Studying the evolution of content providers in IPv4 and IPv6 internet cores

Esteban Carisimo [a,b,*], Carlos Selmo [c], J. Ignacio Alvarez-Hamelin [a,b], Amogh Dhamdhere [d]

[a] *Universidad de Buenos Aires, Argentina*
[b] *CONICET, Argentina*
[c] *Instituto Tecnológico de Buenos Aires, Argentina*
[d] *CAIDA, San Diego Supercomputer Center, UC San Diego, United States*

Keywords:
Content providers
CDNs
Topology
k-cores

A B S T R A C T

There is recent evidence that the core of the Internet, which was formerly dominated by large transit providers, has been reshaped after the transition to a multimedia-oriented network, first by general-purpose CDNs and now by private CDNs. In this work we use *k*-cores, an element of graph theory, to define which ASes compose the core of the Internet and to track the evolution of the core since 1999. Specifically, we investigate whether large players in the Internet content and CDN ecosystem belong to the core and, if so, since when. In addition, we examine differences between the IPv4 and IPv6 cores. We further investigate regional differences in the evolution of large content providers. Finally, we show that the core of the Internet has incorporated an increasing number of content ASes in recent years. To enable reproducibility of this work, we provide a website to allow interactive analysis of our datasets to detect, for example, "up and coming" ASes using customized queries.

1. Introduction

The structure of the Autonomous System (AS) network has been changing over the years driven by disruptive changes on the Internet [1]. In the NSFNET era, the Internet had a monolithic backbone deployed in the U.S. to interconnect research and educational institutions [2]. After the US government decommissioned the NSFNET, the interdomain network moved onto a Transit era where the network had a hierarchical structure [1,3]. More recently, the Internet has transformed into multimedia network, driven by high bandwidth demands and low latency requirements, resulting in a Content era [4].

Content Delivery Networks (CDNs) have played a decisive role in the evolution towards a multimedia network [5] and the resulting *flattening* of the Internet [1,6]. CDNs are decentralized serving infrastructures that provide front-ends close to users to reduce latency, maximize the throughput and avoid delivering packets through long routes, which increase latency and can be congested [7]. CDNs typically establish a large number of peering agreements with ASes hosting customers of their content ("eyeballs"). It is not necessary that every Content Provider (CP) needs to deploy its own CDN. A number of *third-party CDNs* provide hosting services without being content generators, such as Akamai and LimeLight. However, it is apparent that several CPs have transformed into private CDNs with worldwide coverage instead of delivering content through Transit Providers or third-party CDNs due to a range of technical, economic, and legal reasons [8–13].

In addition to CDNs, Internet Exchange Points (IXPs) have been crucial in morphing the hierarchical structure of the AS internetwork, transforming it into a *flat* network [14]. The availability of IXPs is critical to CDNs, which prefer to have direct peering relationships with as many ASes as they can [15]. IXPs too are interested in hosting CDNs to provide a cost-effective way for the IXP members to reach content [16].

More recently, ASes have been slowly incorporating IPv6 reachability to deal with IPv4 address exhaustion. A set of milestones, such as IANA last IPv4 block transfer [17], The World IPv6 Launch [18] and ARIN IPv4 pool total depletion [19] have fostered IPv6 adoption. IPv6 is not backward compatible with IPv4, IPv4 and IPv6 paths between two end-hosts may differ [20].

In this paper we use the term "core" of the network to refer to the subset of ASes that are densely connected. In the past the "core" of the network mostly consisted of tier-1 networks, which were large international transit providers that were connected to all other tier-1 networks with peering links and had no transit providers of their own. CPs, as well as "eyeball" networks that were the destinations of traffic sourced by CPs were on the edge of the network. However, CPs and third-party CDNs have been building intercontinental backbone networks as well as making thousands of peering agreements in recent years. The growing significance of CPs has led to discussion and speculation about whether CPs are now the dominant players in the Internet ecosystem [4].

* Corresponding author at: Universidad de Buenos Aires, Argentina.
E-mail address: carisimo@cnet.fi.uba.ar (E. Carisimo).

Our goal is to investigate what role CPs now play in the Internet ecosystem, and in particular, if CPs are now a part of the “core” of the Internet. Specifically, we motivate this work with the following questions: How can we identify if a CP does or does not belong to the core of the Internet? If the core of the network does indeed include CPs, who are they? As the overall adoption of IPv6 has been slow, do we notice that delay on IPv4 and IPv6 core evolution? As the AS ecosystem has shown striking differences according to geographical regions [15], do we also see geographical differences in the role of CPs and their presence in the “core” of regional Internet structures? Finally, as more CPs deploy their private CDNs, can we detect “up and coming” CDNs that are not currently in the core of the network but are likely to be in the future?

We use the concept of *k-cores* to analyze the structure of the IPv4 AS-level internetwork over the last two decades. We first focus seven of the largest CPs in terms of capacity, geographical footprint and traffic profile, and confirm that they are all currently in the core of the Internet. We then dig deeper into the evolution of these large players to correlate observed topological characteristics with documented business practices which can explain when and why these networks entered the core. Next, we repeat the methodology but using IPv6 dataset to compare and contrast the evolution of CPs in both networks. Based on results, we investigate commercial and technical reasons why CPs started to roll out IPv6 connectivity.

We then take a broader view, characterizing the set of ASes in the core of the IPv4 Internet in terms of business type and geography. Our analysis reveals that an increasing number of CPs are now in the core of the Internet. Finally, we demonstrate that the k-core analysis has the potential to reveal the rise of “up and coming” CPs. To encourage reproducibility of our results, we make our datasets available via an interactive query system at http://cnet.fi.uba.ar/TMA2018/.

2. Related work

The increasing importance of CDNs in the Internet ecosystem has produced a vast literature on this topic, which shares some of the goals of the present article. Several articles studied the internal structure of CDNs [21–24], where the focus was on the economic and technical benefits of CDNs, the need of data replication, techniques for content distribution and cache updates, and cache placement. CDN literature has also acknowledged the rising importance of private CDNs. Indeed, there have been several studies about the largest private CDNs. Google’s CDN has been studied from many points of view: the growth of the serving infrastructure in recent years [25], QoE performance [26], internal load balancing [10], traffic engineering strategy run by its WAN SDN [9] and so on. Facebook’s CDN was studied from the point of view of data replication [27], network administration [28], and Facebook’s SDN [11]. Bottger et al. [29] studied the Netflix serving infrastructure, called Open Connect, due to its remarkably different architecture from other CDNs as well as the importance of Netflix in overall traffic share. Calder et al. analyzed Microsoft’s CDN, known as Azure, as a representative example of an anycast CDN [30].

IXPs have also received a great deal of attention in the research and operational literature during the last decade. During the 2000s, IXPs were in part responsible for a *peering revolution*, offering neutral points for ASes to establish settlement-free peering agreements. IXPs encourage peering in order to keep traffic local and to avoid reaching local neighbors via either paid transit links or longer circuitous routes [3]. A well documented phenomenon is that the proliferation of IXPs has contributed to a *flattening* of the Internet [14], with hundreds of IXPs spread all over the world facilitating connectivity between thousands of co-located networks. In the research literature, a number of papers have studied the anatomy of large IXPs [6] as well as the role of IXPs in developing regions [31,32].

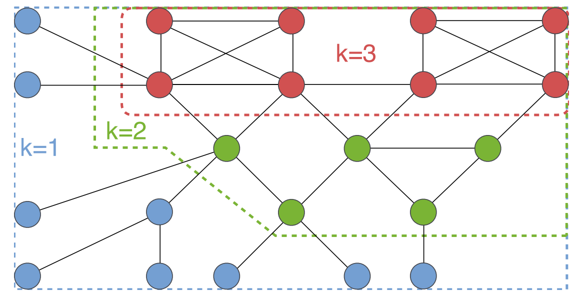


Fig. 1. Example of k-core decomposition of a given graph. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Recently, Geoff Huston observed the wide-ranging effects of the flattening structure of the Internet and the rise of CPs [4]. Huston suggests that these trends are marginalizing the role of Transit Providers, terming this as *“The Death of Transit”*.

IPv6 has gained more attention in recent years due to the exhaustion of the IPv4 address space [17] and the increase of IPv6 adopters [33]. The growth of IPv6 reachability and its incompatibility with IPv4 have encouraged to study differences between both networks such as path lengths, performance and perspectives at the routing system [20].

The foreseeable long-time coexistence of both protocols, in addition to the already scarce number of unallocated IPv4 prefixes, have led to inter-organization IPv4 blocks purchase, allowed by some RIRs, know as the *Transfer market* [34,35]. Despite the large number of transfers that have been signed since 2009, a peculiar pattern has been recently observed — Content Providers are purchasing large address blocks from American universities. This was evidenced when Google and Amazon got transferred IPv4 blocks that previously belonged to Merit (AS237) and MIT (AS3) [36,37]. Averaging 10 dollars paid per address reinforces the importance that IPv4 still has over IPv6.

There is a vast body of previous literature on applying graph theoretic concepts to study the AS graph structure. Some examples of such work are papers that have introduced k-core decomposition to study properties of the network [38–40]. These works mainly take a mathematical perspective about the structure of the AS graph. In this work, we also utilize the k-core decomposition technique from graph theory to study the role specifically of CPs in the Internet over the years. However, we pair the graph-theoretic concept with domain knowledge, insights from other measurement datasets, and documented strategies and actions of the CPs themselves, which gives further context and explanation for the observed phenomenon.

