On the Importance of Being an AS: An Approach to Country-Level AS Rankings

Bradley Huffaker
CAIDA / UC San Diego
San Diego, US
bradley@caida.org

Romain Fontugne
IIJ Research Lab
Tokyo, JP
romain@nii.ac.jp

Alexander Marder[1]
CAIDA / UC San Diego
San Diego, US
amarder@jhu.edu

kc claffy
CAIDA / UC San Diego
San Diego, US
kc@caida.org

ABSTRACT

Recent geopolitical events demonstrate that control of Internet infrastructure in a region is critical to economic activity and defense against armed conflict. This geopolitical importance necessitates novel empirical techniques to assess which countries remain susceptible to degraded or severed Internet connectivity because they rely heavily on networks based in other nation states. Currently, two preeminent BGP-based methods exist to identify influential or market-dominant networks on a global scale—network-level *customer cone* size and path *hegemony*—but these metrics fail to capture regional or national differences.

We adapt the two global metrics to capture country-specific differences by restricting the input data for a country-specific metric to destination prefixes in that country. Although conceptually simple, our study required tackling methodological challenges common to most Internet measurement research today, such as geolocation, incomplete data, vantage point access, and lack of ground truth. Restricting public routing data to individual countries requires substantial downsampling compared to global analysis, and we analyze the impact of downsampling on the robustness and stability of our country-specific metrics. As a measure of validation, we apply our country-specific metrics to case studies of Australia, Japan, Russia, Taiwan, and the United States, illuminating aspects of concentration and interdependence in telecommunications markets. To support reproducibility, we will share our code, inferences, and data sets with other researchers.

CCS CONCEPTS

• **Networks → Network properties**; • **Security and privacy →** *Network security*.

[1] Alexander Marder is currently at John Hopkins.

IMC '23, October 24–26, 2023, Montreal, QC, Canada

ACM ISBN 979-8-4007-0382-9/23/10…$15.00
https://doi.org/10.1145/3618257.3624798

KEYWORDS

asn country security geopolitical topology

ACM Reference Format:
Bradley Huffaker, Romain Fontugne, Alexander Marder[1], and kc claffy. 2023. On the Importance of Being an AS: An Approach to Country-Level AS Rankings. In *Proceedings of the 2023 ACM Internet Measurement Conference (IMC '23), October 24–26, 2023, Montreal, QC, Canada.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3618257.3624798

1 INTRODUCTION

Although the Internet is a global infrastructure, many regional factors drive the evolution of the connectivity graph among its ≈ 80K independent autonomous systems (ASes), often called the Internet's *AS topology*. These factors include local and national regulations, history, and economic forces. For many countries, the network-level (AS) topology, and the physical router topology that manifests it, reflects the connectivity of a few dominant national network providers.

Examining Internet Spheres of Influence. Identifying dominant ASes, and how and where they connect to the rest of the world, allows assessments of a country's dependence on other countries' networks. The Internet is critical infrastructure throughout much of the world, with countries now relying on the Internet as the dominant mode of commerce and information. Adversarial nation-states could weaponize ASes headquartered within their sovereign borders through political or statutory control to monitor, disrupt, or censor traffic into a geographic region. Currently, there is no reliable technique for quantifying Internet sovereignty metrics, such as the extent to which a country's Internet communication relies on networks potentially controlled by adversarial nation-states; e.g., how dependent is Taiwan on Chinese ISPs? These metrics could also predict or evaluate the success of network sanctions, such as the effect of U.S. ISPs de-peering Russian ISPs following Russia's invasion of Ukraine.

The current state of the art for answering these questions is to extract ASes specific to a country from a global AS ranking. But this approach is misleading because it fails to account for the regional influence of ASes. For example, NTT operates a Tier 1 network (2914 a.k.a. NTT America) that appears fourth and fifth in the two canonical global rankings, and NTT also operates AS 4713 specifically for the Japanese market (#201 in ASRank). While both ASes provide service within Japan, AS 4713 is far more critical for Japan's domestic Internet. The same applies in other countries, including Australia, France, Italy, Spain, and Sweden, where usually the incumbent ISP is operating both an international transit network and a large domestic access network. Evaluating the importance of an AS to a given country requires accounting for regional differences in AS topology, something that filtering global rankings cannot do.

A method for analyzing AS-dependence at the country-level can also help answer other questions with geopolitical and regulatory implications. How diverse are a country's dominant ASes, individually or as a set? Are they domestic, foreign, or broadly multinational? Are the dominant ASes in a given country state-owned or privately owned?

BGP Data Can Enable Independent Estimates. Publically available BGP data can provide a useful proxy estimate of topological importance of an AS. Researchers have used two BGP-based metrics to quantify the influence of an AS over the global Internet routing system: (1) AS-level *customer cone* size [30], to estimate the importance of an AS in the Internet transit hierarchy; and (2) *AS hegemony* [20], a measure of the centrality of an AS in the Internet AS-level graph. The customer cone metric uses inferred customer routing relationships to compute the number of other ASes reachable via an AS's customers, including customers of customers. The hegemony metric ranks ASes according to the fraction of a considered set (or subset) of address space reached via that AS, which estimates the likelihood that an AS lies on a path toward that address space.

Our goal is to adapt the global AS Customer Cone and AS Hegemony metrics to country-specific (or region-specific) metrics that can quantify the importance of an AS to a country's (or region's) Internet connectivity. In a country-specific context, the AS customer cone metric captures how much of a country's address space potentially uses the AS for Internet transit, i.e., operators/owners of the address space pay the AS or its customers to send and receive traffic to/from those addresses. A country-specific AS hegemony metric quantifies the centrality of each AS for the country by estimating the likelihood that paths to IP prefixes in the country traverse the AS. Each of these metrics reflects a critical and complementary property of Internet connectivity.

Contributions. We tackle four methodological challenges: incomplete topology data, amplified by downsampling the input (BGP) data to a specific country; bias due to available vantage points; inaccurate IP and AS geolocation; and a general lack of ground truth to validate inferences. In tackling these challenges, this work makes five contributions.

- We adapt the AS customer cone and hegemony ranking metrics to country-specific versions by limiting the input data to observed AS paths to IP addresses geolocated to a target country. We create domestic and international versions of these country-specific metrics (§4).
- We analyze the effect of substantially downsampling public BGP data on the stability of the resulting rankings, and provide a generalizable method to guide use of sample-based rankings in a country (§5).
- Through case studies of Australia, Japan, Russia, and the U.S., we show that our metrics reveal important characteristics of each country's topologically dominant ASes that global metrics cannot capture (§6).
- We demonstrate use of our metrics to analyze the effects of geopolitics on country-level AS topology: (1) the negligible observed change in carriers for Russia's international Internet traffic following Russia's February 2022 invasion of Ukraine (§7.1); (2) Taiwan's clear self-reliance in its Internet connectivity and surprising independence from Chinese ISPs (§7.2); and (3) the predominant influence of the U.S. on the global Internet through U.S. transnational ISPs (§7.3).
- We provide a public dataset with the country-inferred AS Rankings, set of AS paths used as input for the inferences, collector geolocations, and Internet eXchange Point (IXP) data.

2 AS TOPOLOGY METRICS

AS topology metrics generally use BGP routing updates and tables collected and published by the Routeviews [34] and RIPE RIS [32] projects. Each collector receives routes from many different vantage points (VPs), which are individual network devices (BGP peers) within an AS. Every BGP announcement that a VP shares with a collector includes an IP prefix, the sequence of ASes that the announcement traversed to reach the VP (*AS path*), and the VP's IP address. While incomplete, these collections provide the best available public global view of the Internet's topology. The two best-known rankings—CAIDA's AS customer cone ranking (§2.1) and IIJ's AS Hegemony (§2.2)—use these public data sources to rank ASes according to their connectivity relationships with other ASes in the topology. For each of these two metrics (AH and CC), we create two country-level versions—national and international (§4.2)—that reflect the relative importance of an AS for reaching address space in a country. The *national* views (CCN and AHN) capture how a country reaches itself and the *international* views (CCI and AHI) capture how the outside world reaches the country.

2.1 Customer Cone Metric (CC)

At an AS-level, a *customer cone* (CC) is the set of ASes that an AS can reach using only customer relationships, as opposed to peer and provider relationships. The customer cone reflects the scope of an AS's transit service; i.e., the ASes (or alternatively, prefixes or IP addresses) reachable through all of an AS's observed customer paths (§2).

