NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

AUTONOMOUS SYSTEM CHOKE POINTS IN COUNTRY-LEVEL NETWORK TOPOLOGY

by

Eric T. Regnier

March 2021

Thesis Advisor: Mathias N. Kolsch
Second Reader: Thomas J. Krenc

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** March 2021	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE
AUTONOMOUS SYSTEM CHOKE POINTS IN COUNTRY-LEVEL NETWORK TOPOLOGY

5. FUNDING NUMBERS

6. AUTHOR(S) Eric T. Regnier

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release. Distribution is unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT (maximum 200 words)

Internet traffic choke points within country-level logical networks exist at the Autonomous System (AS) level, with consequences and implications for country-level network topology and vulnerability to network disruption or surveillance. This thesis introduces the concept of such "Gateway ASs," which serve to connect the logical interior of a given country's network to the larger internet, and further demonstrates it to be a well-defined and useful concept. By fully characterizing the prevalence and nature of these Gateway ASes across the internet as a whole, this study demonstrates that the internet remains highly hierarchical at the country-level, despite the internet's evolutionary trend towards a "flattened" topology. Further, this conception and characterization of country-level network topology is leveraged to map vast portions of the logical internet landscape to physical country borders, but ultimately fails to provide an accurate and complete heuristic for internet infrastructure geolocation based upon logical AS classification. Finally, this study provides an assessment of the countries most susceptible to censorship events based upon the structure of their network topology, and quantifies an upper bound (by percentage of available IP space within the geographic confines of the country) for the effectiveness of such censorship schemes to fully sever network connectivity with the larger internet.

14. SUBJECT TERMS
networking, autonomous systems, internet censorship, internet topology, border gateway protocol

15. NUMBER OF PAGES
119

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

AUTONOMOUS SYSTEM CHOKE POINTS IN COUNTRY-LEVEL NETWORK TOPOLOGY

Eric T. Regnier
Lieutenant Commander, United States Navy
BS, United States Naval Academy, 2008

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
March 2021

Approved by: Mathias N. Kolsch
Advisor

Thomas J. Krenc
Second Reader

Gurminder Singh
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Internet traffic choke points within country-level logical networks exist at the Autonomous System (AS) level, with consequences and implications for country-level network topology and vulnerability to network disruption or surveillance. This thesis introduces the concept of such "Gateway ASs," which serve to connect the logical interior of a given country's network to the larger internet and further demonstrates it to be a well-defined and useful concept. By fully characterizing the prevalence and nature of these Gateway ASes across the internet as a whole, this study demonstrates that the internet remains highly hierarchical at the country level, despite the internet's evolutionary trend toward a "flattened" topology. Furthermore, this conception and characterization of country-level network topology is leveraged to map vast portions of the logical internet landscape to physical country borders but ultimately fails to provide an accurate and complete heuristic for internet infrastructure geolocation based upon logical AS classification. Finally, this study provides an assessment of the country's most susceptible-to-censorship events based upon the structure of their network topology and quantifies an upper bound (by percentage of available Internet Protocol [IP] space within the geographic confines of the country) for the effectiveness of such censorship schemes to fully sever network connectivity with the larger internet.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

List of Acronyms and Abbreviations

AS Autonomous System

ASN Autonomous System Number

BGP Border Gateway Protocol

CAIDA Center for Applied Internet Data Analysis

CIDR Classless Inter-Domain Routing

CSV Comma Separated Values

CDF Cumulative Distribution Function

C2P Customer to Provider

HE Hurricane Electric

IANA Internet Assigned Numbers Authority

ICMP Internet Control Message Protocol

IETF Internet Engineering Task Force

IXP Internet Exchange Point

IP Internet Protocol

ISP Internet Service Provider

JSON JavaScript Object Notation

NPS Naval Postgraduate School

PCH Packet Clearing House

PDB PeeringDB

P2P Peer to Peer

P2C Provider to Customer

RIR Regional Internet Registry

RFC Request for Comments

RIPE NCC Réseaux IP Européens Network Coordination Center

RIB Routing Information Base

RIS Routing Information Service

Acknowledgments

First and foremost, I recognize and appreciate my advisors for their guidance and counsel throughout this project. Their experience and perspective were instrumental in developing both the initial conception and final result, and they were extraordinarily generous with their time. I am also grateful to the NPS Computer Science Department faculty and NPS Administration for supporting me in my efforts to complete a master's degree while on staff, for which this thesis serves as the final requirement.

As with all things, I am profoundly grateful to my wife, Kasie, for her constant and unwavering support through the early mornings, the late nights, and the long weekends required to complete this thesis and the degree as a whole.

And finally, I express my greatest gratitude to our two miniature wiener dogs, Maddy and Nora, my constant desk companions throughout these long months of coronavirus quarantine. They protected me from all manner of mail deliverymen and -women, gave me bellies to rub when bellies needed rubbing, and were always cheerful to wake from the deepest slumber to provide a supportive slurp when required.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:
Introduction

If there were any doubt about the internet's centrality to the global economy, daily life of the world's inhabitants, and the administration of governments, the coronavirus pandemic laid those doubts to rest. At the outset of the pandemic, government and self-imposed restrictions designed to control the spread of the virus had the effect of migrating many aspects of work, economic, and social activities to the internet. The resulting surge in internet traffic created an unprecedented strain upon the logical and physical structure of the internet at a time when it is needed most. The present crisis will surely pass, but this increasing reliance upon the internet for governmental, economic, and personal activities will certainly endure, if not expand.

Yet despite its importance, the research community struggles to understand many features of the internet, including some foundational questions regarding the topology and organization of the various components that make up the internet as a whole. The most pernicious form of system vulnerabilities are those that remain unknown until intentionally exploited by a malicious actor. Consequently, this thesis endeavors to further the community's understanding of internet topology, with a focus on topological features that may pose security or availability vulnerabilities for the global network.

1.1 Motivation

In the military sciences, geographic choke points provide unique vulnerabilities and opportunities for both defensive and offensive operations, and the strategic value of such choke points has been used to great effect throughout history. The concept and implications of the choke point transcends geography, however, and similar features appear within the internet landscape as they do upon the land and sea. Many of these internet choke points are already well-known in the form of Internet Exchange Points (IXP), which serve to connect vast portions of the internet together in one logical location [1]. The impact of these IXPs on the internet's overall topology is an active area of research and will be discussed further in Chapter 2.

This study seeks to prove that other choke points exist within the internet landscape (formally introduced as "Gateway ASs" in Section 3.1), when the internet is viewed at the country-level of granularity. While the internet largely operates without regard for international borders, the larger internet arises from the connectivity of thousands of individual components called Autonomous Systems (AS), that themselves fall within the physical jurisdiction and regulatory cognizance of specific countries and territories. Within this context, this study attempts to group and analyze these thousands of individual internet components by country affiliation in order to determine whether unique logical topological features exist within the "territorial limits" of any given country's logical internet landscape. Since any Gateway ASs that do exist could be exploited to target individual countries, as opposed to general internet traffic, such features would provide even greater strategic importance than the IXPs currently known. The identification and exploration of such features is the motivation of this study.

1.2 Research Plan

The connections made between individual components of the larger internet (the logical, vice physical network) are implemented and facilitated by the Border Gateway Protocol (BGP), and data sets detailing these BGP connections are publicly accessible from various research services around the globe. This study endeavors to mine these data sets and map the connections contained therein at the country-level of granularity. Additionally, since BGP information is retained by the aforementioned research services, historical snapshots of the internet are also available to apply this analysis and identify differences in a given country's logical internet topology during significant events, such as internet censorship or attack. In this way, ground truth data, both past and present, will inform the results of this study.

1.3 Thesis Outline

The remaining chapters are organized as follows: Chapter 2 provides general information on the BGP, historical evolution of the internet's topology, economic considerations resulting from AS and BGP operation, and the role of BGP in internet censorship events. Chapter 3 discusses the objectives and hypothesis of this study, further proposing new terms and concepts to the field that support this analysis. Chapter 4 details the methodology and experimental design used to conduct the study. Chapter 5 provides results and initial observations of the experiment. Chapter 6 provides an in-depth analysis of the experimental results and implications for the field. And finally, Chapter 7 summarizes the contribution of this study and suggests avenues for follow on work.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 2:
Background

This chapter discusses the fundamentals of the internet's construction at the macro-level, to include the constituent units from which the larger internet arises, the protocol through which they communicate, economic considerations arising from their operation, evolution of the logical internet, and ways in which the BGP has been exploited to effect internet censorship in the past.

2.1 Structure of the Internet

The internet is a globally connected network of networks that communicate with one another through the specialized BGP, discussed in Section 2.2. Each of the constituent networks within the internet are called ASs, which are defined as, "a connected group of one or more Internet Protocol (IP) prefixes run by one or more network operators which has a single and clearly defined routing policy" [2]. As such, ASs range in type, size, and function from small Internet Service Providers (ISP) that provide connectivity to regional customers to massive networks that serve to connect ASs themselves together across the globe.

ASs are identified by globally unique 16 or 32 bit numbers assigned by the Internet Assigned Numbers Authority (IANA) via the AS's cognizant Regional Internet Registry (RIR), with constituent areas of responsibility detailed per Figure 2.1 [3]. RIRs make these assignments according to their own policies and in accordance with the criteria delineated in the Internet Engineering Task Force (IETF) Request for Comments (RFC) 1930 [2]. Due to differences in the size, degree of economic development, etc. of countries around the world, the distribution of Autonomous System Numbers (ASN) assigned to individual countries ranges from 1 (as with Andorra, for example) to 2,500 (as with the United States) [4].

While ASs themselves may operate internationally, each AS is registered by their parent RIR to one specific country at the time of ASN assignment, and the majority of ASs operate exclusively within their country of registration (discussed further in Section 4.4) [5]. Therefore, in order for ASs to provide their customers with "reachability" to customers serviced by a different AS (either within the same country or internationally), they must have a mechanism to exchange internet traffic amongst themselves, the BGP.

REGISTRY	AREA COVERED
AFRINIC	Africa Region
APNIC	Asia/Pacific Region
ARIN	Canada, USA, and some Caribbean Islands
LACNIC	Latin America and some Caribbean Islands
RIPE NCC	Europe, the Middle East, and Central Asia

Figure 2.1. Description of RIR Geographic Coverage. Source: [6].

2.2 The Border Gateway Protocol

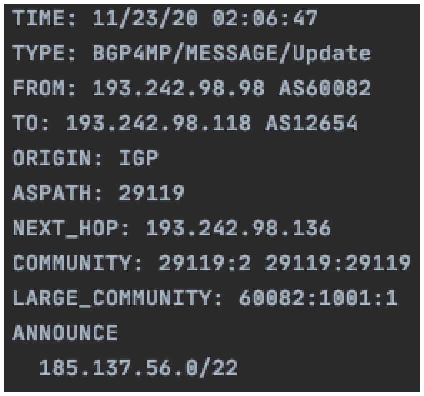
Internet routing is enabled via the BGP, which links ASs together and promulgates these associations throughout all ASs on the internet, such that traffic may route from the origin to the intended destination. Under the BGP, ASs advertise "routes" to one another, defined as "a unit of information that pairs a set of destinations with the attributes of a path to those destinations," via border routers at the logical "edges" of their networks [7]. Routes are stored locally within each receiving AS border router in a Routing Information Base (RIB), and RIBs in turn form a road map of sorts to inform the router of where internet traffic must be sent in order to reach any other destination on the internet [7].

As shown in Figure 2.2, amongst other fields RIB table entries include PREFIX, ASPATH, and NEXT_HOP attributes which provide all the information required to route traffic to its destination. RFC 8430 specifies the meaning and use of these attribute fields with the PREFIX attribute describing which IP address range (expressed in Classless Inter-Domain Routing (CIDR) format) is reachable via the corresponding entry in the RIB table. The ASPATH attribute provides a listing of ASs through which internet traffic must pass to reach IP addresses within the corresponding PREFIX, beginning with the first AS traffic must pass through and ending with the AS that owns the advertised prefix. And finally, the NEXT_HOP attribute lists the actual address (either fully resolved or unresolved) which the router must use to send traffic to the first AS listed in the ASPATH [8]. Therefore, when a border router running BGP receives network traffic destined for an IP address its AS does not possess, the border router consults its RIB table to identify the entry containing the appropriate PREFIX for the destination IP address of the outbound traffic and sends the traffic to the NEXT_HOP address listed in the entry. The receiving border router at the next AS forwards the internet traffic through the AS to another border router at the "other" end of its network (if required), which in turn performs the same operation as the original border router to send traffic to the next AS in the ASPATH.

```
TIME: 09/13/20 08:00:00
TYPE: TABLE_DUMP_V2/IPV4_UNICAST
PREFIX: 203.153.32.0/24
SEQUENCE: 842935
FROM: 198.32.242.254 AS52320
ORIGINATED: 09/09/20 11:00:49
ORIGIN: INCOMPLETE
ASPATH: 52320 9498 24186
NEXT_HOP: 198.32.242.254
MULTI_EXIT_DISC: 110
AGGREGATOR: AS24186 172.31.31.136
COMMUNITY: 52320:21311
```

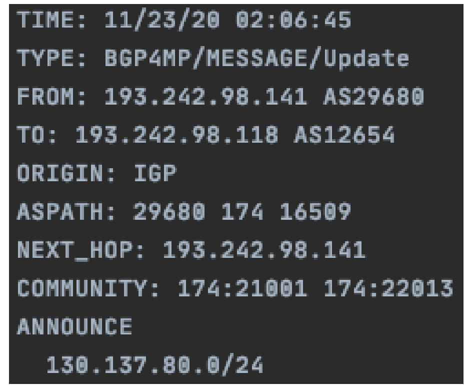
Figure 2.2. BGP RIB Table Entry

As detailed in RFC 4271, routes are advertised between ASs via BGP update messages which serve to provide information regarding new routes as well as to withdraw existing routes from service. As shown in Figure 2.3a, when an originating AS broadcasts a new route, it sends a BGP update message containing the PREFIX which the AS desires to advertise, an ASPATH consisting of the originating ASN, and the NEXT_HOP address to which traffic must be sent from receiving ASs in order to reach the originating AS (amongst other fields). The receiving ASs subsequently make a determination based upon their own internal routing policies to either propagate the update message further to its own AS connections or not. If so, they prepend their ASN to the AS_PATH, change the NEXT_HOP attribute to reflect the address of its own border router, and promulgate their own BGP update messages to ASs they themselves connect with (see Figure 2.3b) [7]. In promulgating route information from AS to AS through the entire network of ASs, all internet connected devices become reachable from any other device.



```
TIME: 11/23/20 02:06:47
TYPE: BGP4MP/MESSAGE/Update
FROM: 193.242.98.98 AS60082
TO: 193.242.98.118 AS12654
ORIGIN: IGP
ASPATH: 29119
NEXT_HOP: 193.242.98.136
COMMUNITY: 29119:2 29119:29119
LARGE_COMMUNITY: 60082:1001:1
ANNOUNCE
  185.137.56.0/22
```

(a) Prefix Announcement from Originating AS



```
TIME: 11/23/20 02:06:45
TYPE: BGP4MP/MESSAGE/Update
FROM: 193.242.98.141 AS29680
TO: 193.242.98.118 AS12654
ORIGIN: IGP
ASPATH: 29680 174 16509
NEXT_HOP: 193.242.98.141
COMMUNITY: 174:21001 174:22013
ANNOUNCE
  130.137.80.0/24
```

(b) Prefix Announcement Propagated Via Multiple ASs

Figure 2.3. Example BGP Update Entries

Figure 2.4 illustrates this process whereby a BGP update for prefix “X” is promulgated through the network shown and how this information is used to subsequently route internet traffic. AS1 advertises prefix “X” via BGP update message to AS2, which in turn determines via its own internal routing policy to propagate the route to AS3. Consequently, when border router 3b receives traffic from within its own AS bound for an IP address “A” within prefix “X,” border router 3b sends the traffic to the NEXT_HOP address advertised by AS2. Border router 2d receives and forwards the traffic within AS2 to border router 2b, which repeats the process performed by AS3 to deliver the traffic to AS1 border router 1a. Finally, border router 1a forwards the traffic internally within AS1 to reach the destination IP address A within prefix X.

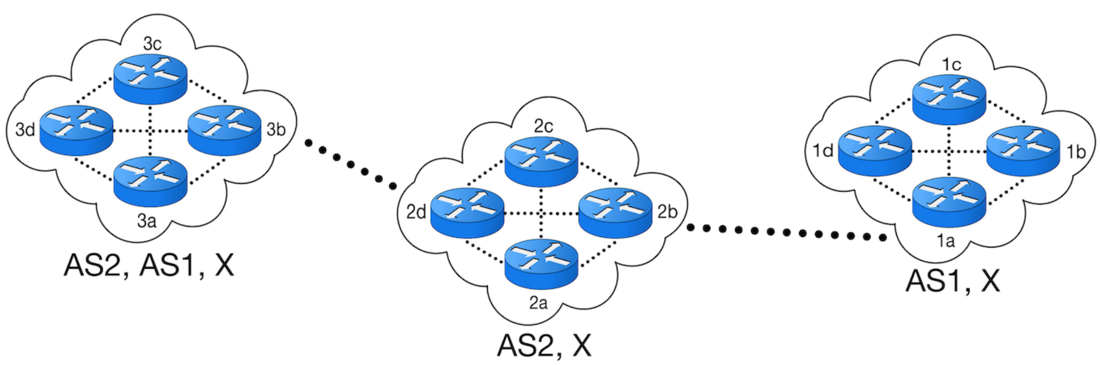


Figure 2.4. BGP Operation Between ASs

2.3 Border Gateway Protocol Route Collectors

BGP RIB tables provide extremely useful information to infer details of internet topology and have therefore been widely used by the research community to study the structure and dynamics of the internet. However, because ASs are operated by private or government entities, individual AS BGP RIB tables are typically not made available in any public venue. As a result, several research projects have established BGP route "collectors" which serve to collect BGP update messages from participating ASs and form a single, aggregated RIB table for public use [9].

It is crucial to understand that there is no single "truth" to the structure of the internet, and route collectors provide snapshots of the logical internet from the perspective of the participating ASs only. Just as a sculpture will show different aspects to each person standing in a crowded art gallery, with some details in sharper focus to those closest to the sculpture and some aspects hidden entirely from any one person, logical features of the internet present differently based upon an AS's logical position within the network of networks. In his doctoral dissertation, Improta further explicated this issue as follows. Because collector projects receive BGP update messages from a finite number of feeders, their "view" of the internet only reflects those of their feeders. Consequently, details of connections between non-feeder ASs may remain hidden to the collector projects. Exacerbating this problem, the ASs that agree to feed collector projects are typically very large ISPs with global reachability. As such, their BGP RIB tables lack many of the connections formed between medium and small ASs and these details remain hidden from the research community [5]. This "vantage point" issue remains a problematic, unavoidable aspect of internet research and the results of this study are similarly qualified by this limitation (discussed further in Section 4.3).

2.4 Border Gateway Protocol Customer Relationships

One important function of an AS is to provide global connectivity to its customers, a task that no single AS can perform in isolation. Because the internet functions as a network of networks, all ASs rely upon other ASs to carry traffic from their own customers to customers belonging to another AS. Furthermore, each AS functions as an independent organization, with its own business and resource considerations [2]. Consequently, the arrangements and agreements that ASs enter into with one another to facilitate this exchange of internet traffic are complex, and their analysis provides insight into the size, dependencies, and relative importance of a given AS on the regional and global scale.

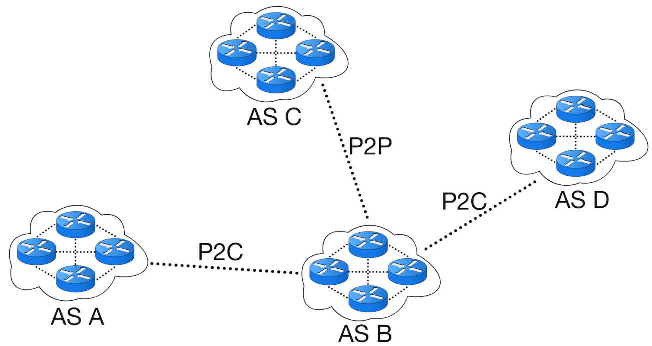
In their seminal work on the subject, Dimitropoulos et al. [10] provide insight into these customer arrangements and relationships as summarized in this paragraph. They describe that in general, inter-AS business relationships may be categorized as Provider to Customer (P2C) or Peer to Peer (P2P). Under a P2C customer relationship the customer AS pays the providing AS for reachability to all ASs available from the providing AS (including the providing AS itself), whereas under a P2P customer relationship the ASs entering into the agreement mutually share their AS reachability with the other at no cost. As independent business organizations, ASs are financially incentivized to enter into peering relationships that benefit themselves. Therefore, P2P arrangements are normally established to avoid a more expensive P2C relationship providing reachability to desired ASs and/or when two ASs desire access to the other and neither is sufficiently incentivized to purchase access from the other [10].

Building upon these concepts, the AS customer cone for AS “X” is defined as “the set of ASs that X can reach using P2C links; for AS X the customer cone includes X’s customers, as well as X’s customers’ customers, and so on” [11]. One particularly useful interpretation of the customer cone is as a metric to describe the “size” of a given AS with regards to logical reachability of other ASs, vice a summation of infrastructure components or geographic footprint of coverage. For example, as reflected on the Center for Applied Internet Data Analysis (CAIDA) repository of customer cone data, AS3356 (one of the few large “Tier-1” ASs that form the backbone of the internet from a reachability perspective) possesses a customer cone size of 45,771 ASs, well more than half of the approximately 68,000 ASs operating today [12]. In contrast, nearly 96 percent of all ASs possess 10 or fewer ASs in their customer cones and 84 percent have no other ASs in their customer cones at all [12]. Consequently, customer cone size represents an important metric of individual AS size and reachability since AS size and number of other ASs capable of being reached from the AS in question are directly related.

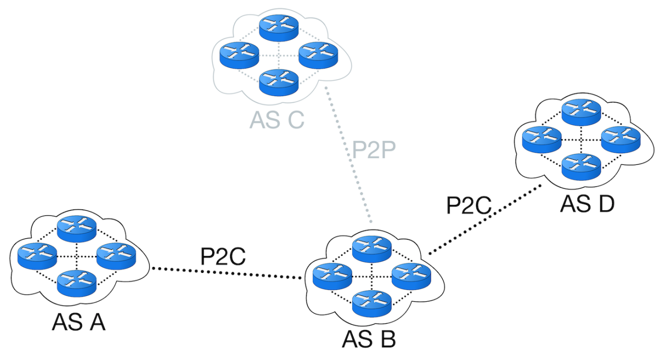
One additional usage of the customer cone size is in the prediction of P2P relationships amongst ASs. Intuitively, CAIDA finds that ASs with a large disparity in customer cone size are less likely to enter into P2P relationships with one another, whereas those with similar customer cone sizes are more likely to initiate P2P relationships [13]. Within the context of this study, customer cone size will play a crucial role in explicating the size and influence of any choke points identified in country-level network topology.

The final economic aspect with implications for this study is the "valley-free" principle, introduced by Gao. Under this principle, "after traversing a P2C or P2P edge, the AS path cannot traverse a Customer to Provider (C2P) or P2P edge," and largely describes routing behavior that tends to minimize cost and maximize revenue for a given AS to route traffic (here, "edge" refers to a BGP connection between ASs of the specified type) [14]. The implication of this principle is that ASs tend to announce all routes to their own customers (whether obtained from a provider, peer, or a customer), whereas they tend to only announce customer routes to peers and providers [15]. For example, consider Figure 2.5 where AS A is a provider to AS B, AS B is a provider to AS D, and AS B is a peer with AS C. By the valley-free principle and as shown in Figure 2.5a, AS D (as a customer of AS B) will receive all available routes from AS B ("ASPATH: AS C, AS B" and "ASPATH: AS A, AS B"). However, as shown in Figure 2.5b, AS A (as a provider to AS B) will only receive a route advertisement from AS B to AS B's customer ("ASPATH: AS B, AS D"), not AS C as a peer to AS B. Similarly, as shown in Figure 2.5c, AS C (as a peer to AS B) will only receive a route advertisement from AS B to to AS D ("ASPATH: AS B, AS D").

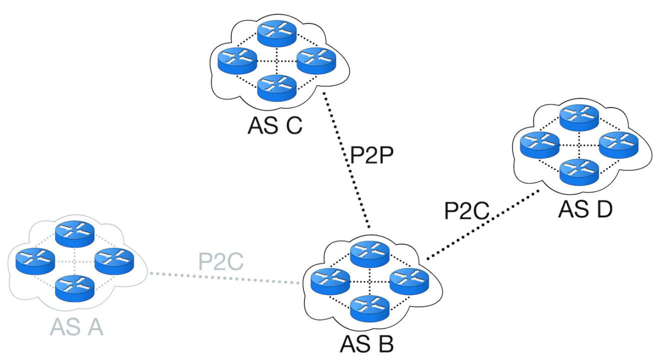
This is an important principle to consider within the context of the discussion on BGP collectors in Section 2.3. Because collector feeders tend to be large ASs with large customer cones, the majority of their routes are obtained via P2C links and they will consequently not provide visibility into many customer AS P2P connections [5]. (e.g. As shown in Figure 2.5b, AS A's RIB table is not likely to contain routes to AS C even though a BGP connection does exist between AS B and AS C.) This exacerbates the vantage point issue for the RIB data provided by the collector projects and remains an unresolved limitation within the internet research community.



