country list

Subregion	Countries
C. Asia	Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan
E. Asia	China, Hong Kong, Japan, Macao, Mongolia, North Korea, South Korea
Oceania	American Samoa, Australia, Cook Islands, Fiji, French Polynesia, Guam, Kiribati, Marshall Islands, Micronesia, Nauru, New Caledonia, New Zealand, Niue, Norfolk Island, Northern Mariana Islands, Palau, Papua New Guinea, Pitcairn, Samoa, Solomon Islands, Tokelau, Tonga, Tuvalu, Vanuatu, Wallis and Futuna Islands
SE Asia	Brunei, Cambodia, East Timor, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, Vietnam
S Asia	Afghanistan, Bangladesh, Bhutan, India, Iran, Maldives, Nepal, Pakistan, Sri Lanka
W Asia	Armenia, Azerbaijan, Bahrain, Cyprus, Georgia, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, State of Palestine, Syrian Arab Republic, Turkey, United Arab Emirates, Yemen

Table A.3: ARIN country list

Subregion	Countries
N. America	Bermuda, Canada, Greenland, Saint Pierre and Miquelon, United States of America

Table A.4: LACNIC country list

Subregion	Countries
C. America	Anguilla, Antigua and Barbuda, Aruba, Bahamas, Barbados, Belize, Bonaire Sint Eustatius and Saba, British Virgin Islands, Cayman Islands, Costa Rica, Cuba, Curacao, Dominica, Dominican Republic, El Salvador, Grenada, Guadeloupe, Guatemala, Haiti, Honduras, Jamaica, Martinique, Mexico, Montserrat, Nicaragua, Panama, Puerto Rico, Saint Barthelemy, Saint Kitts and Nevis, Saint Lucia, Saint Martin, Saint Vincent and the Grenadines, Trinidad and Tobago, Turks and Caicos Islands, United States Virgin Islands
S. America	Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Falkland Islands, French Guiana, Guyana, Paraguay, Peru, Suriname, Uruguay, Venezuela

Table A.5: RIPE country list

Subregion	Countries
E. Europe	Belarus, Bulgaria, Czech Republic, Hungary, Poland, Republic of Moldova, Romania, Russian, Slovakia, Ukraine
N. Europe	Aland Islands, Channel Islands, Denmark, Estonia, Faeroe Islands, Finland, Guernsey, Iceland, Ireland, Isle of Man, Jersey, Latvia, Lithuania, Norway, Sark, Svalbard and Jan Mayen, Sweden, United Kingdom of Great Britain and Northern Ireland
S. Europe	Albania, Andorra, Bosnia and Herzegovina, Croatia, Gibraltar, Greece, Holy See, Italy, Malta, Montenegro, Portugal, San Marino, Serbia, Slovenia, Spain, The former Yugoslav Republic of Macedonia
W. Europe	Austria, Belgium, France, Germany, Liechtenstein, Luxembourg, Monaco, Netherlands, Switzerland

APPENDIX B

RIPE ADDITIONAL DATA

This appendix provides addition information for region RIPE. Figure B.1 and figure B.3 provide information about the types of ASes in RIPE, while the table provides information per country.

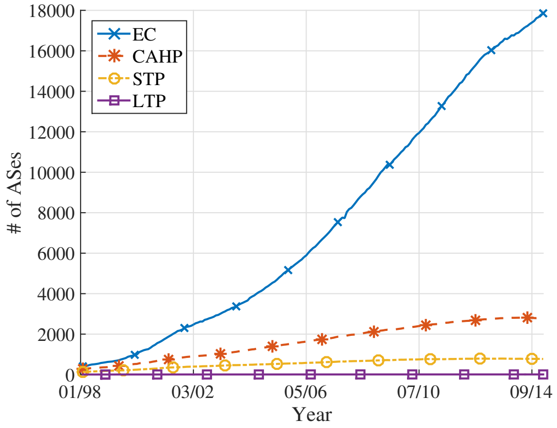


Figure B.1: Number of ASes per Type in RIPE.

Table B.1: RIPE Types per Country 2015

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Aland Islands	1	1	0	0	0
Albania	34	29	0	5	0
Andorra	1	1	0	0	0
Austria	416	340	67	9	0
Belarus	82	77	2	3	0
Belgium	193	152	35	6	0
Bosnia and Herzegovina	28	23	1	4	0
Bulgaria	514	462	29	23	0
Croatia	98	88	3	7	0

Continued on next page

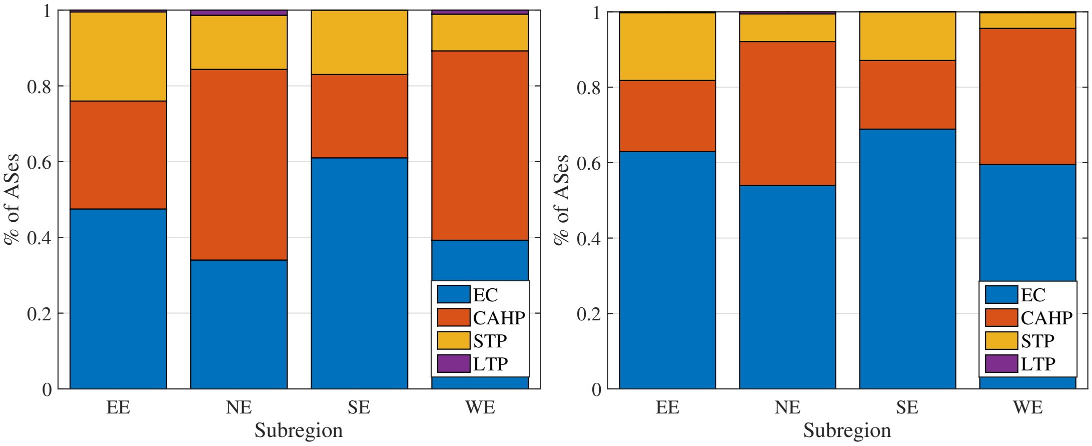
Table B.1 – *Continued from previous page*

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Czech Republic	484	389	89	6	0
Denmark	226	198	24	4	0
Estonia	66	62	3	1	0
Faroe Islands	2	2	0	0	0
Finland	190	172	9	9	0
France	885	647	228	10	0
Germany	1407	1109	275	21	2
Gibraltar	14	12	2	0	0
Greece	133	120	5	8	0
Guernsey	2	2	0	0	0
Hungary	177	159	12	6	0
Iceland	45	38	4	3	0
Ireland	140	106	33	1	0
Isle of Man	6	6	0	0	0
Italy	650	514	122	14	0
Jersey	5	3	1	1	0
Latvia	203	185	7	11	0
Liechtenstein	14	10	4	0	0
Lithuania	107	99	2	6	0
Luxembourg	56	36	17	3	0
Macedonia	28	24	0	4	0
Malta	26	22	1	3	0
Moldova	71	65	4	2	0
Monaco	1	0	1	0	0
Montenegro	14	12	0	2	0
Netherlands	613	363	243	7	0
Norway	194	148	43	3	0

Continued on next page

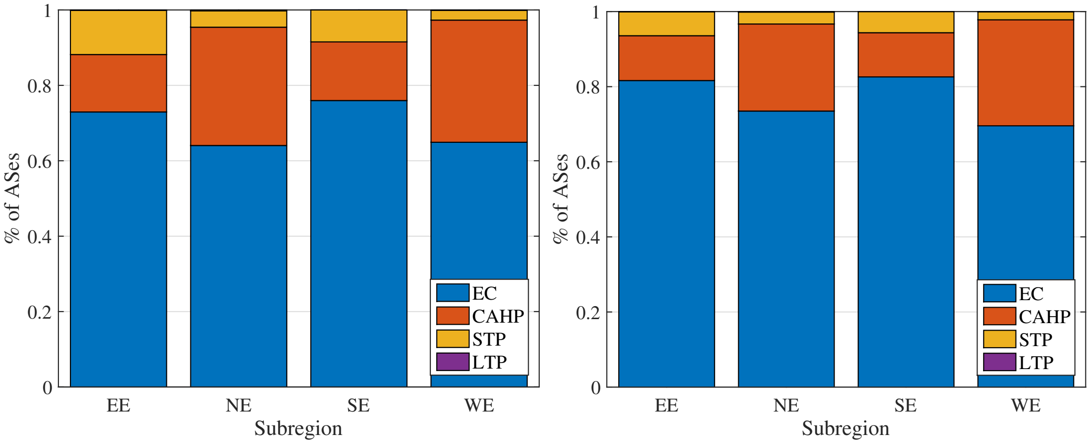
Table B.1 – *Continued from previous page*