In 2013, Luckie *et al.* [30] proposed CC as a framework for comparing large Internet transit providers based on their customer footprint, as part of a broader algorithm to infer business relationships (customer/provider or peering) between ASes. The method for inferring the customer cone is predicated on the valley-free assumption that ASes do not generally provide transit for free, so customer prefixes are the only prefixes an AS will propagate to peers and providers. The core of the Luckie *et al.* method is identifying a peer↔peer or provider→customer relationship in a BGP AS path, and inferring that all subsequent relationships in the path are provider→customer in accordance with the valley-free assumption (Figure 1). To avoid inflating cone size, CC does not compute the cone recursively from observed provider→customer links, but only includes AS B in AS A's customer cone if an observed path

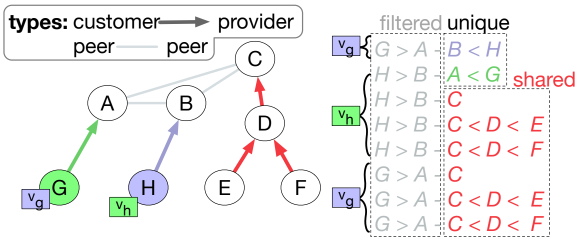
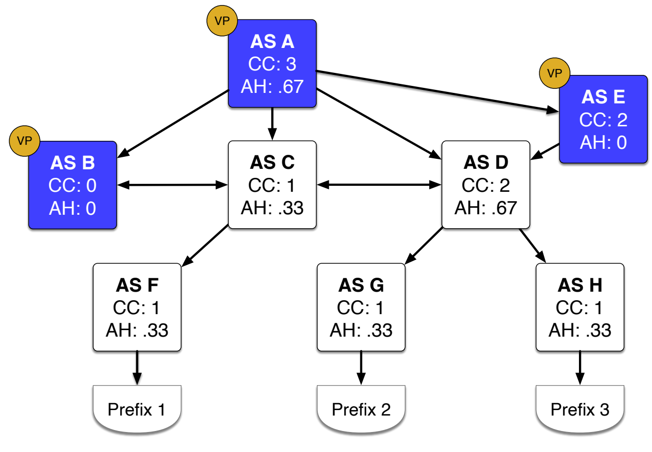


Figure 1: Computing an AS's customer cone requires examining the paths the AS announces to its peers and providers, and for each path removing the segment up to and including the first (inferred) peer↔peer or provider-to-customer link, then using the remaining customer-to-provider links to infer the customer cone [8]. Above, C is a provider of D (C<D). D is a provider of E (D<E) and F (D<F). A, B, and C are peers of each other (A-B,A-C,B-C). A is a provider of G (A<G) and B is a provider of H (B<H). The algorithm drops the (inferred) customer-to-provider segments of each path (gray). Of the remaining path segments, both VPs v_g and v_h share visibility of C<D<E and C<D<F (red). Each VP contributes only a single other segment: B<H from v_g (blue) and A<G from v_h (green).



	A	B	C	D	E	F	G	H
A	~~1~~	~~0~~	~~.33~~	~~.67~~	~~0~~	~~.33~~	~~.33~~	~~.33~~
B	.67	~~1~~	~~.33~~	~~.67~~	0	~~.33~~	~~.33~~	~~.33~~
E	~~.33~~	0	.33	.67	~~1~~	.33	.33	.33
	.67	0	.33	.67	0	.33	.33	.33

Figure 2: Contrasting the calculations for Customer Cone (CC) and AS Hegemony (AH) for an AS topology with provider-to-customer (→) and peer-to-peer (↔) relationships. For simplicity, all prefixes contain the same number of IP addresses. The CC calculation sums the number of addresses reached through provider-to-customer edges on observed paths. The AS Hegemony (AH) calculation first computes for each VP the fraction of advertised AS paths to a set of addresses where the path contains the AS; e.g., AS A receives the three scores 1, 0.67, and 0.33. (2) AH then averages the per-VP scores after removing the top and bottom 10%. For AS A, this leaves only the 0.67 score.

contained B downstream from A. An AS might announce a subset of its paths to a provider (a *complex relationship* [23]), so including all of a customer's ASes in the provider's cone can inflate its size. CAIDA's AS Rank [9] uses this algorithm.

A prefix-level CC approximates the IP addresses that an AS can reach through its customers (Figure 2). The prefix CC for an AS includes every prefix that an AS in its customer cone announced into BGP. We adapt the idea of a prefix CC to create country-specific CCs that capture how much of a country's address space is downstream of each transit provider (§4).

2.2 AS Hegemony metric (AH)

The *AS Hegemony (AH)* metric estimates the likelihood that an AS lies on a path toward a set of prefixes. In 2017, Fontugne *et al.* [20] designed this metric to approximate the *betweenness centrality* of an AS; i.e., the fraction of observed paths crossing that AS, weighted by the number of addresses reached by each path (Figure 2). Unlike CC, AH incorporates peer-to-peer and customer-to-provider portions of the AS paths in its computation. AH uses a two-step approach to reduce the bias toward ASes very near or very far from VPs. (1) For each VP it computes the betweenness centrality of all ASes (seen across all VPs) using only the paths collected by the VP. This produces multiple betweenness centrality values for each AS, where high and low values are typically obtained with VPs that are respectively close to and far from the AS. (2) The final AH score of an AS is obtained by removing the highest and lowest 10% betweenness values and averaging the remaining values. We adapt this metric by restricting the input to AS paths for prefixes geolocated in the target country, which allows us to rank ASes based on how much of a country's address space (not necessarily exclusively) is reached through each AS.

2.2.1 IHR Country-level network dependency (AHC). Since late 2020, IIJ has included a country-level hegemony ranking in their Internet Health Report (IHR) [19, 27]. We call this metric *AHC* to distinguish it from our four metrics described above (AHI, AHN, CCI, CCN). AHC is a simplified approximation to a country-level ranking, which relies on AH values computed daily per origin AS, and average values across all ASes located in the same country [19]. This averaging process includes two weighting schemes: AH values are either weighted by the number of origin ASes or number of users per AS as estimated by APNIC [3]. Since our focus is on infrastructure that serves a country rather than its population per se, we use the metric weighted by the number of ASes (§6.1.1). The primary computational difference with the original AH metric is that the AHC metric disregards AS size, i.e., it weights all ASes of a country equally.

Our approach (§4) has three notable differences from IHR's current computation of a country-level hegemony value (AHC). (1) We use a more precise set of AS paths to produce results better tailored to a given country. Our metrics include only AS paths to *prefixes* geolocated in a country, whereas IHR includes all paths to *ASes* registered in a country, regardless of where the AS originates its prefixes. (2) We clean and filter the set of input paths to reduce the effects of poisoned and looped paths on our calculations. IHR uses all available paths. (3) We introduce two views—a national (domestic) view that uses only VPs within the country, and an international view that uses only VPs outside the country—reflecting the reality that different ASes might dominate a country's international and domestic Internet paths.

2.3 Other AS Ranking metrics

In 2013, Wagner *et al.* [7] ranked ASes according to their BGP-observable provision of transit service to suspicious actors. They integrated a previously developed metric [17] that ranked ASes according to the fraction of the AS' own addresses associated with a blacklist, i.e., IP addresses inferred or reported as suspicious.

In 2018, Nur *et al.* [2] ranked ASes based on their *IP spatial path stress* metric, which estimates traffic crossing a path by assuming that every IP address originated by the first AS in the path sends traffic to every IP address originated by the last AS in the path. Specifically, they used the observed AS paths to construct a topology, but then ignored those observed paths, and assumed shortest paths across the topology, weighting the paths by number of addresses BGP-originated by the source and destination ASes. They then used these weighted paths to calculate betweenness.

IP spatial path stress, CC, and AH identify large transit providers, but do not capture the rich connectivity of many content providers. Carisimo *et al.* [10] used *k-core* graphs to track the evolution of CDNs into the Internet core. Arnold *et al.* [4] proposed the hierarchy-free reachability metric to compare the number of prefixes reachable through an AS's customers or peers, using traceroute-observed AS interconnections to supplement BGP-inferred AS relationships.

Most recently, Gamero-Garrido *et al.* [21] combined some of the filtering aspects of the CC and AH metrics to estimate the fraction of a country's IP address space that depends on a given AS for international transit. Their *Country-level Transit Influence (CTI)* metric is a modified betweenness value, similar to AH but considering transit only. The metric computes, for each AS and country, the fraction of paths from an external VP to prefixes originated in the country that use that AS *for transit.* When scoring an individual AS on a path, CTI scores by the number of addresses in the path's prefix, weighted in reverse path order (0, 1/1, 1/2, .., 1/k) where k is the number of ASes from the origin. This results in lower CTI scores relative to CC and AH, all else equal, for ASes originating large prefixes (AOLP), and higher scores for ASes adjacent to those AOLPs. Thus CTI inflates scores (vs. CC and AH) for ASes directly adjacent to the edge AS, but not for the edge AS itself. Furthermore, CTI considers only the transit portion of the AS path, i.e., provider→customer AS links (similar to CC). Thus, all else equal, ASes who get much of their country's connectivity through peering will have a higher AH than CTI or CC score. Similar to AH, the CTI approach tries to mitigate VP-bias by dropping the top 10% and bottom 10% VP scores for each AS. The combined effect of these differences is that for a given AS and country, the CTI score will generally fall between CC and AH, but sacrifices the distinction properties of the two metrics capture.

3 BGP TOPOLOGY DATA CHALLENGES

Creating country-specific metrics amplifies preexisting challenges for inferring topological properties from BGP announcements: incomplete and biased topology data; inaccurate IP and AS geolocation; and a lack of ground truth to validate inferred rankings.

Incomplete and biased public BGP data. A VP learns only the subset of paths that its BGP-peering routers choose as their best path to each prefix, leaving the VP with an incomplete view of the full AS-level topology. No single router, nor even a large subset of routers, has enough information to build the full AS-level topology, so even the large set of VPs provided by the public route collector projects [32, 34] observe only a subset of available AS paths. For instance, region-specific announcements are only visible within a local region of the router that announced them. Thus any single BGP peer (VP) will have a set of AS paths that favor ASes topologically and geographically close to the peer. Restricting a set of VPs to those within a single country can amplify this sampling problem, since many countries have only a few VPs, and most countries have none at all. ASes with an abundance of VPs could also distort the results, although our April 2021 dataset does not have this issue (Figure 10). We develop methods for building both national (domestic) and international (from outside the country) views for each country (§4.2) that produce stable outputs despite changes in the available set of VPs and AS paths.