(a) BGP Routes Advertised to AS D



(b) BGP Routes Advertised to AS A



(c) BGP Routes Advertised to AS C

Figure 2.5. Illustration of Valley-Free Principle

2.5 Autonomous System Topology and the “Flattening” of the Internet

The mechanisms and resulting topologies by which ASs peer with one another are complex and evolving, but the trend since the inception of the internet is for increased peering and increased inter-connectiveness between ASs. As discussed by Böttger et al. [16], historically, a small number of very large ASs (Tier-1 ASs) sat atop the internet hierarchy aggregating traffic from within their respective customer cones for exchange between the Tier-1 ASs themselves to provide global reachability. They further detail how these large Tier-1 ASs remain today, yet their dominance atop the internet hierarchy is diminished by the proliferation of IXPs, which enable many ASs to establish their own peering relationships on a common subnet without the need to traverse a Tier-1 AS. With traffic freed from the traditional constraints of the Tier-1 AS customer cones, the internet hierarchy has indeed “flattened” to an extent. However, the diversion of traffic away from Tier-1 ASs in favor of IXPs has had its own centralizing effect upon inter-AS traffic bound for IXP sites [16]. Consequently, while the proliferation of IXPs has served to flatten the original internet hierarchy, in doing so IXPs have given rise to new hierarchies within the logical internet topology. These changes represent important steps in the evolution of the internet, with implications for economic and resiliency aspects of internet operation as traffic redistributes across a changing topological terrain. This study endeavors to identify similar, undiscovered topologies and explicate their significance within the internet as a whole.

2.6 The Role of the Border Gateway Protocol in Internet Censorship

The BGP has been and continues to be exploited by governments to facilitate internet censorship, surveillance, or complete disconnection from the internet. While the internet as a whole remains remarkably robust and resilient, due in large part to the independent nature of the constituent ASs which make up the network of networks, the BGP-enabled censorship events described in this section reveal a vulnerability in this system for disruption. When a single organizational entity is able to exert control or influence over a grouping of ASs, they are able to effect regional disruptions to the flow of internet traffic [17]. Such is the case when governments exert control or influence over the ASs registered to and operating within its territorial borders, as is discussed next.

Dainotti et al. [17] provide a comprehensive examination of the Arab Spring censorship events in Egypt and Libya, which effected near-complete disconnection from the internet as a way of censoring their population during a time of political unrest. Figure 2.6 shows the visibility of prefixes advertised from the main Egyptian ASs during the censorship event on January 27, 2011. Their research showed that by leveraging the control and influence it possessed over the ASs registered within its borders, the Egyptian regime was able to force these ASs to issue near-simultaneous BGP route withdraw update messages and effectively sever its internet space from the global internet [17]. In doing so, the Egyptian regime effectively weaponized the BGP itself to effect their own aims. As such, this mechanism was not a "bug," but rather a feature of the BGP that they repurposed for nefarious ends. Later that same year, Libya experienced a very similar BGP-enabled country-wide internet disconnection, further demonstrating the efficacy of BGP-enabled government censorship of its citizens.

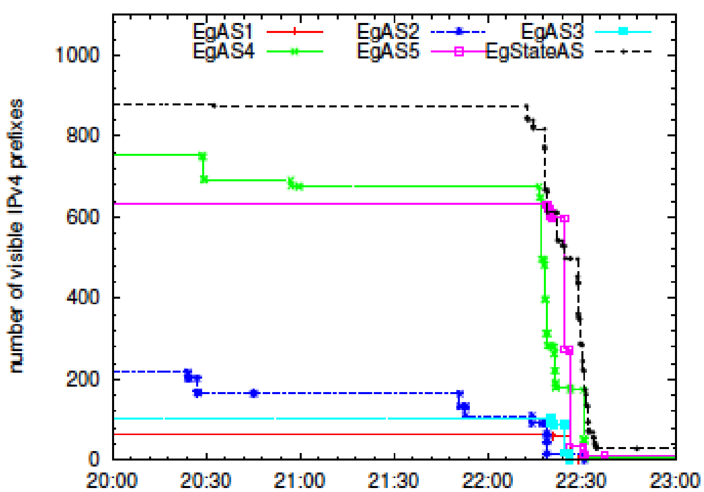


Figure 2.6. Visibility of Egyptian AS Prefixes During the Arab Spring Censorship Event. Source: [17].

A lesser known and researched censorship event exploiting the BGP is purported to have been tested by Russia. In December 2019, Russian authorities reported to have successfully demonstrated country-wide disconnection of the Russian internet space from the larger internet, maintaining an operational internal "Sovereign RuNet" [18]. A government spokesman explained that the test was conducted in response to a newly passed "Sovereign Internet" law, and was intended to demonstrate Russia's capability to respond and defend against external threats [19]. Technical details for how such a feat was accomplished are not publicly available (if it was indeed accomplished at all), but such country-wide disconnection from the larger internet while preserving an internal RuNet would necessarily have involved BGP, either directly or indirectly. If accomplished directly, Russian ASs which peer with non-Russian ASs would have had to withdraw all Russia-internal route advertisements to external ASs and withdraw all Russia-external route advertisements from Russia-internal ASs. If accomplished indirectly, the same grouping of ASs would have had to turn off their border routers running BGP and/or physically disrupt the network connections running from their routers to the larger internet. Since this set of ASs which interface with Russia-external ASs are a product of the BGP, BGP is what fundamentally enables this scenario for physical disconnection of the RuNet. In either case, BGP would necessarily have played a central role in the undertaking and the possibility of BGP-enabled country wide internet disconnection in the future should not be discounted.

2.7 Geolocation of BGP Prefixes

The final aspect of BGP operation discussed in this section concerns the physical, geographic boundaries within which ASs and their constituent BGP prefixes provide network connectivity (hereafter referred to as "range"). There is no restriction on the geographic range of any AS, with the result that ASs can (and many do) possess range within countries to which they are not registered. For these ASs, some subset of their total prefix allotment range outside the geographic confines of their registered country, with the remainder of their prefixes ranging within [5]. This characteristic substantially complicates the task of "mapping" the logical internet topology to physical boundaries, with many implications for the findings of this study and past censorship events. For instance, Egypt and Libya may have been able to effect internet shutdown via BGP for the ASs registered to them (therefore under their influence), but their regimes would not have been able to similarly disrupt internet connectivity from country-external ASs with range inside Libyan and Egyptian geographic boundaries by this method. The RuNet experiment would be similarly affected, with some relatively small subset of users within the country maintaining internet connectivity via country-external ASs with range inside Russia. These logical observations are explored further in Sections 5.8 and 6.2.

It must be noted that the geolocation of internet connected devices is a developing field, with varying approaches and levels of accuracy depending upon the granularity of geolocation desired (precise location, city-level, country-level, etc.). Wang et al. [20] discuss the various methodologies for approaching this problem to include active, passive, and hybrid techniques. Under the active approach, measured time delays in Internet Control Message Protocol (ICMP) "ping" queries from multiple known physical locations, along with physical properties for the transmission of electronic information, develop an area of uncertainty within which the target host must be located. Under passive geolocation, database products and data-mining techniques are used to correlate specific IP addresses with likely physical locations. And finally, hybrid approaches combine these techniques in an effort to improve the accuracy of the result produced [20]. Many factors impact the ultimate accuracy of these approaches to include the fact that internet traffic routing does not proceed in a straight line from physical point to physical point, internet traffic normally passes through many physical devices (which require time to process and forward the traffic and artificially inflates the overall transit time), data-driven clues for geolocation may provide inaccurate or outdated information, etc. Researchers are developing clever methodologies to improve these tools, but at present the accuracy of individual IP address geolocation remains notoriously poor [21].

In contrast, the geolocation accuracy of IP addresses at the *country*-level of granularity is substantially improved, with a 2011 CAIDA study revealing >95 percent agreement between the major geolocation services at the country-level [22]. A related, separate product provided by these services builds upon *individual* IP address geolocation discussed thus far to provide geolocation data for entire BGP *prefixes*. Yet nuances in BGP prefix geolocation data may degrade the overall accuracy of country-level BGP prefix geolocation as compared to constituent IP address geolocation, as detailed by a 2019 CAIDA study by Winter et al. [23]. In this work, they address BGP prefix geolocation specifically and identify "ambiguous" prefixes within the data available from the geolocation services. They determine ambiguous prefixes to be those where a smaller prefix geolocates to a different location than a larger prefix subsuming the smaller, with current methodologies incapable of definitively resolving the discrepancy. They further detail the extent of prefix ambiguities within the available databases, showing 0.3 percent of all prefixes (representing 8 percent of the overall IP address space) is affected at the country-level of granularity for the MaxMind database [23]. The implications of these findings for the purposes of this study is discussed in Section 4.1.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 3:
Objectives, Hypothesis, and Research Questions

This chapter discusses terms and definitions, the central hypothesis, and research questions that scope the study.

3.1 Introduction of the Gateway AS and Gateway Factor as Features of Country-Level Logical Network Topology

Several new classifications of ASs are defined in this section to support the study's analysis of choke points in country-level logical network topology. To enable these classifications, logical boundaries in internet space are first defined as follows. A given country's logical internet "territory" represents the network topology arising exclusively from those ASs registered to the same country. As a result, a country's logical network "border" represents the BGP connections between ASs registered to the country by its parent RIR and ASs that are not. Consequently, this boundary represents the logical partition between a given country's logical network and the larger internet.

Within this construct, Figure 3.1 depicts the four classifications of ASs introduced by this study. An "External AS" is one that that is *not* registered to the target country and possesses at least one BGP connection to an AS that *is* registered to the target country. An "Internal AS" is one that is registered to the target country and only possesses BGP connections to ASs that share the same country registration. "Gateway ASs" represent the logical choke points in country-level network topology discussed thus far, and are defined to be those which possess at least one BGP connection to an External AS *and* an AS sharing its country registration. And lastly, an "Outpost AS" is one that possesses BGP connections to *only* External ASs. Note that these definitions are mutually exclusive, they represent the minimum criteria for AS classification, and each type of AS may possess numerous BGP connections to various ASs within the restrictions of its classification. (e.g., One Gateway AS may possess BGP connections to many External ASs and many Internal ASs...)

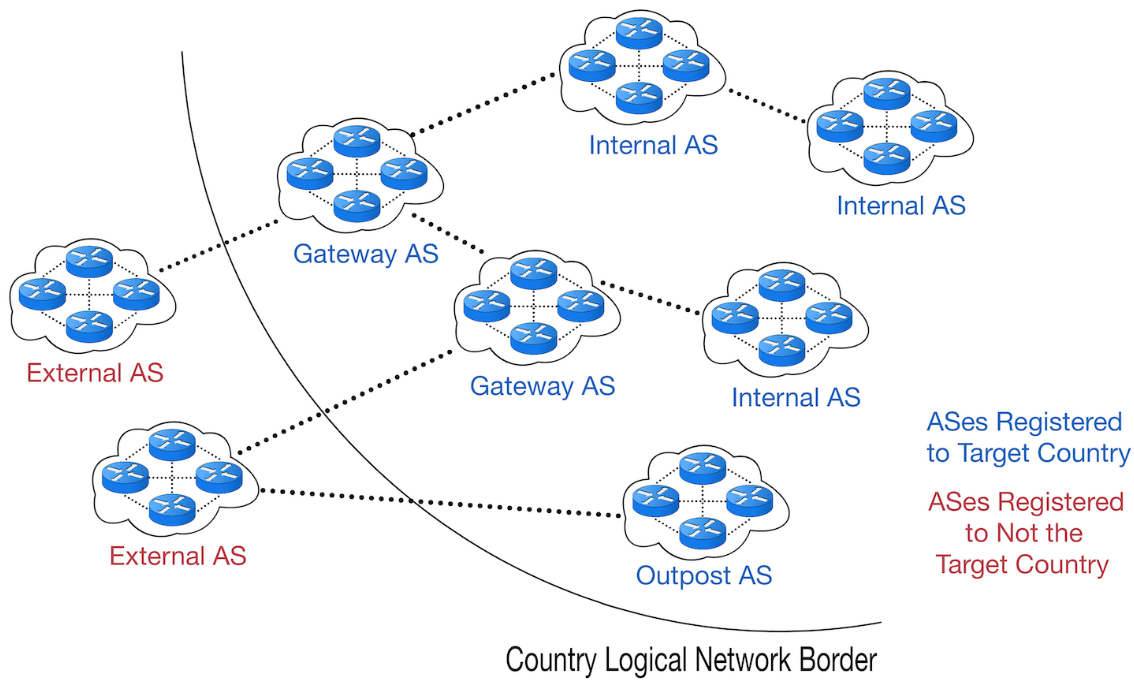


Figure 3.1. Country-Level Logical Network AS Classifications

Finally, the "Gateway Factor" is introduced as a metric to describe the proportion of a given country's logical network that depends upon Gateway ASs for reachability to the larger internet. As shown in Equation 3.1, the Gateway Factor is a simple ratio between the number of Internal ASs and total visible ASs for a given country, with a higher Gateway Factor describing a higher degree of traffic constraint for the country in question. By this definition, a Gateway Factor value of "0" represents a country with no Gateway ASs, whereas a Gateway Factor approaching "1" represents a country with most of its ASs falling logically behind the country's Gateway AS(s). Consequently, another interpretation for the Gateway Factor is how meshed the given country's logical network is with the larger internet. Within this context, the Gateway Factor will be utilized to address research questions pertaining to internet hierarchy as discussed in Section 3.3.

$$Gateway\ Factor = \frac{\#\ Internal\ ASs}{\#\ Internal\ ASs\ +\ \#\ Gateway\ ASs\ +\ \#\ Outpost\ ASs} \tag{3.1}$$

3.2 Hypothesis

The hypothesis motivating this study is as follows: Constrained points of entry into a target country's logical internet routing network exist at the AS level, and these Gateway ASs have implications for understanding hierarchies within country-level logical network topology, for the mapping of a country's logical network to its physical borders, and for the identification of vulnerabilities within a country's logical network.

This study anticipates that BGP connections between ASs sharing a common country registration will give rise to Gateway ASs, which aggregate network traffic in a manner similar to IXPs. Consequently, these Gateway ASs may functionally serve the same role as an IXP, connecting a relatively large set of origin and destination ASs that must traverse the AS in question. If true, Gateway ASs would serve to improve interconnectedness and provide a benefit for internet traffic routing, but they would also aggregate network traffic and may facilitate surveillance activities or unintentionally create vulnerabilities for disruption as do IXPs [1]. In contrast to IXPs, however, Gateway ASs constitute constrained point of entry into a given country's logical network and therefore represent gateways into an internet environment with a common country-affiliation and regulatory framework. Such operational subordination has been exploited by governments in the past (as observed with Libya and Egypt per Section 2.6). Consequently, this study's concept and investigation of Gateway ASs forms the primary area of research. As the first AS registered to a specified country encountered by internet traffic inbound from the larger internet, Gateway ASs would represent a unique feature of internet routing due to their centralizing effect on internet traffic, their necessity for the flow of traffic into and out of a country's logical borders, and their potential susceptibility to influence or directly control from the host country.

3.3 Research Questions

The research questions that support exploration of the study's hypothesis are as follows:

1. Do ASs registered to the same country form BGP topologies with one another that give rise to Gateway ASs?
2. Is the identification of Gateway ASs dependent on BGP collector vantage point?

3. Are there commonalities in country-level AS network topology between nations of similar governmental structures (authoritarian vs. non-authoritarian) or IANA Region?
4. Are there commonalities in country-level AS network topology between countries that have experienced, or continue to experience censorship events?
5. Are some Gateway ASs disproportionally important for their country's overall connectivity to the larger internet?
6. Is the Gateway AS analysis useful in mapping a country's logical internet to its physical borders?

CHAPTER 4:
Methodology and Experimental Design

This chapter discusses the data sources utilized, experimental design, and experimental execution upon which this study is constructed. Limitations of the study are also introduced along with an initial discussion of their impact upon the overall study results and analysis.

4.1 Data Sources

This study employs data from the following sources for the usages indicated:

- Per Table 4.1, the Routing Information Service (RIS) Réseaux IP Européens Network Coordination Center (RIPE NCC) maintains active BGP RIB collectors at 21 separate locations around the world, publishes the resulting BGP RIB tables and RIB update messages, and maintains a historical repository of this data for use and analysis by the research community. For the study's analysis of the current state of the internet presented in Chapter 5, the 0800 November 23, 2020, BGP RIB tables and preceding eight hours of BGP update messages were downloaded from each of the 21 collector locations for use in the main algorithm upon which this study is constructed [24]. The rationale for this data selection scheme as well as the compilation and usage of this data is further discussed in Section 4.3. To investigate how country-level network topology has evolved over time per Section 6.1, BGP RIB tables and preceding eight hours of BGP update messages were also downloaded from 0800 November 23 in the years 2005, 2010, and 2015.

- As discussed in Section 2.1, ASN assignments are made by the IANA via the respective RIRs. These RIR country to ASN assignments were utilized by this study to ensure that the study utilized ground truth data provided by the IANA, vice any third-party association data. Numerous services and databases exist to parse and map this data in various ways, and this study obtained IANA ASN country registration information via ipinfo.io [4].

Table 4.1. RIPE NCC Collector Details. Source: [24].

Collector Number	Collector Location	Number of Feeders
rrc00	Amsterdam, Netherlands	142
rrc01	London, United Kingdom	142
rrc03	Amsterdam, Netherlands	153
rrc04	Geneva, Switzerland	20
rrc05	Vienna, Austria	61
rrc06	Otemachi, Japan	8
rrc07	Stockholm, Sweden	36
rrc10	Milan, Italy	67
rrc11	New York, USA	45
rrc12	Frankfurt, Germany	150
rrc13	Moscow, Russia	36
rrc14	Palo Alto, USA	28
rrc15	Sao Paulo, Brazil	66
rrc16	Miami, USA	35
rrc18	Barcelona, Spain	26
rrc19	Johannesburg, South Africa	61
rrc20	Zurich, Switzerland	65
rrc21	Paris, France	76
rrc22	Bucharest, Romania	30
rrc23	Singapore	35
rrc24	Montevideo, Uruguay	21

- CAIDA compiles, maintains, and publishes an inferred listing of ASN customer relationships and customer cones based upon the work of Luckie et al. [11]. The 01 November, 2020, version of this data table was downloaded and utilized for data analysis in this study [25].

- The private company MaxMind produces and maintains a publicly available IP geolocation service at various levels of granularity. The MaxMind GeoLite2 country-level BGP prefix geolocation database was downloaded on 24 November, 2020, for use in this study [26]. The usage of MaxMind data is discussed further in Section 4.4.

- CAIDA utilizes machine learning techniques to compile a database of AS classifications based upon business type (content, transit, or enterprise). The 01 November, 2020, version of this database is utilized for data analysis in this study [27].

4.2 Experimental Methodology and Design

Fundamentally, this study analyzes ASPATH and prefix information resident within BGP RIB tables to identify logical topological features of the internet. More specifically, the algorithm employed by this study operates on each entry within a given BGP RIB table to identify which country the originating AS is registered to. The algorithm subsequently maps the upstream AS relationships revealed by the entry's ASPATH field up to the first External AS. Each AS in the ASPATH that is both within the set of those assigned to the target country and makes BGP connections only with another AS's within the set of those assigned to the target country is identified as an Internal AS. In contrast, the single AS within the ASPATH that is both within the set of those assigned to the target country and makes BGP connections only with an AS not within the set of those assigned to the target country is identified as a Gateway AS. ASs registered to the target country that make BGP connections with one or more External ASs, but do not make BGP connections with any Internal AS or Gateway AS are identified as an Outpost AS. Through the application of this process to each entry in all BGP RIB tables analyzed by the study, a complete country-level logical network topology is determined for each target country under analysis. Additionally, each entry's prefix is geolocated via the GeoLite2 database discussed in Section 4.1 to track which prefixes range within their registered country, which do not, and which country those that do not range domestically actually range within. Longest prefix matching is employed when querying the GeoLite2 database, as the product does not contain entries for all prefixes subsumed by the longest matching.

The output of this main algorithm includes various Comma Separated Values (CSV) files capturing metrics and observations and a series of JavaScript Object Notation (JSON) files for each BGP RIB table and target country which captures the network topology. A subsequent algorithm aggregates the data and network topologies identified from each of the individual BGP RIB tables to provide a comprehensive view of the target country's logical network as if viewed from all collector locations simultaneously. Having determined country-level logical network topologies in this way, subsequent data analysis involving the CAIDA customer relationship databases, AS classifications database, etc. give rise to determinations and analysis described in Chapter 5. Figure 4.1 provides a process diagram for the algorithm's operation and the full algorithm code is provided in Appendix C.

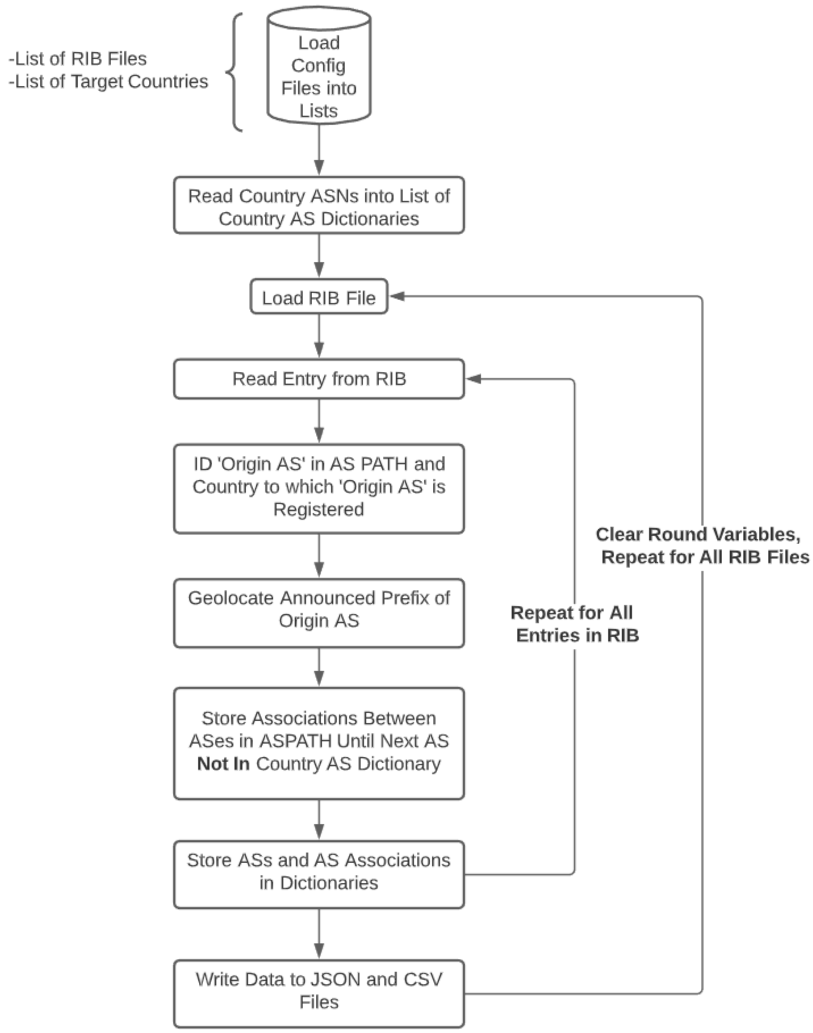


Figure 4.1. Main Algorithm Process Diagram

Finally, a slight modification of the aforementioned experimental design is implemented in order to enable analysis of censorship events and the evolving state of country-level logical internet topology. The algorithm discussed thus far involves analysis of BGP RIB tables obtained from different collector locations but with the same date and time of BGP RIB table capture, enabling analysis of a common snap shot of the internet across multiple collector locations. By obtaining BGP RIB tables from the same collector location but with differing collection times, this study was able to identify and analyze differences in country-level logical internet topology over time.

4.3 Limitations of BGP RIB Table Accuracy and Completeness

By utilizing the ASPATH field within each relevant BGP RIB table entry, the algorithm employed by this study operates on the same information used by AS BGP routers to actually send and receive internet traffic. Consequently, the data generated from this study will be accurate and complete insofar as the BGP RIB tables themselves are accurate and complete. Unfortunately, BGP RIB tables are well-known to provide an incomplete representation of the logical internet for two reasons. First, as discussed Section 2.2, ASs may choose to use and propagate BGP advertisements they receive based upon their own routing policies and business relationships, or not. As such, an AS's BGP RIB table does not list many of the routes available for service, as these additional routes are not consistent with the AS's routing policies and are therefore not captured in the BGP RIB table. Fortunately, these additional routes may be identified by analysis of the BGP update messages that individual feeders use to construct the BGP RIB tables themselves.

The RIPE NCC collector project publishes full BGP RIB tables from its constituent feeders every eight hours, with a full listing of BGP update messages received by those same feeders published every five minutes in between. These update messages contain the additional routes discarded by individual feeders as discussed thus far in this section. And so, by aggregating all BGP update messages published between full BGP RIB table snapshots and combining them with a full BGP RIB table, this study ensures that *all* available routes are analyzed by the main algorithm detailed in Section 4.2. This mitigates the first cause for incompleteness in BGP RIB table data, to the extent that it can be.