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Poland	1725	1387	288	50	0
Portugal	71	58	8	5	0
Romania	1134	1056	29	49	0
Russia	4471	3774	506	188	3
San Marino	7	6	1	0	0
Serbia	135	122	4	9	0
Slovakia	116	103	8	5	0
Slovenia	246	233	4	9	0
Spain	492	446	33	13	0
Sweden	460	383	72	4	1
Switzerland	522	396	122	4	0
Ukraine	1639	1475	51	112	1
United Kingdom	1502	1105	372	23	2
Vatican	1	1	0	0	0



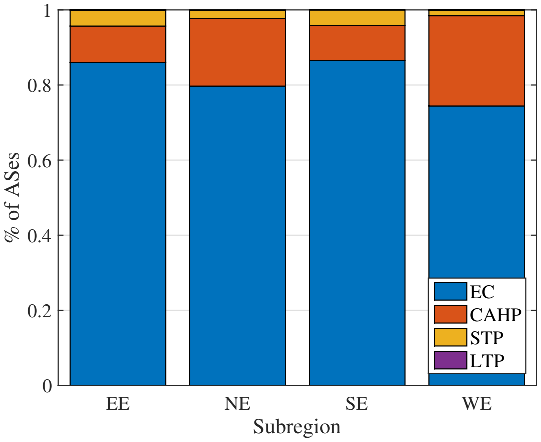
(a) 1998

(b) 2002



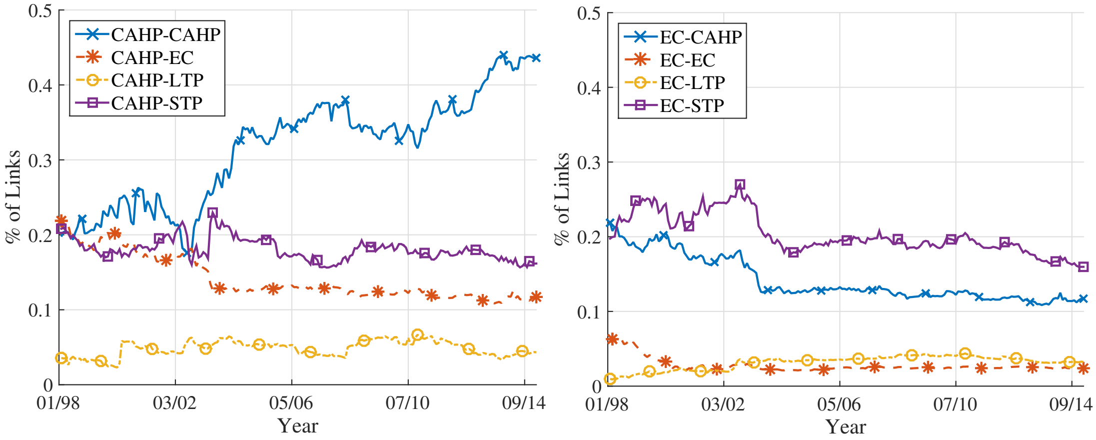
(c) 2006

(d) 2010



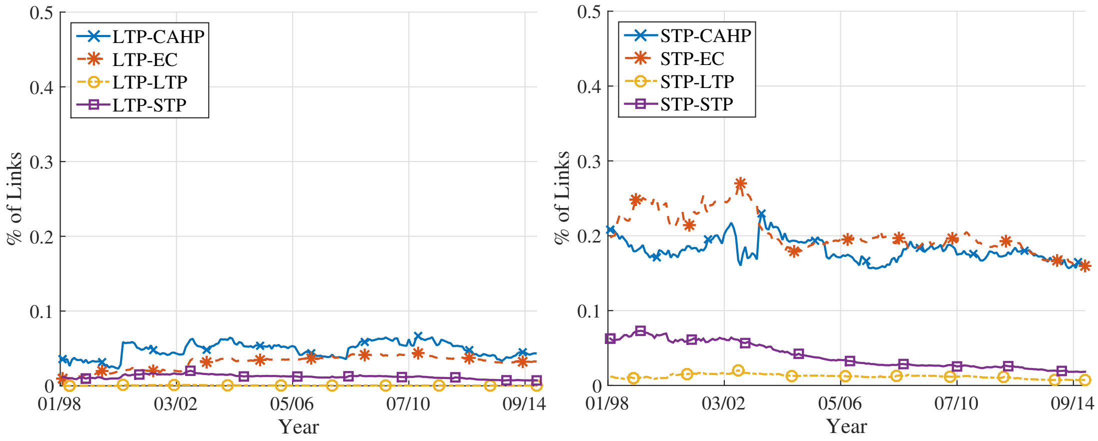
(e) 2015

Figure B.2: Proportion of ASes in each region per type



(a) CAHP

(b) EC



(c) LTP

(d) STP

Figure B.3: Percentage of links between diffent types of ASes out of the total number of links connecting two AS in RIPE

APPENDIX C

ARIN ADDITIONAL DATA

This appendix provides addition information for region ARIN. Figure C.1 and figure C.2 provide information about the types of ASes in ARIN, while the table provides information per country.

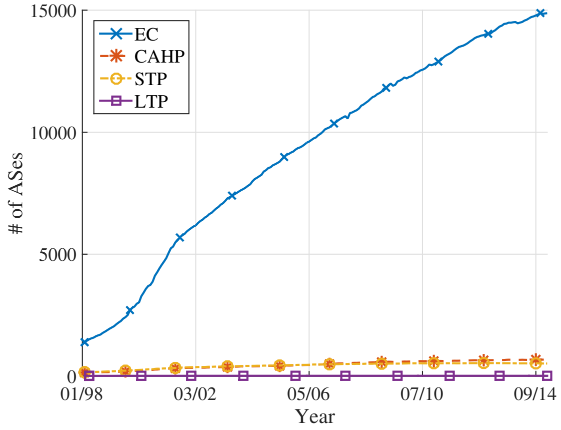
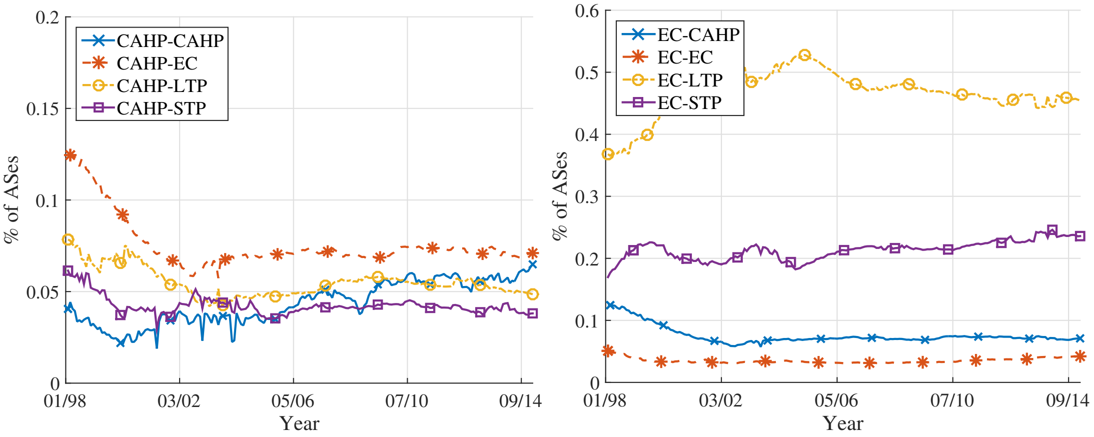