3. Methodology

Our goal is to study changes in the structure of the AS-level Internet ecosystem from the perspective of content providers and CDNs, specifically, whether large CPs are now part of the core of the network, and the historical evolution of when such a transition occurred. For this purpose, it is necessary to define a methodology to determine which ASes are part of the core of the network.

Since we look to prove that CPs have become as densely connected as Transit providers, the chosen methodology needs to determine AS connectivity based on the number of AS links and its neighbors links. If the methodology is capable of doing so, and CPs are in fact as densely connected as Transit providers, values must be equal for both kind of ASes.

3.1. Definitions

To begin with, we examined a set of graph-theoretical metrics to determine which is capable indicating which ASes have the same level of connectivity. The studied metrics were node degree, Transit degree, Average nearest-neighbor degree (AND) and the k-core shell-index. The mathematical and computational complexity of these metrics is fairly different and the meaning of the metric as well.

Node degree. It is the simplest graph metric to evaluate the connectivity of a node. However, this metric does not take into account the properties of a node's neighbors. Furthermore, Faloutsos et al. [41] showed empirically that the node degree distribution of the AS-level ecosystem can be modeled using a random variable having a power-law distribution. With a power-law degree distribution, a non-negligible set of nodes will have large node degree while a large number of nodes will have much smaller degrees.

Transit degree. It is defined by the number of unique neighbors that appear on either side of an AS [42]. In other words, since the routing structure of the Internet is given by the AS-PATHS on BGP announcements, Transit degree counts in how many unique triples a node is, for all the observed AS-PATHS. Due to this definition, the Transit degree is a metric that measures the relevance of *intermediation* of an AS.

AND. As it is defined by Pastor-Satorras et al. [43], the AND is computed as the average degree of the neighbors of a AS. Compared to *Node degree*, this metric does take into account the relevance of the neighbors but it might be sensitive to the node degree distribution.

k-core shell-index. A k-core of a graph $\mathcal{G}$ is the maximum induced subgraph[1] in which all the vertices have at least degree k (see [44]). A vertex or node that belongs to a k-core has at least k neighbors which all have degree at least k. Moreover, a node that belongs to core k also belongs to any core $j < k$, thus the shell-index is given by the maximum core that a node belongs to. Fig. 1 displays k-cores using a small graph example where nodes are colored to indicate their shell-index. As the figure shows, the shell-index (or simply "core") is given by the degree of the node as well as the degree of the neighbors in the induced graph. This can be seen in the example where some four-degree nodes are in core 2 while nodes of degree 3 are in core 3. Furthermore, AS graphs are *core-connected* [45], which means that there are k different paths between two ASes of the same k-core.

3.2. Evaluation on the AS ecosystem

Having defined a set of candidate metrics to determine which ASes are the most densely-connected in the Internet, we next analyze how these metrics perform. To do such analysis, we picked ASes that are presumably densely connected but have different business purpose. Among those ASes, we included the TOP10 ASes in CAIDA's AS-RANK [46] in March 2018, which are *TIER-1 Transit Providers* and seven popular Content Providers that corresponds to the top 7 *hypergiant* ASes identified by Bottger et al. [47]. Transit ASes are Level3 (AS3356), Telia (AS1299), NTT (AS2914), GTT (AS3257), Telecom Italia (AS6762), HE (AS6939), TATA (AS6453), PCCW (AS3491) and Level3 (formerly GBLX) (AS3549). The selected Content Providers are Akamai (AS20940), Amazon (AS16509), Apple (AS714), Facebook (AS32934), Google (AS15169), Netflix (AS2906) and Yahoo! (AS10310). For now on, we refer to the latter group as the *Big Seven*.

Table 1 displays the four metrics for the 17 ASes under study — 10 Transit ASes and 7 CPs, for the IPv4 AS graph. This table does not contain information about IPv6 because we just focus on analyzing what properties of the network these metric can capture rather than comparing results in IPv4 and IPv6 networks.

This table indicates that Node degree significantly varies among the TOP10 Transit ASes in the AS-RANK as well as the *Big Seven*. For instance, while the observed node degree for Level3 (AS3356) is 4924, for Telecom Italia (AS6762) is 488. Moreover, node degree for CPs is by far smaller than for Transits since many peering links are often invisible in the dataset where these metrics were calculated [48].

Transit degree was meant to measure transit intermediation, therefore it is expected that Transit Providers have the highest Transit degree. Comparing both tables, Transit Degree for Transit ASes is usually an order of magnitude longer than for Content Providers. Content Providers are usually on the end of the AS-PATH, thus Transit Degree tends to be fairly small.

AND is fairly different between Transit and Content Provider ASes according Table 1. Graph edges between Transits and their customers are always visible. In addition, a large fraction of Transits customers are ASes that have a low node degree (1 or 2), thus AND tends to be low for Transit providers. On the other hand, Content Providers also peer with a large number of ASes of degree 1 or 2, however, those links are likely to be peering links and would remain invisible in our dataset [48,49]. Furthermore, CPs have no customers that affect their AND value. Therefore, CPs are just visible through a small subset of TIER-1 Transit providers, which lead CPs to have a fairly large AND value.

Table 1 shows that Large Content Providers as well as TIER-1 Transit Provider have the same (or almost the same) shell-index. Even though AND and k-core definitions are apparently similar, they are actually not. AND is defined by the average degree of the neighbors of a node while the shell-index of a node says that a node has k neighbors of degree k in the induced subgraph. Using k-cores we can see that CPs and TIER-1 are both densely connected because the ones on k=82 have al least 82 peers with ASes that have at least 82 peers with other networks. Thus, according to Table 1 , shell-index is the only indicator capable of reflecting connectivity equally for CPs and Transit providers.

3.3. Proposed methodology

To sum up the analysis of proposed metrics, the shell-index of the k-core decomposition is the only metric among the ones that we looked at that indicated similar values for Content Providers and Transit Providers. This is due to the definition of the k-core decomposition sets that an AS is densely-connected *if and only if* it is connected to ASes that are as connected as it is. This restriction is so strict that only large CPs and TIER-1 Transits can fulfill, and therefore we will use this definition to compute our analysis of the evolution of Content Providers. We are going to refer to the *core* of the network as the subset of ASes that are densely connected.

Applying the k-core decomposition, the central part of the network is made of ASes that belong to the maximum core k_{max}. In our analysis we study the evolution of cores of the CPs. However, the k_{max} as well as the k-indices of the AS graph change over time. For this reason, we normalize k in each snapshot by its k_{max} index, which leads to a normalized k with values between 0 and 1, referred to as k^*. For now on, TOPcore will refer to $k^* = 1$. To calculate k-core decomposition on each snapshot of an AS graph we used two tools, LaNet-vi [45], which also provides network visualization, and NetworkX, a python library.

4. Dataset

To apply the above k-core decomposition methodology on the Internet graph longitudinally, we need periodic historical snapshots of the IPv4 and IPv6 Internet's AS-level topology.

We rely on publicly available AS topology snapshots from CAIDA. CAIDA curates AS topology data from both BGP and traceroute-derived sources. The BGP AS relationship IPv4 dataset[2] is derived from BGP dumps taken from RouteViews and RIPE RIS collectors [42] from 1998 to present, and contains AS links observed at the BGP collectors along

[1] Let $G = (V, E)$ be any graph and W subset of vertices $W \subset V$. An *induced graph* is a subgraph of G whose nodes are given by W and its edges are the ones that have both endpoints in W.

[2] CAIDA's BGP serial-1 dataset: http://data.caida.org/datasets/as-relationships/serial-1/.

Table 1
Transit degree, node degree, AND and k-core for the TOP10 ASes in the AS-RANK and seven well-known Content Providers in the IPv4 graph. Transit Degree and AS-RANK were taken from March 2018 snapshot which is different from the one in which we calculated the k-core decomposition and node degree, that was taken from October 2017. We assume these parameters did not vary significantly during that lapse.

	AS	ASN	AS Rank	Transit Degree	Degree	AND	K-Core
Transit ASes	Level 3 Comm.	3 356	1	5130	4924	38	82
	Telia Company AB	1 299	2	1972	2240	81	82
	Cogent Comm.	174	3	5718	6041	32	82
	NTT America	2 914	4	2068	2438	75	82
	GTT Comm.	3 257	5	1641	1577	86	81
	TELECOM ITALIA	6 762	6	828	488	191	79
	Hurricane Electric, Inc.	6 939	7	699	6792	53	82
	TATA Comm.	6 453	8	2596	772	147	81
	PCCW Global, Inc	3 491	9	352	636	186	81
	Level 3 Comm.	3 549	10	1433	2648	33	72
Content providers	Apple	714	5668	185	161	892	82
	Netflix	2 906	4389	213	200	860	82
	Yahoo!	10 310	687	278	291	621	82
	Google	15 169	1397	237	205	834	82
	Amazon	16 509	3054	173	203	900	82
	Akamai	20 940	2679	285	364	563	82
	Facebook	32 934	4417	227	202	905	82

with inferred business relationships. The BGP AS relationship IPv6 dataset[3] was created with exactly the same inputs but based on BGPv6 announcements [50] from 2004 to present.