The second reason BGP RIB tables are incomplete is due to the collector vantage point issues discussed in Section 2.3. BGP RIB tables are constructed from collectors which receive peering information from a grouping of participating, or feeder ASs. Within the context of the full internet, the number of feeder ASs into each BGP collector is minuscule and feeder ASs tend to be large ASs. As a result of feeder AS size and reachability, the peering relationships they themselves enter into are predominantly P2C and therefore, due to the valley-free principle BGP RIB tables tend not to reveal P2P connections made between medium and small ASs [9]. Consequently, the state of the internet captured by BGP collectors is inherently skewed and incomplete.

The final factor to consider with regards to BGP RIB data addresses inaccuracies in BGP RIBs themselves. The BGP itself is susceptible to data corruption due to route hijacking and configuration errors as described by Shi et al. [28], which can introduce false ASPATHS into BGP RIB table entries. However, such attacks and anomalies are shown to be negligible within the context of the larger internet and relatively short-lived. They are therefore disregarded for the purposes of this study.

The aforementioned BGP incompleteness problem is unavoidable with the tools currently available to internet researchers, yet research proceeds with due caution for the inferences and conclusions drawn from such data. This study endeavors to do the same and will exercise appropriate restraint when drawing conclusions in Chapter 6. To qualify the impact that BGP incompleteness is anticipated to present to this study, however, it is important to emphasize that the Gateway AS analysis endeavors to identify *major* choke points into a country's logical internet topology as viewed from the larger internet. The methodology presented in this chapter will certainly fail to identify some number of peering connections between smaller, regional ASs (either between Gateway ASs and/or Internal ASs, or between External ASs and Gateway ASs, and/or Internal ASs, and/or Outpost ASs), yet the methodology is expected to show utility for identifying the major internet through-fares into a target country's network. Gregori et al. [9] likened the act of performing research based upon BGP RIB tables to, "analyzing a road map of a given country where highways are known, but most of the secondary roads are not shown." For the purposes of this study, it is the highways themselves which are the routes of interest.

4.4 Exclusivity of Country Registered AS Operation within Geographic Bounds of Its Registered Country

This study seeks to explore how the boundaries of a country's logical network correlates with its physical boundaries, and how Gateway ASs constrain traffic into these logical and physical regions alike. It is anticipated that the reality of inter-country AS ranging will complicate this task. In his 2013 doctoral publication, Improta introduced an AS level geolocation methodology based upon BGP prefix advertisements for each AS and the MaxMind GeoIPLite database, identifying that 7.99 percent of all ASs operate in two countries and 3.2 percent operate in more than two [5]. He additionally identified that the probability of a country-registered AS operating across national borders is correlated with the number of ASs registered to the country, as shown in Figure 4.2. These findings are expected to have particular importance for this study, since the number of registered ASs within the distribution of countries used in this analysis ranges from 1 to 2,500.

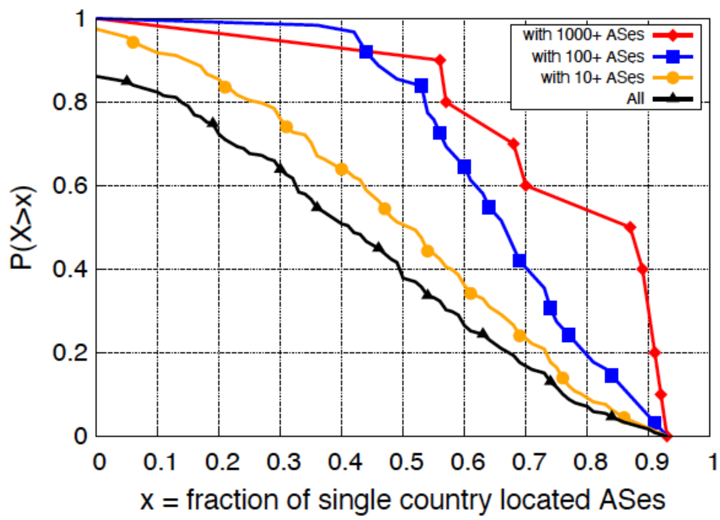


Figure 4.2. CDF of AS Operating in One Country. Source: [5].

Since it is known that some ASs operate across physical country borders, this study also seeks to characterize the degree to which ASs not registered to the target country under analysis range within the borders of the target country. To enable this analysis, each prefix announced in the BGP RIB table included in this analysis is geolocated utilizing the MaxMind GeoIP2 database and catalogued to its AS, the AS's registered country, and the country where the prefix ranges. This data will inform assertions and conclusions presented in Chapter 6 about the dependence of individual country's logical network on their Gateway ASs to provide reachability to the larger internet, a key objective of this study.

4.5 Experimental Execution

The algorithm discussed in Section 4.2 is executed against all 21 BGP RIB tables and preceding 8 hours of BGP update messages collected from RIPE NCC (as discussed in Sections 4.1 and 4.3), analyzing each of the 230 countries and territories to which the IANA has made AS assignments via their respective RIRs.

4.6 Related Work

Böttger et al. [16] discuss the internet's evolutionary trend towards less hierarchical networks, specifically within the context of IXP proliferation in recent years. In this work, IXPs are shown to facilitate the bypassing of internet traffic around the Tier-1 ASs that formerly sat atop the internet hierarchy. In doing so, IXPs indeed "flatten" the overall internet topology, yet their work reveals that additional hierarchies remain. The traffic that formerly flowed through Tier-1 ASs is shown to instead aggregate within a relatively small number of internet transit providers. While these new hierarchies are less severe than what existed before the growth of IXPs, their work shows that the internet remains hierarchical in some respects [16].

As introduced in Section 2.6, Dainotti et al. [17] provide a detailed examination of how BGP was exploited to facilitate the Egyptian and Libyan Arab Spring censorship events. Through analysis of BGP update messages sent to withdraw Egyptian and Libyan BGP prefixes from service, they show how these countries' logical internet space effectively disappeared from the larger internet at the direction of the Egyptian and Libyan regimes. While all of the ASs involved in these events were state-owned, this study demonstrated the ability of a state actor to leverage BGP to effect internet censorship of its citizens [17].

The vantage point dependency for BGP RIB table analysis, and the inherent incompleteness therein, is well-studied within the field. Oliveira et al. [29] endeavored to create a "ground truth" view of the logical internet by aggregating data from both proprietary and non-proprietary sources for Tier-1 and Tier-2 networks. Through comparison of their aggregated, ground truth logical network view with that obtained from BGP RIB table analysis, they validate the incomplete view of the internet afforded by BGP RIB table analysis [29]. Gregory et al. [9] advanced this understanding of BGP incompleteness by developing a novel methodology for determining which ASs would provide a *more* complete view of the internet to the research community, were they to act as feeders to the collector projects.

CHAPTER 5:
Initial Results and Observations

This chapter presents the experimental results obtained per the methodology detailed in Chapter 4. Congruent with the objectives and hypotheses proposed in Chapter 3, these results address the existence of Gateway ASs as features of country-level logical network topology, the extent to which Gateway AS visibility is dependent on collector vantage point, customer relationships formed by Gateway ASs between upstream and downstream ASs, hierarchies identified within country-level network topologies, and the utility of the Gateway AS analysis for mapping a country's logical network to its physical borders.

5.1 Existence, Prevalence, and Extent of Gateway ASs within Country-Level Logical Network Topology

This study clearly demonstrates the existence and prevalence of Gateway ASs within country-level logical network topology. Appendix A provides the output of the study's main algorithm for the aggregated collector aspect, and analysis of this data table reveals that 77.8 percent of all countries and territories possess at least one Gateway AS (for countries without a Gateway AS, each of their registered ASs possess BGP connections to External ASs and are therefore fully meshed with the larger internet). One explanation for why some countries' logical network topologies result in the formation of Gateway ASs whereas others do not is that countries with a larger number of ASs produce Gateway AS, an assertion that is supported by the data. The statistical breakdown for the number of visible ASs in the set of countries with Gateway ASs and the set of those without Gateway ASs is presented in Table 5.1, and the single tailed T-test comparison between these two sets produces a t-value of 3.35 and a corresponding p-value of 0.0009 (well below 0.05, the threshold for statistical significance in a T-test). Consequently, a statistical correlation exists to show that countries with a larger number of visible ASs (at least 4 or more) result in the formation of Gateway ASs within their country-level network topology.

Having established this explanation for the formation of Gateway ASs within country-level network topology, Table 5.2 shows the statistical breakdown of Appendix A data for the

Table 5.1. Number of Visible ASs Between Countries With and Without Gateway ASs

	# Visible ASs in Countries Without Gateway ASs	# Visible ASs in Countries with Gateway ASs
mean	2.89	262.09
std	4.45	511.31
min	1.00	2.00
max	30.00	2498.00

Table 5.2. Summary of Gateway AS Statistics

	# Gateway ASs	# Visible Country ASs	# Registered ASs	Gateway Factor
mean	26.15	262.09	364.84	0.53
std	59.31	511.31	653.62	0.23
min	1.00	2.00	4.00	0.05
max	451.00	2498.00	2500.00	0.95

subset of countries that possess Gateway ASs. This data summary shows that for countries with Gateway ASs, on average 53 percent of their total ASs fall behind a Gateway AS with substantial variability in this figure country to country. This finding presents substantial follow-on questions and implications for country-level network topology, vulnerabilities in country-level networks, etc. and is discussed further in Section 6.1.

The significance of Gateway ASs within country-level logical network topology is further demonstrated through visualization of target country logical network topologies. Figure 5.1 shows the country-level logical network topology chord view for Bolivia. This figure shows not only how Gateway ASs aggregate traffic inbound and outbound from the target country (thereby producing a hierarchical country-level logical network topology), but the figure also reveals that some Gateway ASs are more "important" than others by virtue of the number of BGP connections they establish with External ASs and Internal ASs. This feature is common to all countries that possess Gateway ASs, and is further explored in Section 5.4.

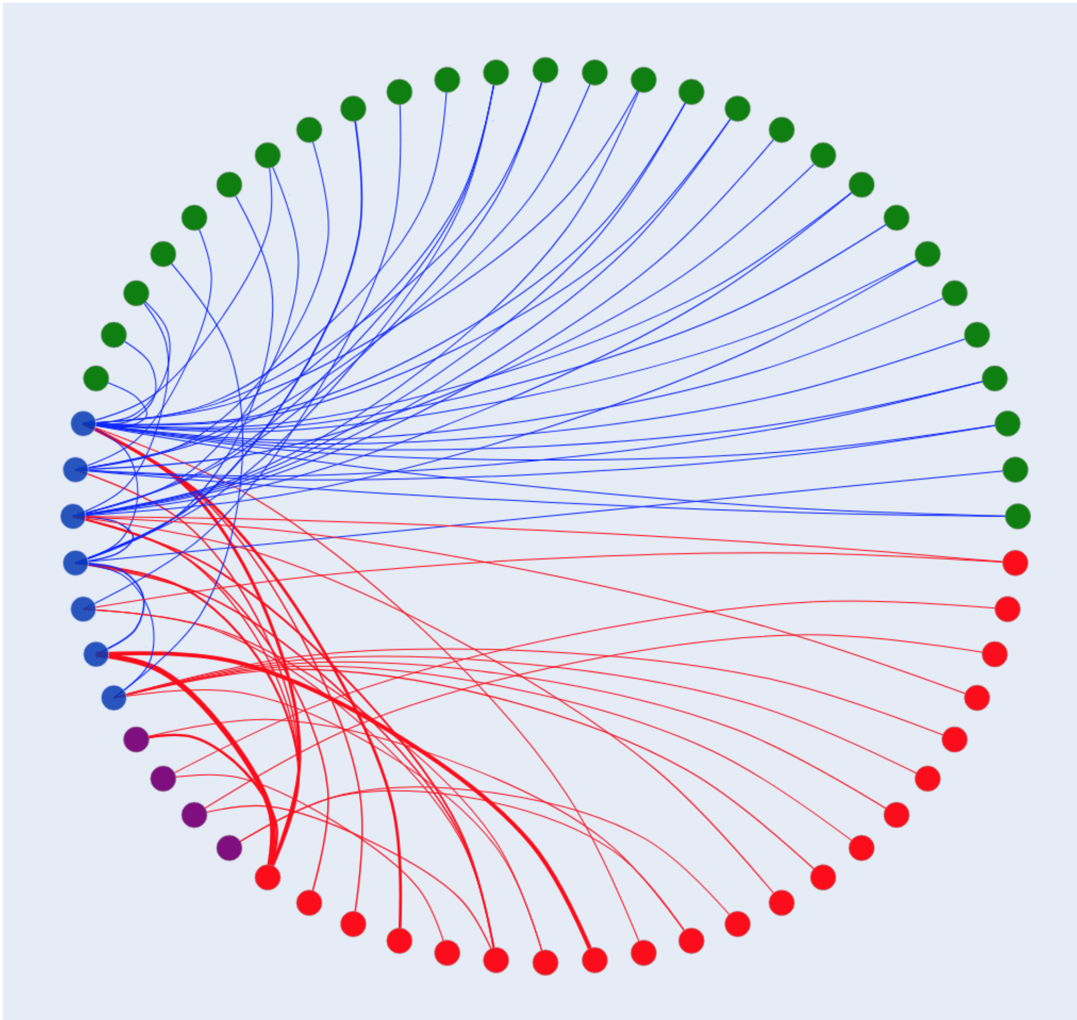


Figure 5.1. Country-Level Logical Network Topology: Bolivia

Figure 5.2 shows the country-level logical network topology chord view for Iran, a larger and more complex network than Bolivia's. In addition to reinforcing the characteristics discussed thus far in this section, inspection of this topology graph reveals multiple layers of traffic aggregation within the country's network topology. That is, a relatively small number of Gateway ASs connect the larger internet to a relatively small number of Internal ASs, which themselves serve to provide connectivity to a much larger population of Internal ASs. This characteristic amplifies the importance of Gateway ASs that support such country-internal choke points, since all traffic distributed by these internal choke points to the myriad Internal ASs must necessarily travel through an upstream Gateway AS. Inspection of these two graphs also reveals that some countries possess greater “depth” in their logical network topology hierarchy, with Internal ASs having to traverse more Internal ASs before reaching a Gateway AS and the larger internet. This characteristic is discussed further in Section 5.6.

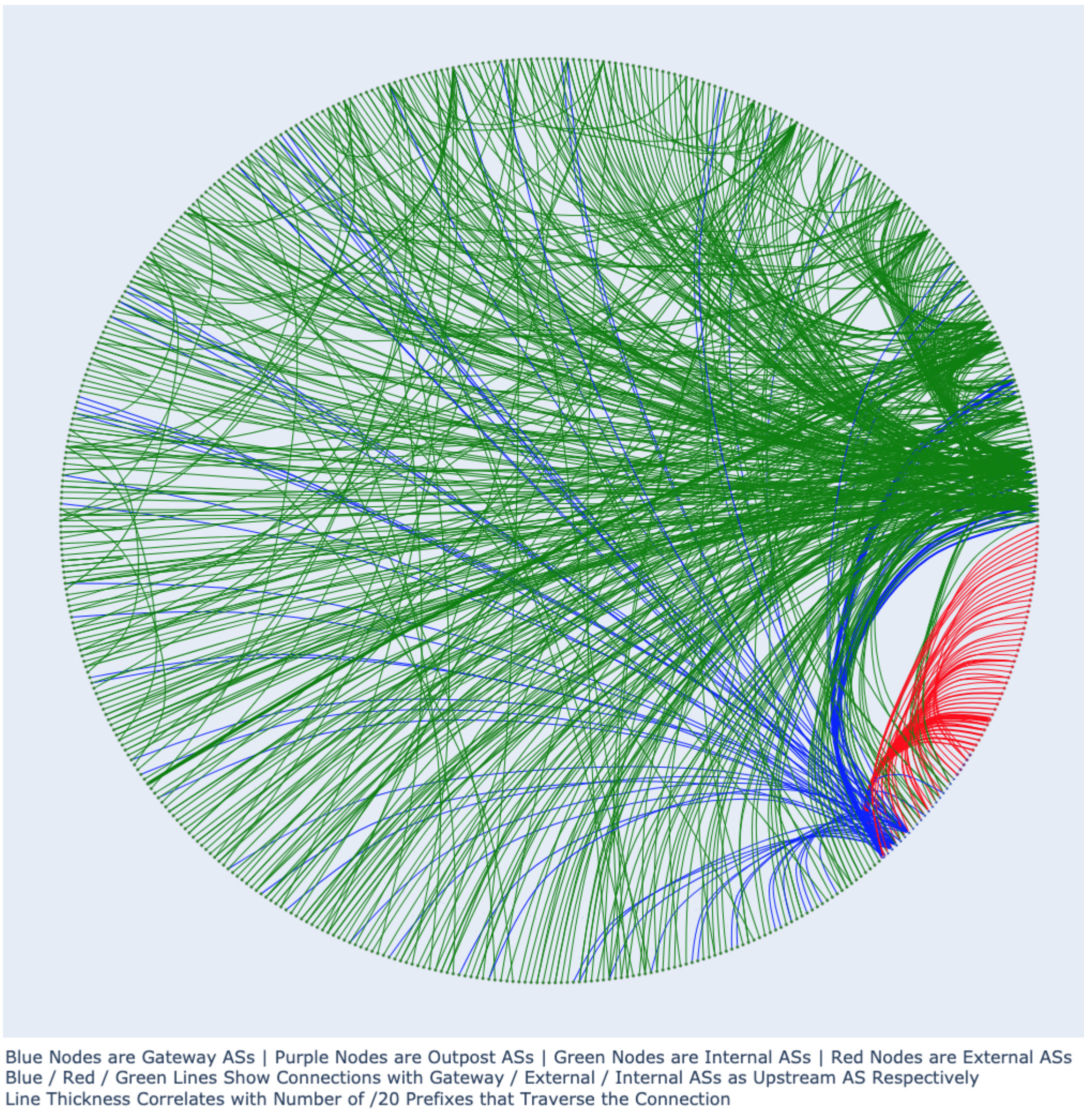


Figure 5.2. Country-Level Logical Network Topology: Iran

5.2 On BGP RIB Collector Vantage Point and Gateway AS Visibility

Analysis of the individual BGP RIB collector algorithm output data clearly demonstrates that countries present different logical network topologies to different collector vantage points. Figure 5.3 is constructed from the individual BGP RIB collector algorithm output data to show the fraction of a given country's total Gateway ASs that are detected from the indicated BGP RIB collector vantage point (with a random selection of all countries analyzed by this study included in the plot). Inspection of this figure reveals the following two features of BGP RIB collectors as it pertains to the identification of Gateway ASs: No BGP RIB collector is able to view all Gateway ASs for all countries, and some countries' Gateway ASs are not completely visible from any single BGP RIB collector. These findings reveal a collector vantage point dependency with regards to the structure of a given country's logical network topology, as well as substantive differences in BGP RIB collector efficacy to fully identify a given country's Gateway ASs.

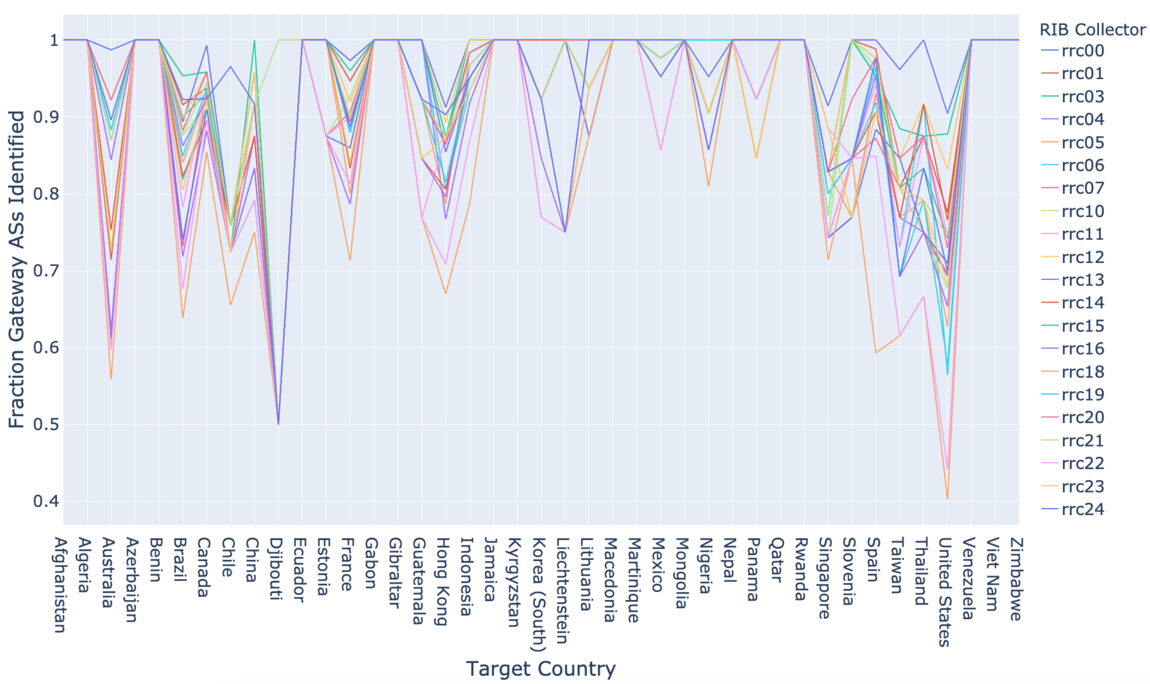


Figure 5.3. BGP RIB Collector Vantage Point Efficacy for Gateway AS Visibility

The collector vantage point dependency for Gateway AS visibility is further explored via Figure 5.4, which presents the proportion of all country Gateway ASs that are completely visible in BGP data by the number of BGP RIB collector locations involved in the analysis. (Note that this figure is produced from a “greedy set cover” analysis of the individual BGP RIB collector algorithm output data, specifically the listing of Gateway ASs visible from each collector. Consequently, this figure does not represent a traditional Cumulative Distribution Function (CDF), as the constituent collectors which contribute to complete view-ability of Gateway ASs are not interchangeable with other collectors.) As shown, 67 percent of all countries’ Gateway ASs are completely view-able from one RIPE NCC collector location (specifically rcc00 at Amsterdam), with the following seven collector locations required to obtain a complete view of all countries’ Gateway ASs:

1. rrc00 at Amsterdam, Netherlands
2. rrc03 at Amsterdam, Netherlands
3. rrc04 at Geneva, Switzerland
4. rrc14 at Palo Alto, USA
5. rrc15 at Sao Paulo, Brazil
6. rrc18 at Barcelona, Spain
7. rrc19 at Johannesburg, South Africa

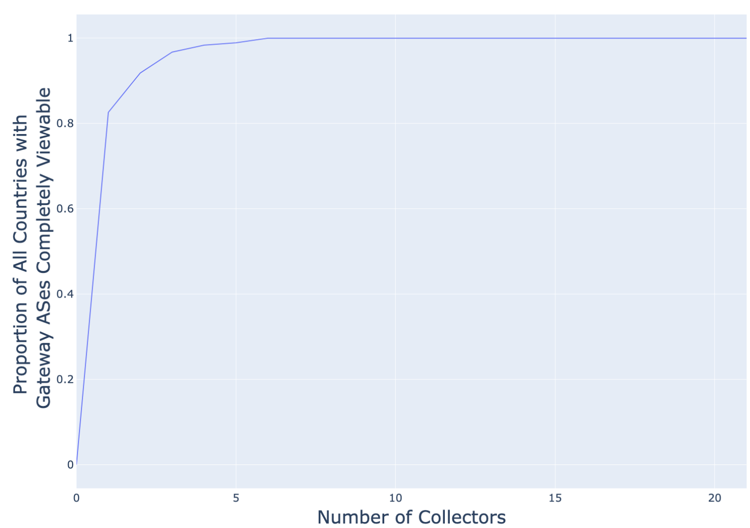


Figure 5.4. Minimum Number of Collectors Required for Full Gateway AS Visibility

Because Gateway ASs represent choke points into and out of their country's logical network, they are foundational to the country's network topology itself. Therefore, the consequence of this collector vantage point dependency for Gateway AS visibility is that countries present different logical network topologies based upon where in the logical internet they are being viewed from, and at best any one single RIPE NCC collector is only able to completely view the Gateway ASs for two thirds of all countries and territories. These findings are further discussed in Chapter 6.

As with Gateway AS visibility, each country's Gateway Factor is similarly vantage point-dependent. Recall from Section 3.1 that a country's Gateway Factor describes how meshed the given country's logical network is with the outside internet and therefore describes an important aspect of network topology itself. Within this context, Figure 5.5 visualizes the Gateway Factor collector vantage point dependency for the same set of countries presented in Figure 5.3. Inspection of this figure reveals that as with the identification of Gateway ASs themselves, no single BGP RIB collector possesses a comprehensive view of logical network topology for all countries, and also shows substantial variation in the network topologies presented to various BGP RIB collectors by different countries.