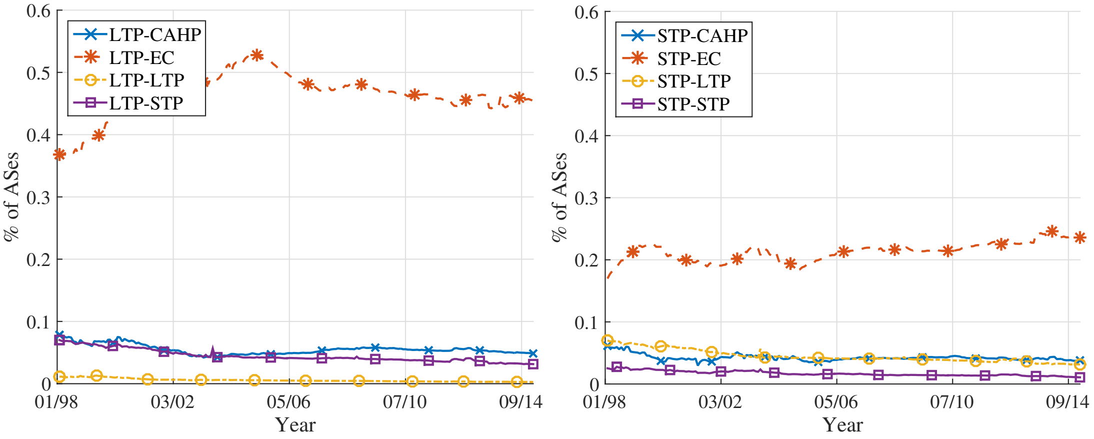
Figure C.1: Number of ASes in ARIN by type

Table C.1: Arin Types per Country 2015

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Bermuda	10	8	1	1	0
Canada	1004	839	123	42	0
Greenland	1	1	0	0	0
Saint Pierre and Miquelon	1	1	0	0	0
United States	14955	13830	585	522	18



(a) CAHP (b) EC



(c) LTP (d) STP

Figure C.2: Percentage of links between different types of ASes out of the total number of links connecting two AS in Arin. Notice the scale is diffrenent in CAHP

APPENDIX D

APNIC ADDITIONAL DATA

This appendix provides addition information for region APNIC. Figure D.1 and figure D.2 provide information about the types of ASes in APNIC, while the table provides information per country.

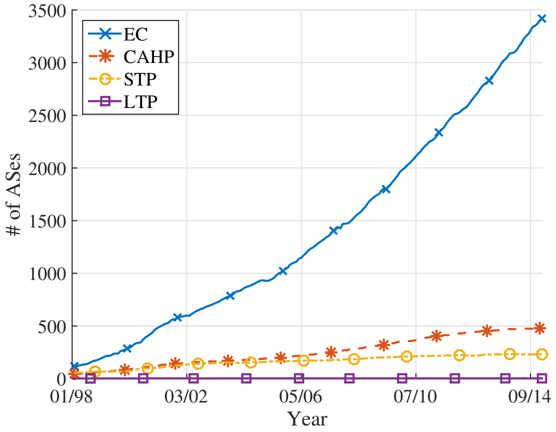


Figure D.1: Number of ASes per Type in RIPE

Table D.1: APNIC Types per Country 2015

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Afghanistan	37	36	0	1	0
American Samoa	2	2	0	0	0
Armenia	53	46	0	7	0
Australia	1116	845	242	29	0
Azerbaijan	35	33	1	1	0
Bahrain	19	17	1	1	0
Bangladesh	200	180	3	17	0
Bhutan	5	4	1	0	0
Brunei	6	6	0	0	0

Continued on next page

Table D.1 – *Continued from previous page*

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Cambodia	44	35	4	5	0
China	284	246	18	20	0
Cook Islands	1	1	0	0	0
Cyprus	67	56	8	3	0
East Timor	2	2	0	0	0
Fiji	6	6	0	0	0
French Polynesia	2	2	0	0	0
Georgia	56	50	2	4	0
Guam	6	4	2	0	0
Hong Kong	356	241	96	18	1
India	737	710	4	23	0
Indonesia	636	510	74	52	0
Iran	309	283	1	25	0
Iraq	42	36	2	4	0
Israel	202	191	3	8	0
Japan	576	447	96	33	0
Jordan	25	22	1	2	0
Kazakhstan	93	83	2	8	0
Kuwait	48	42	3	3	0
Kyrgyzstan	29	24	0	5	0
Laos	10	9	0	1	0
Lebanon	58	54	0	4	0
Macao	4	3	1	0	0
Malaysia	134	117	11	6	0
Maldives	3	2	1	0	0
Marshall Islands	1	1	0	0	0
Micronesia	4	4	0	0	0

Continued on next page

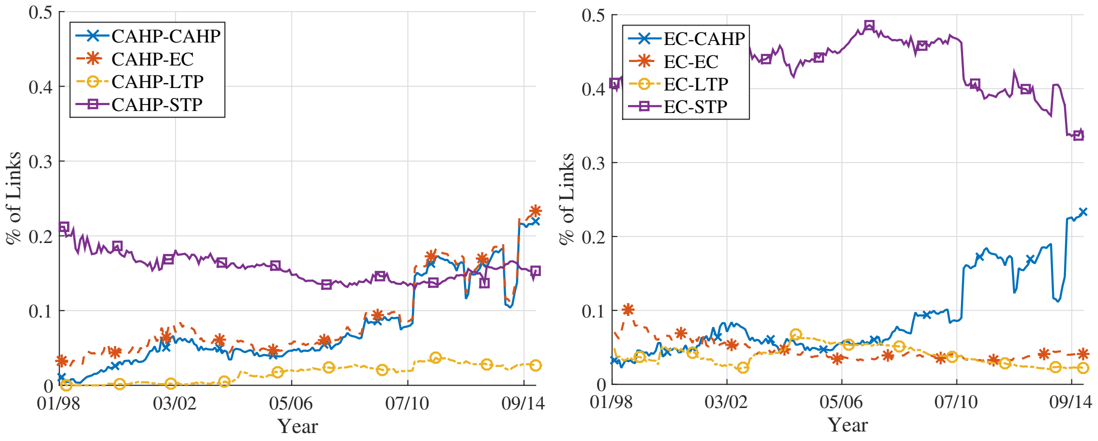
Table D.1 – *Continued from previous page*

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Mongolia	35	32	1	2	0
Myanmar	8	8	0	0	0
Nauru	1	1	0	0	0
Nepal	34	28	1	5	0
New Caledonia	8	7	0	1	0
New Zealand	292	267	10	15	0
Norfolk Island	1	1	0	0	0
North Korea	1	1	0	0	0
Oman	7	6	1	0	0
Pakistan	74	68	2	4	0
Palau	2	2	0	0	0
Palestinian Territory	34	32	0	2	0
Papua New Guinea	6	6	0	0	0
Philippines	205	189	7	9	0
Qatar	9	5	3	1	0
Samoa	4	4	0	0	0
Saudi Arabia	108	95	2	11	0
Singapore	238	194	30	14	0
Solomon Islands	2	2	0	0	0
South Korea	665	633	6	24	2
Sri Lanka	15	13	1	1	0
Syria	3	3	0	0	0
Taiwan	127	105	10	12	0
Tajikistan	6	5	0	1	0
Thailand	269	240	9	20	0
Tonga	2	1	1	0	0
Turkey	331	316	0	15	0

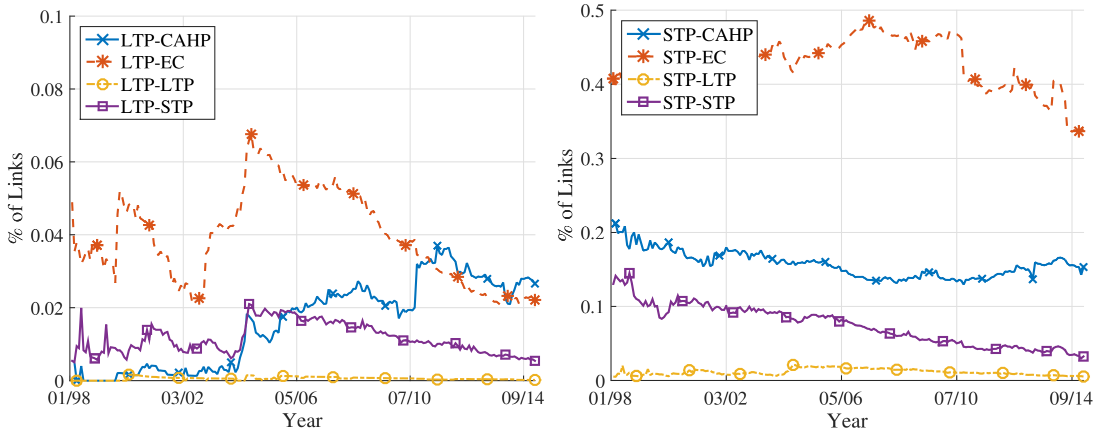
Continued on next page

Table D.1 – *Continued from previous page*

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Turkmenistan	3	3	0	0	0
United Arab Emirates	50	45	3	2	0
Uzbekistan	30	30	0	0	0
Vanuatu	5	3	2	0	0
Vietnam	156	149	7	0	0
Wallis and Futuna	1	1	0	0	0
Yemen	1	1	0	0	0



(a) CAHP (b) EC



(c) LTP (d) STP

Figure D.2: Percentage of links between diffrent types of ASes out of the total number of links connecting two AS in APNIC

APPENDIX E

LACNIC ADDITIONAL DATA

This appendix provides addition information for region LACNIC. Figure E.1 and Figure E.2 provide information about the types of ASes in LACNIC, while the table provides information per country.