Inaccurate IP address geolocation. Our approach for mitigating topological and geographic bias requires geolocating prefixes and VPs to construct views of the topology specific to individual countries. Commercial services typically accurately geolocate end host IP addresses at the country granularity; we rely on such a service for prefix geolocation (§4.2.1). Accurately geolocating infrastructure IP addresses, such as the VPs, remains a long-standing challenge [22, 26]. Knowing the collector's location is insufficient, since some VP ASes peer remotely with collectors. As such we only use VPs at IXPs that do not support multi-hop (§4.2.2).

Lack of ground truth. Validating our inferences presents another challenge due to a lack of available ground truth. We validate our inferences through two evaluations: (1) we assess the stability of our inferences as we artificially vary the set of available VPs, ensuring that the metrics capture properties of the underlying topology rather than properties specific to a given sample of VPs (§5); and (2) we explore the efficacy of our inferences by comparing them with external economic knowledge about the countries and their telecommunications infrastructure, including earnings and subscription numbers in corporate annual reports (§6).

4 CONSTRUCTING GEO-SPECIFIC RANKINGS

We prepare the input data to create country-specific AS rankings in two phases. First, we sanitize AS paths, and geolocate IP prefixes and VPs (§4.1). Second, we create *national* and *international* views for each country that capture a country's connectivity (§4.2). This allows us to provide the four country-specific ranking metrics: *AS Hegemony National (AHN), Customer Cone National (CCN), AS Hegemony International (AHI),* and *Customer Cone International (CCI).* Finally, in §4.3 we will show how and why these metrics rank ASes differently from the existing Customer Cone Global (CCG),

rejected	74M	30.13%
unstable not seen across all five days	20M	8.06%
unallocated unassigned AS	229K	0.09%
loop nonadjacent duplicates (A C A)	202K	0.08%
poisoned non-top-tier AS between top-tier ASes	10K	0.00%
VP no location VP at multi-hop IX	52M	20.98%
prefix no location geolocated to no or mulitiple countries	2M	0.91%
accepted	173M	69.87%
total	248M	100.00%

Table 1: Filtering paths from our April 2021 data.

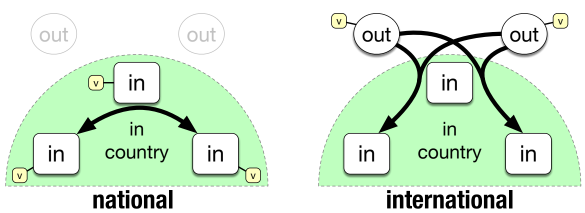


Figure 3: A *national* view captures the set of paths that a country uses to reach itself. The *international* view captures the set of paths the rest of the world uses to reach the country.

AS hegemony Global (AHG), and IHR's current country-level AS Hegemony (AHC) (§6.1.1) metrics.

4.1 Sanitizing and Geolocating Input Data

The first step is preparing the AS paths seen in the BGP announcements for our algorithms and geolocating the advertised IP prefixes. Each BGP announcement recorded by a collector contains three important pieces of information for our method—the VP's IP address, the AS path traversed by the announcement, and the advertised prefix. We use the data curation process depicted in Table 1. We take 5 RIBs, across the first five days of the month, and discard prefixes not seen in all 5 RIBs. We also discard paths that contain AS loops, appear to be poisoned, or include ASes that IANA reports as unassigned. Path poisoning occurs when an AS inserts other ASes into a path to prevent its selection. As in [30], we infer poisoning when two clique ASes are separated by a non-clique AS. In the process of preparing to geolocate prefixes, we remove all prefixes (and paths to them) that are entirely covered by more specific prefixes. We then remove AS paths from VPs or to prefixes that we cannot geolocate (§4.2.1). We clean the remaining AS paths by removing IXP route server ASes and duplicate adjacent ASes.

4.2 Building national vs international views for each metric

For a given country, we create *national* and *international* views. National views use only announcements from VPs located in the country (§4.2.2), capturing a domestic view of that country's AS-level topology. International views use only the remaining BGP

		ASes		prefixes		VPs	
type	**metric**	in	out	in	out	in	out
nation.	AHN,CCN			X		X	
inter.	AHI,CCI			X			X
IHR cty l.	AHC*	X				X	X
global	AHG*			X	X	X	X
	CCG+			X	X	X	X

*Internet Health Report +AS Rank

Table 2: A country's *national* view uses paths from *in*-country VPs to *in*-country prefixes. A country's *international* view uses paths from *out*-of-country VPs to *in*-country prefixes. The table contrasts these views with previous metrics (bottom 3 rows of table) in terms of AS path input data. In particular, IHR's simplified *country-level* AS hegemony (AHC) uses AS paths from all VPs to *in*-country ASes, without consideration of where the AS originates its prefixes (§2.2.1). The *global* views of both customer cone and hegemony metrics use AS paths from all VPs and to all prefixes.

announcements – from VPs outside the country toward destination prefixes inside the country. This view allow us to examine how the rest of the world reaches the country. Creating these views requires geolocating both the IP prefixes announced into BGP and the VPs.

4.2.1 Geolocating Prefixes. Prefix usage might not align with country boundaries; i.e., different IP addresses within a prefix might be assigned to infrastructure in different countries. Before we geolocate the prefixes, we split them into non-overlapping blocks of addresses mapped to their most specific prefix. We then filter (remove from our data set) prefixes that are completely covered by more specifics (1.2% of our April 2021 data set). We then use NetAcuity to geolocate the addresses in the blocks. We filter prefixes for which fewer than 50% of the total addresses from its blocks do not geolocate to the same country (0.2% of prefixes, that collectively held 1.5% of the IP addresses in our data). We ignored the challenge of geolocating anycast prefixes; Sommese et al. recently inferred fewer than 0.4% of /24s were anycast [40].

4.2.2 Geolocating VPs. To geolocate VPs, we first download the reported location of each collector [33, 35, 36] (usually in the form of an IXP location). We divide this list into VPs that peer with collectors labeled as multi-hop and those labeled non-multi-hop. Multi-hop collectors peer with VPs that remotely connect to the IXP [5, 11], and thus the VP could be in a different country from the collector. We do not attempt to geolocate these VPs, and exclude all paths from those VPs in our country-level metrics. In our data from April 2021, we successfully geolocated 806 (91%) VPs to their collector's location, and excluded 74 (8%) multi-hop VPs.

4.3 How the Four Metrics Capture Different Properties

Each of our four metrics—CCI, CCN, AHI, and AHN—may produce different rank orders for a country's ASes, since they capture different properties of a country's connectivity. Both CC metrics try to capture the most important ASes in terms of the traditional Internet transit hierarchy, while the AH metrics estimate which ASes are most often traversed to a given set of destinations, regardless of the nature of the interconnection relationships on the path. ASes at the top of the traditional Internet provider-customer hierarchy reflect a degree of market power, since customers depend on (and pay) them for transit, rather than the other way around. In contrast, the AH metrics capture a different property: the ability of an AS to capture traffic to a certain set of address space. Finally, as described in §4.2, the national and international views capture internal vs external views of these two properties of a country's connectivity, respectively.

We note two caveats to all of these metrics. First, public BGP data reveals only visible portions of topology, and backup paths may only reveal themselves after primary paths are lost. Second, none of these topology metrics necessarily reflect traffic volumes. We are not aware of an accurate technique for inferring traffic volumes.

The next two sections evaluate the metrics by focusing on the *top-ranked ASes (TRAs)*, those considered most important by each metric. In §5, we measure the stability of the rankings of inferred TRAs to ensure that they appear to capture properties of the actual topology rather than sampling artifacts, i.e., properties unique to the VPs included in our samples. Once we establish stability of the metrics, we examine the efficacy of the TRAs by comparing them to expectations based on knowledge of the telecommunications industry and corporate financial reports §6.

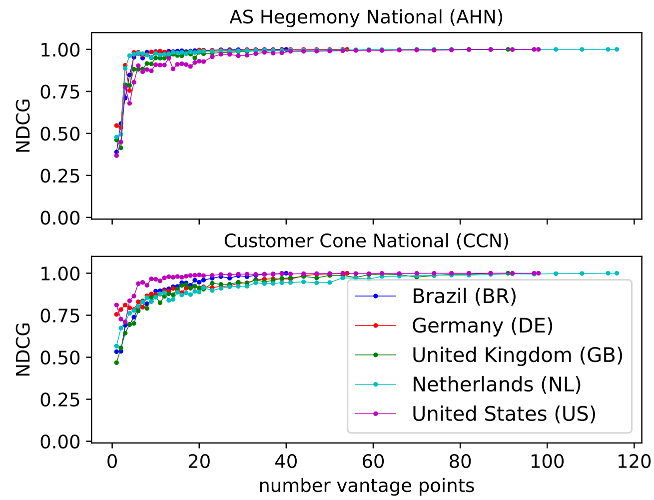
5 EVALUATING RANKING STABILITY

Metrics that capture a property of the underlying topology, rather then a property unique to the sampled topology, should produce a stable ranking despite changes to the VPs included in the sample. To test this stability, we examined the effect of removing VPs from the available set of VPs (*downsampling*) on the CC and AH rankings. If we find that the TRA for a country remains stable once we include at least k VPs, then we conclude the TRA inferred is independent of the specific set of VPs. This evaluation also reveals the minimum number of VPs that produced stable TRAs for our metrics, suggesting opportunities for focused deployment of new VPs to improve the ability to assess a country's Internet connectivity.

5.1 Methodology to compare TRAs: Normalized Discounted Cumulative Gain (NDCG)

To assess stability, we compare sample-based TRAs to the TRA produced by the full set of VPs. We construct each sample by removing VPs (and their associated paths) from the data set, and compute each metric using only the AS paths remaining in the sample.