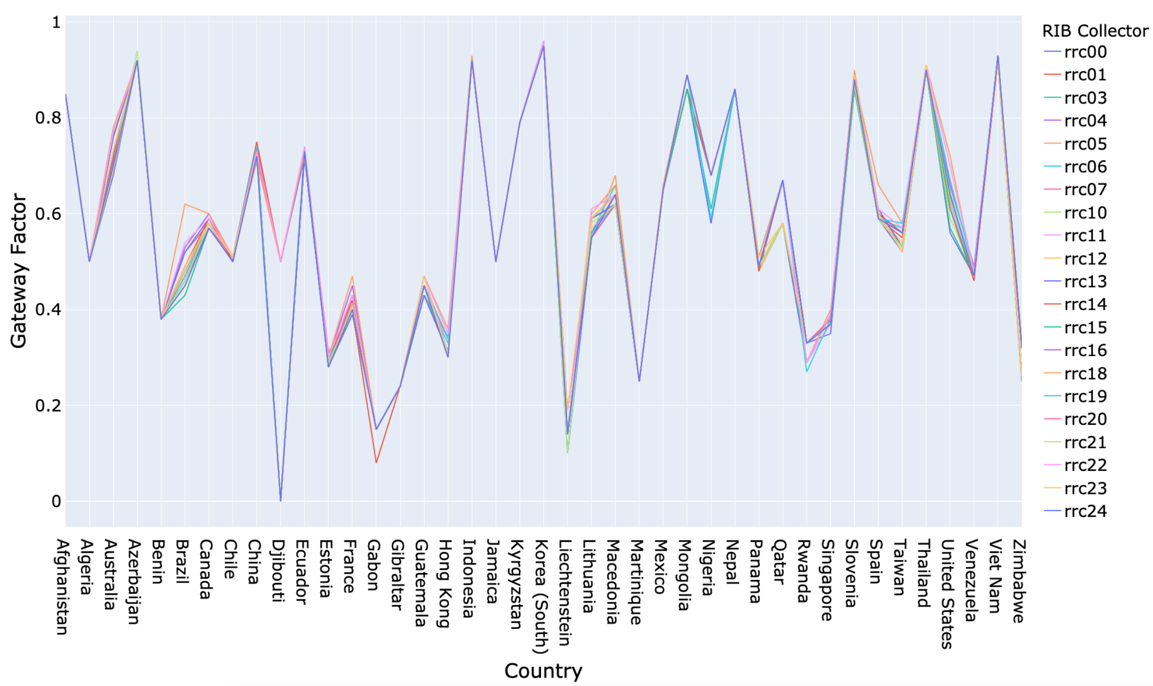


Figure 5.5. Gateway Factor Collector Vantage Point Dependency

5.3 Distribution of Gateway Factor Amongst Countries

Based upon the BGP-enabled censorship events detailed in Section 2.6 and the observation that country ASs operate within the legal and governmental jurisdiction of their registered country, this study hypothesized that countries sharing a common government type would possess similar Gateway Factor values (i.e., democratic countries would possess a lower Gateway Factor while authoritarian countries would possess a higher Gateway Factor). To explore this question, Figure 5.6 presents CDFs of Gateway Factors for countries grouped by government type, using the Economist's Democracy Index to classify each country [30].

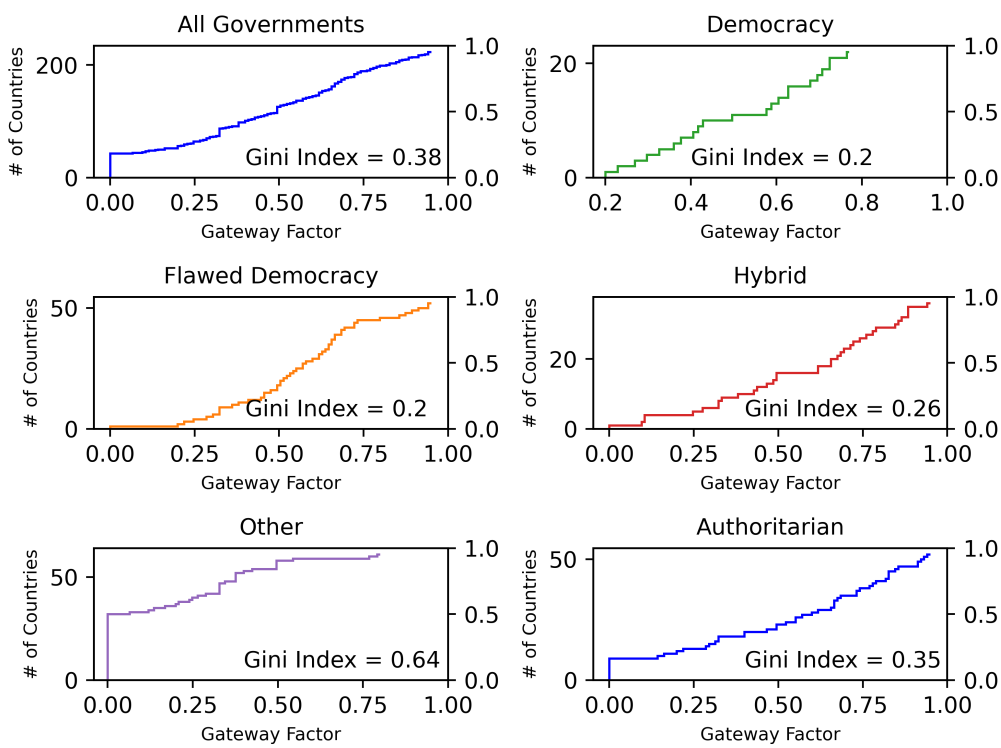


Figure 5.6. CDF for Gateway Factor Distribution by Government Type

From inspection of this figure, it is apparent that this hypothesis is not fully supported by the data. The Gini Index (a measure of inequality within a frequency distribution, with 0 representing perfect equality and 1 representing perfect inequality in the data) reveals near parity between the different government types with respect to the similarity of their Gateway Factor distributions. Governments classified as 'other' (or unclassified) are the outliers, but this discrepancy is explained by the large number of these countries that do not possess any Gateway ASs and therefore reflect a Gateway Factor of 0. The figure does show, however, that democratic country Gateway Factors skew towards lower values (and therefore greater meshedness with the outside internet).

Similarly, the distribution of Gateway Factors across IANA regions is presented in Figure 5.7. As with the previous analysis, there are no clear groupings of Gateway Factor by IANA region, but this analysis does show that ARIN countries trend towards greater network meshedness with the larger internet. With the maximum observed ARIN country Gateway Factor of approximately 0.65 this region stands in contrast to the others, which all include countries with Gateway Factor maximums above 0.8. Consequently, while these findings reveal some limited correlations and similarities in Gateway Factor by these groupings, the data does not bear out substantive conclusions from this analysis.

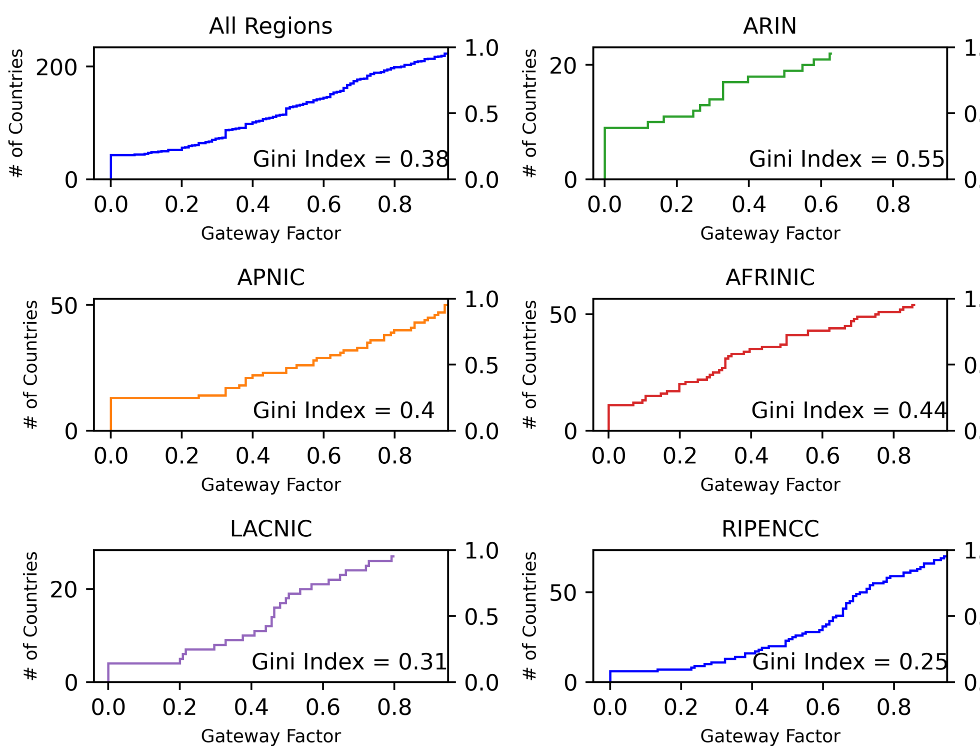


Figure 5.7. CDF for Gateway Factor Distribution by Region

5.4 Gateway AS Node Cardinality

Figures 5.1 and 5.2 revealed that some Gateway ASs have a greater degree of BGP connectivity (or cardinality) than others. Because this subset of country Gateway ASs makes more internal and external connections than other Gateway ASs, they are disproportionately important Gateway ASs with regards to the country's overall logical network topology and traffic flow. In order to quantify this finding, the cardinality of each Gateway AS was utilized to construct Figure 5.8 (where line color is used to differentiate between the countries analyzed by this study). This plot represents the cumulative sum of all BGP connections made by a given country's Gateway ASs, with the list of Gateway ASs sorted from those with the highest cardinality to those with the least. From inspection of this figure, it is evident that a minority of Gateway ASs dominate their country's overall Gateway AS cardinality, with just 20 percent of Gateway ASs contributing between 29 and 85 percent of the total BGP connections made by Gateway ASs within the target country.

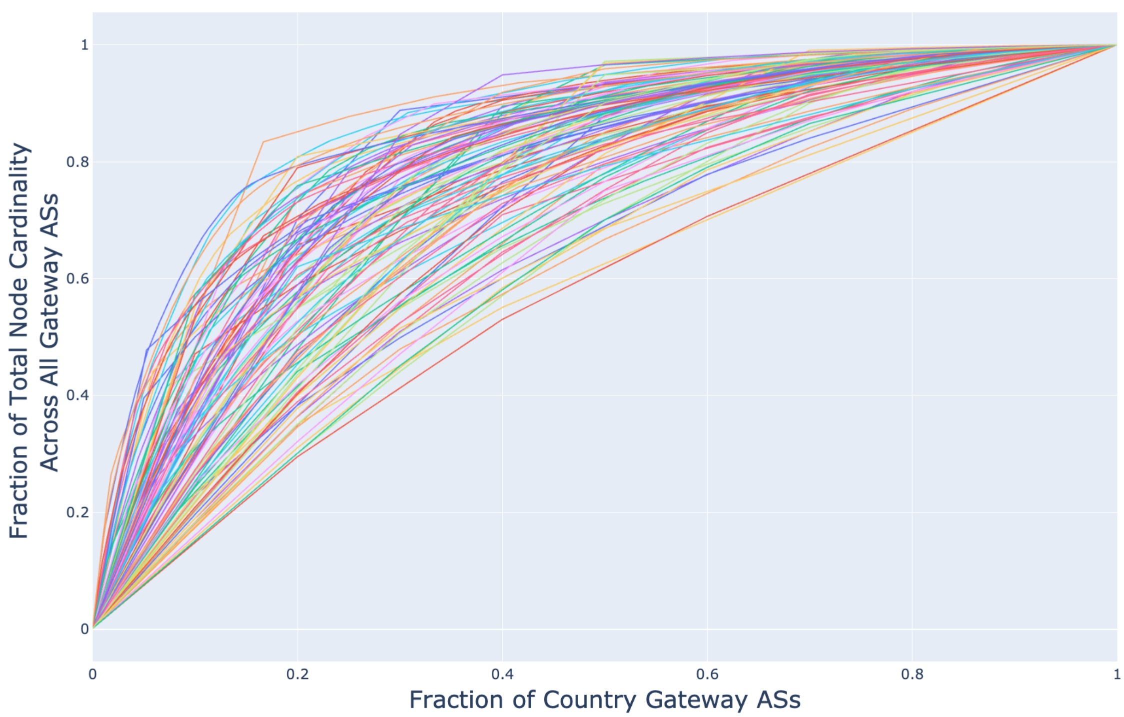


Figure 5.8. Ranked Cumulative Sum Gateway AS Cardinality By Country

This finding reveals that a relatively small subset of Gateway ASs have a disproportionate impact upon network traffic aggregation and suggests that an additional type of hierarchy exists within Gateway ASs themselves. In order to capture and measure this aspect of country-level network hierarchy, the "Cardinality Factor" is defined per Equation 5.1. (Note that the fraction of a country's Gateway ASs that form 66 percent of the total BGP connections is obtained by sorting the Gateway ASs by highest contributing to lowest contributing.) By this definition, a higher Cardinality Factor describes a country where a small subset of country Gateway ASs make the majority (66 percent) of all BGP connections into and out of the country. Table 5.3 lists the 10 countries with the highest Cardinality Factors, and this metric will be utilized for further analysis in Section 6.1.

$$\begin{matrix} Cardinality \\ Factor \end{matrix} = 1 - \begin{matrix} Proportion\ Gateway\ ASs \\ Making\ 66\%\ BGP\ Connections \end{matrix} \tag{5.1}$$

Table 5.3. Top Ten Most Hierarchical Countries by Cardinality Factor

Country	Cardinality Factor	Country	Cardinality Factor
Poland	0.90	Mozambique	0.86
Australia	0.89	Hong Kong	0.85
United States	0.89	Romania	0.85
Palestinian Territory	0.88	Russian Federation	0.85
Macedonia	0.86	India	0.85

Next, Figure 5.9 aggregates the country-level cardinality data discussed thus far to show a single plot for the fraction of total connections made by all Gateway ASs in all countries, and further describes the internal and external cardinality components that give rise to the whole. By this analysis, the top 20 percent of the Gateway ASs by cardinality value on the internet are shown to possess approximately 70 percent of all the BGP connections made by Gateway ASs, and Gateway ASs as a whole possess more BGP connections to external ASs than they do internal ASs.

Finally, a disparity between node cardinalities of Gateway ASs, Internal ASs, and External ASs themselves is shown to exist. As shown in Table 5.4, on average Gateway ASs possess 22.8 times more BGP connections than Internal ASs, and 1.05 times more BGP connections relationships than External ASs. Consequently, in addition to the characteristic that a minority of a country's Gateway ASs makeup the majority of BGP connections made by all

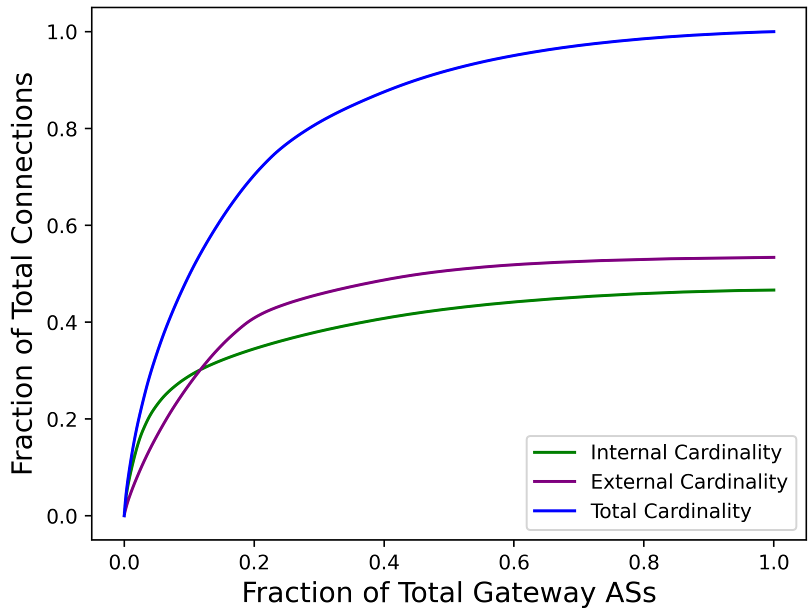


Figure 5.9. Overall Ranked Cumulative Sum Gateway AS Cardinality

Gateway ASs, Gateway ASs are shown to possess substantially more BGP connections in general than either Internal or External ASs. These findings again reinforce the hierarchical nature of country-level logical network topology, as well as the traffic aggregating effect that Gateway ASs have on inbound and outbound internet traffic.

Table 5.4. Average AS Cardinality by Type

	Ave Gateway AS Node Degree	Ave Internal AS Node Degree	Ave External AS Node Degree
mean	36.48	1.60	34.89
std	35.45	0.45	27.94
min	2.00	1.00	5.05
max	173.85	3.08	190.00

5.5 Customer Relationships Between External, Gateway, and Internal ASs

Analysis of country-level network topology customer relationships reveals substantial differences between the types of customer relationships formed between the different categories of ASs proposed by this study. As shown in Table 5.5, by overall percentage customer relationships between Gateway ASs and External ASs are nearly evenly split between P2P and P2C relationships, but the "directionality" of the P2C relationships formed is important and bears further investigation. By an overwhelming proportion, External ASs form P2C connections with Gateway ASs, which in turn form P2C connections with Internal ASs (with Gateway and Internal ASs as the customers respectively). Similarly, External ASs form P2C connections with Outpost ASs by overwhelming proportion. The consequence of these P2C customer relationships is that Internal ASs are almost entirely dependent upon their Gateway ASs for connectivity to the larger internet not just by logical connection, but also by customer relationship. Gateway and Outpost ASs are similarly reliant on External ASs by this measure. This data provides insight into the "directionality" of the dependencies these ASs form with one another, and further demonstrates how Gateway ASs function as choke points within their country's logical network topology.

Table 5.5. Summary of Gateway AS, External AS, Outpost AS, and Internal AS Customer Relationships

	% Gateway to External P2C	% External to Gateway P2C	% Gateway to External P2P	% Outpost to External P2C	% External to Outpost P2C	% Outpost to External P2P
mean	1.69	44.81	53.49	0.28	75.65	24.07
std	5.57	37.95	38.06	1.37	31.51	31.41
min	0.00	0.00	0.00	0.00	8.33	0.00
max	50.00	100.00	100.00	14.29	100.00	91.67

	% Gateway to Internal P2C	% Internal to Gateway P2C	% Gateway to Internal P2P	% Country Internal P2C	% Country Internal P2P
mean	97.43	0.57	2.00	98.05	1.95
std	6.43	2.10	6.06	7.68	7.68
min	39.57	0.00	0.00	20.00	0.00
max	100.00	20.00	60.43	100.00	80.00

Figure 5.10 further leverages customer relationship data to provide insight into the size of Gateway ASs as compared to their counterparts (recall from Section 2.4 that AS customer cone size is a metric of AS size). This figure is constructed from the cumulative sum of total ASs in the specified AS type's customer cones, with the ASs themselves sorted by size (largest to smallest), and reveals a dramatic difference in AS size by type. As a group, Gateway ASs are shown to be vastly larger than either Outpost ASs or Internal ASs. Because cone size is directly related with the "reachability" of a given AS (see Section 2.4) this finding again supports the importance of Gateway ASs within their country network topology as critical enablers for network traffic flow.

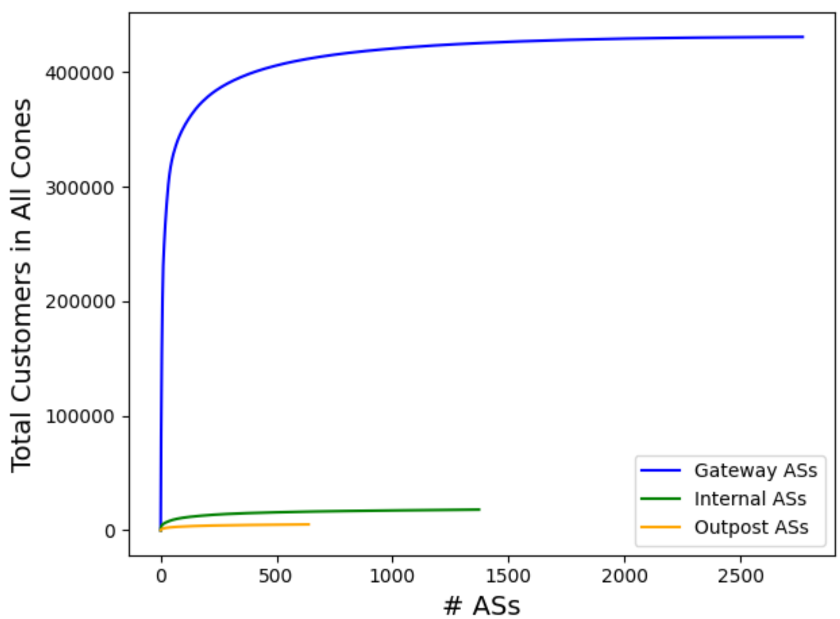


Figure 5.10. Ranked Cumulative Sum Total Customers in Customer Cones by AS Type

5.6 Internal AS Depth Behind Gateway ASs

Recall from Figure 5.2 that Iran's network topology is shown to possess multiple layers of choke points. By inspection of this figure, it is evident that some Internal ASs themselves serve to connect other Internal ASs to Gateway ASs, and so on. This aspect of network topology is described by this study as the "Network Depth" and has bearing on the overall hierarchy of a country's network topology. (i.e., A larger Network Depth describes a more hierarchical network structure with some Internal ASs having to traverse one or more other Internal ASs before reaching a Gateway AS and the larger internet.) Within this context, Figure 5.11 provides a CDF for Iran's Internal AS Network Depth. By inspection of this figure, Iran is shown to possess a very hierarchical network, with 60 percent of its Internal ASs having to traverse three or more Internal ASs before reaching a Gateway AS.

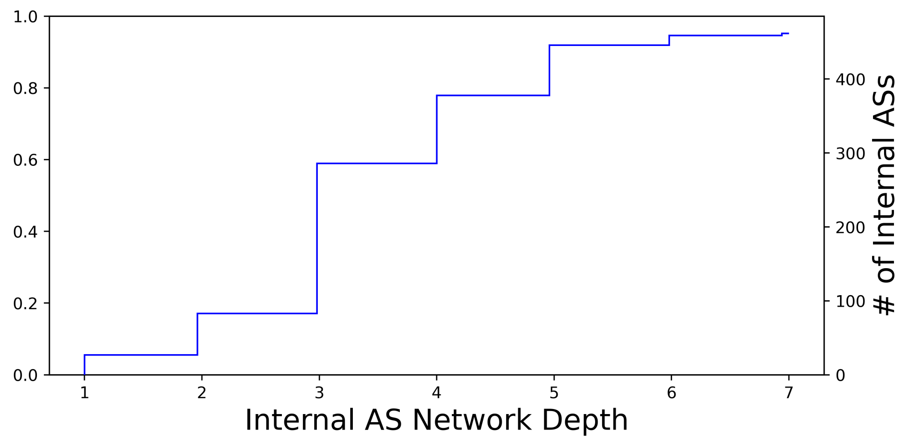


Figure 5.11. CDF of Internal AS Depth within Iran's Network Topology

The Network Depth of Iran's network topology represents an extreme on the internet as a whole, and on average country network topologies are shown to be relatively flat. Figure 5.12 presents a CDF for the overall Network Depth across all countries. Inspection of this figure reveals that over 60 percent of all Internal ASs on the internet directly connect to a Gateway AS with only 10 percent having to traverse two or more ASs to reach a Gateway AS. Analyzed on a country-level (vice whole internet), Figure 5.13 shows a CDF for the *average* Network Depth per country, revealing that 90 percent of countries possess an average Network Depth of 2 or less. These findings have bearing on each country's overall hierarchy and topological structure and will be discussed further in Chapter 6.