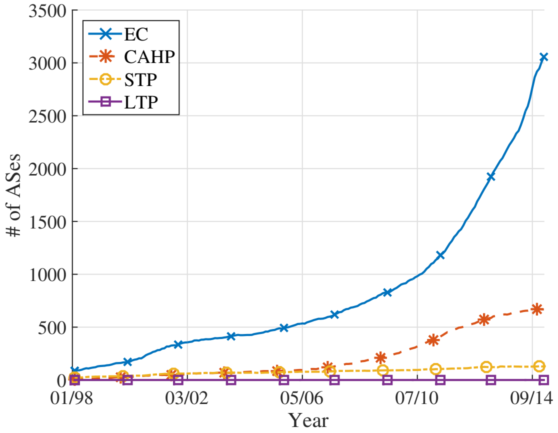


Figure E.1: Number of ASes per Type in LACNIC

Table E.1: LACNIC Types per Country 2015

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Anguilla	2	2	0	0	0
Antigua and Barbuda	4	4	0	0	0
Argentina	338	311	4	23	0
Aruba	2	2	0	0	0
Bahamas	3	3	0	0	0
Barbados	6	5	0	1	0
Belize	6	5	1	0	0
Bolivia	14	12	0	2	0
Bonaire, Saint Eustatius and Saba	4	4	0	0	0

Continued on next page

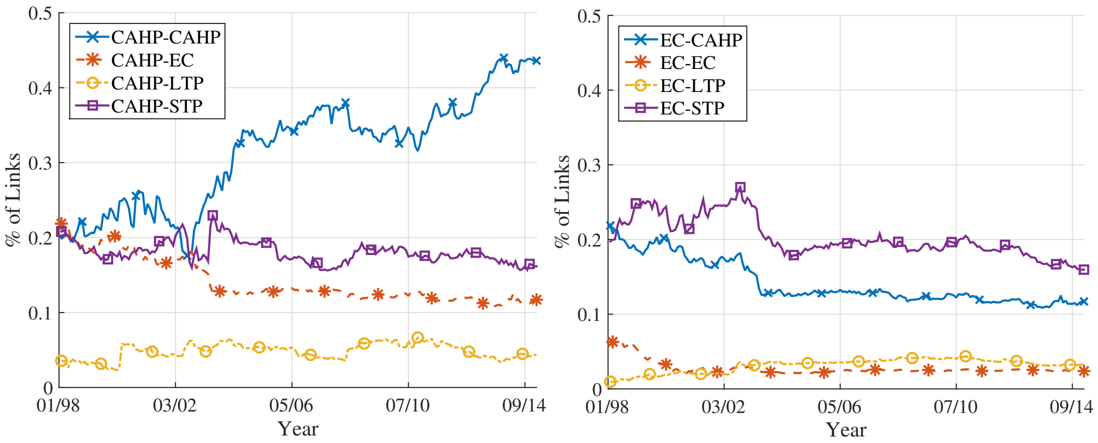
Table E.1 – *Continued from previous page*

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Brazil	2652	1976	645	31	0
British Virgin Islands	5	5	0	0	0
Cayman Islands	5	5	0	0	0
Chile	135	120	6	9	0
Colombia	85	79	2	4	0
Costa Rica	46	43	0	3	0
Cuba	3	3	0	0	0
Curacao	16	13	1	2	0
Dominica	2	1	0	1	0
Dominican Republic	20	20	0	0	0
Ecuador	52	43	3	6	0
El Salvador	15	14	0	1	0
French Guiana	2	2	0	0	0
Grenada	3	2	1	0	0
Guadeloupe	2	1	0	1	0
Guatemala	25	17	0	8	0
Guyana	3	3	0	0	0
Haiti	6	6	0	0	0
Honduras	23	21	0	2	0
Jamaica	8	5	0	3	0
Mexico	194	171	2	21	0
Nicaragua	16	15	0	1	0
Panama	76	71	1	4	0
Paraguay	21	18	0	3	0
Peru	22	21	0	1	0
Puerto Rico	45	38	1	6	0
Saint Kitts and Nevis	3	1	2	0	0

Continued on next page

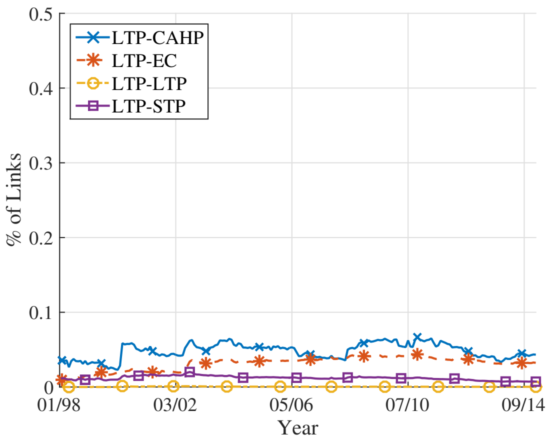
Table E.1 – *Continued from previous page*

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Saint Martin	2	1	0	1	0
Saint Vincent and the Grenadines	2	2	0	0	0
Suriname	2	2	0	0	0
Trinidad and Tobago	8	8	0	0	0
Turks and Caicos Islands	1	1	0	0	0
U.S. Virgin Islands	7	7	0	0	0
Uruguay	21	16	2	3	0
Venezuela	44	39	2	3	0



(a) CAHP

(b) EC



(c) STP

Figure E.2: Percentage of links between diffent types of ASes out of the total number of links connecting two AS in LACNIC

APPENDIX F

AFRINIC ADDITIONAL DATA

This appendix provides addition information for region AFRINIC. Figure F.1 and figure F.2 provide information about the types of ASes in AFRINIC, while the table provides information per country.

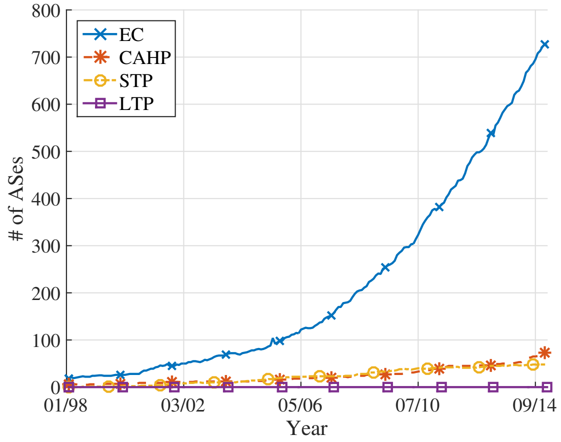


Figure F.1: Number of ASes per Type in RIPE

Table F.1: AFRINIC Types per Country 2015

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Algeria	11	11	0	0	0
Angola	31	27	2	2	0
Benin	7	6	1	0	0
Botswana	13	12	0	1	0
Burkina Faso	4	4	0	0	0
Burundi	10	9	0	1	0
Cameroon	12	11	0	1	0
Central African Republic	2	2	0	0	0