We use a second IPv4 dataset which consists of AS links extracted from traceroutes from CAIDA's Archipelago (Ark) [51] vantage points towards every routed/24 prefix.[4] The two IPv4 datasets can provide somewhat different views of the Internet's AS-level topology. While the number of edges in each BGP data snapshot is larger than in traceroute data snapshots, traceroute often reveals peer-to-peer links which are not seen at BGP collectors [52]. To get the most complete picture of IPv4 AS-level connectivity, we chose to combine data from both the BGP and Ark datasets, which we refer to as the "Ark+ BGP" dataset. This dataset consists of monthly snapshots dating from 1998 to present, which is sufficiently long to detect the evolution of the number of peers of CPs. Unfortunately, we did not have any similar traceroute-based dataset to enlarge the IPv6 BGP AS relationship dataset. To view the *k*-core decomposition using only the BGP dataset or traceroute dataset, we refer the reader to a website with these visualizations.[5]

A limitation of our methodology is that CPs also serve content from caches located within ISPs [12,29], which are not visible as AS links in BGP or traceroute. Even CPs that follow an in-network caching strategy, however, generally need to peer in order to reach ISPs that are not willing to host caches in their networks, to fill the caches, and to serve dynamic content that cannot be cached. In this work we only study the evolution of AS-level connectivity of CPs; even though an analysis of cache infrastructure is important to shed some light about the way content is served, we consider such task out of the scope of this article and we will leave it to future work.

5. A first look into the core evolution of CPs

A well-documented trend in the evolution of the Internet is that the set of ASes responsible for generating most of the traffic has been shrinking; recent studies have shown that only few tens of ASes together generate most of the traffic, while in the past that number was in the thousands [1,53]. Given this trend toward traffic consolidation, we track the core evolution of seven big players, which we refer to as the *Big Seven*: Akamai (AS20940), Amazon (AS16509), Apple (AS714), Facebook (AS32934), Google (AS15169), Yahoo! (AS10310) and Netflix (AS2906). Our selection corresponds to the top 7 *hypergiant* ASes identified by Bottger et al. [47], where the authors ranked *hypergiants* based on port capacity, geographical footprint and traffic profile reported in PeeringDB.

Our *a priori* hypothesis is that all of these CPs currently belong to the TOPcore. We check whether our hypothesis is true, and if so, *when* and *how quickly* they reached the TOPcore. We then attempt to dig deeper into the reasons why we observe these CPs in the TOPcore, and correlate with external factors such as legal disputes, market expansions, QoE improvements, services releases etc. to explain why the CPs appeared in the TOPcore at a certain time.

5.1. Looking at sibling ASes

Organizations, such as CPs, are likely to have multiple ASNs, where ASNs that belong to the same organization are usually called *siblings*. A wide variety of reasons is behind the fact of organizations having multiples ASNs, such as legacy ASN after merges or acquisitions, or having multiple ASNs for different purposes. However, organizations with multiple ASNs tend to have a *primary* ASN, which is presumably more visible than the rest.

We are interested in tracking the evolution of connectivity of Content Providers as organizations, thus we need to find all the ASNs that belong to each of the *Big Seven*. We looked for sibling ASes of the *Big Seven* using CAIDA's AS-to-organization list [54], which is a list based on WHOIS records that binds AS number with `org id`. First we looked for the `org ids` that correspond to the well-known primary ASNs of the *Big Seven* and then we search all the ASNs that match with the previously obtained `org ids`.

After filtering, we found that 39 ASNs belong to exactly the same `org id` as the *Big Seven*, where Akamai has 17 ASNs, Apple 3, Amazon 3, Google 7, Facebook 2, Yahoo! 5 and Netflix 2. Among the siblings we found some ASNs which are popular and frequently mentioned by literature and operators, such as Google's AS36040 (formerly YouTube's ASN) and Apple's AS6185.

We tracked the evolution of the 39 ASNs over the years and we found that there have never been a sibling as relevant in terms of shell-index as the primary ASN in IPv4 graph. Whereas the primary ASNs do belong to the TOPcore, the secondary ASNs have been at most in cores half-way between core 1 and the TOPcore.

After performing this analysis, we can conclude that for IPv4 it is efficient to only track the primary ASN of the *Big Seven*, and then correlate changes on those ASNs with business strategies.

[3] CAIDA's BGP IPv6 AS-REL dataset: http://data.caida.org/datasets/2015-asrank6-data-supplement/.

[4] The Ark dataset was merged with skitter dataset http://data.caida.org/datasets/topology/skitter-aslinks/.

[5] Graph visualization website: http://cnet.fi.uba.ar/TMA2018/.

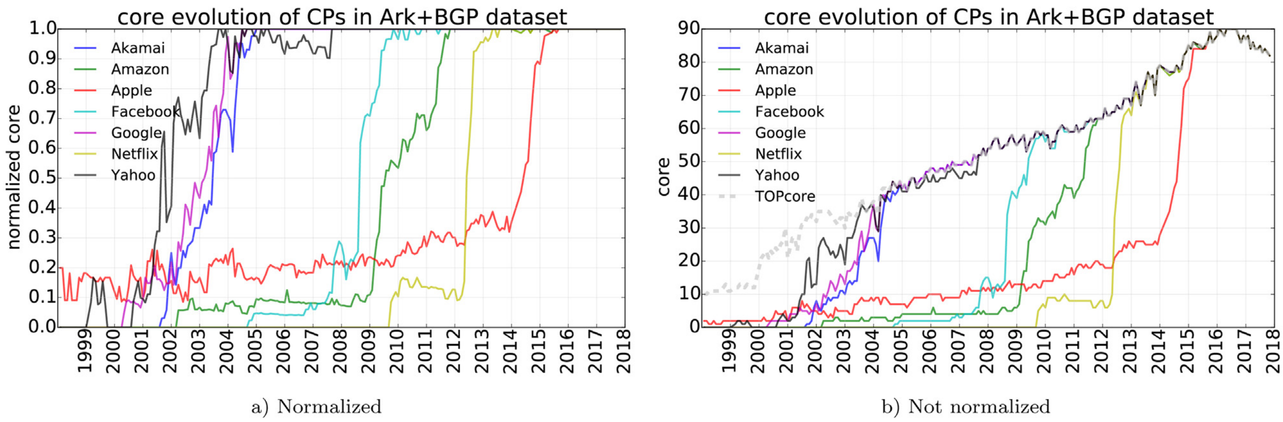


Fig. 2. k-core evolution of the *Big Seven* in IPv4 network. All of these CPs have reached the TOPcore.

5.2. Tracking the evolution of the Big Seven in IPv4

Fig. 2 shows the monthly evolution of the CP-core on the Ark+BGP IPv4 dataset, where Fig. 2a is normalized and Fig. 2b is not. A first observation is that as of the end of 2017, all the studied CPs have joined the TOPcore, indicated by the fact that the normalized core value for each CP is 1.

There appear to be two groups among the studied CPs, one composed of Akamai, Google and Yahoo! which reached the TOPcore by 2005, and another comprising Amazon, Apple, Facebook and Netflix, which became members of the TOPcore between 2010 and 2015. The CPs in the first group are arguably more established, and have been providing a variety of online services for many years. The second group consists of CPs that at some point decided to deploy their own infrastructure and stop serving content using third-party CDNs such as Akamai, as multimedia content began to dominate the Internet traffic share [55]. Moreover, the transition from lower cores to upper cores among the members of the latter group is faster than in the former group. The fast evolution of Amazon, Apple, Facebook and Netflix cores is likely to have been encouraged by the vast number of peering facilities which appeared during the last decade [3,56].

According to Fig. 2b, the TOPcore has been always linearly growing, despite a small decay in late 2017. On the initial snapshot in 1999, there was no CP from the *Big Seven* in the TOPcore, and the maximum core was 10. The TOPcore reached its maximum during 2016, where the value of the core was 90. For example, in 2012 Netflix transitioned 71 cores (from core 6 to 77) to be able to reach the TOPcore. On the other hand, this may not be such a difficulty due to the expansion of peering infrastructure [14].

Next we dig deeper into the evolution of CPs individually. Specifically, we attempt to correlate the topological characteristics of the CPs (their core) with business strategies, acquisitions, or other factors which could explain why the CP entered the TOPcore.

Akamai. Akamai has been in the TOPcore since 2005. Akamai is a *pioneer* in content-delivery, and since its business model relies on providing high-availability low-latency hosting rather than generating content, they have always aimed to have a large number of peers.

Amazon. Amazon's infrastructure deployment appears to have occurred in two steps, according to Fig. 2a, which is corroborated by publicly available information on Amazon's website [57]. In 2009, Amazon established its datacenter in Northern California, which coincides with the first growth. Between 2010 and 2012, Amazon established datacenters in several parts of the world, which coincides with the second growth spurt from 2010 to 2012.

Apple. We find that Apple's AS reached the TOPcore in 2015 after a quick growth. According to public information, Apple has been steadily offloading its content from Akamai onto its own CDN since 2013 [58]. Apple's traffic share has been growing rapidly in recent years due to software updates such as new OS releases [59] and security patches. Further, the company has recently announced that is planning to break into the TV market, producing original television shows, which will be served from Apple's CDN [60].