5.7 On the Nature of Outpost ASs

The existence of Outpost ASs is peculiar since it is unexpected that these ASs would not establish any BGP connections to any Internal ASs, given their logical proximity to one another. One theory is that Outpost ASs serve as content providers (data repositories of sorts)

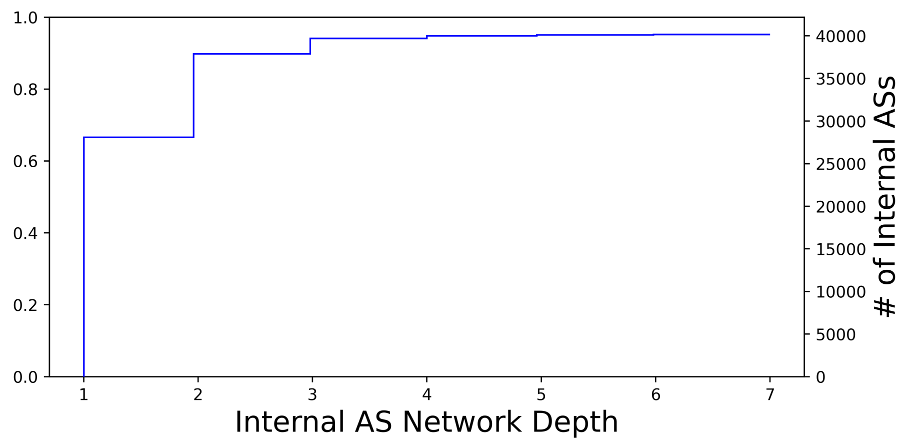


Figure 5.12. CDF of Overall Network Depth Across All Countries

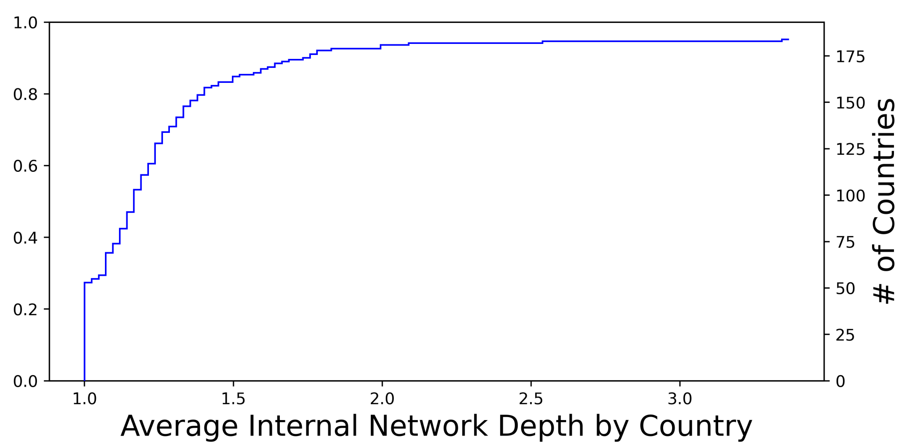


Figure 5.13. CDF of Average Network Depth Per Country

intended to provide service to country external ASs exclusively. Yet this explanation is not supported by the data. Tables 5.6 and 5.7 show the statistical breakdown of AS classification for Gateway and Outpost ASs using the CAIDA classification database described in Section 4.1. These tables show that only 3 percent of Outpost ASs classify as Content ASs and that Outpost ASs and Gateway ASs share a similar classification distribution, with the majority of each group being Transit ASs.

An alternate explanation is for the existence of Outpost ASs is that Outpost ASs are in fact Gateway ASs, but the BGP connections to Internal ASs that would identify them as such are not viewable from the perspective of the BGP collectors. Recall from Section 4.3 that

Table 5.6. Statistical Breakdown of Outpost AS Classification Across Whole Internet

	% Transit	% Content	% Enterprise	% Not in Database
mean	76.40	5.06	9.04	9.50
std	28.81	10.33	13.60	26.66

Table 5.7. Statistical Breakdown of Gateway AS Classification Across Whole Internet

	% Transit	% Content	% Enterprise	% Not in Database
mean	72.69	3.53	2.52	21.26
std	40.50	9.68	12.03	40.61

the network views afforded by BGP collectors are inherently skewed and incomplete, with many connections from smaller ASs not captured in their RIB tables. Further recall that Outpost ASs are shown to be very small within the context of customer cone size per Figure 5.10. These factors, in addition to the demographic similarity between Outpost and Gateway ASs supports the assertion that Outpost ASs are in fact small Gateway ASs with hidden BGP connections to Internal ASs. Yet another explanation for the existence of Outpost ASs is that it may be cheaper for these ASs to purchase access to the larger internet directly from an External AS (via a P2C customer relationship), vice from an Internal AS or Gateway AS. While limitations on BGP collectors and RIB table accuracy discussed in Section 4.3 limit the conclusive resolution of this issue, this study suggests avenue for follow-on work to explore these hypothesis in Section 7.1.

5.8 Geolocation of Announced BGP Prefixes

One of the research questions posed in Section 3.2 endeavored to determine if aspects of country-level network topology map directly to physical geography based upon logical classification alone. Towards that end, Figure 5.14 details the proportion of all prefixes on the internet which geolocate to the same country that their parent ASs are registered to, revealing that nearly 91 percent of all prefixes range within their registered country's logical and physical borders. Of the remaining approximately 8 percent of all prefixes that geolocate internationally, the figure further shows the percentage of these prefixes that originate from Gateway ASs, Outpost ASs, and Internal ASs. By inspection of this figure it is apparent that while Internal ASs are most likely to range entirely within their country of registration, this is not a universal property of Internal ASs or country-level logical network topology. Consequently, these results do not support this aspect of the study's hypothesis, instead reinforcing the complexity and difficulty of mapping the internet to the physical world.

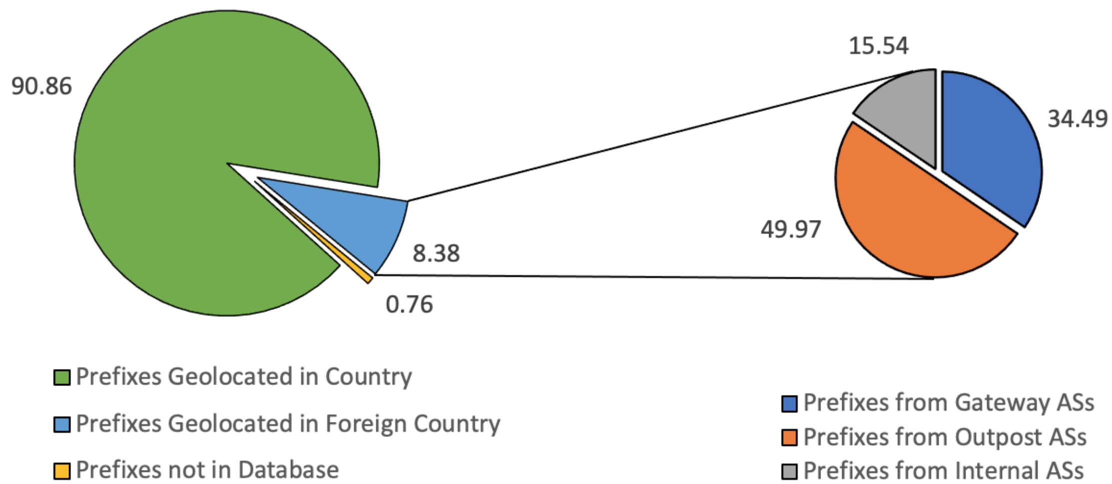


Figure 5.14. Prefix Geolocation to Registered Country

The geolocation analysis discussed thus far explored characteristics of all Gateway ASs, Outpost ASs, and Internal ASs on the internet as a whole, and the following analysis seeks to explore the characteristics of these ASs on a country level. Figure 5.15 shows a CDF for the percentage of a country's total advertised prefixes that range internationally (in this section, "internationally ranging" refers to prefixes or ASs that do not range within their country of registration). As shown, the majority of countries possess ASs with international range, though the percentage of the country's prefixes that range internationally represents small a minority of their total. Figure 5.16 presents a CDF which describes the percentage of all internationally ranging prefixes that correspond to each AS type by country (note that this figure does not include the 45 countries without internationally ranging ASs). This figure shows that for the countries with internationally ranging ASs, Internal ASs contribute a smaller number of internationally ranging prefixes as compared to Gateway ASs or Outpost

ASs (consistent with the whole internet analysis presented at the beginning of this section). Therefore, while this study did not demonstrate that Internal ASs range exclusively within the borders of their registered country, it does show that Internal ASs contribute the least to prefixes that range outside their registered country's physical borders both by number of prefixes advertised and IP address space. These findings have implications for internet infrastructure geolocation and are discussed further in Section 6.2.

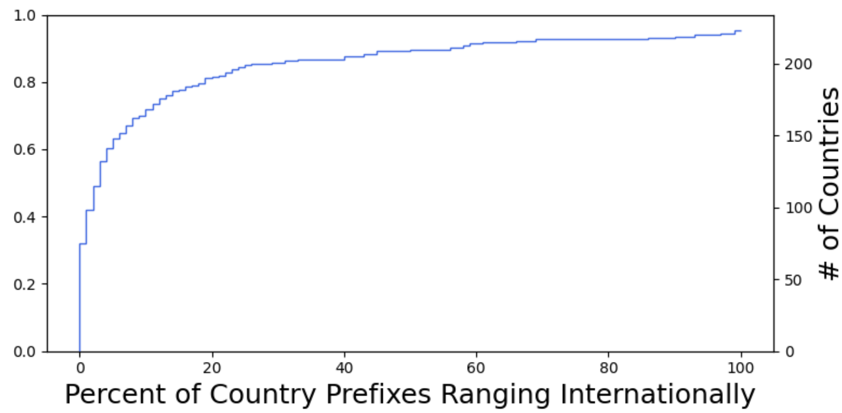


Figure 5.15. Percent of Country Prefixes Ranging Internationally

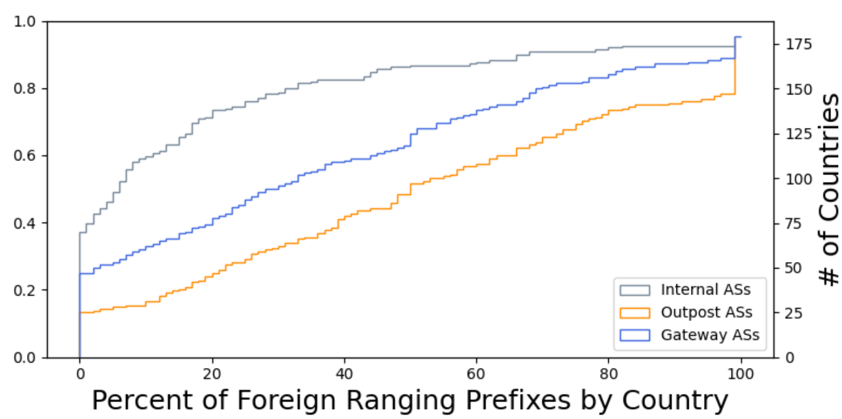


Figure 5.16. Internationally Ranging Prefixes by AS Type

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 6:
Findings and Analysis

This chapter discusses implications of the results presented in Chapter 5 for the community's understanding of country-level network topology, the mapping of logical networks and supporting infrastructure to the physical world, and identifies countries at greatest risk for state-sponsored internet surveillance or network disruption.

6.1 Country-Level Logical Network Topology

Through the identification of Gateway ASs as logical choke points for internet traffic into and out of a given country, this study has identified and proven the existence of BGP hierarchies within country-level logical networks. Further, despite the evolutionary trend towards a flattened internet landscape with regards to AS inter-connectivity, country-level logical networks are becoming *more* hierarchical over time as shown in Figure 6.1. This figure is constructed from analysis of BGP RIB tables from 2005, 2010, 2015, and 2020 by the methodology detailed in Section 4.2, and shows an increase both in the fraction of countries that possess Gateway ASs (and therefore possess Internal ASs which have no BGP connections to External ASs) as well as the average Gateway Factor for these countries on the internet as a whole. Consequently, by this study's analysis, country-level network topologies are becoming more hierarchical over time.

Returning to the current state of the internet, vice the historical analysis discussed thus far in this section, the Network Depth measurement introduced in Section 5.6 provides insight into the extent of these country network hierarchies, their stratification, and the inter-dependence of Internal ASs to reach the outside internet. By analyzing AS organization and topology in this way, this study presents a contribution to the field for the understanding of evolving internet topology and where hierarchies remain with the network of networks. As discussed in Section 4.3, however, missing P2P connections within collector BGP RIB tables affects the completeness of the results presented in this section. Further, with BGP RIB tables from smaller ASs logically "closer" to a given target country, it is impossible to quantify or qualify the full extent of these missing P2P connections. Yet this analysis does reveal

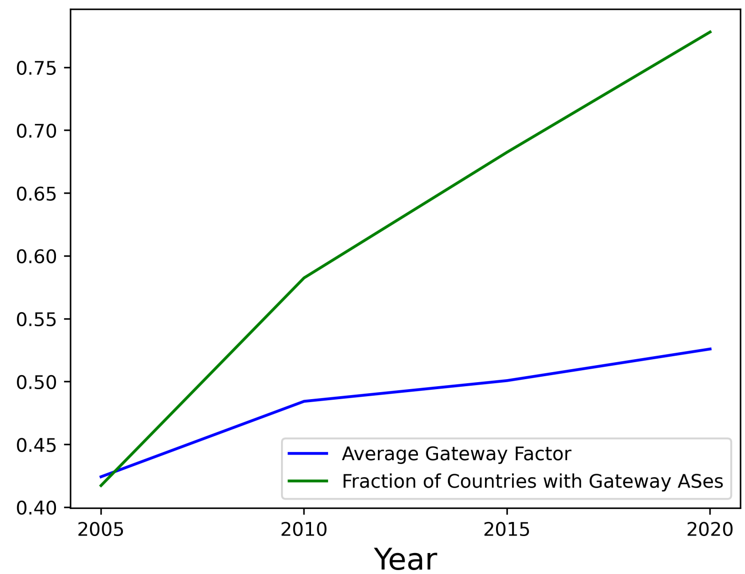


Figure 6.1. Country Level Network Hierarchy Over Time

country-level network topology as viewed from the major feeder ASs which contribute to the collector projects. Consequently, the topologies and hierarchies discussed in this section represent the country network topologies as viewed from all but the small ASs in logical proximity to the country in question.

Beyond the traditional conception of topological hierarchy discussed thus far, this study additionally reveals hierarchies within hierarchies where a small minority of Gateway ASs are shown to provide the majority of a given country's BGP network connections to the larger internet. Such Gateway ASs functionally take on the role of IXPs on a country scale, aggregating inbound and outbound internet traffic for exchange between Internal ASs and External AS destinations. In doing so, they become disproportionally important to the country's overall network topology as compared to other Gateway ASs. Therefore, the Cardinality Factor introduced in Section 5.4 is a critical measurement to quantify how disproportionately important these Gateway ASs are within their country's network, and therefore represents an additional dimension to a country's overall network hierarchy.

With the objective of providing a holistic measure of a country's network topology (as opposed to simply mapping AS connections), these three dimensions of hierarchy (Gateway Factor, Network Depth, and Cardinality Factor) are combined per Equation 6.1. A ranked listing of countries by Hierarchy Rating is subsequently generated and provided in Appendix B, with the top 10 most hierarchical countries reproduced per Table 6.1. Inspection of this table shows substantial country-to-country differences between the constituent factors producing the overall Hierarchy Rating. (e.g., China and Indonesia are ranked 6 and 7 respectively for overall network hierarchy, yet their Gateway Factor and Average Network Depths differ substantially from one another.) Consequently, the Hierarchy Rating is shown to provide a holistic measure of *overall* network hierarchy, as opposed to any single constituent factor.

$$\begin{matrix} Country \\ Hierarchy\ Rating \end{matrix} = \begin{matrix} Gateway \\ Factor \end{matrix} * \begin{matrix} Cardinality \\ Factor \end{matrix} * \begin{matrix} Average\ Internal \\ AS\ Depth \end{matrix} \tag{6.1}$$

Table 6.1. Top Ten Most Hierarchical Country Network Topologies

Country	Gateway Factor	Cardinality Factor	Average Network Depth	Hierarchy Rating
Iran	0.95	0.500000	3.365801	1.598755
Bangladesh	0.94	0.727273	2.008178	1.372863
India	0.95	0.846154	1.634605	1.313971
Thailand	0.90	0.727273	1.785146	1.168459
Uzbekistan	0.94	0.660000	1.803922	1.119153
Iraq	0.78	0.555556	2.545455	1.103030
Armenia	0.87	0.666667	1.847222	1.071389
China	0.75	0.812500	1.743326	1.062340
Indonesia	0.92	0.793103	1.410108	1.028893
Argentina	0.80	0.844444	1.407955	0.951152

Similar to the analysis of Section 5.3, the distribution of Country Hierarchy Ratings are grouped and plotted by government type and IANA region per Figures 6.2 and 6.3. Consistent with the simple Gateway Factor analysis presented in Section 5.3, each government type presents a roughly linear distribution of hierarchies, but by this measure Democratic, Flawed Democratic, and "Other" countries are shown to possess overall less hierarchical networks as compared to their Hybrid and Authoritarian counterparts. Similarly, the LACNIC, ARIN, and AFRINIC IANA regions possess less hierarchical networks than those within APNIC and RIPENCC.

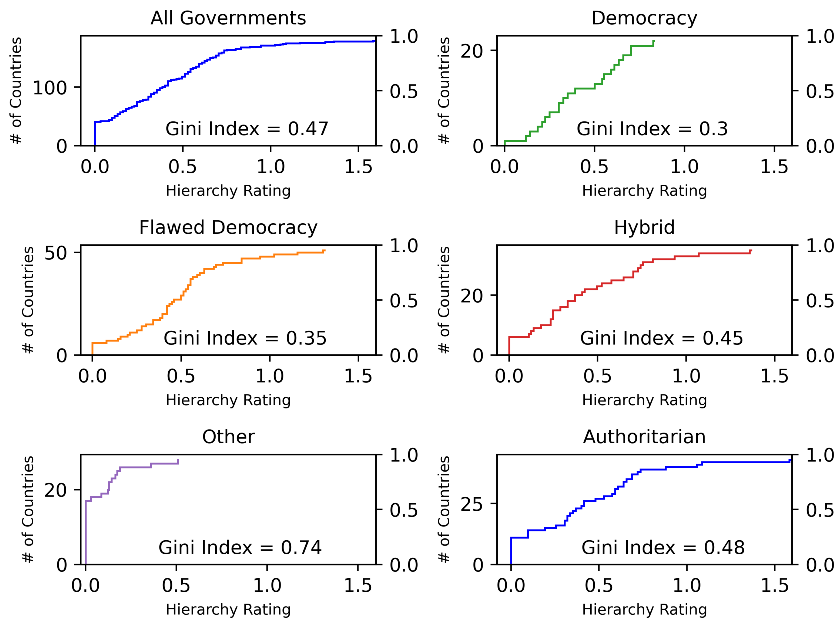


Figure 6.2. CDF for Country Hierarchy Rating by Government Type

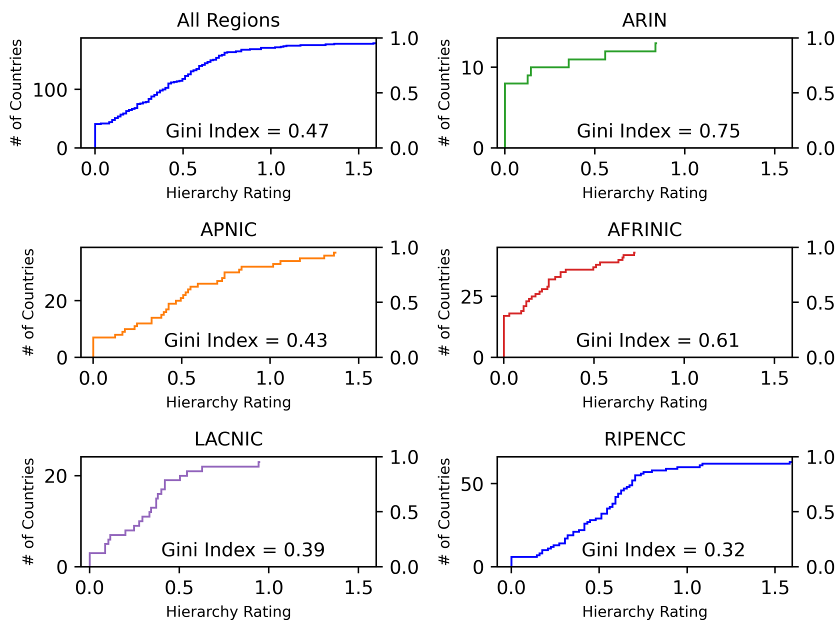


Figure 6.3. CDF for Country Hierarchy Rating by Region

Despite these observations, this study does not assert correlation between government type or IANA region and Hierarchy Rating. The constituent CDFs and (corresponding Gini Factors) do not display the typical “long tailed” distributions present within meaningfully-correlated data sets. However, this study does prove the existence of country-level network hierarchies and suggests the Hierarchy Rating as a means to quantify the extent of these hierarchies as opposed to any one of the constituent factors.

6.2 Mapping of Country-Level Logical Internet to Country-Level Physical Borders

This study also runs into limitations due to the difficulty of geolocating internet infrastructure, specifically at the country level. While *nearly* all Internal AS prefixes geolocate to their registered country, this is not a uniform characteristic of Internal ASs for all countries. Consequently, there is no definitive correlation between AS logical classification as an Internal AS and range exclusivity to the country of registration (as postulated in Section 3.2). As such, the task of mapping any given country’s *overall* logical network topology to physical borders remains a difficult, tedious task.

With regards to the task of mapping *all* of the internet infrastructure and AS range within a given country (not just the country’s own logical network), the task is substantially complicated by the presence of foreign ASs with range inside the country’s physical borders. Figures 6.4 and 6.5 seek to explicate the extent of such international AS operation and its subsequent effect upon network geolocation. Figure 6.4 presents a CDF for the number of foreign countries that have a logical and physical presence within another country’s physical borders. This figure shows that while 50 percent of all countries have a relatively small number of foreign countries with range inside their borders (six or fewer), there are some countries with an extremely significant foreign presence within their geographic boundaries. These tend to be large, well-connected countries like Germany, the United States, and Brazil. Similarly, Figure 6.5 presents a CDF for the percentage of a given country’s total domestic IP address space (both from domestic and international ASs) that is accounted for by foreign ASs. Here, 50 percent of all countries are shown to have 3 percent or less of their overall domestic IP address space provided by foreign countries. In contrast to the previous figure, where outliers were large and well connected ASs, the countries with a very large percentage of its overall IP address space provided by foreign countries are very small, poorly-connected countries like East Timor, Turkmenistan, or territories without indigenous AS assignment of any kind (as with Antarctica).

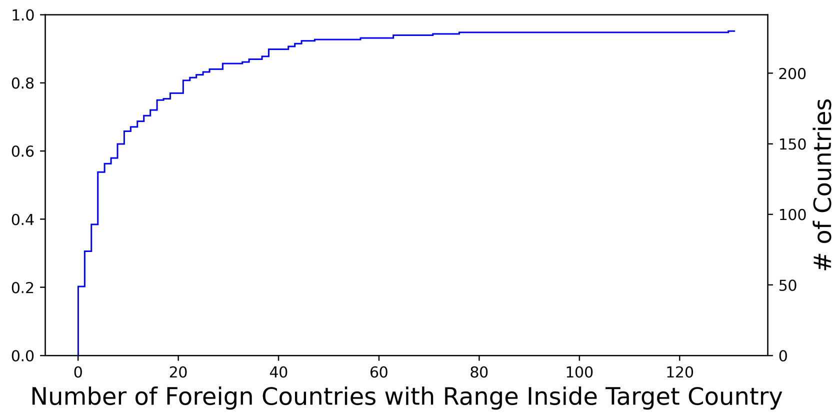


Figure 6.4. CDF for Number of Countries that Range within Another

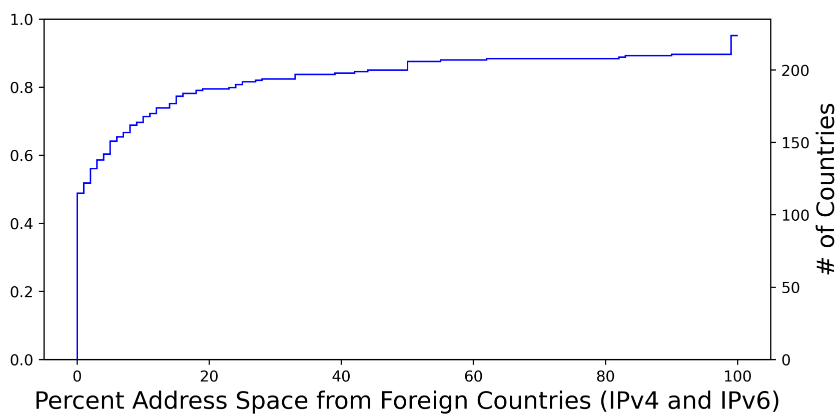


Figure 6.5. CDF for Percentage of Overall Domestic Address Space from External ASs

Given the prevalence of AS ranging across country borders as detailed in this section, the task of thoroughly mapping the logical internet space to the physical world remains a significant challenge to the research community. This characteristic of inter-country AS range has implications beyond the geolocation of internet infrastructure, however, as is discussed in the next section pertaining to internet censorship.

6.3 On Internet Censorship and Vulnerable Countries

This study was motivated by a recognition for the strategic value of logical choke points in internet topology, the vulnerabilities such choke points pose to internet routing, and their potential utility for censorship or surveillance. What differentiates this study from others is the observation that logical choke points within a given country's network could be exploited by the government under which they operate for nefarious ends. This is not to say that this is the inevitable result, but merely to recognize that all business and entities operating within the territorial limits of a country are fundamentally susceptible to government influence or outright control. Consequently, this study's contribution to the discipline is an assessment of countries whose network topologies would be most conducive to state-sponsored control, censorship, and/or surveillance of its own populace.

Within that context, as a purely logistical matter countries with fewer Gateway ASs (and thus a higher Gateway Factor) would find their networks more "controllable," since there are fewer ASs that actually channel traffic into and out of the country in question. These are the ASs where BGP-enabled censorship techniques would need to be implemented in order to achieve country-wide network disconnection from the internet and where less severe censorship or surveillance techniques would have the most impact. Consequently, this study proposes the Gateway Factor as a metric to assess country vulnerability to BGP-enabled exploitation of its Gateway ASs. Yet the question remains for how to differentiate vulnerable countries from the rest within this ranking. In truth, such a determination is not possible since the ranking proposed in this section stands as a *spectrum* of risk, vice a discrete delineation between countries "at risk" and those "not at risk." Therefore, in order to make an assessment of those countries *most* at risk for censorship, this study focuses on the set of countries known to have implemented BGP-enabled censorship or surveillance of its populace (see Section 2.6). Amongst this grouping of countries (China, Egypt, Iran, etc.), China possesses the lowest Gateway Factor at 0.75. Taking this figure as the lower bound for vulnerable countries, a listing of countries most vulnerable to BGP enabled censorship is provided in Table 6.2.