Continued on next page

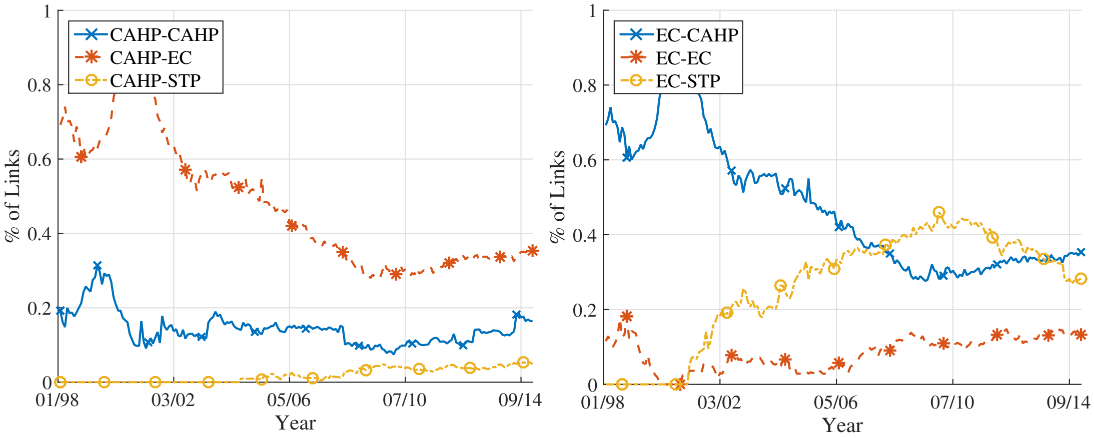
Table F.1 – *Continued from previous page*

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Chad	4	4	0	0	0
Comoros	1	1	0	0	0
Democratic Republic of the Congo	10	10	0	0	0
Djibouti	1	1	0	0	0
Egypt	52	46	0	6	0
Equatorial Guinea	5	4	0	1	0
Ethiopia	1	1	0	0	0
Gabon	9	9	0	0	0
Gambia	5	5	0	0	0
Ghana	35	32	0	3	0
Guinea	6	6	0	0	0
Guinea-Bissau	1	1	0	0	0
Ivory Coast	7	6	0	1	0
Kenya	60	49	6	5	0
Lesotho	5	5	0	0	0
Liberia	5	4	1	0	0
Libya	6	6	0	0	0
Madagascar	4	3	1	0	0
Malawi	6	6	0	0	0
Mali	4	4	0	0	0
Mauritania	2	2	0	0	0
Mauritius	15	12	3	0	0
Mayotte	1	1	0	0	0
Morocco	5	5	0	0	0
Mozambique	19	18	0	1	0
Namibia	9	5	3	1	0
Niger	5	5	0	0	0

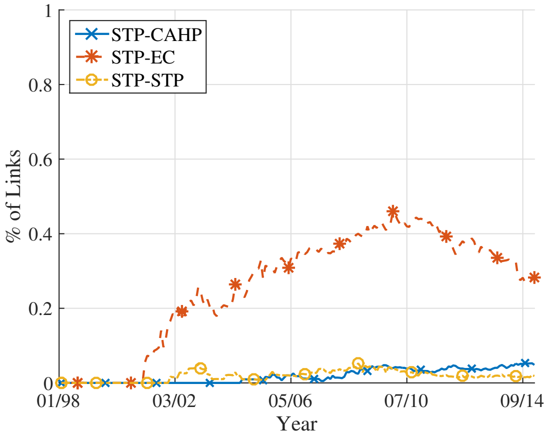
Continued on next page

Table F.1 – *Continued from previous page*

Country	**ASes**	**EC**	**CAHP**	**STP**	**LTP**
Nigeria	106	98	1	7	0
Republic of the Congo	10	10	0	0	0
Reunion	1	0	1	0	0
Rwanda	10	9	0	1	0
Sao Tome and Principe	1	1	0	0	0
Senegal	3	2	0	1	0
Seychelles	6	6	0	0	0
Sierra Leone	8	8	0	0	0
Somalia	10	10	0	0	0
South Africa	201	147	50	4	0
South Sudan	6	6	0	0	0
Sudan	6	4	0	2	0
Swaziland	3	3	0	0	0
Tanzania	41	34	0	7	0
Togo	3	3	0	0	0
Tunisia	7	6	0	1	0
Uganda	27	24	1	2	0
Zambia	14	13	0	1	0
Zimbabwe	16	12	1	3	0



(a) CAHP (b) EC



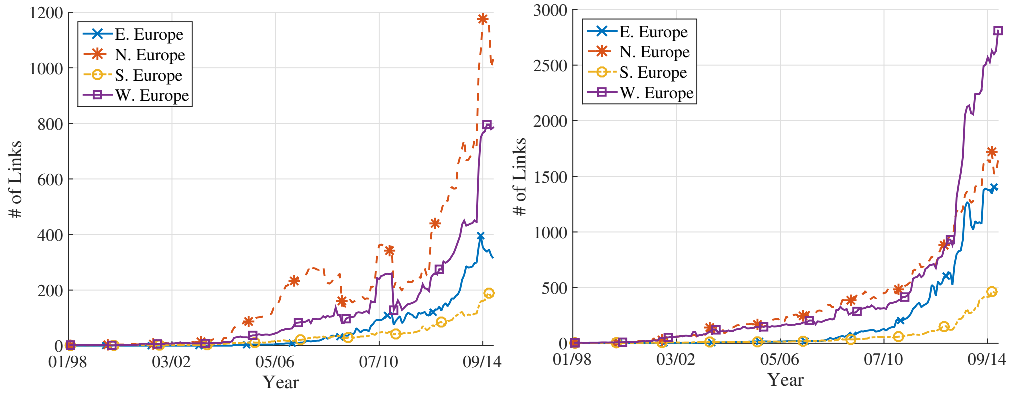
(c) STP

Figure F.2: Percentage of links between diffrent types of ASes out of the total number of links connecting two AS in AFRINIC

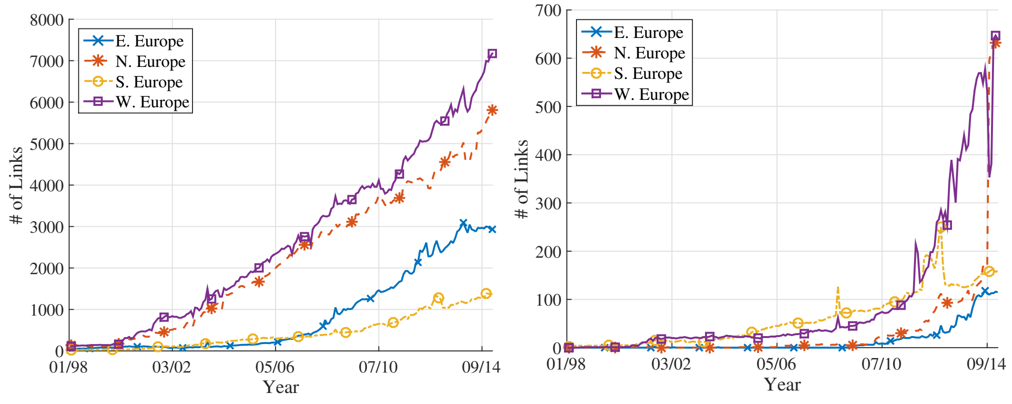
APPENDIX G

CONNECTIONS BETWEEN REGIONS

This appendix provides addition information about the connection between different regions.