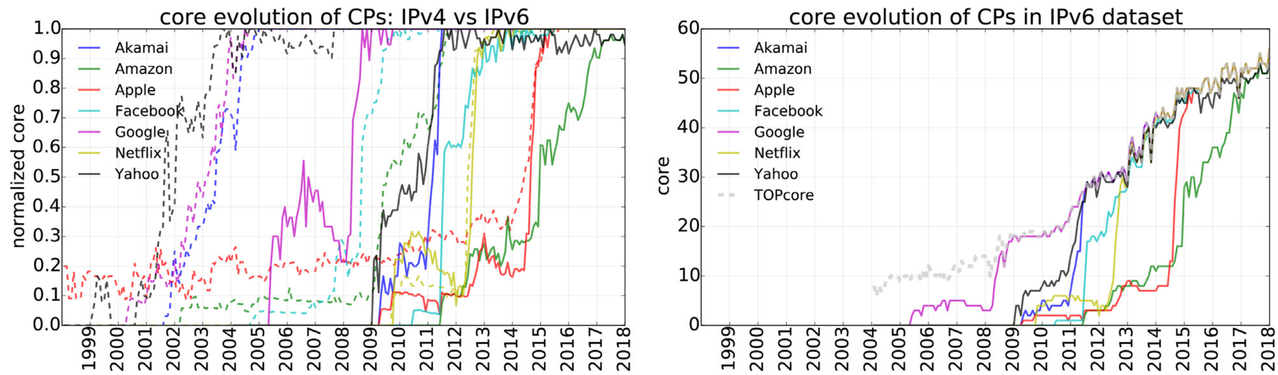
Facebook. Facebook's AS32934 got close to the TOPcore in 2010 after a rapid growth in its normalized core between 2008 and 2010. The number of users on Facebook grew exponentially from 12M in December 2006 to 350M by the end of 2009 [61] which coincides with Facebook's expansion period and rise to the TOPcore. Although Facebook kept growing exponentially since then, the massive growth during that period encouraged Facebook to establish multiple peering agreements that enabled it to reach the TOPcore.

Google. Google was launched in September 1997 and in just a couple of years became the most popular search engine [62]. Over time, as Google started serving large volumes of video traffic via the acquisition of YouTube in 2006 [63], it expanded its CDN to get as close as possible to "eyeball" networks. Even before establishing its CDN, between 1999 and 2003, Google had peering agreements with tier-1 transit providers such as Level3 (AS3549), TATA (AS6453), Telstra (AS4637), NTT (AS2914), Zayo (AS6461), Qwest (AS209), GTT (AS3257) and Cogent (AS174). Links with a number of large Transit Providers resulted in Google becoming part of the same core level as those transit providers.

Yahoo! The dot-com bubble in the early 2000s motivated Yahoo to build their own WAN infrastructure to avoid relying on transits for two reasons: (i) to reduce content delivery dependency on intermediate networks between them and eyeball networks (ii) to reduce operational costs [5]. In fact, Yahoo's core growth coincides with the end of the dot-com bubble in 2002.

Netflix. In 2012, it took Netflix less than a year to move from core $k^* = 0.1$ to the TOPcore. Netflix started offering video streaming in 2007 using third-party CDNs and transit providers. With the growing popularity of the service and increasing traffic volumes, the company moved content to its own Open Connect [64] platform in 2012, which is seen as a sharp increase in its normalized core value between 01/2012 and 09/2012 as shown in Fig. 2a.

In summary, all of the studied CPs moved from third-party CDNs to private CDNs and entered the TOPcore. In particular, Apple, Facebook and Netflix all off-loaded content from Akamai. These changes led to significant loss of revenue for Akamai and a drop in its share price [65]. Despite losing major clients, Fig. 2a shows that Akamai is still in the TOPcore, which means that Akamai's peering agreements do not depend exclusively on these large clients.



a) Normalized k-core evolutions of the *Big Seven*. Dashed lines display IPv4 while solid lines represent IPv6.

b) Not normalized

Fig. 3. k-core evolution of the *Big Seven* in IPv6 network. All of these CPs have reached the TOPcore.

5.3. Tracking the evolution of the Big Seven in IPv6

Fig. 3 shows the monthly evolution of the CP-core on the Ark+BGP IPv6 dataset, where Fig. 3a is normalized and Fig. 3b is not. Fig. 3a also compares the normalized core evolution of the *Big Seven* in IPv4 and IPv6 cores. This figure confirms that all these ASes are currently present in both TOPcores, however, the date of arrival in the IPv6 TOPcore is significantly different than that in the IPv4 TOPcore. Two factors appear to have boosted the IPv6 adoption for these companies: The World IPv6 Launch in 2012 [18] and ARIN's IPv4 pool depletion in 2015 [19].

Even though it has been several years since The World IPv6 Launch and the *Big Seven* reached the IPv6 TOPcore, the size of the IPv6 AS-level topology is still significantly smaller than IPv4 [33,66]. As shown in Figs. 2b and 3b, where the cores are not normalized, the maximum shell-index in the latest snapshot is 83 for IPv4 while it is 52 for IPv6.

Next, we look into the business strategies that fostered the *Big Seven* to rollout IPv6 connectivity.

Akamai. After a soft growth between 2009 and 2011, Akamai quickly reached the IPv6 TOPcore in 2011. This growth matches Akamai's recruiting campaign for early IPv6 adopters in 2011 [67], in which the company selected a subset of customers to start serving their content through dual-stack.

Amazon. Amazon's IPv6 rollout followed an almost identical trend to its IPv4 growth, but a few years delayed as compared to IPv4. Amazon progressively incorporated countries where dual-stack services were available [68,69], similar to its IPv4 worldwide expansion. A notable spurt in Amazon's IPv6 core growth occurred in late 2014. Examining the monthly snapshots from that time, Amazon started peering with a large number of Brazilian over IPv6 at the Brazilian IXP IX.br-SP.

Apple. Apple deployed its own CDN in 2015 and as seen in Fig. 3a, both its IPv4 and IPv6 cores grew at the same pace. Just after Apple deployed its CDN, Apple started implementing IPv6 preference [70], which could be seen as an indicator of why Apple rolled out IPv6 reachability.

Google. This CP was the first among the *Big Seven* to reach the IPv6 TOPcore. Google started testing its IPv6 reachability using the domain `ipv6.google.com` during IETF72 in March 2008. Shortly after, in January 2009, Google became publicly available over IPv6 [71].

Facebook. Facebook reached the IPv6 TOPcore after two large steps, one in 2011 and the other in 2012. Facebook's IPv6 prefix `2a03:2880::/29` was, according to WHOIS records, allocated in August 2011. Then, Facebook was one of the participants of the World IPv6 Launch in June 2012 [72] and during this event the company reached the IPv6 TOPcore.

Yahoo! The company has been endorsing IPv6 adoption, and it joined and sponsored IPv6 World Day and Launch in 2011 and 2012 respectively [18]. Yahoo! reached the IPv6 TOPcore a few months before IPv6 World Day in 2011.

Netflix. Netflix deployed its IPv4 and IPv6 CDN (called Open Connect) simultaneously in 2012. As shown in Fig. 3a, Netflix rapidly climbed in both cores at the same pace. Although the maximum core on IPv4 and IPv6 were different in 2012 (Figs. 2b and 3b), Netflix's aggressive peering strategy allowed them to become densely-connected in both cores at the same time. Moreover, according to Netflix information, video is delivered over IPv6 whenever possible [73,74].

6. Evolution by geographical region of the IPv4 core

We also investigate in whether CPs belong to the IPv4 TOPcore in each geographical region, defined as the Regional Internet Registries (RIR) regions. We repeat the analysis of speed and date of arrival done in Section 5.2 for each CP in every RIR with a focus on detecting differences by region, especially systematic delays in when certain CPs appeared in specific regions.

To determine which regions an AS is present in, we use the NetAcuity [75] geolocation database to geolocate each prefix advertised by an AS in a given snapshot. For this analysis we will just focus on IPv4 core evolution due to the lack of geolocation entries for IPv6 analysis. The (in)accuracy of geolocation databases has been studied extensively [76]. However, previous work has found that the NetAcuity database is mostly reliable for country-level geolocation [77]. We use RIR-level granularity in this work, so we believe that this analysis is not affected by inaccuracies in geolocation. After geolocating ASes, we combine the monthly "Ark+ BGP" snapshots with the mapping between AS and RIR to create monthly RIR subgraphs.

There are two issues with this basic methodology that we need to account for. First, we need AS geolocation information throughout the duration of "Ark+ BGP" dataset. However, CAIDA only had NetAcuity records since November 2011, while our topology dataset starts in January 1998. Second, NetAcuity appears to incur a time lag between when a prefix is active in a new location and when it appears at that location in the database. For example, NetAcuity started reporting the presence of Netflix in the LACNIC region in December 2016, while a June 2015 Wayback Machine snapshot[6] already showed Netflix as a member of a Brazilian IXP. As our goal is to track historical evolution, it is necessary to include an AS in the RIR subgraph when changes are actually happening and not once they have already happened. To account for these issues we made two modifications to the basic methodology.