Given the findings of Section 6.2 however, it is clear that there is an upper bound on the actual effectiveness of BGP-enabled state sponsored censorship schemes, and the internet has evolved to be uncontrollable by any single country actor. For instance, consider the

RuNet purported to isolate Russian internet space from the larger internet. Assuming that Russian authorities were 100 percent successful in exploiting BGP and Russian Gateway ASs to sever its logical network from the global internet, there would still remain a large number of international ASs with range inside of Russia's territorial boundaries. In fact, by this study's analysis, 56 separate countries possess ASs with range inside of Russian geographic borders representing 6.08 percent of the overall IP address space geolocated to Russian soil. Note that this study does not seek to characterize the nature or use of this foreign IP address space (commercial, private, research, etc.) and therefore does not make assertions regarding the potential availability of such foreign IP address space to Russian citizens during a state sponsored censorship event. Instead, the existence of foreign AS range within Russian borders demonstrates the complexity of the modern internet and the challenge of fully controlling the networks operating within Russian geographic boundaries.

In order to capture this upper bound for censorship effectiveness in the list of vulnerable countries proposed thus far, the Censorship Vulnerability Rating is proposed per Equation 6.2 (where the "Fraction of Native IP Space" corresponds to the fraction of the total IP address space ranging within the given country's borders that is provided by domestic ASs). By this measure, the listing of most vulnerable countries presented in Table 6.2 is ranked from most vulnerable to least.

$$\begin{matrix} Country\ Vulnerability \\ Rating \end{matrix} = \begin{matrix} Gateway \\ Factor \end{matrix} \times \begin{matrix} Fraction\ Native \\ IP\ Space \end{matrix} \quad (6.2)$$

Table 6.2. Countries at Risk for BGP-Enabled, State-Sponsored Censorship

Target Country	Number Gateway ASs	Number Foreign Countries Ranging in Target	Gateway Factor	Fraction Domestic IP Space	Censorship Vulnerability Rating
Iran	4	17	0.95	0.9935	0.9438
Korea (South)	13	22	0.95	0.9740	0.9253
Viet Nam	10	9	0.92	0.9987	0.9188
Bangladesh	23	8	0.94	0.9760	0.9174
Azerbaijan	2	5	0.92	0.9965	0.9168
Indonesia	61	16	0.92	0.9905	0.9113
Belarus	3	8	0.92	0.9814	0.9029
Thailand	24	21	0.9	0.9913	0.8922
Uzbekistan	1	5	0.94	0.9453	0.8886
Pakistan	4	15	0.89	0.9979	0.8881
Turkey	18	27	0.89	0.9899	0.8810
Georgia	6	10	0.89	0.9814	0.8734
Armenia	6	7	0.87	0.9911	0.8623
Slovenia	13	16	0.87	0.9901	0.8614
Egypt	6	8	0.86	0.9991	0.8592
Nepal	13	5	0.86	0.9967	0.8572
Mongolia	4	4	0.86	0.9882	0.8499
Equatorial Guinea	1	0	0.83	1.0000	0.8300
Lebanon	3	10	0.85	0.9720	0.8262
Kazakhstan	7	12	0.83	0.9945	0.8254
India	11	43	0.95	0.8633	0.8201
Saudi Arabia	6	15	0.83	0.9879	0.8200
Morocco	3	10	0.82	0.9991	0.8193
Libya	2	2	0.82	0.9966	0.8172
Afghanistan	5	5	0.85	0.9513	0.8086
Micronesia	1	0	0.8	1.0000	0.8000
Kyrgyzstan	6	6	0.79	0.9916	0.7834
Iraq	11	12	0.78	0.9895	0.7718
Argentina	50	24	0.8	0.9620	0.7696
Kuwait	9	9	0.77	0.9967	0.7675
New Caledonia	2	3	0.77	0.9956	0.7666
Maldives	3	3	0.77	0.9932	0.7648
Myanmar	10	5	0.77	0.9865	0.7596
China	23	34	0.75	0.9959	0.7469

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 7: Conclusions

Through the implementation of a novel approach to BGP RIB table analysis, this study proved the Section 3.2 hypothesis for the existence of logical choke points into country-level logical networks. The study further explicated the importance of these choke points for network topology, the geolocation of network infrastructure, and individual country vulnerability to network censorship. The main contributions of this work are as follows: First, a comprehensive, internet-wide characterization of Gateway ASs as choke points into country-level logical network topology is developed and submitted to the community for further work in this area. Second, a novel metric for comprehensive network hierarchy is developed to identify and characterize the extent of country-level network hierarchies. Further, this study shows that country-level network topologies are becoming more *hierarchical* over time, despite the internet's evolutionary trend towards a more "flattened" topological landscape overall. Third, the unique AS classifications proposed by this study are shown to directly map vast portions of the logical internet landscape to physical country borders. And finally, these findings culminate in the identification of countries most vulnerable to BGP-enabled internet censorship due to their topological structures and dominance of domestic IP address space by native ASs. In aggregate, these findings advance the community's knowledge of internet topology with potential for follow-on work as proposed in the next section.

7.1 Opportunities for Follow-On Research

As detailed in Sections 4.1 and 4.2, this study analyzed real-world BGP data upon which internet routing is accomplished. Consequently, the results and conclusions presented in Chapters 5 and 6 should similarly reflect real-world truths. However, this study did not endeavor to provide external validation for these observations to independently substantiate the study's findings. Future works could provide this external validation through the use of internet measurement techniques, ascertainment of an individual country's network topology via government and/or industry contact, or some other means.

The extent of AS range across country borders identified in Section 6.2 raises further questions about the effectiveness of past and current internet censorship efforts. While Egypt was highly successful in severing its own logical network from the larger internet, previous work on this event has not addressed the extent to which foreign ASs with range inside of Egyptian country borders maintained uninterrupted connectivity to the internet. Similarly, the effectiveness of Chinese, Iranian, etc. efforts for ongoing censorship and surveillance is not known in this regard. Dedicated research in this area would help in understanding the true effectiveness of state-sponsored internet censorship.

Through the Gateway AS analysis, this study endeavored to profile country-level network topologies as a whole, vice detailed exploration of individual Gateway ASs themselves. Future works could investigate the most significant Gateway ASs within each country's network topology to explore their dominance of network traffic routing into and out of the country, how dependent Internal ASs are on those specific Gateway ASs, and any business or country-level regulatory factors that might have contributed to their development. For instance, a useful metric in this regard would be what proportion of a country's Internal ASs rely exclusively upon an individual Gateway AS to reach External ASs.

The geolocation of network infrastructure remains an active area of research. One particularly significant aspect of this work would be to determine a method to infer which logical links cross country borders and geolocate the supporting network infrastructure on either side of the border in question. Such an inference tool would enable physical and logical targeting of critical, country-specific network infrastructure with obvious strategic applications in the military and academic domains. The Gateway AS analysis serves as a starting point and road map for further work in this area by identifying the logical points of entry into a country's internet space. Informed by this study's analysis of country-level topology and which small subset of overall ASs serve as Gateway or Outpost ASs, such an inference tool would then be equipped to determine physical link border crossings associated with those ASs only. This approach would reduce the scale of the problem and enable specific targeting of logical network infrastructure to accomplish the objective.

Finally, as detailed in Section 4.1 this study analyzed BGP RIB data from the RIPE NCC collector project. Future work in this area could apply this study's methodology to the other collector projects (RouteViews, Packet Clearing House, etc.) and investigate any topological differences between those analyses and the results presented here. Specifically, given that these other collector projects maintain collectors with different vantage points than RIPE NCC, analysis of this data may reveal some of the hidden P2P relationships discussed in Section 4.3. Insight from this analysis would be particularly useful in exploring the true nature of Outpost ASs, as proposed in Section 5.7.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A:
Algorithm Output - Aggregate View From All Collectors

This appendix details complete country-level network topology data produced by aggregating the main algorithm output from each of the individual RIB tables considered by this study (see Section 4.2). Consequently, the country-level network topologies characterized below are as if viewed from all collector locations simultaneously.

Target Country	# Gateway ASs	# Outpost ASs	# Visible Country ASs	Gateway Factor
Afghanistan	5	3	54	0.85
Aland Islands	0	2	2	0.00
Albania	13	10	73	0.68
Algeria	2	2	8	0.50
American Samoa	0	2	2	0.00
Andorra	0	1	1	0.00
Angola	3	14	44	0.61
Anguilla	0	3	3	0.00
Antigua and Barbuda	2	5	8	0.12
Antarctica	0	0	0	NaN
Argentina	50	140	947	0.80
Armenia	6	4	78	0.87
Australia	77	436	1605	0.68
Austria	63	166	553	0.59
Aruba	0	2	2	0.00
Azerbaijan	2	2	50	0.92
Bosnia and Herzegovina	9	4	38	0.66
Barbados	2	2	6	0.33
Bangladesh	23	30	886	0.94
Belgium	29	161	277	0.31
Burkina Faso	4	1	20	0.75
Bulgaria	55	121	642	0.73

Continued on next page

Target Country	# Gateway ASs	# Outpost ASs	# Visible Country ASs	Gateway Factor
Bahrain	6	2	18	0.56
Burundi	2	2	9	0.56
Benin	4	4	13	0.38
Bermuda	3	6	15	0.40
Brunei Darussalam	3	3	9	0.33
Bolivia	7	4	39	0.72
Brazil	451	995	2498	0.42
Bahamas	3	2	7	0.29
Bhutan	3	1	15	0.73
Botswana	1	8	18	0.50
Belarus	3	6	107	0.92
Belize	3	19	28	0.21
Cambodia	17	24	89	0.54
Cameroon	2	3	16	0.69
Canada	144	447	1372	0.57
Cape Verde	1	0	3	0.67
Central African Republic	0	2	2	0.00
Chad	3	7	12	0.17
Chile	29	113	280	0.49
China	24	126	599	0.75
Colombia	20	93	193	0.41
Congo	3	4	9	0.22
Cote D'Ivoire (Ivory Coast)	3	5	12	0.33
Cook Islands	0	1	1	0.00
Costa Rica	15	49	80	0.20
Croatia (Hrvatska)	19	25	133	0.67
Cuba	1	0	3	0.67
Cyprus	10	45	100	0.45
Czech Republic	70	242	620	0.50
Democratic Republic of the Congo	4	17	35	0.40

Continued on next page

Target Country	# Gateway ASs	# Outpost ASs	# Visible Country ASs	Gateway Factor
Djibouti	2	0	2	0.00
Denmark	23	142	303	0.46
Dominica	0	3	3	0.00
Dominican Republic	11	23	56	0.39
Ecuador	12	24	131	0.73
Egypt	6	3	64	0.86
El Salvador	9	19	36	0.22
East Timor	0	0	0	NaN
Estonia	8	83	127	0.28
Equatorial Guinea	1	0	6	0.83
Eritrea	0	1	1	0.00
Ethiopia	0	1	1	0.00
Finland	17	64	275	0.71
Fiji	3	3	9	0.33
Falkland Islands (Malvinas)	0	1	1	0.00
Federated States of Micronesia	1	0	5	0.80
Faroe Islands	2	1	4	0.25
France	150	649	1303	0.39
French Guiana	0	3	3	0.00
French Polynesia	1	1	4	0.50
Gabon	2	10	13	0.08
Gambia	0	9	9	0.00
Germany	267	973	2093	0.41
Ghana	12	21	71	0.54
Gibraltar	4	12	21	0.24
Great Britain (UK)	231	1117	1958	0.31
Grenada	1	3	6	0.33
Georgia	6	5	99	0.89
Greece	17	40	166	0.66
Greenland	0	1	1	0.00
Guinea	1	3	8	0.50
Guadeloupe	0	4	4	0.00

Continued on next page

Target Country	# Gateway ASs	# Outpost ASs	# Visible Country ASs	Gateway Factor
Guatemala	13	14	48	0.44
Guam	4	0	7	0.43
Guinea-Bissau	0	2	2	0.00
Guyana	1	3	4	0.00
Hong Kong	103	317	590	0.29
Honduras	9	34	81	0.47
Haiti	1	2	8	0.62
Hungary	20	74	219	0.57
Indonesia	61	54	1426	0.92
Ireland	20	133	200	0.23
Israel	15	58	259	0.72
India	13	94	1956	0.95
British Indian Ocean Territory	0	1	1	0.00
Iraq	11	18	130	0.78
Iran	4	20	471	0.95
Italy	121	363	986	0.51
Jamaica	1	3	8	0.50
Jordan	5	9	35	0.60
Japan	89	256	662	0.48
Kenya	13	62	101	0.26
Kyrgyzstan	6	4	47	0.79
Kiribati	0	3	3	0.00
Comoros	0	2	2	0.00
Saint Kitts and Nevis	1	1	3	0.33
Korea (North)	0	1	1	0.00
Korea (South)	13	25	766	0.95
Kuwait	9	7	66	0.76
Cayman Islands	2	6	11	0.27
Kazakhstan	8	15	126	0.82
Laos	6	6	20	0.40
Lebanon	3	18	141	0.85
Saint Lucia	0	4	4	0.00
Liechtenstein	4	16	21	0.05
Sri Lanka	4	5	19	0.53

Continued on next page

Target Country	# Gateway ASs	# Outpost ASs	# Visible Country ASs	Gateway Factor
Liberia	1	7	9	0.11
Lesotho	1	3	6	0.33
Lithuania	16	49	145	0.55
Luxembourg	15	56	91	0.22
Latvia	26	53	231	0.66
Libya	2	0	11	0.82
Macedonia	8	10	47	0.62
Macao	2	4	9	0.33
Madagascar	1	2	4	0.25
Malaysia	30	58	204	0.57
Mali	3	1	7	0.43
Malawi	4	6	15	0.33
Mauritania	0	3	3	0.00
Marshall Islands	0	1	1	0.00
Martinique	1	2	4	0.25
Mauritius	9	10	27	0.30
Mayotte	0	1	1	0.00
Malta	4	16	32	0.38
Mexico	42	78	339	0.65
Morocco	3	1	17	0.76
Monaco	1	1	2	0.00
Moldova	14	20	157	0.78
Mongolia	4	2	44	0.86
Myanmar	10	10	84	0.76
Northern Mariana Islands	0	0	0	NaN
Maldives	3	0	13	0.77
Mozambique	7	7	21	0.33
New Caledonia	2	1	13	0.77
Niger	1	4	5	0.00
Norfolk Island	0	1	1	0.00
Nigeria	21	48	165	0.58
Nicaragua	6	11	23	0.26
Netherlands	133	454	1013	0.42
Norway	29	87	295	0.61
Nepal	13	1	102	0.86

Continued on next page

Target Country	# Gateway ASs	# Outpost ASs	# Visible Country ASs	Gateway Factor
Nauru	0	2	2	0.00
New Zealand (Aotearoa)	38	94	419	0.68
Oman	3	1	16	0.75
Panama	13	35	93	0.48
Peru	12	24	67	0.46
Papua New Guinea	3	4	16	0.56
Philippines	26	77	301	0.66
Pakistan	4	13	160	0.89
Poland	184	664	2131	0.60
Saint Pierre and Miquelon	0	1	1	0.00
Puerto Rico	8	25	71	0.54
Palestinian Territory	9	8	45	0.62
Portugal	14	24	97	0.61
Palau	0	3	3	0.00
Paraguay	9	31	79	0.49
Qatar	3	2	12	0.58
Reunion	1	3	5	0.20
Romania	64	365	1025	0.58
Russian Federation	286	688	2497	0.61
Rwanda	4	7	15	0.27
Saudi Arabia	7	18	144	0.83
Samoa	1	2	6	0.50
Saint Helena	0	0	0	NaN
Saint Vincent and the Grenadines	0	2	2	0.00
San Marino	2	3	8	0.38
Sao Tome and Principe	0	2	2	0.00
Senegal	2	7	10	0.10
Seychelles	3	25	30	0.07
Sierra Leone	1	5	17	0.65
Singapore	35	198	360	0.35

Continued on next page

Target Country	# Gateway ASs	# Outpost ASs	# Visible Country ASs	Gateway Factor
Slovakia	25	41	180	0.63
Slovenia	13	23	256	0.86
Solomon Islands	1	4	8	0.38
Somalia	6	5	17	0.35
South Africa	57	313	453	0.18
Spain	86	293	915	0.59
Sudan	2	3	7	0.29
Suriname	1	1	3	0.33
Sweden	47	201	624	0.60
Switzerland	73	397	700	0.33
Syria	1	0	2	0.50
USSR (former)	0	1	1	0.00
Swaziland	1	3	6	0.33
Taiwan	26	64	186	0.52
Tanzania	12	22	62	0.45
Tajikistan	2	3	19	0.74
Thailand	24	16	398	0.90
Timor-Leste	4	2	8	0.25
Togo	1	3	5	0.20
Tokelau	0	2	2	0.00
Tonga	1	3	4	0.00
Trinidad and Tobago	5	1	14	0.57
Tunisia	4	2	20	0.70
Turkey	18	42	526	0.89
Turkmenistan	2	0	6	0.67
Turks and Caicos Islands	0	2	2	0.00
Tuvalu	0	1	1	0.00
Ukraine	225	431	1820	0.64
Uganda	10	18	41	0.32
United Arab Emirates	6	30	78	0.54
United States	303	843	2498	0.54
United States Minor Outlying Islands	0	1	1	0.00

Continued on next page

Target Country	# Gateway ASs	# Outpost ASs	# Visible Country ASs	Gateway Factor
Uruguay	6	14	32	0.38
Uzbekistan	1	2	52	0.94
Vanuatu	1	5	9	0.33
Vatican City State	0	4	4	0.00
Venezuela	14	37	96	0.47
Virgin Islands (British)	0	30	30	0.00
Virgin Islands (U.S.)	1	4	6	0.17
Viet Nam	10	14	291	0.92
Wallis and Futuna	0	1	1	0.00
Yemen	0	3	3	0.00
Zambia	3	14	19	0.11
Zimbabwe	4	11	20	0.25

APPENDIX B: Country Hierarchy Ranking

This appendix provides a ranked listing of countries by Hierarchy Rating, as detailed in Section 6.1.

Target Country	Gateway Factor	Cardinality Factor	Network Depth	Hierarchy Rating
Iran	0.95	0.500000	3.365801	1.598755
Bangladesh	0.94	0.727273	2.008178	1.372863
India	0.95	0.846154	1.634605	1.313971
Thailand	0.90	0.727273	1.785146	1.168459
Uzbekistan	0.94	0.660000	1.803922	1.119153
Iraq	0.78	0.555556	2.545455	1.103030
Armenia	0.87	0.666667	1.847222	1.071389
China	0.75	0.812500	1.743326	1.062340
Indonesia	0.92	0.793103	1.410108	1.028893
Argentina	0.80	0.844444	1.407955	0.951152
Georgia	0.89	0.833333	1.274725	0.945421
Saudi Arabia	0.83	0.666667	1.598485	0.884495
Korea (South)	0.95	0.750000	1.189081	0.847220
United States	0.54	0.883621	1.771620	0.845338
Australia	0.68	0.884058	1.388850	0.834920
Turkey	0.89	0.764706	1.196850	0.814562
Federated States of Micronesia	0.80	0.660000	1.500000	0.792000
Azerbaijan	0.92	0.660000	1.250000	0.759000
Lebanon	0.85	0.666667	1.338346	0.758396
Nepal	0.86	0.666667	1.307692	0.749744
Viet Nam	0.92	0.600000	1.353357	0.747053
Bulgaria	0.73	0.791667	1.287625	0.744140
Pakistan	0.89	0.500000	1.657895	0.737763
Libya	0.82	0.500000	1.777778	0.728889
Moldova	0.78	0.666667	1.363636	0.709091
Ukraine	0.64	0.835749	1.323270	0.707790
New Zealand (Aotearoa)	0.68	0.714286	1.455764	0.707085

Continued on next page

Target Country	Gateway Factor	Cardinality Factor	Network Depth	Hierarchy Rating
Sweden	0.60	0.780488	1.506352	0.705414
Kazakhstan	0.82	0.625000	1.367521	0.700855
Israel	0.72	0.833333	1.165939	0.699563
Russian Federation	0.61	0.849558	1.349854	0.699535
Slovenia	0.86	0.692308	1.152893	0.686414
Egypt	0.86	0.666667	1.166667	0.668889
Finland	0.71	0.733333	1.276151	0.664449
Equatorial Guinea	0.83	0.660000	1.200000	0.657360
Burkina Faso	0.75	0.750000	1.166667	0.656250
Belarus	0.92	0.333333	2.110000	0.647067
Angola	0.61	0.666667	1.575000	0.640500
Spain	0.59	0.800000	1.349476	0.636953
Ecuador	0.73	0.636364	1.364407	0.633829
Poland	0.60	0.892857	1.182692	0.633585
Kuwait	0.76	0.571429	1.433333	0.622476
Portugal	0.61	0.750000	1.337500	0.611906
Hungary	0.57	0.833333	1.278689	0.607377
Afghanistan	0.85	0.600000	1.176471	0.600000
Norway	0.61	0.785714	1.243028	0.595766
Latvia	0.66	0.750000	1.198020	0.593020
United Arab Emirates	0.54	0.666667	1.647059	0.592941
Palestinian Territory	0.62	0.875000	1.083333	0.587708
Macedonia	0.62	0.857143	1.105263	0.587368
Croatia (Hrvatska)	0.67	0.705882	1.216216	0.575199
Philippines	0.66	0.782609	1.084337	0.560084
Canada	0.57	0.832000	1.177204	0.558277
Romania	0.58	0.851852	1.129736	0.558173
Slovakia	0.63	0.695652	1.266234	0.554941
Samoa	0.50	0.660000	1.666667	0.550000
Austria	0.59	0.698113	1.329897	0.547767
Malaysia	0.57	0.807692	1.182353	0.544337
Mexico	0.65	0.771429	1.082474	0.542784
Tunisia	0.70	0.500000	1.533333	0.536667

Continued on next page

Target Country	Gateway Factor	Cardinality Factor	Network Depth	Hierarchy Rating
Myanmar	0.76	0.555556	1.256757	0.530631
Kyrgyzstan	0.79	0.500000	1.325000	0.523375
Lithuania	0.55	0.800000	1.189189	0.523243
Sierra Leone	0.65	0.660000	1.214286	0.520929
Ghana	0.54	0.800000	1.192308	0.515077
Maldives	0.77	0.666667	1.000000	0.513333
Italy	0.51	0.790000	1.270588	0.511920
New Caledonia	0.77	0.660000	1.000000	0.508200
Chile	0.49	0.727273	1.419689	0.505926
Mongolia	0.86	0.500000	1.170732	0.503415
Nigeria	0.58	0.789474	1.092857	0.500414
Bahrain	0.56	0.600000	1.470588	0.494118
Japan	0.48	0.808219	1.213805	0.470890
Greece	0.66	0.615385	1.118881	0.454438
Cuba	0.67	0.660000	1.000000	0.442200
Cape Verde	0.67	0.660000	1.000000	0.442200
Albania	0.68	0.538462	1.190476	0.435897
Taiwan	0.52	0.666667	1.256410	0.435556
Papua New Guinea	0.56	0.666667	1.142857	0.426667
Brazil	0.42	0.772455	1.312775	0.425905
Paraguay	0.49	0.714286	1.216667	0.425833
Czech Republic	0.50	0.698413	1.215645	0.424511
Bosnia and Herzegovina	0.66	0.555556	1.156250	0.423958
Oman	0.75	0.333333	1.692308	0.423077
Tajikistan	0.74	0.500000	1.133333	0.419333
Cambodia	0.54	0.714286	1.074627	0.414499
French Polynesia	0.50	0.660000	1.250000	0.412500
Haiti	0.62	0.660000	1.000000	0.409200
Hong Kong	0.29	0.865672	1.614286	0.405258
Panama	0.48	0.666667	1.258065	0.402581
Denmark	0.46	0.761905	1.138060	0.398863
Singapore	0.35	0.740741	1.510638	0.391647
Jamaica	0.50	0.660000	1.166667	0.385000
Bolivia	0.72	0.500000	1.064516	0.383226
Guatemala	0.44	0.600000	1.424242	0.376000