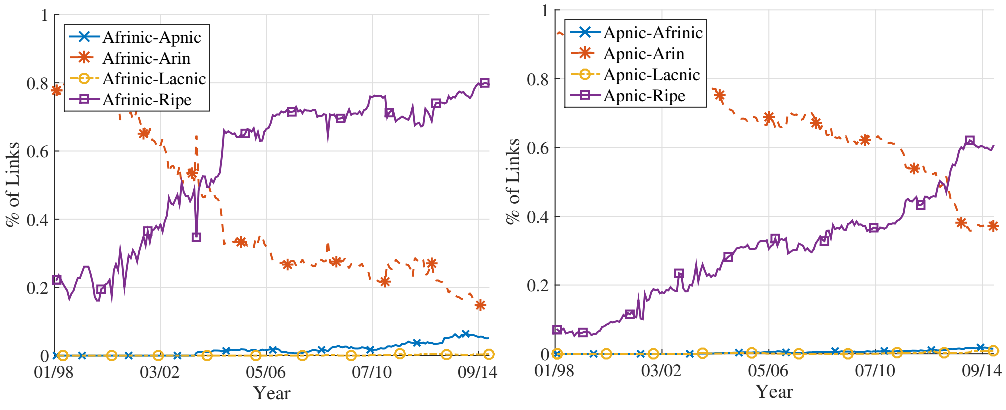


(a) AFRINIC (b) APNIC

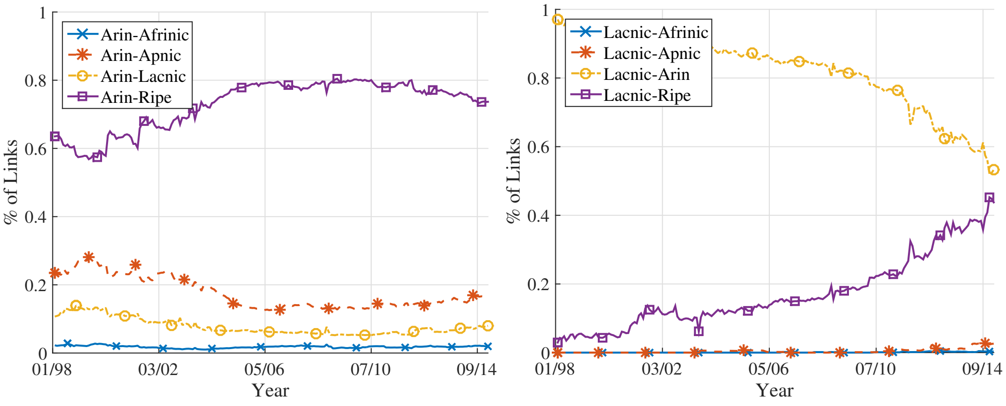


(c) ARIN (d) LACNIC

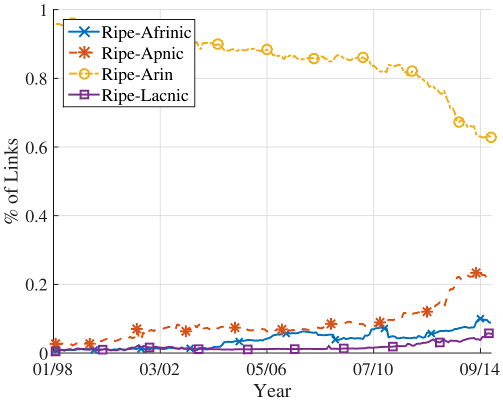
Figure G.1: RIR to RIPE subregions



(a) AFRINIC (b) APNIC

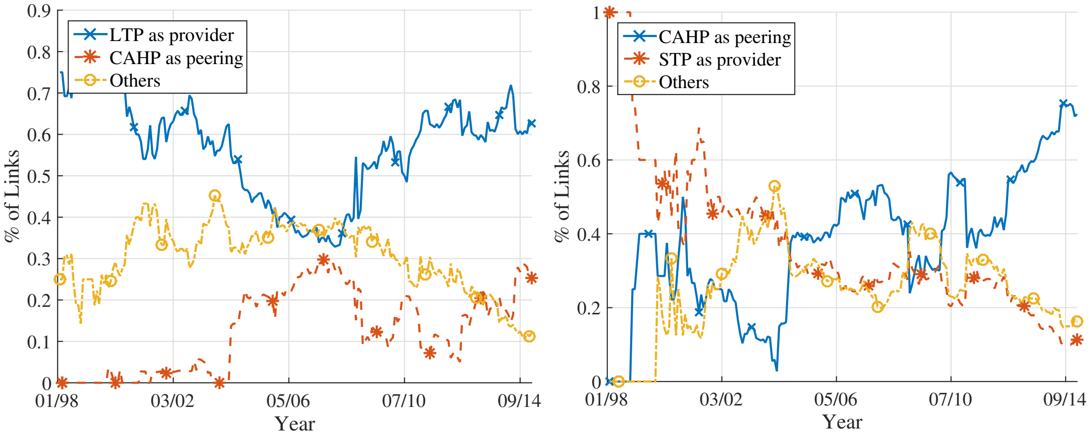


(c) ARIN (d) LACNIC



(e) RIPE

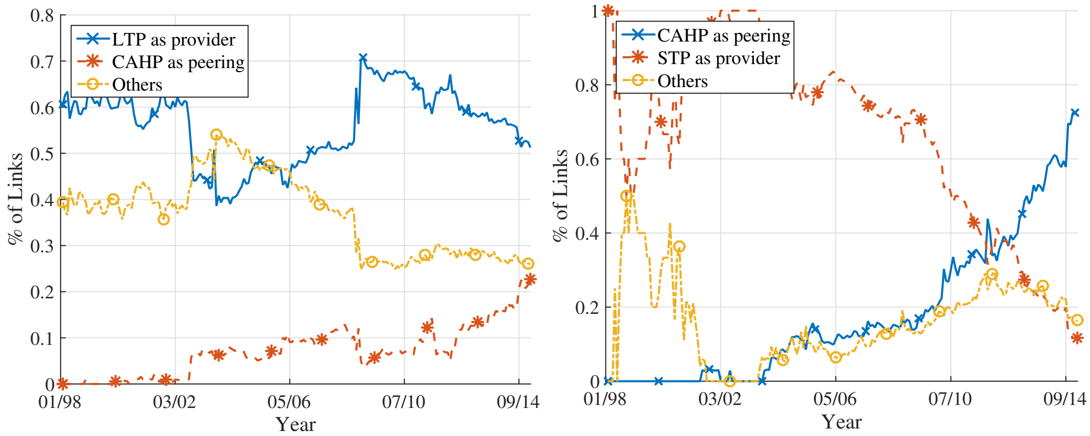
Figure G.2: Preference regions to connect inter-region links



(a) Links from ARIN

(b) Links from RIPE

Figure G.3: Fig (a) percentage of common links that connect ARIN-AFRINIC. Fig(b) percentage of common links that connect RIPE-AFRINIC



(a) Links from ARIN

(b) Links from RIPE

Figure G.4: Fig (a) percentage of common links that connect ARIN-LACNIC. Fig(b) percentage of common links that connect RIPE-LACNIC

REFERENCES

[1] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore. On the bias of traceroute sampling. In *STOC, ACM*, volume 1581139608, page 0005. Citeseer, 2005.

[2] M. B. Akgun and M. H. Gunes. Bipartite internet topology at the subnet-level. In *Network Science Workshop (NSW), 2013 IEEE 2nd*, pages 94–97. IEEE, 2013.

[3] R. Albert and A.-L. Barabási. Topology of evolving networks: local events and universality. *Physical review letters*, 85(24):5234, 2000.

[4] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani. k-core decomposition of internet graphs: hierarchies, self-similarity and measurement biases. *arXiv preprint cs/0511007*, 2005.

[5] B. Augustin, B. Krishnamurthy, and W. Willinger. Ixps: mapped? In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 336–349. ACM, 2009.

[6] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[7] A. Baumann and B. Fabian. Who runs the internet?-classifying autonomous systems into industries. In *WEBIST (1)*, pages 361–368, 2014.

[8] M. Baur, U. Brandes, M. Gaertler, and D. Wagner. Drawing the as graph in 2.5 dimensions. In *Graph Drawing*, pages 43–48. Springer, 2005.

[9] K. Boitmanis, U. Brandes, and C. Pich. Visualizing internet evolution on the autonomous systems level. In *Graph Drawing*, pages 365–376. Springer, 2008.

[10] J. Brodkin. Why YouTube buffers: The secret deals that make-and break-online video. `http://arstechnica.com/information-technology/2013/07/why-youtube-buffers-the-secret-deals-that-make-and-break-online-video/`, July 2013.

[11] T. Bu, N. Duffield, F. L. Presti, and D. Towsley. Network tomography on general topologies. In *ACM SIGMETRICS Performance Evaluation Review*, 2002.

[12] T. Bu and D. Towsley. On distinguishing between internet power law topology generators. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 638–647. IEEE, 2002.

[13] Caida ark project. `http://www.caida.org/projects/ark/`.

[14] Caida datasets. `http://data.caida.org/datasets/`.

[15] Caida. Ipv4 and ipv6 as core: Visualizing ipv4 and ipv6 internet topology at a macroscopic scale in 2014. `http://www.caida.org/research/topology/as_core_network/2014/`, 2014.

[16] H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger. Towards capturing representative as-level internet topologies. *Computer Networks*, 44(6):737–755, 2004.

[17] H. Chang, S. Jamin, and W. Willinger. Internet connectivity at the as-level: an optimization-driven modeling approach. In *Proceedings of the ACM SIGCOMM workshop on Models, methods and tools for reproducible network research*, pages 33–46. ACM, 2003.

[18] H. Chang, S. Jamin, and W. Willinger. To peer or not to peer: Modeling the evolution of the internets as-level topology. *Ann Arbor*, 1001:48109–2122, 2006.