We use the *Normalized Discounted Cumulative Gain (NDCG)* metric to evaluate whether the sample-based TRA contains an equivalent relative ordering to the rank ordering inferred from the full set of VPs, i.e., without sampling. This metric is widely used to evaluate recommender systems, which are effectively learning-to-rank systems [6]. DCGs have been applied to subset ranking tasks [15] such as page ranking and movie recommendation systems. In subset ranking, the objective is to learn a ranking function that approximates the ideal partial ordering of a set of objects based on limited data, and the focus is on the quality of estimates in the top-portion of the ranked list. In order to provide a single value for comparison, we normalized the DCG (NDCG) by the ranking



NDCG	1.00	0.99	0.98	0.95	0.90	0.85	0.80	
CCN	114	87	74	53	25	15	9	VPs
AHN	116	50	40	23	19	6	6	

required to achieve a given NDCG threshold for each metric given available VPs in the five countries.

Figure 4: AHN (top) and CCN (bottom) metrics maintained a relatively stable ranking (NDCG >= .8) with 9 and 6 VPs respectively. AHN was more stable, requiring fewer VPs to achieve the same NDCG.

inferred using all VPs for that view (a non-sampled view, which we call the *full DCG (FDCG)*, The DCG and normalized DCG (NDCG) are defined as:

$$DCG_p = \sum_{p=1}^{10} \frac{rel_p}{log_2(p+1)} \quad (1) \qquad NDCG_p = \frac{DCG_p}{FDCG_p} \quad (2)$$

where p is the object's rank and rel_p is its relevance, which in our case corresponds to the AS's customer cone or hegemony value. The NDCG relies on a fixed set size, so we use the top 10 ASes as the TRA for our analysis.

5.2 Stability of Sampled TRAs

We assessed the stability of national and international views using our April 2021 set of AS paths.

	NL	UK	US	DE	BR
number VPs	141	105	101	73	46

Table 3: These five countries had the highest number of in-country VPs, so we used them to analyze stability of the national-view rankings (CCN and AHN).

5.2.1 *National TRAs.*

In order to explore a large sampling space we need views with many VPs, so that we can create many sample views using subsets of the VPs. For international views this is not a problem, but constructing a national view requires all VPs to be in-country, and only a few countries have enough VPs to support systematic sampling. We computed the sample-based NDCG for the five countries with the most in-country VPs: the Netherlands (141), United Kingdom (105), United States (101), Germany (73), and Brazil (46) (Table 3). Figure 4 shows the NDCG for AHN (hegemony) and CCN (customer cone) for these four countries. The AHN and CCN metrics required 25 and 19 in-country VPs, respectively, to achieve an NDCG above .9; i.e., the ranking was 90% similar to the ranking using all VPs. An NDCG of .8 was possible with only 9 and 6 VPs, respectively. These results indicate a correlation between increased VP deployment and higher fidelity AS rankings.

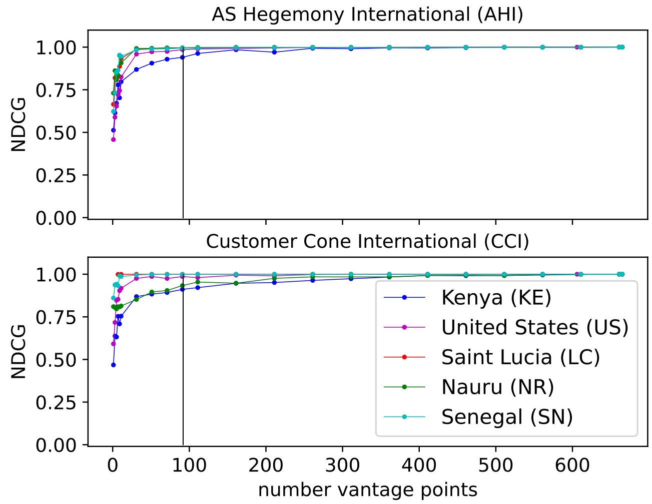
Given this threshold, a researcher wishing to estimate AHN or CCN with an NDCG of 0.9 would have enough VPs in 12 and 13 countries, respectively (§4). This limitation (discussed further in §8) does not apply to international TRAs.

	VP				
	IPs	**ASNs**	**ASNs**	**prefixes**	**addresses**
NL	141	130	1578	10.5 k	40.4 m
GB	105	91	2810	17.2 k	83.8 m
US	101	75	19850	230.2 k	1062.1 m
DE	73	70	2703	20.8 k	122.0 m
BR	46	39	8330	72.5 k	113.9 m
CH	45	41	887	4.0 k	15.0 m
ZA	44	41	642	8.9 k	25.9 m
AT	41	38	690	3.0 k	10.5 m
SG	38	35	778	6.2 k	10.1 m
IT	36	36	1158	14.5 k	55.0 m
FR	35	33	1544	10.7 k	72.9 m
AU	25	18	2060	24.0 k	44.7 m
SE	21	21	750	5.0 k	26.3 m
RU	18	18	5428	40.3 k	43.8 m
ES	14	14	1171	17.5 k	33.7 m
JP	7	7	949	13.2 k	190.6 m

Table 4: Countries with > 7 in-country VPs in our data.

5.2.2 International TRAs. We examined the stability of AH and CC for sample-based international TRAs (AHI and CCI) compared to full international TRAs. As mentioned in §4.2, the international view for a country includes the paths from VPs outside the country to prefixes located inside the country. While most countries do not have enough public in-country VPs to provide stable national rankings, all countries have enough out-of-country VPs for an international ranking.

Figure 5 shows that both AHI and CCI had stable TRAs (≥ 90% NDCG score) in April 2021 for samples that included at least 91 out-of-country VPs. We surmise that international views require more VPs than national views due to the variety of transit providers operating in different regions of the world. This result implies that we can compute a stable international TRA for each country, since the minimum number of VPs available for an individual country (665 VPs for the Netherlands) is more than sufficient. Unlike the national rankings, which require substantial downsampling of the input AS path data, the international metrics retain most of the AS paths, allowing stable rankings using public data.



NDCG	1.00	0.99	0.98	0.95	0.90	0.85	0.80	
CCI	661	411	411	211	91	31	31	**VPs**
AHI	665	261	161	161	51	51	51	

required to achieve a given NDCG threshold

Figure 5: Given at least 411 vantage points (VPs) outside of a given country (which all countries had), both metrics (CCI and AHI) created stable rankings (§5.2.2).

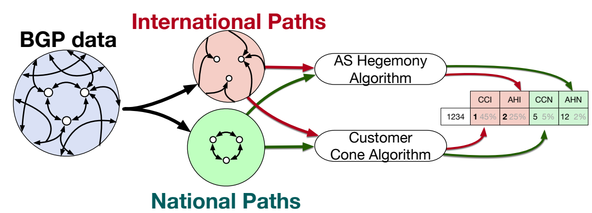


Figure 6: Pipeline to generate tables in §6.

6 EVALUATING RANKING EFFICACY

Producing a stable TRA does not necessarily mean that the TRA has accurately identified the most important ASes. We evaluated the efficacy of the TRA membership by correlating ASes at the top of the rankings with knowledge of a country's telecommunications industry and information contained in annual reports to shareholders. Our analysis confirmed that the ASes ranked highest by our metrics are also considered the most important within those countries by conventional industry metrics.

We examined our four metrics—AS Hegemony national (AHN) and international (AHI), as well as Customer Cone national (CCN) and international (CCI)—in case studies of Australia, Japan, Russia, and the U.S. These countries all had a sufficient number of in-country VPs (at least 7) to generate stable national views. Each case study includes a table containing the top two ASes according to each metric. The tables also indicate if an AS appears in the top ten ASes of the Customer Cone Global (CCG) ranking; e.g., the subscript for AS1299 in Table 5 indicates that it ranked second in CCG.

6.1 Australia

Australia's dominant national transit provider, Telstra, is government-backed (formerly Australia's telecommunication monopoly) and heavily capitalized. Annual reports indicate that Telstra earned $AU 25B dollars of revenue in 2021 [25], compared with just under $AU 9B for its nearest domestic rival SingTel Optus [24], which ranked 3rd and 5th by CCN and AHN respectively.

Australia		CCI	AHI	CCN	AHN
1221	Telstra	7 44%	**1** 40%	**2** 41%	**1** 23%
4826	Vocus	**2** 81%	8 6%	**1** 80%	**2** 16%
1299	Arelion	**1** 83%	10 5%	12 5%	101 0%
4637	Telstra I.	6 49%	**2** 39%	55 0%	140 0%
7474	SingTel O.	12 28%	12 3%	3 26%	5 10%

Table 5: Top two ASes (rank in bold) for each metric in Australia (AU), and corresponding % of AU's address space (CC) or % of paths observed to AU's address space (AH). Arelion's subscript ($_2$) indicates it has the second largest *global* customer cone. (§6.1.)

Telstra's dominance is reflected in the AHI and AHN rankings for Australia. Both AHI and AHN rank Telstra's AS 1221 at the top, i.e., the most important for intra-country communication. But the 2nd highest ranked AS by the AHI metric is quite different. Telstra's AS 4637 (used for international connectivity) has an AHI score barely behind AS 1221 (the domestic backbone AS). But its AHN score is approximately 0, suggesting low domestic use, i.e., for paths with both endpoints in Australia.