Continued on next page

Target Country	Gateway Factor	Cardinality Factor	Network Depth	Hierarchy Rating
Venezuela	0.47	0.692308	1.153846	0.375444
Puerto Rico	0.54	0.625000	1.070175	0.361184
France	0.39	0.736842	1.243927	0.357465
Qatar	0.58	0.500000	1.222222	0.354444
Botswana	0.50	0.660000	1.071429	0.353571
Colombia	0.41	0.764706	1.122137	0.351823
Solomon Islands	0.38	0.660000	1.400000	0.351120
Cameroon	0.69	0.500000	1.000000	0.345000
Sri Lanka	0.53	0.500000	1.294118	0.342941
Honduras	0.47	0.625000	1.155172	0.339332
Turkmenistan	0.67	0.500000	1.000000	0.335000
Guinea	0.50	0.660000	1.000000	0.330000
Syria	0.50	0.660000	1.000000	0.330000
Fiji	0.33	0.500000	2.000000	0.330000
Netherlands	0.42	0.619048	1.259162	0.327382
Burundi	0.56	0.500000	1.142857	0.320000
Mozambique	0.33	0.857143	1.111111	0.314286
Belgium	0.31	0.714286	1.382353	0.306092
Germany	0.41	0.654028	1.140320	0.305779
Jordan	0.60	0.400000	1.266667	0.304000
Uruguay	0.38	0.666667	1.200000	0.304000
Tanzania	0.45	0.600000	1.075000	0.290250
Cyprus	0.45	0.625000	1.018868	0.286557
Trinidad and Tobago	0.57	0.500000	1.000000	0.285000
Vanuatu	0.33	0.660000	1.250000	0.272250
Laos	0.40	0.600000	1.083333	0.260000
Morocco	0.76	0.333333	1.000000	0.253333
Peru	0.46	0.500000	1.100000	0.253000
Switzerland	0.33	0.666667	1.144161	0.251715
Malawi	0.33	0.666667	1.142857	0.251429
Algeria	0.50	0.500000	1.000000	0.250000
Kenya	0.26	0.700000	1.350000	0.245700
Bhutan	0.73	0.333333	1.000000	0.243333
Great Britain (UK)	0.31	0.636364	1.153617	0.227577
Malta	0.38	0.500000	1.166667	0.221667

Continued on next page

Target Country	Gateway Factor	Cardinality Factor	Network Depth	Hierarchy Rating
Saint Kitts and Nevis	0.33	0.660000	1.000000	0.217800
Swaziland	0.33	0.660000	1.000000	0.217800
Lesotho	0.33	0.660000	1.000000	0.217800
Grenada	0.33	0.660000	1.000000	0.217800
Suriname	0.33	0.660000	1.000000	0.217800
Barbados	0.33	0.660000	1.000000	0.217800
Mauritius	0.30	0.666667	1.071429	0.214286
Dominican Republic	0.39	0.500000	1.027778	0.200417
Democratic Republic of the Congo	0.40	0.500000	1.000000	0.200000
Sudan	0.29	0.660000	1.000000	0.191400
San Marino	0.38	0.500000	1.000000	0.190000
Ireland	0.23	0.750000	1.082353	0.186706
Uganda	0.32	0.500000	1.125000	0.180000
Guam	0.43	0.333333	1.250000	0.179167
Cayman Islands	0.27	0.660000	1.000000	0.178200
Faroe Islands	0.25	0.660000	1.000000	0.165000
Timor-Leste	0.25	0.660000	1.000000	0.165000
Zimbabwe	0.25	0.660000	1.000000	0.165000
Martinique	0.25	0.660000	1.000000	0.165000
Macao	0.33	0.500000	1.000000	0.165000
Madagascar	0.25	0.660000	1.000000	0.165000
Estonia	0.28	0.571429	1.029412	0.164706
Gibraltar	0.24	0.660000	1.000000	0.158400
South Africa	0.18	0.756757	1.132075	0.154207
Luxembourg	0.22	0.600000	1.111111	0.146667
Bahamas	0.29	0.500000	1.000000	0.145000
Mali	0.43	0.333333	1.000000	0.143333
Bermuda	0.40	0.333333	1.000000	0.133333
Brunei Darussalam	0.33	0.333333	1.200000	0.132000
Reunion	0.20	0.660000	1.000000	0.132000
Togo	0.20	0.660000	1.000000	0.132000
Somalia	0.35	0.333333	1.090909	0.127273

Continued on next page

Target Country	Gateway Factor	Cardinality Factor	Network Depth	Hierarchy Rating
Benin	0.38	0.333333	1.000000	0.126667
Costa Rica	0.20	0.500000	1.193548	0.119355
Virgin Islands (U.S.)	0.17	0.660000	1.000000	0.112200
Chad	0.17	0.660000	1.000000	0.112200
Nicaragua	0.26	0.400000	1.071429	0.111429
Cote D'Ivoire (Ivory Coast)	0.33	0.333333	1.000000	0.110000
Congo	0.22	0.500000	1.000000	0.110000
Rwanda	0.27	0.333333	1.111111	0.100000
Belize	0.21	0.333333	1.300000	0.091000
Zambia	0.11	0.660000	1.250000	0.090750
El Salvador	0.22	0.400000	1.000000	0.088000
Antigua and Barbuda	0.12	0.660000	1.000000	0.079200
Liberia	0.11	0.660000	1.000000	0.072600
Senegal	0.10	0.660000	1.000000	0.066000
Gabon	0.08	0.660000	1.000000	0.052800
Seychelles	0.07	0.500000	1.000000	0.035000
Liechtenstein	0.05	0.660000	1.000000	0.033000

APPENDIX C: Network Topology Identification Algorithm

This appendix provides the main algorithm developed by this study to identify country network topology from individual BGP collectors and to aggregate individual collector views into a single master topology for each country.

```python

import pandas as pd
import ast
import numpy as np
import SubnetTree


def rib_parser(rib, target_countries, directory_prefix, full_country_list):

    # parse specified rib files for country network topologies, generate json file of AS network associations,
    # generate statistics files

    # initialize variables common to all rib tables
    country_asns = [{} for i in range(len(target_countries))]
    master_line_counter = 0
    geo_dict = {}
    rosetta_dict = {}
    name_change_dict = {}

    # load country asn assignments from file into list of dictionaries
    for i in range(len(target_countries)):

        # read in country ASNs into country_asns list
        fd = open(directory_prefix + 'country ASNs/' + target_countries[i] + ".txt", "r")

        for line in fd:
            index = line.find("\t")
            country_asns[i][line[2:index]] = 1

        fd.close()

    # load configuration file to resolve country name discrepancies with geolocation database
    fchange = open(directory_prefix + 'Gateway AS Project/geo data/country_changes.csv',
                   'r')

    for line in fchange:
        name_change_dict[line[:line.index(",")]] = line[line.index(",") + 1:-1]
    name_change_dict['DR Congo'] = 'Congo'

    # load geolocation databases
    fprefixes = open('/Users/ericregnier/PycharmProjects/Gateway AS Project/geo data/GeoLite2-Country-Blocks-IPv4.csv',
                     'r')
    fprefixes.readline()

    for line in fprefixes:
        linelist = line.split(',')
        if linelist[1] == '':
            continue
        else:
            geo_dict[linelist[0]] = linelist[1]
    fprefixes.close()

    fprefixes = open('/Users/ericregnier/PycharmProjects/Gateway AS Project/geo data/GeoLite2-Country-Blocks-IPv6.csv',
                     'r')
    fprefixes.readline()

    for line in fprefixes:
        linelist = line.split(',')
        if linelist[1] == '':
            continue
        else:
            geo_dict[linelist[0]] = linelist[1]

    fcountry = open('/Users/ericregnier/PycharmProjects/Gateway AS Project/geo data/GeoLite2-Country-Locations-en.csv',
```

```
                    'r')
    fcountry.readline()

    # reformat geolocation databases to allow prefix queries to return country name
    for line in fcountry:
        linelist = line.split(',')
        rosetta_dict[linelist[0]] = linelist[5].replace('"', '')
        if linelist[5].replace('"', '') not in target_countries_list:
            if linelist[5].replace('"', '') in name_change_dict:
                rosetta_dict[linelist[0]] = name_change_dict[linelist[5].replace('"', '')]
        else:
            rosetta_dict[linelist[0]] = linelist[5].replace('"', '')

    for key in geo_dict:
        newkey = geo_dict[key]
        geo_dict[key] = rosetta_dict[newkey]

    for line in fprefixes:
        linelist = line.split(',')
        geo_dict[linelist[0]] = linelist[4]

    fcountry.close()
    fprefixes.close()
    fchange.close()

    # load geolocation database prefixes into subnettree object to enable longest prefix matching
    t = SubnetTree.SubnetTree()
    for key in geo_dict:
        t[key] = geo_dict[key]

    # begin processing rib files
    for n in range(len(rib)):

        # initialize rib round variables
        round_master_asn_dict = [{} for i in range(len(target_countries))]
        round_gateway_asn_dict = [{} for i in range(len(target_countries))]
        round_external_asn_dict = [{} for i in range(len(target_countries))]
        round_outpost_asn_dict = [{} for i in range(len(target_countries))]
        round_asn_association_dict = [{} for i in range(len(target_countries))]
        round_internal_asn_dict = [{} for i in range(len(target_countries))]
        internal_asn_depth_dict = [{} for i in range(len(target_countries))]
        asns_id = [{} for i in range(len(target_countries))]
        path_info_dict = [{} for i in range(len(target_countries))]
        as_path_dict = [{} for i in range(len(target_countries))]
        external_asn_degree_dict = [{} for i in range(len(target_countries))]
        ctr_prefix_in_country_dict = [{} for i in range(len(target_countries))]
        foreign_prefix_in_country_dict = [{} for i in range(len(full_country_list))]
        ctr_prefix_not_in_country_dict = [{} for i in range(len(target_countries))]
        ctr_prefix_not_in_database_dict = [{} for i in range(len(target_countries))]
        unique_prefix_dict = [{} for i in range(len(target_countries))]
        data = []
        data_prefixes = []
        prefix_list = []
        network_size_list = []
        line_counter = 0
        counter = 0
        index3 = 0
        path_flag = 0
        RIB_flag = 0
        Update_flag = 0
        Loop_flag = 0
        Announce_Flag = 0

        # open rib file
        rib_file = open(directory_prefix + 'RIBs/' + rib[n] + ".txt", encoding='latin-1')

        # read in lines from BGP table
        for line in rib_file:
```

```

            line_counter += 1
            counter += 1

            if counter == 500000:
                print("RIB Location:", rib[n], "| Number of lines processed: ", line_counter)
                counter = 0

            # identify rib table entry, store prefix
            if line[:7] == "PREFIX:":

                prefix_list.append(line[8:-1])
                network_size_list.append(line[-3:-1])
                RIB_flag = 1
                if network_size_list[-1][0] == '/':
                    network_size_list[-1] = network_size_list[-1][1]

            # identify BGP update entry, store ASPATH
            if line[:7] == "ASPATH:" and RIB_flag == 0:
                line_string = line[:-1]
                line_list = line_string.split(" ")
                if len(line_list) > 2:
                    Update_flag = 1

            # store ASPATH for rib table entry
            if RIB_flag == 1:
                if line[:7] == "ASPATH:":
                    line_string = line[:-1]
                    Loop_flag = 1

            # store prefix for BGP update entry
            if Announce_Flag == 1:
                if line[:2] == '  ':

                    prefix_list.append(line[2:-1])
                    network_size_list.append(line[-3:-1])
                    if network_size_list[-1][0] == '/':
                        network_size_list[-1] = network_size_list[-1][1]

            if line[:8] == 'ANNOUNCE' and Update_flag == 1:
                Announce_Flag = 1

            # identify end of rib table or update entry, reset markers, enter data processing loop
            if line == '\n':
                if Update_flag == 1:
                    Loop_flag = 1
                RIB_flag = 0
                Update_flag = 0
                Announce_Flag = 0

            # process data from rib table or BGP udpate entry
            if Loop_flag == 1:

                Loop_flag = 0

                # load ASPATH into list, identify origin AS
                line_list_index = -1
                line_list = line_string.split(" ")
                origin_as = line_list[line_list_index]
                origin_anchor = origin_as

                # identify registered country for origin AS
                for i in range(len(target_countries)):

                    for key in country_asns[i]:

                        # identify if origin AS is in the country of interest
                        if origin_anchor == key:

```

```python
                            outpost_flag = 1
                            depth_ctr = 0

                            # geolocate prefix via longest prefix matching, store location
                            for v in range(len(prefix_list)):

                                if prefix_list[v] not in unique_prefix_dict[i]:
                                    unique_prefix_dict[i][prefix_list[v]] = 1

                                try:

                                    if t[prefix_list[v][:-3]] == target_countries[i]:
                                        ctr_prefix_in_country_dict[i][prefix_list[v]] = network_size_list[v]

                                    else:

                                        ctr_prefix_not_in_country_dict[i][prefix_list[v]] = network_size_list[v]

                                        if t[prefix_list[v][:-3]] in full_country_list:
                                            foreign_prefix_in_country_dict[
                                                full_country_list.index(t[prefix_list[v][:-3]])][prefix_list[v]] = \
                                            target_countries[i]

                                # store prefixes not in geolocation database
                                except KeyError as err:
                                    ctr_prefix_not_in_database_dict[i][prefix_list[v]] = network_size_list[v]

                            if origin_anchor not in round_master_asn_dict[i]:
                                round_master_asn_dict[i][origin_anchor] = len(round_master_asn_dict[i])

                            # store list of ASs identified
                            if origin_as not in asns_id[i]:
                                asns_id[i][origin_as] = 1

                            # identify upstream AS from origin AS
                            next_as = line_list[line_list_index - 1]

                            # process upstream ASs that are also within the country of interest
                            while next_as in country_asns[i]:

                                # ignore AS prepending
                                if next_as != origin_as:

                                    depth_ctr += 1
                                    outpost_flag = 0

                                    if next_as not in round_master_asn_dict[i]:
                                        round_master_asn_dict[i][next_as] = len(round_master_asn_dict[i])

                                    # store list of ASs identified
                                    if origin_as not in asns_id[i]:
                                        asns_id[i][origin_as] = 1

                                    if origin_as not in round_internal_asn_dict[i]:
                                        round_internal_asn_dict[i][origin_as] = 1

                                    # store upstream AS and Original AS associations
                                    association = '{"source":' + str(round_master_asn_dict[i][next_as]) + \
                                                  ',"target":' + str(round_master_asn_dict[i][origin_as]) + ',"value":'
```

```python

                                    # store link prefix size as \20 equivolent
                                    if association not in round_asn_association_dict[i]:
                                        round_asn_association_dict[i][association] = 2 ** 20 / 2 ** int(
                                            network_size_list[0])
                                        if len(network_size_list) > 1:
                                            for l in range(len(network_size_list) - 1):
                                                round_asn_association_dict[i][association] += 2 ** 20 / 2 ** int(
                                                    network_size_list[l + 1])

                                    else:
                                        for l in range(len(network_size_list)):
                                            round_asn_association_dict[i][association] += 2 ** 20 / 2 ** int(
                                                network_size_list[l])

                                # identify next upstream AS from origin AS
                                line_list_index -= 1
                                origin_as = line_list[line_list_index]
                                next_as = line_list[line_list_index - 1]

                                path_flag = 1

                            # process upstream AS not within country of interest, store upstream AS (external to country)
                            # and downstream AS (internal to country) pairs
                            if next_as not in country_asns[i]:

                                # discard improper reading of upstream AS paths from the recursive function above
                                if next_as == "ASPATH:":
                                    continue

                                # store upstream AS information
                                if next_as not in round_master_asn_dict[i]:
                                    round_master_asn_dict[i][next_as] = len(round_master_asn_dict[i])

                                if next_as not in asns_id[i]:
                                    asns_id[i][next_as] = 1

                                # store country external ASs
                                if next_as not in round_external_asn_dict[i]:
                                    round_external_asn_dict[i][next_as] = 1

                                if outpost_flag == 1:

                                    # store outpost ASs
                                    if origin_anchor not in round_gateway_asn_dict[i]:

                                        if origin_anchor not in round_outpost_asn_dict[i]:
                                            round_outpost_asn_dict[i][origin_anchor] = 1

                                else:

                                    # store gateway ASs
                                    if origin_as not in round_gateway_asn_dict[i]:
                                        round_gateway_asn_dict[i][origin_as] = 1

                                    if origin_as in round_outpost_asn_dict[i]:
                                        round_outpost_asn_dict[i].pop(origin_as)

                                # store gateway/outpost to upstream AS Associations
                                association = '{"source":' + str(
```

```
round_master_asn_dict[i][next_as]) + ',"target":' \
                                              + str(round_master_asn_dict[i][
origin_as]) + ',"value":'

                                if association not in round_asn_association_dict[i]:
                                    round_asn_association_dict[i][association] = 2
 ** 20 / 2 ** int(
                                        network_size_list[0])
                                    if len(network_size_list) > 1:
                                        for l in range(len(network_size_list) - 1):
                                            round_asn_association_dict[i][
association] += 2 ** 20 / 2 ** int(
                                                network_size_list[l + 1])

                                else:
                                    for l in range(len(network_size_list)):
                                        round_asn_association_dict[i][association
] += 2 ** 20 / 2 ** int(
                                            network_size_list[l])

                                origin_as = line_list[line_list_index]
                                External_AS = line_list[line_list_index - 1]

                                # store external AS association
                                association = External_AS + origin_as
                                if association not in external_asn_degree_dict[i]:
                                    external_asn_degree_dict[i][association] = 1
                                    round_external_asn_dict[i][next_as] += 1

                            network_size_list = []
                            prefix_list = []

                            # store origin as depth within country topology, store
lowest depth value found in data
                            if depth_ctr != 0:
                                if depth_ctr not in internal_asn_depth_dict[i]:
                                    internal_asn_depth_dict[i][origin_anchor] =
depth_ctr
                                else:
                                    if depth_ctr < internal_asn_depth_dict[i][
origin_anchor]:
                                        internal_asn_depth_dict[i][origin_anchor] =
depth_ctr
                            break

        master_line_counter += line_counter
        print('Total Lines Processed: ', master_line_counter)
        rib_file.close()


        # write data to JSON file for visualizer method
        for i in range(len(target_countries)):

            # initialize data processing variables
            counter1 = 0
            counter2 = 0
            gateway_ases = []
            outpost_ases = []
            prefixesincountry = []
            prefixesnotincountry = []
            prefixesnotindatabase = []
            master_asn_list = []

            # write AS topology to JSON file for visualization and data processing
            master_asn_list = sorted(round_master_asn_dict[i].keys(), key=
round_master_asn_dict[i].get)
            rib_parsed_json = open(
                directory_prefix + 'working files/RIB_Parsed_JSON_' +
target_countries[i] + '_' + rib[n] + '.json',
                'w+')
            rib_parsed_json.write('{"nodes":[')
```

```python

            round_internal_asn_dict = [{} for i in range(len(target_countries))]
            round_internal_asn_dict[i] = round_master_asn_dict[i].copy()

            for key in round_external_asn_dict[i]:
                round_internal_asn_dict[i].pop(key)
            for key in round_outpost_asn_dict[i]:
                round_internal_asn_dict[i].pop(key)
            for key in round_gateway_asn_dict[i]:
                round_internal_asn_dict[i].pop(key)

            for m in range(len(master_asn_list)):
                write_string = ('{"name":"ASN' + master_asn_list[m] + '","group":')
                rib_parsed_json.write(write_string)
                if master_asn_list[m] in round_gateway_asn_dict[i]:
                    rib_parsed_json.write('2},')
                elif master_asn_list[m] in round_external_asn_dict[i]:
                    rib_parsed_json.write('3},')
                elif master_asn_list[m] in round_outpost_asn_dict[i]:
                    rib_parsed_json.write('4},')
                else:
                    rib_parsed_json.write('1},')

            rib_parsed_json.seek(0, 2)
            rib_parsed_json.seek(rib_parsed_json.tell() - 2, 0)
            rib_parsed_json.truncate()
            rib_parsed_json.write('}],"links":[')

            for key in round_asn_association_dict[i]:
                write_string = (key + str(round_asn_association_dict[i][key]) + '},')
                rib_parsed_json.write(write_string)

            rib_parsed_json.seek(0, 2)
            rib_parsed_json.seek(rib_parsed_json.tell() - 2, 0)
            rib_parsed_json.truncate()
            rib_parsed_json.write('}]}')
            rib_parsed_json.close()

            # store external AS node degree for data processing and visualization
            f = open(directory_prefix + 'working files/ROUND_EXTERNAL_ASN_DICT_' + target_countries[i] + '_' + rib[
                n] + '.txt', 'w')
            f.write(str(round_external_asn_dict[i]))
            f.close()

            # generate round statistics and write data to outputs

            # identify how many ASs from country listing were identified in the BGP table
            for key in country_asns[i]:
                if key not in asns_id[i]:
                    counter1 += 1
                else:
                    counter2 += 1

            # store round data for entry into dataframe
            for key in round_gateway_asn_dict[i]:
                gateway_ases.append(key)

            for key in round_outpost_asn_dict[i]:
                outpost_ases.append(key)

            for key in ctr_prefix_in_country_dict[i]:
                prefixesincountry.append(key)

            for key in ctr_prefix_not_in_country_dict[i]:
                prefixesnotincountry.append(key)

            for key in ctr_prefix_not_in_database_dict[i]:
                prefixesnotindatabase.append(key)
```

```
            if len(ctr_prefix_in_country_dict[i]) + len(
ctr_prefix_not_in_country_dict[i]) != 0:
                percent_in_country_geo = str(len(ctr_prefix_in_country_dict[i]) / (
                            len(ctr_prefix_in_country_dict[i]) + len(
ctr_prefix_not_in_country_dict[i])))
                percent_not_in_country_geo = str(len(ctr_prefix_not_in_country_dict[
i]) / (
                            len(ctr_prefix_in_country_dict[i]) + len(
ctr_prefix_not_in_country_dict[i])))

            else:
                percent_in_country_geo = np.NaN
                percent_not_in_country_geo = np.NaN

            if len(unique_prefix_dict[i]) != 0:
                percent_total_in_country_geo = str(len(ctr_prefix_in_country_dict[i
]) / len(unique_prefix_dict[i]))
                percent_total_not_in_country_geo = str(
                    len(ctr_prefix_not_in_country_dict[i]) / len(unique_prefix_dict[
i]))

            else:
                percent_total_in_country_geo = np.NaN
                percent_total_not_in_country_geo = np.NaN

            if counter2 == 0:

                data.append([target_countries[i],
                             rib[n],
                             str(len(round_gateway_asn_dict[i])),
                             str(len(round_outpost_asn_dict[i])),
                             str(counter2),
                             len(country_asns[i]),
                             np.NaN,
                             str(len(ctr_prefix_in_country_dict[i])),
                             str(len(ctr_prefix_not_in_country_dict[i])),
                             str(len(ctr_prefix_not_in_database_dict[i])),
                             str(len(unique_prefix_dict[i])),
                             percent_in_country_geo,
                             percent_not_in_country_geo,
                             percent_total_in_country_geo,
                             percent_total_not_in_country_geo,
                             gateway_ases,
                             outpost_ases])

            else:

                data.append([target_countries[i],
                             rib[n],
                             str(len(round_gateway_asn_dict[i])),
                             str(len(round_outpost_asn_dict[i])),
                             str(counter2), len(country_asns[i]),
                             str('%.2f' % ((counter2 - len(round_gateway_asn_dict[i
]) - len(
                                 round_outpost_asn_dict[i])) / counter2)),
                             str(len(ctr_prefix_in_country_dict[i])),
                             str(len(ctr_prefix_not_in_country_dict[i])),
                             str(len(ctr_prefix_not_in_database_dict[i])),
                             str(len(unique_prefix_dict[i])),
                             percent_in_country_geo,
                             percent_not_in_country_geo,
                             percent_total_in_country_geo,
                             percent_total_not_in_country_geo,
                             gateway_ases,
                             outpost_ases])

            data_prefixes.append([foreign_prefix_in_country_dict[full_country_list.
index(target_countries[i])],
                                  prefixesincountry,
                                  prefixesnotincountry,
```

```
                                   prefixesnotindatabase])

        df = pd.DataFrame(data, columns=['Target Country',
                                          'RIB Collector',
                                          '# Gateway ASs',
                                          '# Outpost ASs',
                                          '# Visible Country ASs',
                                          '# Registered ASs',
                                          'Gateway Factor',
                                          '# Prefixes Geolocated in Country',
                                          '# Prefixes Not Geolocated in Country',
                                          '# Prefixes Not in Database',
                                          '# Total Prefixes in RIB',
                                          '% Prefixes in Database, Geolocated in Country',
                                          '% Prefixes in Database, Not Geolocated in Country',
                                          '% Total Prefixes Geolocated in Country',
                                          '% Total Prefixes Not Geolocated in Country',
                                          'Gateway AS Listing',
                                          'Outpost AS Listing'])

        # generate dataframes, store to file
        df_prefixes = pd.DataFrame(data_prefixes,
                                    columns=['Prefixes Geolocated in Country Registered to Foreign Country',
                                             'Prefixes Geolocated in Country',
                                             'Prefixes Not Geolocated in Country',
                                             'Prefixes Not in Database'])

        df.to_csv(directory_prefix + 'Output/stats/Country_ASN_Mapper_Out_' + rib[n] + '.csv')
        df_prefixes.to_csv(directory_prefix + 'Output/stats/Country_ASN_Mapper_Out_Prefixes_' + rib[n] + '.csv')

        # write rib variables to file for data processing and aggregation
        f = open(directory_prefix + 'working files/path_info_dict_' + rib[n] + '.txt', 'w')
        f.write(str(path_info_dict))
        f.close()

        f = open(directory_prefix + 'working files/round_gateway_asn_dict_' + rib[n] + '.txt', 'w')
        f.write(str(round_gateway_asn_dict))
        f.close()

        f = open(directory_prefix + 'working files/round_external_asn_dict_' + rib[n] + '.txt', 'w')
        f.write(str(round_external_asn_dict))
        f.close()

        f = open(directory_prefix + 'working files/round_asn_association_dict_' + rib[n] + '.txt', 'w')
        f.write(str(round_asn_association_dict))
        f.close()

        f = open(directory_prefix + 'working files/round_outpost_asn_dict_' + rib[n] + '.txt', 'w')
        f.write(str(round_outpost_asn_dict))
        f.close()

        f = open(directory_prefix + 'working files/round_master_asn_dict_' + rib[n] + '.txt', 'w')
        f.write(str(round_master_asn_dict))
        f.close()

        f = open(directory_prefix + 'working files/round_internal_asn_dict_' + rib[n] + '.txt', 'w')
        f.write(str(round_internal_asn_dict))
        f.close()
```

```python

        f = open(directory_prefix + 'working files/foreign prefixes' + rib[n] + '.txt', 'w')
        f.write(str(foreign_prefix_in_country_dict))
        f.close()

        f = open(directory_prefix + 'working files/in country prefixes' + rib[n] + '.txt', 'w')
        f.write(str(ctr_prefix_in_country_dict))
        f.close()

        f = open(directory_prefix + 'working files/not in country prefixes' + rib[n] + '.txt', 'w')
        f.write(str(ctr_prefix_not_in_country_dict))
        f.close()

        f = open(directory_prefix + 'working files/foreign_prefixes_in_country' + rib[n] + '.txt', 'w')
        f.write(str(foreign_prefix_in_country_dict))
        f.close()

        f = open(directory_prefix + 'working files/prefixes_not_in_database' + rib[n] + '.txt', 'w')
        f.write(str(ctr_prefix_not_in_database_dict))
        f.close()

        f = open(directory_prefix + 'working files/unique_prefix_dict' + rib[n] + '.txt', 'w')
        f.write(str(unique_prefix_dict))
        f.close()

        f = open(directory_prefix + 'working files/internal_asn_depth_dict' + rib[n] + '.txt', 'w')
        f.write(str(internal_asn_depth_dict))
        f.close()


def rib_aggregator(rib, target_countries, directory_prefix):

    # initialize variables
    data_master = []
    data_prefixes = []
    country_asns = [{} for i in range(len(target_countries))]

    # load as country registration
    for i in range(len(target_countries)):

        # read in country ASNs into country_asns list
        fd = open(directory_prefix + 'country ASNs/' + target_countries[i] + ".txt", "r")

        for line in fd:

            index = line.find("\t")
            country_asns[i][line[2:index]] = 1

        fd.close()

    # load first rib table variables as 'master' variables
    n=0
    print(rib[n])

    f = open(directory_prefix + 'working files/round_gateway_asn_dict_' + rib[n] + '.txt','r')
    master_gateway_asn_dict = ast.literal_eval(f.read())
    f.close()

    f = open(directory_prefix + 'working files/round_external_asn_dict_' + rib[n] + '.txt','r')
    master_external_asn_dict = ast.literal_eval(f.read())
    f.close()
```

```
    f = open(directory_prefix + 'working files/round_asn_association_dict_' + rib[n] + '.txt','r')
    master_asn_association_dict = ast.literal_eval(f.read())
    f.close()

    f = open(directory_prefix + 'working files/round_outpost_asn_dict_' + rib[n] + '.txt','r')
    master_outpost_asn_dict = ast.literal_eval(f.read())
    f.close()

    f = open(directory_prefix + 'working files/round_master_asn_dict_' + rib[n] + '.txt','r')
    master_asn_dict = ast.literal_eval(f.read())
    f.close()

    f = open(directory_prefix + 'working files/in country prefixes' + rib[n] + '.txt', 'r')
    master_ctr_prefix_in_country_dict = ast.literal_eval(f.read())
    f.close()

    f = open(directory_prefix + 'working files/not in country prefixes' + rib[n] + '.txt', 'r')
    master_ctr_prefix_not_in_country_dict = ast.literal_eval(f.read())
    f.close()

    f = open(directory_prefix + 'working files/foreign_prefixes_in_country' + rib[n] + '.txt', 'r')
    master_foreign_prefix_in_country_dict = ast.literal_eval(f.read())
    f.close()

    f = open(directory_prefix + 'working files/prefixes_not_in_database' + rib[n] + '.txt', 'r')
    master_ctr_prefix_not_in_database_dict = ast.literal_eval(f.read())
    f.close()

    f = open(directory_prefix + 'working files/unique_prefix_dict' + rib[n] + '.txt', 'r')
    master_unique_prefix_dict = ast.literal_eval(f.read())
    f.close()

    f = open(directory_prefix + 'working files/internal_asn_depth_dict' + rib[n] + '.txt', 'r')
    master_internal_asn_depth_dict = ast.literal_eval(f.read())
    f.close()

    # load subsequent rib table variables and aggregate into master variables
    for i in range(len(rib)-1):

        print('Processing RIB Table Variables from: ',rib[i+1])

        mapper_dict = [{} for i in range(len(target_countries))]

        f = open(directory_prefix + 'working files/round_master_asn_dict_' + rib[i+1] + '.txt', 'r')
        round_master_asn_dict = ast.literal_eval(f.read())
        f.close()

        for n in range(len(target_countries)):
            for key in round_master_asn_dict[n]:
                if key not in master_asn_dict[n]:
                    master_asn_dict[n][key] = len(master_asn_dict[n])

        for n in range(len(target_countries)):
            mapper_dict[n] = dict((v, k) for k, v in round_master_asn_dict[n].items())

        f = open(directory_prefix + 'working files/round_gateway_asn_dict_' + rib[i+1] + '.txt','r')
        round_gateway_asn_dict = ast.literal_eval(f.read())
        f.close()
```

```

        for n in range(len(target_countries)):
            for key in round_gateway_asn_dict[n]:
                if key not in master_gateway_asn_dict[n]:
                    master_gateway_asn_dict[n][key] = 1

        f = open(directory_prefix + 'working files/round_external_asn_dict_' + rib[i+1] + '.txt','r')
        round_external_asn_dict = ast.literal_eval(f.read())
        f.close()

        for n in range(len(target_countries)):
            for key in round_external_asn_dict[n]:
                if key not in master_external_asn_dict[n]:
                    master_external_asn_dict[n][key] = round_external_asn_dict[n][key]
                else:
                    master_external_asn_dict[n][key] += round_external_asn_dict[n][key]

        f = open(directory_prefix + 'working files/round_outpost_asn_dict_' + rib[i+1] + '.txt','r')
        round_outpost_asn_dict = ast.literal_eval(f.read())
        f.close()

        for n in range(len(target_countries)):
            for key in round_outpost_asn_dict[n]:
                if key not in master_outpost_asn_dict[n]:
                    master_outpost_asn_dict[n][key] = 1

        f = open(directory_prefix + 'working files/in country prefixes' + rib[i+1] + '.txt', 'r')
        round_ctr_prefix_in_country_dict = ast.literal_eval(f.read())
        f.close()

        for n in range(len(target_countries)):
            for key in round_ctr_prefix_in_country_dict[n]:
                if key not in master_ctr_prefix_in_country_dict[n]:
                    master_ctr_prefix_in_country_dict[n][key] = 1

        f = open(directory_prefix + 'working files/not in country prefixes' + rib[i+1] + '.txt', 'r')
        round_ctr_prefix_not_in_country_dict = ast.literal_eval(f.read())
        f.close()

        for n in range(len(target_countries)):
            for key in round_ctr_prefix_not_in_country_dict[n]:
                if key not in master_ctr_prefix_not_in_country_dict[n]:
                    master_ctr_prefix_not_in_country_dict[n][key] = 1

        f = open(directory_prefix + 'working files/foreign_prefixes_in_country' + rib[i+1] + '.txt', 'r')
        round_foreign_prefix_in_country_dict = ast.literal_eval(f.read())
        f.close()

        for n in range(len(target_countries)):
            for key in round_foreign_prefix_in_country_dict[n]:
                if key not in master_foreign_prefix_in_country_dict[n]:
                    master_foreign_prefix_in_country_dict[n][key] = round_foreign_prefix_in_country_dict[n][key]

        f = open(directory_prefix + 'working files/prefixes_not_in_database' + rib[i+1] + '.txt', 'r')
        round_ctr_prefix_not_in_database_dict = ast.literal_eval(f.read())
        f.close()

        for n in range(len(target_countries)):
            for key in round_ctr_prefix_not_in_database_dict[n]:
                if key not in master_ctr_prefix_not_in_database_dict[n]:
                    master_ctr_prefix_not_in_database_dict[n][key] = 1