[19] H. Chang and W. Willinger. Difficulties measuring the internet's as-level ecosystem. In *Information Sciences and Systems, 2006 40th Annual Conference on*, pages 1479–1483. IEEE, 2006.

[20] K. Chen, D. R. Choffnes, R. Potharaju, Y. Chen, F. E. Bustamante, D. Pei, and Y. Zhao. Where the sidewalk ends: Extending the internet as graph using traceroutes from p2p users. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, pages 217–228. ACM, 2009.

[21] Q. Chen, H. Chang, R. Govindan, and S. Jamin. The origin of power laws in internet topologies revisited. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 608–617. IEEE, 2002.

[22] R. Cohen and D. Raz. The internet dark matter-on the missing links in the as connectivity map. In *INFOCOM*, 2006.

[23] 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*, pages 183–196. ACM, 2008.

[24] A. Dhamdhere and C. Dovrolis. The internet is flat: modeling the transition from a transit hierarchy to a peering mesh. In *Proceedings of the 6th International COnference*, page 21. ACM, 2010.

[25] A. Dhamdhere and C. Dovrolis. Twelve years in the evolution of the internet ecosystem. *IEEE/ACM Transactions on Networking (ToN)*, 19(5):1420–1433, 2011.

[26] G. Di Battista, M. Patrignani, and M. Pizzonia. Computing the types of the relationships between autonomous systems. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 1, pages 156–165. IEEE, 2003.

[27] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, G. Riley, et al. As relationships: Inference and validation. *ACM SIGCOMM Computer Communication Review*, 37(1):29–40, 2007.

[28] X. Dimitropoulos, D. Krioukov, G. Riley, and K. Claffy. Classifying the types of autonomous systems in the internet. *SIGCOMM Poster*, 151, 2005.

[29] B. Donnet and T. Friedman. Internet topology discovery: a survey. *Communications Surveys & Tutorials, IEEE*, 9(4):56–69, 2007.

[30] S. N. Dorogovtsev and J. F. Mendes. *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press, 2013.

[31] N. Economides. The economics of the internet backbone. *NYU, Law and Economics Research Paper*, (04-033):04–23, 2005.

[32] B. Edwards, S. Hofmeyr, G. Stelle, and S. Forrest. Internet topology over time. *arXiv preprint arXiv:1202.3993*, 2012.

[33] Eu telecom's rules. `https://ec.europa.eu/digital-agenda/en/telecoms-rules/`.

[34] A. Fabrikant, E. Koutsoupias, and C. H. Papadimitriou. Heuristically optimized trade-offs: A new paradigm for power laws in the internet. In *Automata, languages and programming*, pages 110–122. Springer, 2002.

[35] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *ACM SIGCOMM computer communication review*, volume 29, pages 251–262. ACM, 1999.

[36] D. Fay, H. Haddadi, A. Thomason, A. W. Moore, R. Mortier, A. Jamakovic, S. Uhlig, and M. Rio. Weighted spectral distribution for internet topology analysis: theory and applications. *Networking, IEEE/ACM Transactions on*, 18(1):164–176, 2010.

[37] S. E. Fienberg. A brief history of statistical models for network analysis and open challenges. *Journal of Computational and Graphical Statistics*, 21(4):825–839, 2012.

[38] E. Fox Keller. Revisiting scale-free networks. *BioEssays*, 27(10):1060–1068, 2005.

[39] V. Fuller and T. Li. Classless Inter-domain Routing (CIDR):The Internet Address Assignment and Aggregation Plan. RFC 4632, RFC Editor, August 2006.

[40] L. Gao. On inferring autonomous system relationships in the internet. *IEEE/ACM Transactions on Networking (ToN)*, 9(6):733–745, 2001.

[41] L. Gao and J. Rexford. Stable internet routing without global coordination. *IEEE/ACM Transactions on Networking (TON)*, 9(6):681–692, 2001.

[42] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. The flattening internet topology: Natural evolution, unsightly barnacles or contrived collapse? In *Passive and Active Network Measurement*, pages 1–10. Springer, 2008.

[43] V. Giotsas, M. Luckie, B. Huffaker, et al. Inferring complex as relationships. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 23–30. ACM, 2014.

[44] V. Giotsas and S. Zhou. Improving the discovery of ixp peering links through passive bgp measurements. In *Computer Communications Workshops (INFOCOM WKSHPS), 2013 IEEE Conference on*, pages 121–126, April 2013.

[45] V. Giotsas and S. Zhou. Improving the discovery of ixp peering links through passive bgp measurements. In *Computer Communications Workshops (INFOCOM WKSHPS), 2013 IEEE Conference on*, pages 121–126. IEEE, 2013.

[46] G. Goth. New internet economics might not make it to the edge [news & trends]. *Internet Computing, IEEE*, 14(1):7–9, 2010.

[47] R. Govindan and A. Reddy. An analysis of internet inter-domain topology and route stability. In *INFOCOM'97. Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Driving the Information Revolution., Proceedings IEEE*, volume 2, pages 850–857. IEEE, 1997.

[48] R. Govindan and H. Tangmunarunkit. Heuristics for internet map discovery. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1371–1380. IEEE, 2000.

[49] E. Gregori, A. Improta, L. Lenzini, and C. Orsini. The impact of ixps on the as-level topology structure of the internet. *Computer Communications*, 34(1):68–82, 2011.

[50] E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani. Bgp and inter-as economic relationships. In *NETWORKING 2011*, pages 54–67. Springer, 2011.

[51] E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani. On the incompleteness of the as-level graph: A novel methodology for bgp route collector placement. In *Proceedings of the 2012 ACM Conference on Internet Measurement Conference*, pages 253–264, 2012.

[52] E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani. On the incompleteness of the as-level graph: a novel methodology for bgp route collector placement. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, pages 253–264. ACM, 2012.

[53] E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani. Discovering the geographic properties of the internet as-level topology. *Networking Science*, 3(1-4):34–42, 2013.

[54] E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani. A novel methodology to address the internet as-level data incompleteness. 2014.

[55] H. Haddadi, D. Fay, S. Uhlig, A. Moore, R. Mortier, and A. Jamakovic. Mixing biases: Structural changes in the as topology evolution. In *Traffic Monitoring and Analysis*, pages 32–45. Springer, 2010.

[56] H. Haddadi, M. Rio, G. Iannaccone, A. Moore, and R. Mortier. Network topologies: inference, modeling, and generation. *Communications Surveys & Tutorials, IEEE*, 10(2):48–69, 2008.

[57] H. Haddadi, S. Uhlig, A. Moore, R. Mortier, and M. Rio. Modeling internet topology dynamics. *ACM SIGCOMM Computer Communication Review*, 38(2):65–68, 2008.

[58] J. Hawkinson and T. Bates. Guidelines for creation, selection, and registration of an Autonomous System. RFC 1930, RFC Editor, March 1996.

[59] Y. He, G. Siganos, M. Faloutsos, and S. V. Krishnamurthy. A systematic framework for unearthing the missing links: Measurements and impact. In *NSDI*, 2007.

[60] C. Housley and C. Huston. The Internet Numbers Registry System. RFC 7020, RFC Editor, August 2013.

[61] B. Huffaker, A. Dhamdhere, M. Fomenkov, et al. Toward topology dualism: improving the accuracy of as annotations for routers. In *Passive and Active Measurement*, pages 101–110. Springer, 2010.

[62] B. Huffaker, M. Fomenkov, and K. Claffy. Internet topology data comparison. 2012.

[63] B. Huffaker, M. Fomenkov, D. Moore, et al. Macroscopic analyses of the infrastructure: Measurement and visualization of internet connectivity and performance. 2001.

[64] G. Huston. Interconnection, peering, and settlements. In *proc. INET*, volume 9, 1999.

[65] Internet2. `http://www.internet2.edu/`.

[66] Internet routing registries. `http://www.irr.net/`.

[67] N. Kamiyama, T. Mori, R. Kawahara, S. Harada, and H. Hasegawa. Analyzing influence of network topology on designing isp-operated cdn. *Telecommunication Systems*, 52(2):969–977, 2013.

[68] C. Kintzel, J. Fuchs, and F. Mansmann. Monitoring large ip spaces with clockview. In *VIZSEC*, 2011.

[69] D. Krioukov, M. Fomenkov, F. Chung, A. Vespignani, W. Willinger, et al. The workshop on internet topology (wit) report. *ACM SIGCOMM Computer Communication Review*, 37(1):69–73, 2007.