[6] Wayback Machine snapshot of members of Brazilian IXP. 06/2015: http://web.archive.org/web/20150617231252/http://ix.br/particip/sp.

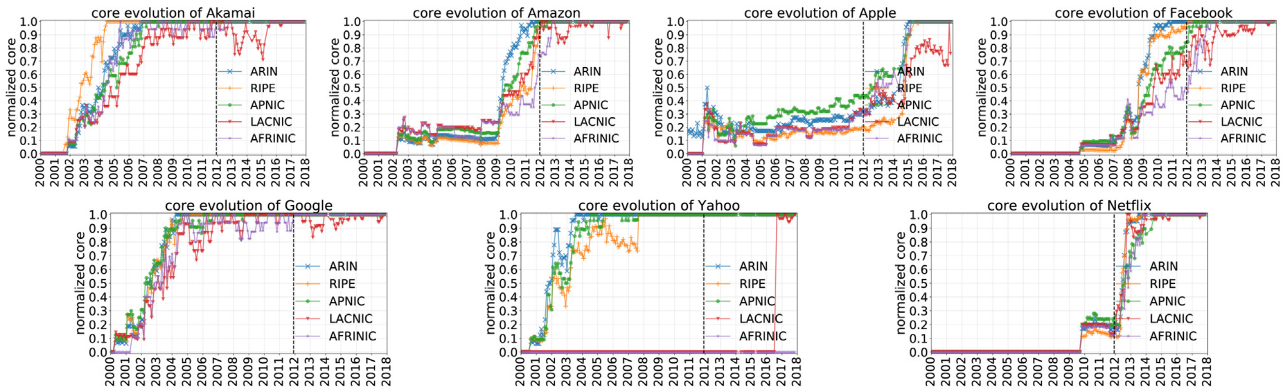


Fig. 4. k-core evolution of the *Big Seven* in each RIR. The dashed line displays the beginning of NetAcuity geolocation database.

1. We assume that the 7 CPs we study have always had a presence in every RIR. While building the RIR subgraph, however, we only include observed connectivity between the CPs and other ASes geolocated to the RIR.
2. We assume that prior to November 2011 (the start of our Netacuity dataset), ASes had the same locations that they had in November 2011.

While this methodology allows us to create RIR subgraphs, we cannot infer where the connection between two ASes actually happens when those ASes have presence in multiple RIR subgraphs. For instance, Google and Level3, which are currently present in every RIR subgraph, may not have a physical link in each RIR.

6.1. Geographical evolution of the Big Seven in IPv4

Fig. 4 shows the evolution of each CP by RIR. We find that all CPs have reached the IPv4 TOPcore in every RIR although the arrival date varies by CP and RIR.

Amazon and Facebook show differences between RIRs in their growth in the late 2000s and early 2010s. Amazon first established datacenters and PoPs in the US before 2009, then expanded to Singapore (APNIC) in 2010, Brazil (LACNIC) in 2011, and several locations in Europe (RIPE) in 2011 [57]. Fig. 4 shows that Amazon's core trends follow its documented infrastructure deployment. Facebook, which has been part of the worldwide IPv4 TOPcore since 2009, lagged in APNIC, LACNIC and AFRINIC, where it got to the TOPcore several years after ARIN and RIPE. Facebook got to the IPv4 TOPcore in ARIN in August 2010, APNIC in August 2012, LACNIC in August 2013 and in AFRINIC in March 2013. In RIPE, Facebook has been in the upper cores ($k^* \geq 0.9$) since early 2010, however, it finally reached the IPv4 TOPcore in January 2012. Facebook publicly acknowledged its lack of presence in developing regions and took steps to correct in order to improve user QoE in those regions [78].

Since the *Big Seven* are all U.S. based companies, one might expect that they first reached the IPv4 TOPcore in ARIN, and later expanded to developing regions such as LACNIC and AFRINIC. We investigate this hypothesis next, while noting that the analysis that follows is specific to these companies and may not generalize to other content providers or regions. Fig. 4 shows, however, that Akamai and Google showed negligible differences across RIRs in the early 2000s, which does not match documented information about their CDN deployment. For instance, Google established a PoP in Argentina only in 2011 [79]. The reason for this discrepancy is that Akamai and Google had peering links with tier-1 transit providers present in those regions, which caused the CPs to be in the TOPcore of those regions as well. A look at peering relationships in the early 2000s confirms this hypothesis — Google was not present in the LACNIC region, however, it peered with Level3 (AS3549), TATA (AS6453) and Qwest (AS209), which were present in LACNIC. We confirmed that the tier-1 ASes were present in LACNIC because they peered with the two largest Argentinian ISPs, Cablevision (AS10318) and TASA (AS4926), which were only present in Argentina at the time.

Table 2
Percentage of local peers in each region.

	%	ARIN	RIPE	APNIC	LACNIC	AFRINIC
Akamai	2007	0.33	0.75	0.21	0.0	0.05
	2012	0.45	0.74	**0.45**	0.11	0.0
	2017	0.41	0.71	0.47	**0.56**	**0.23**
Amazon	2007	1.0	0	0	0	0
	2012	*0.49*	**0.75**	**0.24**	**0.35**	0.0
	2017	0.40	0.68	0.37	0.53	0.03
Apple	2007	0.73	0	0.40	0	0
	2012	0.60	0.15	0.29	0	0
	2017	*0.42*	**0.67**	0.4	**0.17**	**0.07**
FB	2007	0.51	0.68	0.21	0.11	0.11
	2012	0.49	0.75	0.37	**0.36**	0.05
	2017	0.44	0.73	0.37	0.51	**0.14**
Google	2007	1.0	0	0	0	0
	2012	*0.43*	**0.81**	0.27	0.07	0.0
	2017	0.39	0.70	**0.38**	**0.56**	**0.07**
Yahoo!	2007	0.7	0.45	0.15	0	0
	2012	0.57	0.72	0.44	0	0
	2017	0.53	0.73	0.46	**0.6**	0
Netflix	2007	0	0	0	0	0
	2012	**0.86**	0.14	0	0	0
	2017	*0.39*	**0.77**	**0.39**	**0.57**	**0.10**

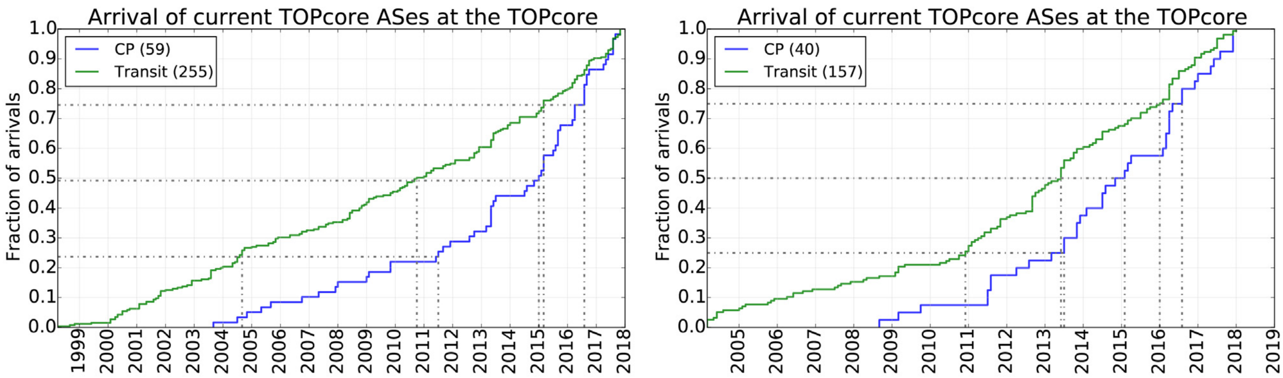
Table 3
Origin according to WHOIS for TOPcore ASes.

		ARIN	RIPE	APNIC	LACNIC	AFRINIC	Unknown
COREv4	Content	36	20	3	0	0	–
	Transit	35	165	38	3	8	6
COREv6	Content	25	14	1	0	0	–
	Transit	20	108	15	3	6	7

Similar Google and Akamai, Yahoo had similar core evolution trends in ARIN, RIPE and APNIC in early 2000s. However, Yahoo had a significant delay in LACNIC region where the company reached the TOPcore in 2016 when it joined the Brazilian IXP in Sao Paulo[7]

Netflix and Apple were the latest to enter the worldwide IPv4 TOPcore as well as the IPv4 TOPcore of each RIR. Netflix was in the lower cores ($k^* < 0.3$) in every RIR in January 2012. By January 2014 it moved to the IPv4 TOPcore in every RIR. Apple's growth was similar — in June 2014 it was in cores lower than 0.5. One year later it was

[7] Wayback Machine snapshot of members of Brazilian IXP. 09/2016 http://web.archive.org/web/20160904012004/http://ix.br/particip/sp.



a) IPv4 b) IPv6

Fig. 5. Date of first arrival at the TOPcore for ASes which currently compose the TOPcore.

in the IPv4 TOPcore of every RIR except LACNIC where it reached the TOPcore in Jan 2017.