```

```
        f = open(directory_prefix + 'working files/unique_prefix_dict' + rib[i+1] + '.txt', 'r')
        round_unique_prefix_dict = ast.literal_eval(f.read())
        f.close()

        for n in range(len(target_countries)):
            for key in round_unique_prefix_dict[n]:
                if key not in master_unique_prefix_dict[n]:
                    master_unique_prefix_dict[n][key] = 1

        f = open(directory_prefix + 'working files/round_asn_association_dict_' + rib[i+1] + '.txt', 'r')
        round_asn_association_dict = ast.literal_eval(f.read())
        f.close()

        f = open(directory_prefix + 'working files/internal_asn_depth_dict' + rib[i+1] + '.txt', 'r')
        round_internal_asn_depth_dict = ast.literal_eval(f.read())
        f.close()

        for key in round_internal_asn_depth_dict[n]:
            if key in master_internal_asn_depth_dict[n]:
                if round_internal_asn_depth_dict[n][key] < master_internal_asn_depth_dict[n][key]:
                    master_internal_asn_depth_dict[n][key] = round_internal_asn_depth_dict[n][key]

            if key not in master_internal_asn_depth_dict[n]:
                master_internal_asn_depth_dict[n][key] = round_internal_asn_depth_dict[n][key]

        for n in range(len(target_countries)):

            for key in round_asn_association_dict[n]:
                string = key
                source = int(string[string.index('e":') + 3 : string.index(',"t')])
                target = int(string[string.index('t":') + 3 : string.index(',"v')])

                association = '{"source":' + str(master_asn_dict[n][mapper_dict[n][source]]) + ',"target":' \
                              + str(master_asn_dict[n][mapper_dict[n][target]]) + ',"value":'

                if association in master_asn_association_dict[n]:
                    master_asn_association_dict[n][association] += round_asn_association_dict[n][key]

                else:
                    master_asn_association_dict[n][association] = round_asn_association_dict[n][key]

    # generate master JSON file from aggregated network topologies, write to file
    for i in range(len(target_countries)):

        master_asn_list = sorted(master_asn_dict[i].keys(), key=master_asn_dict[i].get)
        rib_parsed_json = open(
            directory_prefix + 'working files/RIB_Parsed_JSON_' + target_countries[i] + '_Master.json',
            'w+')
        rib_parsed_json.write('{"nodes":[')

        for m in range(len(master_asn_list)):
            write_string = ('{"name":"ASN' + master_asn_list[m] + '","group":')
            rib_parsed_json.write(write_string)
            if master_asn_list[m] in master_gateway_asn_dict[i]:
                rib_parsed_json.write('2},')
            elif master_asn_list[m] in master_external_asn_dict[i]:
                rib_parsed_json.write('3},')
            elif master_asn_list[m] in master_outpost_asn_dict[i]:
                rib_parsed_json.write('4},')
```

```
            else:
                rib_parsed_json.write('1},')

        rib_parsed_json.seek(0, 2)
        rib_parsed_json.seek(rib_parsed_json.tell() - 2, 0)
        rib_parsed_json.truncate()
        rib_parsed_json.write('}],"links":[')

        for key in master_asn_association_dict[i]:
            write_string = (key + str(master_asn_association_dict[i][key]) + '},')
            rib_parsed_json.write(write_string)

        rib_parsed_json.seek(0, 2)
        rib_parsed_json.seek(rib_parsed_json.tell() - 2, 0)
        rib_parsed_json.truncate()
        rib_parsed_json.write('}]}')
        rib_parsed_json.close()

        f = open(
            directory_prefix + 'working files/EXTERNAL_ASN_DICT_' + target_countries[i] + '_MASTER.TXT',
            'w')
        f.write(str(master_external_asn_dict[i]))
        f.close()

    # save aggregated variables for data processing
    f = open(directory_prefix + 'working files/internal_asn_depth_dict_Master.txt', 'w')
    f.write(str(master_internal_asn_depth_dict))
    f.close()

    f = open(directory_prefix + 'working files/foreign_prefix_in_country_dict_Master.txt', 'w')
    f.write(str(master_foreign_prefix_in_country_dict))
    f.close()

    f = open(directory_prefix + 'working files/unique_prefix_dict_Master.txt', 'w')
    f.write(str(master_unique_prefix_dict))
    f.close()

    f = open(directory_prefix + 'working files/prefix_in_country_dict_Master.txt', 'w')
    f.write(str(master_ctr_prefix_in_country_dict))
    f.close()

    for i in range(len(target_countries)):

        # generate round statistics and write data to outputs
        f = open(directory_prefix + 'working files/EXTERNAL_ASN_DICT_' + target_countries[i] + '.txt',
                 'w')
        f.write(str(master_external_asn_dict[i]))
        f.close()

        # store round data for entry into dataframe
        master_gateway_ases = []
        master_outpost_ases = []
        counter1 = 0
        counter2 = 0

        # identify how many ASs from country listing were identified in the BGP table
        for key in country_asns[i]:
            if key not in master_asn_dict[i]:
                counter1 += 1
            else:
                counter2 += 1

        for key in master_gateway_asn_dict[i]:
            if key in master_outpost_asn_dict[i]:
                master_outpost_asn_dict[i].pop(key)
```

```python


        # store round data for entry into dataframe
        for key in master_gateway_asn_dict[i]:
            master_gateway_ases.append(key)

        for key in master_outpost_asn_dict[i]:
            master_outpost_ases.append(key)

        prefixesincountry = []
        prefixesnotincountry = []
        prefixesnotindatabase = []
        gateway_ases = []
        outpost_ases = []

        for key in master_gateway_asn_dict[i]:
            gateway_ases.append(key)

        for key in master_outpost_asn_dict[i]:
            outpost_ases.append(key)

        for key in master_ctr_prefix_in_country_dict[i]:
            prefixesincountry.append(key)

        for key in master_ctr_prefix_not_in_country_dict[i]:
            prefixesnotincountry.append(key)

        for key in master_ctr_prefix_not_in_database_dict[i]:
            prefixesnotindatabase.append(key)

        if len(master_ctr_prefix_not_in_country_dict[i]) + len(
master_ctr_prefix_not_in_country_dict[i]) != 0:
            percent_in_country_geo = str(len(master_ctr_prefix_in_country_dict[i])/
                            (len(master_ctr_prefix_in_country_dict[i]) + len(
master_ctr_prefix_not_in_country_dict[i])))
            percent_not_in_country_geo = str(len(
master_ctr_prefix_not_in_country_dict[i]) /
                            (len(master_ctr_prefix_in_country_dict[i]) + len(
master_ctr_prefix_not_in_country_dict[i])))

        else:
            percent_in_country_geo = np.NaN
            percent_not_in_country_geo = np.NaN

        if len(master_unique_prefix_dict[i]) != 0:
            percent_total_in_country_geo = str(len(
master_ctr_prefix_not_in_country_dict[i]) /
                                               len(master_unique_prefix_dict[i]))
            percent_total_not_in_country_geo = str(len(
master_ctr_prefix_not_in_country_dict[i]) /
                                                   len(master_unique_prefix_dict[i
]))

        else:
            percent_total_in_country_geo = np.NaN
            percent_total_not_in_country_geo = np.NaN

        if counter2 == 0:

            data_master.append([target_countries[i],
                         str(len(master_gateway_asn_dict[i])),
                         str(len(master_outpost_asn_dict[i])),
                         str(counter2),
                         len(country_asns[i]),
                         np.NaN,
                         str(len(master_ctr_prefix_in_country_dict[i])),
                         str(len(master_ctr_prefix_not_in_country_dict[i])),
                         str(len(master_ctr_prefix_not_in_database_dict[i])),
                         str(len(master_unique_prefix_dict[i])),
                         percent_in_country_geo,
                         percent_not_in_country_geo,
```

```
                         percent_total_in_country_geo,
                         percent_total_not_in_country_geo,
                         gateway_ases,
                         outpost_ases])

        else:

            data_master.append([target_countries[i],
                         str(len(master_gateway_asn_dict[i])),
                         str(len(master_outpost_asn_dict[i])),
                         str(counter2), len(country_asns[i]),
                         str('%.2f' % (
                                     (counter2 - len(master_gateway_asn_dict[i]) - len(master_outpost_asn_dict[i])) /
                                     counter2)),
                         str(len(master_ctr_prefix_in_country_dict[i])),
                         str(len(master_ctr_prefix_not_in_country_dict[i])),
                         str(len(master_ctr_prefix_not_in_database_dict[i])),
                         str(len(master_unique_prefix_dict[i])),
                         percent_in_country_geo,
                         percent_not_in_country_geo,
                         percent_total_in_country_geo,
                         percent_total_not_in_country_geo,
                         gateway_ases,
                         outpost_ases])

        data_prefixes.append([master_foreign_prefix_in_country_dict[full_country_list.index(target_countries[i])],
                              prefixesincountry,
                              prefixesnotincountry,
                              prefixesnotindatabase])

        df_master = pd.DataFrame(data_master, columns=['Target Country',
                                         '# Gateway ASs',
                                         '# Outpost ASs',
                                         '# Visible Country ASs',
                                         '# Registered ASs',
                                         'Gateway Factor',
                                         '# Prefixes Geolocated in Country',
                                         '# Prefixes Not Geolocated in Country',
                                         '# Prefixes Not in Database',
                                         '# Total Prefixes in RIB',
                                         '% Prefixes in Database, Geolocated in Country',
                                         '% Prefixes in Database, Not Geolocated in Country',
                                         '% Total Prefixes Geolocated in Country',
                                         '% Total Prefixes Not Geolocated in Country',
                                         'Gateway AS Listing',
                                         'Outpost AS Listing'])

        df_prefixes = pd.DataFrame(data_prefixes, columns=['Prefixes Geolocated in Country Registered to Foreign Country',
                                                           'Prefixes Geolocated in Country',
                                                           'Prefixes Not Geolocated in Country',
                                                           'Prefixes Not in Database'])

    # save output to file
    df_master.to_csv(directory_prefix + 'Output/stats/Country_ASN_Mapper_Out_Master.csv')
    df_prefixes.to_csv(directory_prefix + 'Output/stats/Country_ASN_Mapper_Out_Prefixes_Master.csv')


def configuration_file_importer(filename, directory_prefix):

    # read in configuration files passed into lists, return lists
```

```python
    file_in = []
    fd = open(directory_prefix + filename, 'r')

    for line in fd:
        file_in.append(line[:-1])

    fd.close()

    return file_in


#-----------------------------------------------------------------------------------------------------------------------
# specify list of target countries and RIBs to parse in their respective configuration files located in the
# parent directory
#-----------------------------------------------------------------------------------------------------------------------

# open config files, read into lists
directory_prefix = '/Users/ericregnier/PycharmProjects/Gateway AS Project/'
rib_config_list = configuration_file_importer('RIB_Config.txt', directory_prefix)
target_countries_list = configuration_file_importer('target_countries_list.txt', directory_prefix)
full_country_list = configuration_file_importer('full_target_countries_list.txt', directory_prefix)

# pass target countries and RIBs to main parser method to identify country network topologies
rib_parser(rib_config_list, target_countries_list, directory_prefix, full_country_list)

# pass target countries and RIB list for network topology aggregation for all collector locations
# rib_aggregator(rib_config_list, target_countries_list, directory_prefix)
```

List of References

[1] G. Nomikos and X. Dimitropoulos, “Traixroute: detecting Internet Exchange Point (IXP)s in traceroute paths,” in *Passive and Active Measurement* (Lecture Notes in Computer Science). Cham: Springer International Publishing, 2016, pp. 346–358.

[2] J. Hawkinson and T. Bates, “Guidelines for creation, selection, and registration of an autonomous system,” RFC 1930, Mar. 1996. [Online]. Available: https://tools.ietf.org/html/rfc1930

[3] R. Housley, “Internet numbers registries,” RFC 7249, May 2014. [Online]. Available: https://tools.ietf.org/html/rfc7249

[4] ipinfo.io. “ASNs by country report,” Oct 18, 2020. [Online]. Available: https://ipinfo.io/countries

[5] A. Improta, “On the analysis of the internet from a geographic and economic perspective via BGP raw data,” Ph.D. dissertation, School of Engineering Research, University of Pisa, Pisa, Italy, 2013.

[6] Internet Assigned Numbers Authority. “Number resources,” 03 Jan., 2021. [Online]. Available: https://www.iana.org/numbers

[7] Y. Rekhter, T. Li, and S. Hares, “A border gateway protocol 4,” RFC 4271, Jan. 2006. [Online]. Available: https://tools.ietf.org/html/rfc4271

[8] N. Bahadur, S. Kini, and J. Medved, “RIB information model,” RFC 8430, 2018. [Online]. Available: https://tools.ietf.org/html/rfc8430

[9] E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani, “A novel methodology to address the internet Autonomous System (AS)-level data incompleteness,” *IEEE/ACM Transactions on Networking*, vol. 23, no. 4, pp. 1314–1327, 2015.

[10] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, K. Claffy, and G. Riley, “AS relationships: inference and validation,” *ACM SIGCOMM Computer Communication Review (CCR)*, vol. 37, no. 1, pp. 29–40, Jan 2007.

[11] M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas, and K. Claffy, “AS relationships, customer cones, and validation” (IMC ’13). ACM, 2013, pp. 243–256.

[12] Center for Applied Internet Data Analysis. “ASRank,” Dec 09, 2020. [Online]. Available: https://asrank.caida.org/

[13] Center for Applied Internet Data Analysis. "AS relationships," 01 Nov, 2020. [Online]. Available: https://www.caida.org/data/as-relationships

[14] L. Gao, "On inferring autonomous system relationships in the internet," *IEEE/ACM transactions on networking*, vol. 9, no. 6, pp. 733–745, 2001.

[15] V. Giotsas and S. Zhou, "Valley-free violation in internet routing — analysis based on BGP community data," Jun 2012, pp. 1193–1197.

[16] T. Böttger, G. Antichi, E. Fernandes, R. Di Lallo, M. Bruyere, S. Uhlig, G. Tyson, and I. Castro, "Shaping the internet: 10 years of IXP growth," 2018.

[17] A. Dainotti, C. Squarcella, E. Aben, K. Claffy, M. Chiesa, M. Russo, and A. Pescapè, "Analysis of country-wide internet outages caused by censorship," in *ACM Internet Measurement Conference (IMC)*, Nov 2011, pp. 1–18.

[18] British Broadcasting Company. "Russia 'successfully tests' its unplugged internet," Dec 24, 2019. [Online]. Available: https://www.bbc.com/news/technology-50902496

[19] Forbes. "Russia cuts off its internet, with mixed results," Dec 24, 2019. [Online]. Available: https://www.forbes.com/sites/emmawoollacott/2019/12/24/russia-cuts-off-its-internet-with-mixed-results/?sh=710d9a06619d

[20] Y. Wang, D. Burgener, M. Flores, A. Kuzmanovic, and C. Huang, "Towards street-level client-independent IP geolocation," in *Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation* (NSDI'11). USA: USENIX Association, 2011, p. 365–379.

[21] Y. Shavitt and N. Zilberman, "A geolocation databases study," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 10, pp. 2044–2056, 2011.

[22] B. Huffaker, M. Fomenkov, and K. Claffy, "Geocompare: a comparison of public and commercial geolocation databases - technical report ," Cooperative Association for Internet Data Analysis (CAIDA), Tech. Rep., May 2011.

[23] P. Winter, R. Padmanabhan, A. King, and A. Dainotti, "Geo-locating BGP prefixes," in *Network Traffic Measurement and Analysis Conference (TMA)*, Jun 2019.

[24] RIPE Network Coordination Center. "RIS raw data," 23 Nov, 2020. [Online]. Available: https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/ris-raw-data

[25] Center for Applied Internet Data Analysis. "The CAIDA AS relationships dataset," 7 Oct, 2020. [Online]. Available: http://www.caida.org/data/active/as-relationships/

[26] MaxMind. "GeoIP2 databases," Nov 24, 2020. [Online]. Available: https://www.maxmind.com/en/geoip2-databases

[27] Center for Applied Internet Data Analysis. "AS classification," 01 Nov, 2020. [Online]. Available: https://www.caida.org/data/as-classification/

[28] X. Shi, Y. Xiang, Z. Wang, X. Yin, and J. Wu, "Detecting prefix hijackings in the internet with argus" (IMC '12). ACM, 2012, pp. 15–28.

[29] "The (in)completeness of the observed internet AS-level structure," *IEEE/ACM transactions on networking*, vol. 18, no. 1, pp. 109–122, 2010.

[30] The Economist. "Democracy index 2019," Nov, 2020. [Online]. Available: https://www.eiu.com/public/topical_report.aspx?campaignid=democracyindex2019

THIS PAGE INTENTIONALLY LEFT BLANK

Initial Distribution List

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California