The CCI metric contrasts with the AS hegemony rankings, in that CCI prefers the large ISPs that provide Internet transit to Australia. CCI ranks the global ISP Arelion (1299, formerly Telia) first, with the large domestic ISP—and Arelion customer—Vocus (4826) a close second. As intended, customer cone highlights the importance of Vocus for both international and domestic paths to Australia: Vocus provides transit to ≈ 80% of Australia's IP address space. A side effect of the customer cone metric is that it tends to inflate the ranking of large providers (Figure 1). For example, while Arelion appears important to Australia's international paths, since paths following the Internet's traditional hierarchy would transit Arelion to reach Vocus, most of the ranking comes from transitively including Vocus' customer cone in Arelion's cone.

Considering both customer cone and AS hegemony metrics reveals a richer view of AS importance. An AS that is highly ranked by customer cone and AS hegemony appears on hierarchical paths along provider-customer relationships, and lies along paths that other ASes frequently select as best paths. Both metrics rank Vocus and Telstra highly, consistent with the importance of both ASes in Australia.

6.1.1 Inadequacy of global rankings metrics. Now that we have reviewed our country-specific rankings for a specific example, we can use this example to demonstrate that global rankings do not capture these properties. Table 9 illustrates how trying to cull the subset of Australian ISPs from global rankings would yield an inaccurate ranking. Above we showed that Telstra's domestic AS (1221) and Vocus (4826) are critical to Australia's domestic topology, but both CCG and AHG would have ranked Telstra's international AS (4637) above them. Furthermore, filtering out the non-Australian ASes would disregard the critical role that multi-national ISPs (grey font in table 9) play in the country's international connectivity. We can also use our metrics to explore dominant networks across multi-country regions (§7.3).

6.1.2 Limitations of previous country-level ranking metrics. We can also now illuminate why a finer-grained analysis yields more insight than that provided by the existing country-specific AS hegemony rankings (§2.2.1). IHR's country-level ranking (AHC) does include both domestic and multi-national ISPs but it does not distinguish the roles of these networks. For example, in Table 9, AHC's six highest-ranked ASes consist of the top 4 ASes ranked by AHI and the top 2 ASes ranked by AHN, confounding the facts that Telstra's domestic AS (1221) and Vocus (4826) are important domestic networks, while multi-national ISPs (Telstra's AS 4637 and Hurricane's AS 6939) provide substantial international connectivity. Another noticeable difference is the presence of Amazon (16509) in Australia's AHN ranking but not in the AHC ranking. Amazon originates many prefixes in Australia, but Amazon's AS 16509 is registered in the U.S. Our AHN metric captures fine-grained prefix geolocation, but the original IHR country-level AHC estimate did not, which explains Amazon's presence in our ranking but not in AHC's.

6.2 Japan

Japan		CCI	AHI	CCN	AHN
2516	KDDI	4 50%	**2** 21%	**1** 28%	**1** 29%
2914	NTT A.	**1** 87%	**1** 25%	8 5%	20 1%
17676	Soft.	6 30%	3 13%	**2** 27%	3 27%
4713	NTT O.	11 22%	5 9%	3 22%	**2** 28%
3257	GTT	**2** 56%	23 1%	123 0%	236 0%

Table 6: Top 2 ASes by each metric in Japan (April 2021 data) (§6.2.)

Like Australia, the former government-owned telecommunication monopoly in Japan (NTT) remains the largest telecommunication company by revenue. In 2021, NTT had annual revenue of 8.9 trillion yen [41] compared to its nearest competitor KDDI (also with state-owned roots), which earned 0.651 trillion yen [16]. NTT operates one of the largest networks in Japan (AS 4713), and bought AS 2914 as part of its acquisition of Verio in 2000. Computing the global customer cone (CCG) for AS paths from that time ranked AS 2914 12th and AS 4713 97th. AS 2914 became NTT America to handle American and international markets, while NTT OCN (AS 4713) provided domestic transit in Japan, a division that continues today. This merger helped NTT become one of the largest Internet transit providers in the world.

NTT ranked in the top three according to all four country-specific metrics, reflecting its importance to the Japanese Internet. NTT America (2914) ranked first for both international metrics, and NTT OCN (4713) ranked second or third in the national rankings. NTT America's CCI score contained more than twice as much of Japan's address space as its AHN score. Represented as a percentage of paths to Japan's address space, it was 87% compared to 34%. Notably, Japan's top two CCI-ranked ASes were also in the top six CCG-ranked ASes: NTT (2914) was 5th and GTT (3257) 3rd.

Their high rankings are consistent with the trend we see elsewhere, where one or two of the ten largest multinational ASes dominates international transit for the country.

6.3 Russia

Russia			CCI	AHI	CCN	AHN
12389	Rostele		7 60%	**1** 32%	3 48%	**1** 20%
1273	Vodafone		5 68%	53 0%	**1** 58%	10 2%
3356	Lumen		**1** 97%	7 6%	30 2%	21 1%
1299	Arelion		**2** 86%	3 11%	4 32%	85 0%
9049	ER-Tele.		20 17%	**2** 11%	17 13%	4 5%
20485	TransTel		6 62%	5 7%	**2** 51%	7 3%
8359	MTS PJSC		19 17%	8 6%	14 15%	**2** 7%

Table 7: Top 2 ASes for each metric in Russia for April 2021 data (§6.3).

Rostelecom (AS 12389), ranked first by AHI and AHN, is majority-owned by the Russian government. In 2020, Rostelecom earned the most revenue among telecommunications companies in Russia: 547B rubles compared to the second-largest company by revenue (MTS Russia at 388B rubles [14]). Our AH metrics capture Rostelecom's importance to Russia's Internet. CC also ranked Rostelecom highly, behind only large multinational ISPs. Only the AHN metric ranked Russian domestic ISP MTS (AS 8359) near the top (2nd). Similar to other countries, multinationals such as Arelion (1299) play a major role in Russia's international connectivity, despite Russia stated goal in 2019 to decrease Russian dependence on non Russian actors [42]. (See also §7.1).

6.4 United States

United States			CCI	AHI	CCN	AHN
3356	Lumen		**1** 64%	**2** 15%	**1** 46%	**1** 11%
6939	Hurricane		9 19%	**1** 18%	11 17%	3 7%
1299	Arelion		3 35%	7 4%	**2** 34%	12 2%
7018	AT&T		7 22%	4 12%	6 22%	**2** 8%
3257	GTT		**2** 39%	17 2%	7 22%	22 1%

Table 8: Top 2 ASes for each metric in U.S. for April 2021 data set. Hurricane (AS 6939) is known to engage in a high degree of international peering (§6.4.)

Unsurprisingly, Lumen's primary AS 3356 dominated U.S. rankings. AS 3356's customer cone contained 64% or 46% of address space that geolocated to the U.S., when considering non-U.S. (CCI) or only U.S. (CCN) VPs, respectively. With respect to AS Hegemony, the international (AHI) and national (AHN) metrics showed that 15% or 11% of the AS paths to U.S. address space included AS 3356. The U.S. is the only country in our case studies where the AHN's top-ranked AS is a large multi-national, not surprising given the United States is Lumen's home market. Our data indicated that Lumen is the dominant international carrier in the world, and also the dominant carrier in the U.S. domestic market, appearing first in all U.S. rankings except AHI. Hurricane's top AHI rank derives from its famously liberal peering policy.

Internationally, the highest ranked non-U.S. autonomous system was Arelion (1299) in third place. AS 1299 has the second-largest *global* customer cone (CCG). Unlike other countries where the richest telecommunication firm is also the dominant Internet player, Lumen had annual revenue in 2021 of US$19B [1] compared to AHN's 4th-ranked AT&T's with 2021 revenue of US$168B [13]. However, U.S. networks are typically quite horizontally and vertically integrated with other profitable lines of business. We are less confident using aggregate revenue as calibration in the U.S.

Notably, the top-ranked ASes in the U.S. by all metrics are neither government-backed nor the most heavily capitalized. But AT&T has its roots as a monopoly telecom provider in the U.S., and has high national rankings similar to such providers in other countries. Interestingly, the prefix coverage percentage values of all metrics are lower in Table 8, suggesting a less concentrated U.S. market.

6.5 Summary of country-specific analysis

The topologically dominant AS in CCI rankings tended to be a large multi-national transit provider. Domestic carriers ranked more highly in the CCN rankings, consistent with their role, and with the goal of this metric. In contrast, the AS Hegemony metrics tend to rank large national carriers above the multinational carriers. Also note that the largest national carrier will often have two ASes, one for international and one for domestic transit. We can infer these roles from AS hegemony rankings: the international AS will rank higher in AHI, and the national AS will rank higher in AHN. This large national carrier is normally a current or former government-run utility, often with the largest annual telecommunications revenue in the country. The U.S. is an exception; its dominant transit provider (Lumen) is not a former government monopoly, nor does it use use different ASes for international and national transit. Lumen started out as a small telephone company but has experienced tremendous consolidation over the last two decades via mergers and acquisitions.

7 ANALYZING TEMPORAL AND SPATIAL INSIGHTS OF METRICS

After gaining confidence in the stability and efficacy of our country-specific rankings, we use them to analyze important ASes in three case studies. First, we find negligible changes in Russia's top-ranked ASes following the Russian military's invasion of Ukraine and associated economic sanctions. Second, we show that Taiwan's Internet is observably independent from China. Third, we examine AS rankings for each continent, finding that U.S. ASes continue to be the most prominent networks around the world. (Note that our metrics capture properties of transit and core peering infrastructure, so the hypergiant CDNs, though clearly important edge networks, will not generally dominate these rankings.)