6.2. Local peers

The analysis of the previous section showed that core evolution does not necessarily reflect the geographical expansion of CPs. Here we present a complementary analysis. Table 2 shows the percentage of peers of a CP in a region that are registered in that region (according to WHOIS records). For example, Google had 38% of local peers in APNIC in 2017, meaning that 38% of Google's links with ASes present in APNIC were with ASes registered in APNIC, while the remaining 62% were with ASes present in APNIC but registered elsewhere. This metric provides information about when a CP first arrived in a region, as that would intuitively lead to an increase in the local peering metric.

Table 2 shows that Akamai, Google and Yahoo! significantly increased the number of local peers in Latin America (LACNIC) in the last five years. APNIC has also shown a growth in the number of local peers, but slower than in LACNIC. In contrast to Fig. 4 where all of the CPs belong to every TOPcore, Table 2 shows a fairly low number of local peers of these CPs in AFRINIC. As of 2017, Akamai had the largest fraction with 0.23, Facebook second with 0.14 and all the rest were under 0.10.

While the percentage of local peers of CPs increases over the years in regions where they initially had a small fraction of local peers, ARIN shows the opposite trend. This is likely because the studied CPs are U.S. companies. Consequently, their number of local peers in ARIN saturates, while the number of non-local peers increases as companies outside the U.S. deploy infrastructure in ARIN and peer with the CPs.

7. The TOPcore beyond the Big Seven

We conclude our analysis by looking at other networks in the TOPcore. Specifically, we investigate four aspects related to this set of ASes: (i) Composition of the TOPcores (ii) Evolution of Dual-Stack adopters (iii) Time required to reach the TOPcore (iv) Trends of some other remarkable CPs that were not included in the *Big Seven*.

To identify ASes in the TOPcore, we use the criterion that an AS must be in $k^* > 0.975$ at any point in time, and in $k^* \geq 0.95$ during the last six months of our dataset (Mar-2017 to Oct-2017). Note that this definition of the TOPcore is broader than that used in the previous section where the criterion for belonging to the TOPcore was $k^* = 1$.

7.1. Composition of the TOPcores

We would like to investigate how many networks are in the TOPcore, what type of networks they are (transit or content), and what fraction of the TOPcore networks is accounted for by content networks.

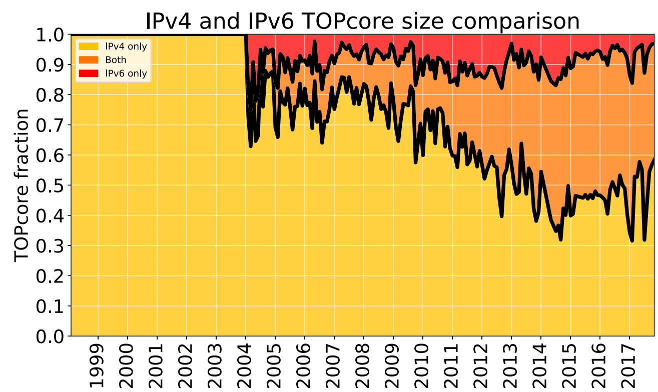


Fig. 6. Evolution of dual-stack ASes among members of both TOPcores.

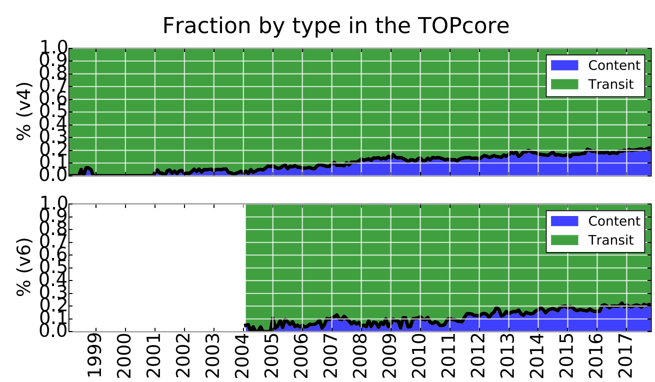
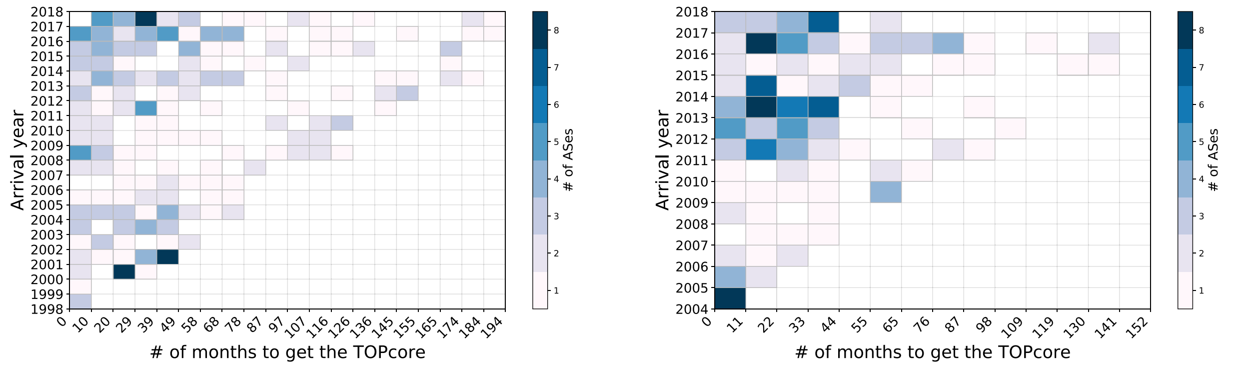


Fig. 7. Monthly evolution of the fraction of CPs and Transit in the TOPcore.

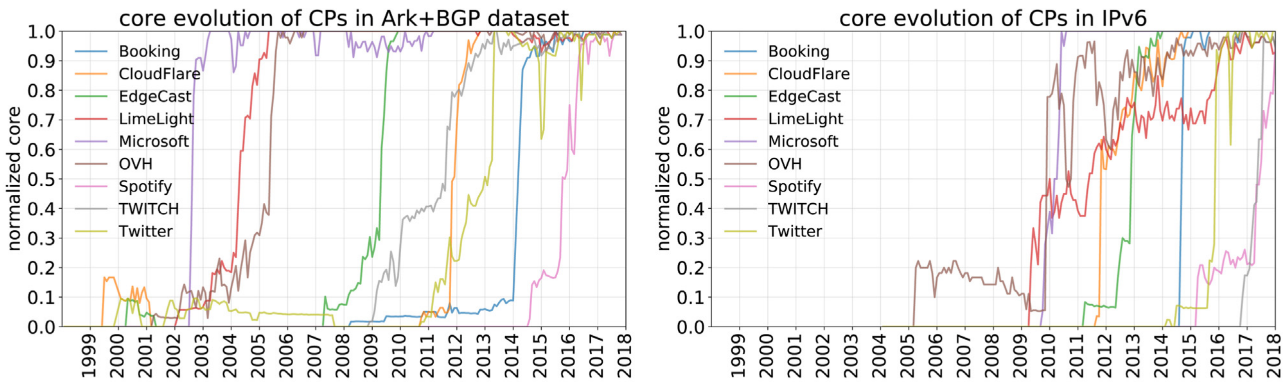
By the TOPcore definition, we had 314 ASes in the IPv4 TOPcore — 59 Content Providers and 255 Transit/Access Providers according to CAIDA's AS classification [80]. In the IPv6 TOPcore we found 197 ASes, where 40 are Content Providers and 157 Transit/Access providers. We refer to the set of ASes in IPv4 and IPv6 TOPcore as COREv4 and COREv6, respectively.

Fig. 5a shows the fraction of COREv4 (separated into Content and Transit) that first reached the IPv4 TOPcore over time. This plot clearly shows that over time, more CPs have been joining the TOPcore. Interestingly, 75% of the CPs in the studied set first entered the TOPcore after 2011. Moreover, we see two distinct phases in the CP curve — the rate at which CPs arrived in the TOPcore has increased since 2011. The arrival of Transit Providers, on the other hand, appears steady over the years.



(a) IPv4 (b) IPv6

Fig. 8. Correlation between speed of growth and date of arrival at the TOPcore.



a) IPv4 b) IPv6

Fig. 9. k-core evolution of CPs other than the *Big Seven*.

Fig. 5b displays the same analysis as in Fig. 5a but for the IPv6 TOPcore. While the trend for Transits in Fig. 5a linearly increased during the years, Fig. 5b shows an inflection point in early 2011 when IANA announced the allocation of its last /8 to the RIR [17]. With respect to Content Providers, 75% of CPs in COREv4 reached TOPcore after 2011, while more than 90% of CPs in COREv6 reached the TOPcore in the same period. The arrival of Transit and Content in the IPv6 TOPcore show an acceleration, especially for CPs, after ARIN announced its IPv4 pool reached zero in September 2015 [19].

Table 3 shows the geographical distribution of ASes in the TOPcores. We see that CPs in COREv4 and COREv6 are mostly from ARIN and RIPE (with the exception of 3 and 1 from APNIC in COREv4 and COREv6 respectively). However, among Transit Providers, RIPE has significantly more ASes in COREv4 as well as COREv6 than other regions. AFRINIC and LACNIC have negligible or no presence in either category. APNIC has a considerable number of Transit Providers in COREv4 and COREv6 but few CPs. Comparing the geographical composition of in COREv4 and COREv6 by category, both have exactly the same distribution. Therefore, the geographical distribution of densely-connected ASes is invariant to changes on the IP protocol.