[70] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. *ACM SIGCOMM Computer Communication Review*, 41(4):75–86, 2011.

[71] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2, 2007.

[72] M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas, et al. As relationships, customer cones, and validation. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 243–256. ACM, 2013.

[73] D. Magoni and J. J. Pansiot. Analysis of the autonomous system network topology. *ACM SIGCOMM Computer Communication Review*, 31(3):26–37, 2001.

[74] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat. Systematic topology analysis and generation using degree correlations. In *ACM SIGCOMM Computer Communication Review*, volume 36, pages 135–146. ACM, 2006.

[75] P. Mahadevan, D. Krioukov, M. Fomenkov, X. Dimitropoulos, A. Vahdat, et al. The internet as-level topology: three data sources and one definitive metric. *ACM SIGCOMM Computer Communication Review*, 36(1):17–26, 2006.

[76] M. Z. Masoud, X. Hei, and W. Cheng. A graph-theoretic study of the flattening internet as topology. In *Networks (ICON), 2013 19th IEEE International Conference on*, pages 1–6. IEEE, 2013.

[77] W. B. Norton. The evolution of the us internet peering ecosystem. *Equinix white papers*, 2004.

[78] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. The (in) completeness of the observed internet as-level structure. *IEEE/ACM Transactions on Networking (ToN)*, 18(1):109–122, 2010.

[79] R. Oliveira, W. Willinger, B. Zhang, et al. Quantifying the completeness of the observed internet as-level structure. 2008.

[80] R. V. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. In search of the elusive ground truth: the internet's as-level connectivity structure. In *ACM SIGMETRICS Performance Evaluation Review*, volume 36, pages 217–228. ACM, 2008.

[81] R. V. Oliveira, B. Zhang, and L. Zhang. Observing the evolution of internet as topology. In *ACM SIGCOMM Computer Communication Review*, volume 37, pages 313–324. ACM, 2007.

[82] C. Orsini, E. Gregori, L. Lenzini, and D. Krioukov. Evolution of the internet k-dense structure. *IEEE/ACM Transactions on Networking (TON)*, 22(6):1769–1780, 2014.

[83] S.-T. Park, D. M. Pennock, and C. L. Giles. Comparing static and dynamic measurements and models of the internet's as topology. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 3, pages 1616–1627. IEEE, 2004.

[84] R. Pastor-Satorras and A. Vespignani. *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press, 2004.

[85] Packet clearing house. `https://www.pch.net/`.

[86] L. L. Peterson and B. S. Davie. *Computer networks: a systems approach*. Elsevier, 2007.

[87] Y. Rekhter and T. Li. A Border Gateway Protocol 4 (BGP-4). RFC 1654, RFC Editor, January 2006.

[88] The ripe routing information services. `http://www.ris.ripe.net/`.

[89] Regional internet registry afrinic. `ftp://ftp.afrinic.net/pub/stats/afrinic/`.

[90] Regional internet registry apnic. `ftp://ftp.apnic.net/pub/stats/apnic/`.

[91] Regional internet registry arin. `ftp://ftp.arin.net/pub/stats/arin/`.

[92] Regional internet registry lacnic. `ftp://ftp.lacnic.net/pub/stats/lacnic/`.

[93] Regional internet registry ripe ncc. `ftp://ftp.ripe.net/pub/stats/ripencc/`.

[94] M. Roughan, S. J. Tuke, and O. Maennel. Bigfoot, sasquatch, the yeti and other missing links: what we don't know about the as graph. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, pages 325–330, 2008.

[95] M. Roughan, W. Willinger, O. Maennel, D. Perouli, and R. Bush. 10 lessons from 10 years of measuring and modeling the internet's autonomous systems. *Selected Areas in Communications, IEEE Journal on*, 29(9):1810–1821, 2011.

[96] The route views project. `http://www.antc.uoregon.edu/route-views/`.

[97] A. Sallaberry, C. Muelder, and K.-L. Ma. Clustering, visualizing, and navigating for large dynamic graphs. In *Graph Drawing*, pages 487–498. Springer, 2013.

[98] M. Á. Serrano, M. Boguná, and A. Diaz-Guilera. Modeling the internet. *The European Physical Journal B-Condensed Matter and Complex Systems*, 50(1-2):249–254, 2006.

[99] S. Shakkottai, M. Fomenkov, R. Koga, D. Krioukov, and K. C. Claffy. Evolution of the internet as-level ecosystem. *The European Physical Journal B*, 74(2):271–278, 2010.

[100] Y. Shavitt and E. Shir. Dimes: Let the internet measure itself. *ACM SIGCOMM Computer Communication Review*, 35:71–74, 2005.

[101] G. Siganos, M. Faloutsos, and C. Faloutsos. The evolution of the internet: Topology and routing. 2002.

[102] L. Subramanian, S. Agarwal, J. Rexford, and R. H. Katz. Characterizing the internet hierarchy from multiple vantage points. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 618–627. IEEE, 2002.

[103] H. Tangmunarunkit, J. Doyle, R. Govindan, W. Willinger, S. Jamin, and S. Shenker. Does as size determine degree in as topology? *ACM SIGCOMM computer communication review*, 31(5):7–8, 2001.

[104] Internet AS-level topology archive. `http://irl.cs.ucla.edu/topology/`.

[105] Un: Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings. `http://unstats.un.org/unsd/methods/m49/m49regin.htm`.

[106] Verizon. Unbalanced peering, and the real story behind the Verizon/Cogent dispute. `http://publicpolicy.verizon.com/blog/entry/unbalanced-peering-and-the-real-story-behind-the-verizon-cogent-dispute`, June 2013.

[107] X. Wang and D. Loguinov. Wealth-based evolution model for the internet as-level topology. In *INFOCOM*, 2006.

[108] G. XiaJ. Ontheevaluationofasrelationshipinferences. *ProceedingsofIEEE-GLOBECOM*, page r1377, 2004.

[109] K. Xu, Z. Duan, Z.-L. Zhang, and J. Chandrashekar. On properties of internet exchange points and their impact on as topology and relationship. In *Networking 2004*, pages 284–295. Springer, 2004.

[110] S.-H. Yook, H. Jeong, and A.-L. Barabási. Modeling the internet's large-scale topology. *Proceedings of the National Academy of Sciences*, 99(21):13382–13386, 2002.

[111] M. Yu, W. Jiang, H. Li, and I. Stoica. Tradeoffs in cdn designs for throughput oriented traffic. In *Proceedings of the 8th international conference on Emerging networking experiments and technologies*, pages 145–156. ACM, 2012.

[112] S. Zarifzadeh, M. Gowdagere, and C. Dovrolis. Range tomography: combining the practicality of boolean tomography with the resolution of analog tomography. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, pages 385–398, 2012.

[113] B. Zhang, R. Liu, D. Massey, and L. Zhang. Collecting the internet as-level topology. *ACM SIGCOMM Computer Communication Review*, 35(1):53–61, 2005.

[114] Y. Zhang, R. Oliveira, Y. Wang, S. Su, B. Zhang, J. Bi, H. Zhang, and L. Zhang. A framework to quantify the pitfalls of using traceroute in as-level topology measurement. *Selected Areas in Communications, IEEE Journal on*, 29(9):1822–1836, 2011.

[115] Y. Zhang, R. Oliveira, H. Zhang, and L. Zhang. Quantifying the pitfalls of traceroute in as connectivity inference. In *Passive and Active Measurement*, pages 91–100. Springer, 2010.

[116] Y. Zhang, Z. Zhang, Z. M. Mao, C. Hu, and B. MacDowell Maggs. On the impact of route monitor selection. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 215–220. ACM, 2007.

[117] Z. Zhao and J. Bi. Characterizing and analysis of the flattening internet topology. In *Computers and Communications (ISCC), 2013 IEEE Symposium on*, pages 000219–000225. IEEE, 2013.

[118] S. Zhou. Characterising and modelling the internet topologythe rich-club phenomenon and the pfp model. *BT Technology Journal*, 24(3):108–115, 2006.

[119] S. Zhou and R. J. Mondragón. Accurately modeling the internet topology. *Physical Review E*, 70(6):066108, 2004.

[120] S. Zhou and R. J. Mondragón. The rich-club phenomenon in the internet topology. *Communications Letters, IEEE*, 8:180–182, 2004.