7.1 Effect of Russian Invasion of Ukraine

In the immediate aftermath of Russia's invasion of Ukraine in February 2022, there was considerable pressure to disconnect from or stop supporting Russian industry [38]. In March 2022, Lumen and Cogent announced plans to stop doing business in Russia [31, 39]. Importantly, they did not commit to stop connecting with Russian ISPs outside of Russia. We compared the top-ranked ASes for Russia by CCI and AHI in April 2021 and March 2023, and observed modest effects of political events on these metrics (Table 10). Cogent

Customer Cone (CC)			AS Hegemony (AH)				
CCI	CCG*		AHI	AHG+	AHC+	AHN	
1	2	🇸🇪 1299 Arelion	**1**	131	**6**	**1**	🇦🇺 **1221 Telstra**
2	36	🇦🇺 **4826 Vocus**	**2**	24	**4**	**140**	🇦🇺 **4637 Telstra**
3	8	🇺🇸 6461 Zayo	3	4	1	56	🇺🇸 6939 Hurricane
4	1	🇺🇸 3356 Lumen	**4**	94	**3**	**7**	🇦🇺 **7545 TPG**
5	6	🇺🇸 3257 GTT	5	41	7	223	🇸🇬 7473 Singapore Tel.
6	14	🇦🇺 **4637 Telstra**	6	30	-	10	🇺🇸 16509 Amazon
7	53	🇦🇺 **1221 Telstra**	**7**	1176	**1014**	**6**	🇦🇺 **4804 SingTel**
8	7	🇺🇸 6939 Hurricane	**8**	62	**2**	**2**	🇦🇺 **4826 Vocus**
9	9	🇺🇸 6453 TATA	9	9	9	317	🇺🇸 6461 Zayo
10	27	🇷🇺 3216 Vimpelcom	10	2	11	101	🇸🇪 1299 Arelion

*AS Rank ranking +Internet Health Report ranking

Table 9: Top 10 ASes in Australia by Customer Cone International (CCI) and and AS Hegemony International (AHI) (§4.2) compared with those same AS's rankings in AS Rank's CCG and Internet Health Report's AHG and AHC. Selecting the global ranking by ASes in a specific country (Australian ASes in bold) yields a different relative ranking for those ASes than our country-specific metrics yield. Our country-specific metrics capture connectivity into and within a country, more accurately reflecting the roles of international and national carriers with respect to that country. Also, our metrics operate at a prefix granularity, allowing them to capture the fact that Amazon originated some of its address space from Australia. (RouteViews and RIPE data for 1 April 2021.)

	cone							hegemony						
	20210401			20230301				20210401			20230301			
1	Lumen 3356	🇺🇸	97%	3356	🇺🇸	**0**	-1.2%	Rostele 12389	ru	32%	12389	ru	**0**	+0.5%
2	Arelion 1299	🇸🇪	86%	1299	🇸🇪	**0**	+1.5%	ER-Telecom 9049	ru	11%	1299	🇸🇪	**+1**	+2.2%
3	Telecom Ita. 6762	🇮🇹	80%	6762	🇮🇹	**0**	-1.4%	Arelion 1299	🇸🇪	11%	31133	ru	**+3**	+3.3%
4	GTT 3257	🇺🇸	70%	174	🇺🇸	**+6**	+31	Vimpel 3216	ru	9%	9049	ru	**-2**	-2.2%
5	Vodafone 1273	🇬🇧	68%	1273	🇬🇧	**0**	-1.7%	TransTel 20485	ru	7%	3356	🇺🇸	**+2**	+2.5%
6	TransTel 20485	ru	62%	12389	ru	**+1**	+1.0%	MegaFon 31133	ru	6%	8359	ru	**+2**	+1.9%
7	Rostele 12389	ru	60%	3491	🇺🇸	**+1**	-4.2%	Lumen 3356	🇺🇸	6%	3216	ru	**-3**	-1.8%
8	PCCW 3491	🇺🇸	51%	5511	🇫🇷	**+13**	+24	MTS PJSC 8359	ru	6%	9002	🇬🇧	**+1**	+2.3%
9	NTT Ame. 2914	🇺🇸	37%	9002	🇬🇧	**+3**	+2.1%	RETN 9002	🇬🇧	5%	20485	ru	**-4**	-0.1%
10	Cogent 174	🇺🇸	36%	3216	ru	**+1**	-4.9%	Vimpel 8402	ru	5%	8402	ru	**0**	-0.2%

Table 10: While a few ranks changed, Russia's dependence on foreign transit ISPs has not decreased since 2021.

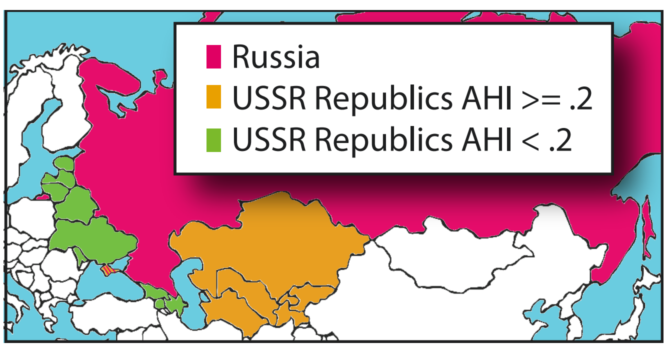


Figure 7: Russia's AHI hegemony over former Soviet block countries (April 2021)

(AS 174) jumps in CCI rank, and one AS (3257 GTT US) disappears while another (5511 Orange FR) joins the top 10. But unchanged over the last several years is the dominance of non-Russian ISPs for international transit into the country, in contrast to other countries in our case studies.

We found five countries for which Russian ASes had a significant AHI (> 20%): Turkmenistan, Russia, Tajikistan, Kazakhstan, and Kyrgyzstan. These are former Soviet bloc countries, whose networks still rely heavily on Russian network infrastructure. However, the nine Western and Central former republics do not (Figure 7). Fontugne *et al.* observed that in the three years following Russian's 2014 annexation of Crimea [19], Crimea transformed from reliance on local and Western ASNs to reliance on Russian ASNs. On the other hand, a more recent study found continued and surprising dependence of domains in the Russian Federation on non-Russian naming, hosting and certificate infrastructure [28].

7.2 Region of Taiwan

The pre-Mao government fled to Taiwan in 1949 and established self-rule. The legal status of Taiwan is disputed to this day. China views Taiwan as one of its provinces. This tenuous circumstance makes Taiwan an interesting case study. Our metrics show (Table 11) that Taiwan has achieved an impressive level of Internet infrastructure independence from China, with the CCI ranking of China Telecom (4134) dropping from 7th to 77th between April 2021 and March 2023. Additionally, 7 of 10 ISPs in the top AHI

	cone						hegemony					
	20210401			20230301			20210401			20230301		
1	Lumen 3356	🇺🇸	98%	3356	🇺🇸	**0** +1.3%	Chunghwa 3462	tw	48%	3462	tw	**0** -4.3%
2	Arelion 1299	🇸🇪	95%	1299	🇸🇪	**0** +1.7%	Data Comm. 9680	tw	23%	9680	tw	**0** -5.2%
3	Chunghwa 9505	tw	88%	9505	tw	**0** -4.5%	Digital United 4780	tw	22%	1659	tw	**+3** +8.8%
4	Chunghwa 3462	tw	87%	174	🇺🇸	**+2** +1.8%	Far EastTone 9674	tw	17%	17717	tw	**+4** +10
5	Sprint 1239	🇺🇸	80%	6461	🇺🇸	**+5** +0.1%	RETN 9002	🇬🇧	9%	4780	tw	**-2** -8.7%
6	Cogent 174	🇺🇸	80%	4637	🇦🇺	**+5** -0.4%	Education B. 1659	tw	8%	174	🇺🇸	**+4** +7.1%
7	China Tel. 4134	🇨🇳	64%	9680	tw	**+5** +0.8%	Telstra 4637	🇦🇺	8%	4637	🇦🇺	**0** +4.5%
8	Lumen 3549	🇺🇸	51%	3462	tw	**-4** -41.4%	Minstry Edu. 17717	tw	6%	9924	tw	**+1** +3.6%
9	GTT 3257	🇺🇸	49%	7018	🇺🇸	**+5** +0.8%	Taiwan Fixed 9924	tw	5%	3491	🇺🇸	**+4** +2.2%
10	Zayo 6461	🇺🇸	47%	701	🇺🇸	**+3** +0.4%	Cogent 174	🇺🇸	5%	1299	🇸🇪	**+5** +2.5%

Table 11: Taiwan and U.S. ISPs dominated Taiwan's top-ranked ASes (using the international CC (CCI) and AH (AHI) metrics) from April 2021 to March 2023. China Telecom (4134) dropped out of Taiwan's top 10 rank by CCI.

ranking were Taiwanese. It will be interesting to watch trends in these metrics as conflict continues in Ukraine, which has prompted speculative comparisons to Taiwan [18].

7.3 Dominant networks across regions

Finally, we examined the most important networks across continents. Table 12 shows the top 12 countries that have ASes who provide international connectivity for other countries; i.e. the AHI score was > 0.1. Decades after the U.S. launched the Internet [12], the U.S. continues to dominate international transit. Many globally top-ranked networks are U.S.-based with a large role in international connectivity: 76% of the world's countries are served (not exclusively) by a U.S. ISP. However, 29% and 27% of countries in Europe and Asia, respectively, did not appear to heavily depend on U.S. ISPs in our April 2021 data set. Sweden, represented by Arelion's AS 1299, observably supports many countries in Europe, North America, Asia, and Oceania.