7.2. Dual-stack in the TOPcore

Next, we analyze the fraction of ASes that belong simultaneously to both TOPcores in each snapshot since 1999. Fig. 6 displays the fraction of ASes in the IPv4 TOPcore, IPv6 TOPcore and in both. Since 2004 when IPv6 data starts, the fraction of ASes that only belong to the IPv6 TOPcore has been fluctuating around 10%. However, since then more and more ASes have been incorporating dual stack technology, which is reflected on the increase of ASes that belong to both TOPcores and the reduction of member that exclusively belong to the IPv4 TOPcore. In March 2018, the network indicates that roughly 50% the TOPcore ASes are reachable via IPv6. This figure lets us conclude that densely-connected ASes, which already are in the IPv4 TOPcore, are rolling out IPv6 but it is fairly rarely to find ASes that only belong to the IPv6 TOPcore.

We next investigate the composition of ASes in the TOPcore over time. In Fig. 7, we applied the TOPcore criterion to determine which ASes belong to the TOPcore every month, and then classified the ASes in the TOPcore as Content or Transit. We find that the fraction of CPs in both TOPcores has been steadily increasing; as of the October 2017 snapshot, 22% of ASes in both TOPcores were CPs. Note that the absolute number of ASes in the TOPcores has been increasing as well, which implies that both TOPcores have been incorporating more CPs than Transit ASes over time.

7.3. Speed to reach the TOPcore

We are interested in analyzing *how quickly* networks reached the TOPcore.

Figs. 8a and 8b show a heatmap of the number of ASes that arrived at the TOPcore at a certain time and at a certain *speed*. We define *speed* as the number of months to move from $k^* = 0.3$ to $k^* = 0.975$, and this definition is based on the transitions from lower to upper cores seen in Fig. 2a. Fig. 8a shows that 172 ASes from COREv4 joined the TOPcore between 2011 and 2018 and most of them moved from lower cores in just a few months, where the average time required for joining COREv4 was 61 months. Fig. 8b shows the counterpart for IPv6 where

154 ASes from COREv6 joined the TOPcore between 2011 and 2018. The average time required for transitioning from lower cores to the TOPcore in IPv6 was on average 35 months, which is smaller than IPv4 network. This fast evolution of the IPv4 TOPcore in recent years can be possibly explained by the growth of the number of peering facilities and participants at those facilities in this time frame.

7.4. Other remarkable CPs in the TOPcore

Finally, we study the core evolution of nine other remarkable CPs that belong to the TOPcore but were not included in the *Big Seven*. Seven of the nine selected ASes are the remaining ASes in Bottger et al.'s [47] TOP15 list, except Hurricane Electric (AS6939) which we do not consider as a CP since it is labeled as Transit/Access in CAIDA's AS classification [80]. These seven ASes are OVH (AS16276), LimeLight (AS22822), Microsoft (AS8075), Twitter (AS13414), Twitch (AS46489), CloudFlare (AS13335) and EdgeCast (AS15133). The other two ASes are Booking.com (AS43996) and Spotify (AS8403). Interestingly, Booking.com or Spotify are not normally considered among the top CPs, however, they are in both TOPcores.

Figs. 9a and 9b show the evolution of nine CPs that have joined the IPv4 and IPv6 TOPcores in recent years (different from the Big Seven). The figures also indicate that many rapidly transitioned from lower to upper cores.

Twitch is another remarkable CP in this list, which may not be as known as the *Big Seven* are, however, it is extremely popular among the gamer community. Twitch is a video streaming platform that allows its users to live stream what they are currently playing. The service is responsible for being the fourth traffic source of peak traffic in the US [81] and its audience is even larger than traditional media broadcasters [82]. Live streaming video is exclusively served by Twitch serving infrastructure (AS46489) that spreads over 21 airport codes and 12 countries [83]. Furthermore, looking at Twitch's records in PeeringDB, the CP peers at 47 IXPs all over the world [84]. Twitch's IPv4 CDN deployment is clearly evidenced in Fig. 9a, where it rapidly reached the TOPcore in 2014. Twitch is also present in COREv6 as shown in Fig. 9b. It is worth noting that according to this figure, Twitch IPv6 rollout happened in 2017.

We found trends similar trends in Figs. 9 and 3a (*Big Seven*). To begin with, ASes that reached the IPv4 TOPcore in early 2000s, such as LimeLight in Fig. 9 or Akamai in Fig. 3a, postponed IPv6 rollout. On the other hand, we also notice ASes that deployed their CDN in recent years are the ones that have less or no delay between the IPv4 and IPv6 core evolution. While Netflix evidences this pattern in Fig. 3a, so does Booking.com in Fig. 9.

8. Conclusions

In this work we demonstrated that CPs have taken a decisive role in the AS ecosystem, where seven large companies in the Internet content market have moved towards the core of the network. By analyzing the evolution of the cores of the CPs, we were able to identify possible reasons related to business practices, strategies, and geographical expansion that explain the rise of these networks to the top core. Furthermore, we showed the core of the network has been rapidly incorporating content ASes over time.

We also showed that most of the CPs as well as Transits reached the IPv6 TOPcore several years after reaching IPv4 TOPcore, which coincides with the fact that many ASes postponed IPv6 rollout. However, ASes were faster to reach IPv6 TOPcore since the physical infrastructure was already available by then.

We believe that analysis of core evolution can be a possible tool to identify ASes that are increasing in significance, the so-called "up and coming" CPs. We refer the reader to the following website to replicate our results: http://cnet.fi.uba.ar/TMA2018/

Acknowledgments

This work was partially founded by UBACyT, Argentina 2018 (20020170100421BA) and National Science Foundation, United States grant CNS-1513847. Esteban Carisimo acknowledges CONICET Argentina for a Ph.D. fellowship.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, F. Jahanian, Internet inter-domain traffic, in: Proceedings of the ACM SIGCOMM 2010 Conference, in: SIGCOMM '10, ACM, New York, NY, USA, 2010, pp. 75–86, http://dx.doi.org/10.1145/1851182.1851194, URL http://doi.acm.org/10.1145/1851182.1851194.

[2] k. claffy, G. Polyzos, H. Braun, Traffic characteristics of the T1 NSFNET backbone, in: IEEE Conference on Computer Communications (INFOCOM), Vol. 2, 1993, pp. 885–893.

[3] N. Chatzis, G. Smaragdakis, A. Feldmann, W. Willinger, There is more to ixps than meets the eye, SIGCOMM Comput. Commun. Rev. 43 (5) (2013) 19–28, http://dx.doi.org/10.1145/2541468.2541473, URL http://doi.acm.org/10.1145/2541468.2541473.

[4] G. Huston, The death of Transit, https://blog.apnic.net/2016/10/28/the-death-of-transit/ (2016).

[5] P. Gill, M. Arlitt, Z. Li, A. Mahanti, The flattening internet topology: Natural evolution, unsightly barnacles or contrived collapse? in: M. Claypool, S. Uhlig (Eds.), Passive and Active Network Measurement, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 1–10.

[6] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, W. Willinger, Anatomy of a large european ixp, in: Proceedings of the ACM SIGCOMM 2012 Conference, in: SIGCOMM '12, ACM, New York, NY, USA, 2012, pp. 163–174, http://dx.doi.org/10.1145/2342356.2342393, URL http://doi.acm.org/10.1145/2342356.2342393.

[7] T. Leighton, Improving performance on the internet, Commun. ACM 52 (2) (2009) 44–51, http://dx.doi.org/10.1145/1461928.1461944, URL http://doi.acm.org/10.1145/1461928.1461944.

[8] Wired, Google and Netflix Make Land Grab On Edge Of Internet, https://www.wired.com/2012/06/cdn/ (2016).

[9] S. Jain, A. Kumar, S. Mandal, J. Ong, L. Poutievski, A. Singh, S. Venkata, J. Wanderer, J. Zhou, M. Zhu, J. Zolla, U. Hölzle, S. Stuart, A. Vahdat, B4: Experience with a globally-deployed software defined wan, in: Proceedings of the ACM SIGCOMM 2013 Conference, in: SIGCOMM '13, ACM, New York, NY, USA, 2013, pp. 3–14, http://dx.doi.org/10.1145/2486001.2486019, URL http://doi.acm.org/10.1145/2486001.2486019.

[10] D.E. Eisenbud, C. Yi, C. Contavalli, C. Smith, R. Kononov, E. Mann-Hielscher, A. Cilingiroglu, B. Cheyney, W. Shang, J.D. Hosein, Maglev: A fast and reliable software network load balancer, in: 13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16), USENIX Association, Santa Clara, CA, 2016, pp. 523–535, URL https://www.usenix.org/conference/nsdi16/technical-sessions/presentation/eisenbud.

[11] B. Schlinker, H. Kim, T. Cui, E. Katz-Bassett, H.V. Madhyastha, I. Cunha, J. Quinn, S. Hasan, P. Lapukhov, H. Zeng, Engineering egress with edge fabric: Steering oceans of content to the world, in: Proceedings of the ACM SIGCOMM 2017 Conference, in: SIGCOMM '17, ACM, New York, NY, USA, 2017, pp. 418–431, http://dx.doi.org/10.1145/3098822.3098853, URL http://doi.acm.org/10.1145/3098822.3098853.