The next three rows in Table 12 – UK, France, and Italy - have ASes that serve many African countries, consistent with historical international relations: the importance of French ISP Orange (5511) in North and West Africa (e.g.. Morocco and Ivory Coast), Telecom Italia (6762) in North Africa (e.g. Tunisia), and Liquid Telecom (30844) in previous British colonies (e.g. Kenya and Uganda). Spain's Telefonica (12956) has high AHI scores for Spanish-speaking South American countries. Africa is the continent with the most diverse set of dominant international transit providers.

The metrics also show how some ASes play an important role in neighboring countries. Table 12 offers several insights: Australian ISPs are the most dominant in Oceania; ASes from South Africa and Mauritius are dominant in multiple African countries; and Russian networks are important for former Soviet bloc countries in Central Asia (§6.3).

8 LIMITATIONS AND OPPORTUNITIES

In addition to the geolocation challenges we described (§3) and navigated (§4.2), we note other limitations of our data.

Limits of BGP data. BGP data has several limitations for our work. First, aggregation of customer prefixes may affect the accuracy of geolocation. Second, the existence of a BGP interconnection does not imply it is heavily (or at all) used §4.3. The difficulty of access to traffic data limits the ability to validate inferred correlations [37]. Third, public BGP data does not reveal backup paths, so it cannot reliably support resilience assessments. Future work could attempt to infer backup paths, perhaps using active path measurements, including historical traceroute archives. One could also use such data to weight ASes according to how frequently packets traverse them, or to inform geolocation inference.

Limited vantage points. Our national rankings suffer low-fidelity for countries with too few VPs (§5). We also removed 20% of paths in our data set because they came from a multi-hop collector VP, which we could not reliably geolocate. Future work could tackle this open challenge. RIPE RIS and RouteViews are aware of the visibility gaps of their data platforms, and seek new peers in less observed locations. We hope this work supports identification of opportunities and incentives for new collectors and peers, and recording of peer geolocation. Expanded data collection would facilitate exploration of how VP-proximity bias affects the two metrics. Another future direction would be to develop a metric that characterizes paths *out of* a country rather than only into or within a country.

9 CONCLUSION

We created metrics to identify and rank the most important ASes from a connectivity perspective for countries around the world. We adapted the two most-used AS Ranking metrics to country-specific versions, and navigated the challenges of incomplete BGP data coverage and geolocation. We analyzed our country-specific metrics through two prisms: international, which considered inbound paths to the country; and national, for paths starting and ending within the country. The *Customer Cone International* metric identifies transit providers commonly used outside a country to reach addresses in that country, while the *Customer Cone National* metric identifies the top ISPs in the domestic transit market. The corresponding *AS Hegemony* metrics capture dominant providers without regard for whether links are transit (customer) or peering. We showed that the metrics are consistent with geopolitical and economic knowledge about the ranked networks and countries. The metrics also confirm the dominant role the U.S. still plays in global telecommunications infrastructure, with at least one dominant U.S carrier (Lumen) providing international transit for 81% of countries.

	America								
	No. 42	So. 14	Eu. 56	Af. 59	As. 52	Oc. 27	total 255	top in country	
US	40 95%	13 92%	31 55%	51 86%	38 73%	22 81%	196 76%	6939 Hurricane	48%
SE	18 42%	2 14%	14 25%	1 1%	10 19%	11 40%	56 21%	1299 Arelion$_1$	78%
NL	9 21%	1 7%	4 7%	8 13%		4 14%	26 10%	20940 Akamai..	30%
FR	4 9%	1 7%	1 1%	14 23%	4 7%	1 3%	25 9%	5511 Orange	64%
UK		1 7%	5 8%	12 20%	5 9%		23 9%	30844 Liquid..	43%
IT	2 4%	2 14%	3 5%	9 15%	2 3%		18 7%	6762 Telecom..	61%
AU					2 3%	13 48%	15 5%	4637 Telstra	53%
ZA				15 25%			15 5%	16637 MTN SA	66%
ES	2 4%	7 50%	1 1%	5 8%			15 5%	12956 Telefo..	60%
MU				14 23%			14 5%	37662 West I..	71%
EU	1 2%		5 8%	5 8%	1 1%	1 3%	13 5%	1273 Vodafone	30%
SG				2 3%	2 3%	5 18%	9 3%	7473 Singapo..	33%
	key:	+60%		59%-30%		29%-14%			

Table 12: Countries in each continent with an ISP with a AHI value > 0.1. The "top in country" is the AS from that country that matched that criteria in the most countries. Hurricane (6939) was the U.S.-based AS that exceeded the 0.1 threshold in the largest number (94) of North American countries.

These metrics can help analyze network concentration and interdependence in the face of multiple forces driving the Internet's evolution: globalization, industry consolidation, and cyber-sovereignty concerns. Researchers can also use these rankings to track evolution of a network's geographic or topological scope. They complement other rankings that political scientists and economists use [29]. With geolocation caveats, these metrics can generalize to analyze network infrastructure sovereignty in any given region of the world. To support reproducibility, we will share our artifacts, including rankings and VP geolocations.

ACKNOWLEDGMENTS

This material is based on research sponsored by the National Science Foundation (NSF) grants CNS-2120399 and OAC-2131987. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NSF.

REFERENCES

[1] 2021. LUMEN annual revenue. https://ycharts.com/companies/LUMN/revenues_annual. (December 2021).
[2] Nur A and Tozal M. 2018. Identifying Critical Autonomous Systems in the Internet. In *The Journal of Supercomputing*.
[3] APNIC. 2022. Population per Prefix. https://labs.apnic.net/?p=526. (2022).
[4] Todd Arnold, Jia He, Weifan Jiang, Matt Calder, Italo Cunha, Vasileios Giotsas, and Ethan Katz-Bassett. 2020. Cloud Provider Connectivity in the Flat Internet. In *ACM Internet Measurement Conference (IMC)*.
[5] Rui Bian, Shuai Hao, Haining Wang, Amogh Dhamdere, Alberto Dainotti, and Chase Cotton. 2019. Towards Passive Analysis of Anycast in Global Routing: Unintended Impact of Remote Peering. *SIGCOMM Comput. Commun. Rev.* (Nov. 2019). https://doi.org/10.1145/3371927.3371930
[6] R. Busa-Fekete, G. Szarvas, T. Elteto, and B. Kegl. 2012. An apple-to-apple comparison of Learning-to-rank algorithms in terms of Normalized Discounted Cumulative Gain. In *European Conference on Artificial Intelligence*.
[7] Wagner C, Francios J, State R, Dulaunoy A, Engel T, and Masseen G. 2013. ASMATRA: Ranking ASes providing transit service to malware hosters. In *Integrated Network Management, IFIP/IEEE*.
[8] CAIDA. March 2018, April 2021, March 2022. AS Relationships. http://www.caida.org/data/as-relationships/. (March 2018, April 2021, March 2022).
[9] CAIDA. March 2018, February 2021, March 2022. AS Rank. http://as-rank.caida.org. (March 2018, February 2021, March 2022).
[10] E. Carisimo, C. Selmo, J. Alvarez-Hamelin, and A. Dhamdhere. 2019-09. Studying the Evolution of Content Providers in IPv4 and IPv6 Internet Cores. *Elsevier Computer Communications Journal* 145 (2019-09).
[11] Ignacio Castro, Juan Camilo Cardona, Sergey Gorinsky, and Pierre Francois. 2014. Remote Peering: More Peering without Internet Flattening. In *Proceedings of the 10th ACM International on Conference on Emerging Networking Experiments and Technologies (CoNEXT '14)*. Association for Computing Machinery, New York, NY, USA, 185–198. https://doi.org/10.1145/2674005.2675013
[12] Online Library Learning Center. 2022. A Brief History of the Internet. https://www.usg.edu/galileo/skills/unit07/internet07_02.phtml. (2022).
[13] Y Chart. 2021. AT&T annual revenue. https://ycharts.com/companies/T/revenues_annual. (December 2021).
[14] Y Charts. 2021. 2020: A Mixed Year for Russia's Telcos. https://blog.telegeography.com/2020-a-mixed-year-for-russias-telco-finances. (December 2021).
[15] D. Cossock and T. Zhang. 2008. Statistical analysis of Bayes optimal subset ranking. In *IEEE Transactions on Information Theory*.
[16] Statista Research Department. 2022. Profit of KDDI Corporation. https://www.statista.com/statistics/807192/kddi-corporation-net-income/. (December 2022).
[17] A. Dulaunoy. BGP Ranking. http://www.terena.org/activities/tf-csirt/meeting32/dulaunoy-bgpranking.pdf. In *32nd TF-CSIRT Meeting*.
[18] Economist. 2022. How to deter China from attacking Taiwan. https://www.economist.com/leaders/2022/04/23/how-to-deter-china-from-attacking-taiwan. (April 2022).
[19] Romain Fontugne, Ksenia Ermoshina, and Emile Aben. 2020. The Internet in Crimea: a Case Study on Routing Interregnum. In *2020 IFIP Networking Conference (Networking)*. IEEE, 809–814.
[20] Romain Fontugne, Anant Shah, and Emile Aben. 2017. The (thin) Bridges of AS Connectivity: Measuring Dependency using AS Hegemony. In *Passive and Active Measurement Conference (PAM)*.
[21] Alex Gamero-Garrido. 2022. Transit Influence of Autonomous Systems: Country-Specific Exposure of Internet Traffic. (2022).
[22] M Gharaibeh, A Shah, B Huffaker, H Zhang, R Ensafi, and C Papadopoulos. 2017. A Look at Router Geolocation in Public and Commercial Databases. In *undefined*. https://doi.org/10.1145/3131365.3131380
[23] Vasileios Giotsas, Matthew Luckie, Bradley Huffaker, and kc claffy. 2014. Inferring Complex AS Relationships. In *Proceedings of the 2014 Conference on Internet Measurement Conference (IMC '14)*. ACM, New York, NY, USA, 23–30. https://doi.org/10.1145/2663716.2663743
[24] L. Granwal. 2022. Optus annual revenue. https://www.statista.com/statistics/1128051/optus-total-revenue/. (April 2022).
[25] L. Granwal. 2022. TELSTRA annual revenue. https://www.statista.com/statistics/769691/australia-telstra-s-retail-market-share-of-fixed-broadband-services/. (April 2022).
[26] B. Huffaker, M. Fomenkov, and k. claffy. 2011. Geocompare: a comparison of public and commercial geolocation databases. (May 2011). Network Mapping and Measurement Conference (NMMC).
[27] IIJ. 2022. Internet Health Report (IHR). https://ihr.iijlab.net. (October 2022).
[28] Mattijs Jonker, Gautam Akiwate, Antonia Affinito, kc Claffy, Alessio Botta, Geoffrey M. Voelker, Roland van Rijswijk-Deij, and Stefan Savage. 2022. Where .Ru? Assessing the Impact of Conflict on Russian Domain Infrastructure. In *22nd ACM Internet Measurement Conference (IMC '22)*. 159–165. https://doi.org/10.1145/3517745.3561423
[29] Kelley, Judith and Simmons, Beth. 2021. Governance by Other Means: Rankings as Regulatory Systems. *International Theory* 13, 1 (2021).

filtered prefixes

case studies							most filtered			
RU	TW	UA	US	AU	JP	..	IM	GG	MQ	NA
0	0	0.0	0.0	0.0	0.1	..	1.0	1.2	1.3	1.4

Table 13: Percentage of each country's prefixes filtered by the 50% threshold. The first four countries are used in the case studies. The last four are the countries with the highest percentage of prefixes filtered.

[30] Matthew Luckie, Bradley Huffaker, Amogh Dhamdhere, Vasileios Giotsas, and kc claffy. 2013. AS Relationships, Customer Cones, and Validation. In *ACM SIGCOMM Internet Measurement Conference (IMC)*.

[31] Lumen. 2022. Lumen's Readiness to Meet Global Events. https://news.lumen.com/RussiaUkraine. (March 2022).

[32] RIPE NCC. April 2021, March 2022. Routing Information Service (RIS). http://www.ripe.net/projects/ris/rawdata.html. (April 2021, March 2022).

[33] RIPE NCC. March 2018, April 2021, March 2022. collector locations. http://www.ris.ripe.net/peerlist/all.shtml. (March 2018, April 2021, March 2022).

[34] RouteViews Project. 2021. http://www.routeviews.org/. (April 2021).

[35] RouteViews Project. March 2018, April 2021, March 2022. collector locations. http://www.routeviews.org/routeviews/index.php/collectors/. (March 2018, April 2021, March 2022).

[36] RouteViews Project. March 2018, April 2021, March 2022. RouteViews Peering Status Report. http://www.routeviews.org/peers/peering-status.html. (March 2018, April 2021, March 2022).

[37] Mario Sanchez, Fabian Bustamante, Balachander Krishnamurthy, Walter Willinger, Georgios Smaragdakis, and Jeffrey Erman. 2014. Inter-Domain Traffic Estimation for the Outsider. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC)*.

[38] Adam Satariano and Valerie Hopkins. 2022. Russia, Blocked From the Global Internet, Plunges Into Digital Isolation. https://www.nytimes.com/2022/03/07/technology/russia-ukraine-internet-isolation.html. (March 2022).

[39] Simon Sharwood. 2022. Internet backbone provider Lumen quits Russia. https://www.theregister.com/2022/03/09/lumen_quits_russia/. (March 2022).

[40] R Sommese, L Bertholdo, G Akiwate, M Jonker, R Van Rijswijk-Deij, A Dainotti, k claffy, and A Sperotto. 2020. MAnycast2 - Using Anycast to Measure Anycast. In *ACM Internet Measurement Conference (IMC)*. https://doi.org/10.1145/3419394.3423646

[41] Comms Update. 2022. NTT Corp sees 24% growth in 9M net income. https://www.commsupdate.com/articles/2022/02/07/ntt-corp-sees-24-growth-in-9m-net-income/. (February 2022).

[42] Jane Wakefield. 2019. BBC: Russia unplugged. https://www.bbc.com/news/technology-50902496. (December 2019).

A ETHICS CONSIDERATIONS

We analyzed network relationships and dependencies based on public BGP data. We did not examine Internet identifiers at a granularity specific enough to identify individual users. There are political sensitivities around cyber-sovereignty of infrastructure in certain countries, who may not want their infrastructure relationships public, We will consider this risk in our data-sharing model, but in general we believe the risk is outweighed by the benefits the work offers to an expanded understanding of global critical infrastructure and how to improve its reliability and resilience.

B FILTERING PATHS TO ALLOW GEOLOCATION

We filtered 0.2% of prefixes due to a lack of consensus on their geolocation. Our geolocation process, summarized in §4.2.1 is as follows. We assign a prefix to a country if the number of addresses in the prefix that geolocate to that country is the majority of addresses in that prefix *and* is above the chosen threshold. We choose a 50%

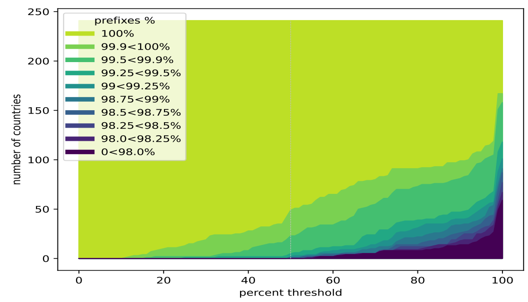


Figure 8: Number of countries for which a given percentage range of prefixes (depicted by color) geolocated to that country, as a function of the threshold (of IPs in the prefix that must geolocate to that country) used for prefix geolocation.

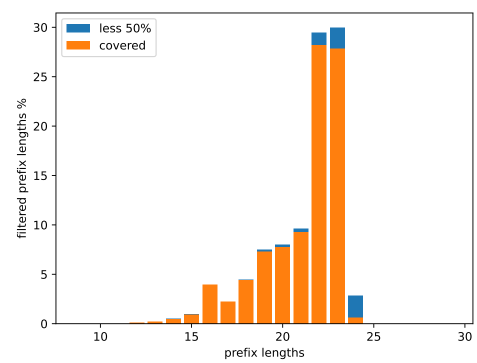


Figure 9: Prefix lengths for prefixes that we filtered, 85% of which we filtered because they were covered by by more specifics. The other 15% we dropped due to a lack of (50%) consensus on their geolocation.

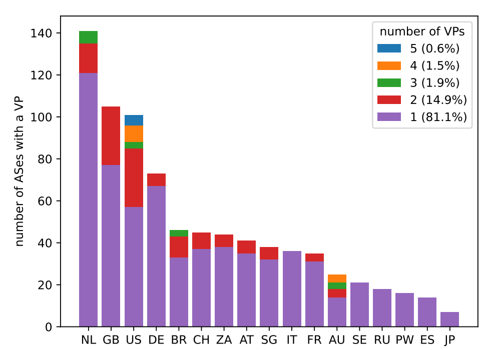


Figure 10: VP distribution across ASes, by country. The percentage in the legend is the percentage of all ASes with that number of VPs.

filered addresses										
case studies							most filtered			
US	RU	TW	UA	JP	AU	..	AF	HR	IN	LT
0	0	0	0.2	3.0	7.6	..	15	15	16	18

Table 14: Percentage of each country's addresses filtered by the 50% threshold. The first four are the countries used in the case studies. The last four are the countries with the highest percentage of addresses filtered.

threshold for observed prefixes. As an example, we would geolocate a prefix to France if over half of the IP addresses in the prefix geolocated to France. If the threshold were 30%, then so long as no more than 30% of addresses in the prefix geolocated to another country, and at least 30% geolocated to France, we would geolocate the prefix to France.

Before we geolocate the prefixes, we split them into non-overlapping blocks of addresses mapped to their most specific prefix. We then filter (remove from our data set) prefixes that are completely covered by more specifics (1.2% of our April 2021 data set).

Of the remaining 98.8% of prefixes, we filter an additional 0.2% of prefixes because they failed to geolocate based on our 50% threshold. This filters 1.5% of the routed IP addresses in our data set. To illustrate the minimal impact on our data set, Tables 13 and 14 show the percentage of prefixes (addresses, respectively) filtered by the 50% threshold for individual countries. While countries such as Afghanistan and Lithuania saw up to 18% of their geolocated addressed filtered, countries examined in our case studies saw minimal loss of addresses, and at most 0.1% of their prefixes. Future work could use an alternative method such as breaking up the prefixes into /24s to increase the number of IP addresses geolocated to the country.

Figure 8 shows the number of countries for which a given percentage of their total prefixes pass the majority threshold, as a function of the percentage threshold. Only Guernsey, Martinique, and Namibia have more then 1% of their majority prefixes not passing this 50% threshold.

C VANTAGE POINT (VP)

Figure 10 examined the concentration of VPs within individual ASes. 81% of VPs are in a single ASN. 96% of VPs are in ASes with two or fewer VPs. Most countries – 15 out of 17 – have over 93% of their VPs sharing an ASN with 1 or fewer VPs Only Australia and the U.S. had greater concentrations of VPs, with 74we do not think AS concentration impaired the accuracy of our results.