[12] A.-J. Su, D.R. Choffnes, A. Kuzmanovic, F.E. Bustamante, Drafting behind akamai: Inferring network conditions based on cdn redirections, IEEE/ACM Trans. Netw. 17 (6) (2009) 1752–1765, http://dx.doi.org/10.1109/TNET.2009.2022157.

[13] N. Economides, J. Tåg, Network neutrality on the internet: A two-sided market analysis, Inf. Econ. Policy 24 (2) (2012) 91–104.

[14] A. Dhamdhere, C. Dovrolis, The internet is flat: Modeling the transition from a transit hierarchy to a peering mesh, in: Proceedings of Conference on Emerging Networking EXperiments and Technologies, in: CoNEXT '10, ACM, New York, NY, USA, 2010, pp. 21:1–21:12, http://dx.doi.org/10.1145/1921168.1921196, URL http://doi.acm.org/10.1145/1921168.1921196.

[15] A. Dhamdhere, C. Dovrolis, Ten years in the evolution of the internet ecosystem, in: Proceedings of the 2008 Internet Measurement Conference, in: IMC '08, ACM, New York, NY, USA, 2008, pp. 183–196, http://dx.doi.org/10.1145/1452520.1452543, URL http://doi.acm.org/10.1145/1452520.1452543.

[16] P. Faratin, Economics of overlay networks: An industrial organization perspective on network economics, in: Proceedings of the NetEcon+ IBC Workshop, 2007, p. 1.

[17] IANA, The IANA IPv4 Address Free Pool is Now Depleted, https://www.arin.net/vault/announcements/2011/20110203.html (2011).
[18] Internet Society, World IPv6 Launch, www.worldipv6launch.org/ (2012).
[19] ARIN, ARIN IPv4 Free Pool Reaches Zero, https://www.arin.net/vault/announcements/2015/20150924.html. (2015).
[20] A. Dhamdhere, M. Luckie, B. Huffaker, k. claffy, A. Elmokashfi, E. Aben, Measuring the deployment of ipv6: Topology, routing and performance, in: Proceedings of the 2012 Internet Measurement Conference, in: IMC '12, ACM, New York, NY, USA, 2012, pp. 537–550, http://dx.doi.org/10.1145/2398776.2398832, URL http://doi.acm.org/10.1145/2398776.2398832.
[21] J. Dilley, B. Maggs, J. Parikh, H. Prokop, R. Sitaraman, B. Weihl, Globally distributed content delivery, IEEE Internet Comput. 6 (5) (2002) 50–58, http: //dx.doi.org/10.1109/MIC.2002.1036038.
[22] G. Pallis, A. Vakali, Insight and perspectives for content delivery networks, Commun. ACM 49 (1) (2006) 101–106, http://dx.doi.org/10.1145/1107458.1107462, URL http://doi.acm.org/10.1145/1107458.1107462.
[23] C. Huang, A. Wang, J. Li, K.W. Ross, Understanding hybrid cdn-p2p: Why limelight needs its own red swoosh, in: Proceedings of the 18th International Workshop on Network and Operating Systems Support for Digital Audio and Video, in: NOSSDAV '08, ACM, New York, NY, USA, 2008, pp. 75–80, http://dx.doi.org/10.1145/1496046.1496064, URL http://doi.acm.org/10.1145/1496046.1496064.
[24] M. Pathan, R. Buyya, A. Vakali, Content delivery networks: State of the art, insights, and imperatives, Content Deliv. Netw. (2008) 3–32.
[25] M. Calder, X. Fan, Z. Hu, E. Katz-Bassett, J. Heidemann, R. Govindan, Mapping the expansion of google's serving infrastructure, in: Proceedings of the 2013 Internet Measurement Conference, in: IMC '13, ACM, New York, NY, USA, 2013, pp. 313–326, http://dx.doi.org/10.1145/2504730.2504754, URL http://doi.acm.org/10.1145/2504730.2504754.
[26] P. Casas, A. D'Alconzo, P. Fiadino, A. Bär, A. Finamore, T. Zseby, When youtube does not work?analysis of qoe-relevant degradation in google cdn traffic, IEEE Trans. Netw. Serv. Manag. 11 (4) (2014) 441–457, http://dx.doi.org/10.1109/TNSM.2014.2377691.
[27] Q. Huang, K. Birman, R. van Renesse, W. Lloyd, S. Kumar, H.C. Li, An analysis of facebook photo caching, in: Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles, in: SOSP '13, ACM, New York, NY, USA, 2013, pp. 167–181, http://dx.doi.org/10.1145/2517349.2522722, URL http://doi.acm.org/10.1145/2517349.2522722.
[28] Y.-W.E. Sung, X. Tie, S.H. Wong, H. Zeng, Robotron: Top-down network management at facebook scale, in: Proceedings of the 2016 ACM SIGCOMM Conference, in: SIGCOMM '16, ACM, New York, NY, USA, 2016, pp. 426–439, http://dx.doi.org/10.1145/2934872.2934874, URL http://doi.acm.org/10.1145/2934872.2934874.
[29] T. Böttger, F. Cuadrado, G. Tyson, I. Castro, S. Uhlig, Open connect everywhere: A glimpse at the internet ecosystem through the lens of the netflix cdn, SIGCOMM Comput. Commun. Rev. 48 (1) (2018) 28–34, http://dx.doi.org/10.1145/3211852.3211857, URL http://doi.acm.org/10.1145/3211852.3211857.
[30] M. Calder, A. Flavel, E. Katz-Bassett, R. Mahajan, J. Padhye, Analyzing the performance of an anycast cdn, in: Proceedings of the 2015 Internet Measurement Conference, in: IMC '15, ACM, New York, NY, USA, 2015, pp. 531–537, http://dx.doi.org/10.1145/2815675.2815717, URL http://doi.acm.org/10.1145/2815675.2815717.
[31] A. Gupta, M. Calder, N. Feamster, M. Chetty, E. Calandro, E. Katz-Bassett, Peering at the internet's frontier: A first look at isp interconnectivity in africa, in: M. Faloutsos, A. Kuzmanovic (Eds.), Passive and Active Measurement, Springer International Publishing, Cham, 2014, pp. 204–213.
[32] R. Fanou, P. Francois, E. Aben, On the diversity of interdomain routing in africa, in: J. Mirkovic, Y. Liu (Eds.), Passive and Active Measurement, Springer International Publishing, Cham, 2015, pp. 41–54.
[33] Geoff Huston, What Drives IPv6 Deployment?, https://labs.ripe.net/Members/gih/what-drives-ipv6-deployment (2018).
[34] I. Livadariu, A. Elmokashfi, A. Dhamdhere, k. claffy, A first look at ipv4 transfer markets, in: Proceedings of Conference on Emerging Networking EXperiments and Technologies, in: CoNEXT '13, ACM, New York, NY, USA, 2013, pp. 7–12, http://dx.doi.org/10.1145/2535372.2535416, URL http://doi.acm.org/10.1145/2535372.2535416.
[35] I. Livadariu, A. Elmokashfi, A. Dhamdhere, On ipv4 transfer markets: Analyzing reported transfers and inferring transfers in the wild, Comput. Commun. 111 (2017) 105–119.
[36] Internet Society, Google buys a /12 IPv4 Address Block, https://www.internetsociety.org/blog/2017/05/google-buys-a-12-ipv4-address-block/ (2017).
[37] Internet Society, MIT Goes on IPv4 Selling Spree, https://www.internetsociety.org/blog/2017/05/mit-goes-on-ipv4-selling-spree/ (2017).
[38] J.I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, A. Vespignani, K-core decomposition of Internet graphs: hierarchies, self-similarity and measurement biases, Netw. Heterogeneous Media 3 (2) (2008) 371–393.
[39] S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, K-core organization of complex networks, Phys. Rev. Lett. 96 (4) (2006) 040601.
[40] C. Orsini, E. Gregori, L. Lenzini, D. Krioukov, Evolution of the internet k-dense structure, IEEE/ACM Trans. Netw. 22 (6) (2014) 1769–1780, http://dx.doi.org/10.1109/TNET.2013.2282756.
[41] M. Faloutsos, P. Faloutsos, C. Faloutsos, On power-law relationships of the internet topology, SIGCOMM Comput. Commun. Rev. 29 (4) (1999) 251–262.
[42] M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas, k. claffy, As relationships, customer cones, and validation, in: Proceedings of the 2013 Internet Measurement Conference, in: IMC '13, ACM, New York, NY, USA, 2013, pp. 243–256, http://dx.doi.org/10.1145/2504730.2504735, URL http://doi.acm.org/10.1145/2504730.2504735.
[43] R. Pastor-Satorras, A. Vázquez, A. Vespignani, Dynamical and correlation properties of the internet, Phys. Rev. Lett. 87 (25) (2001) 258701.
[44] V. Batagelj, M. Zaveršnik, Fast algorithms for determining (generalized) core groups in social networks, Adv. Data Anal. Classif. 5 (2) (2011) 129–145.
[45] M.G. Beiró, J.I. Alvarez-Hamelin, J.R. Busch, A low complexity visualization tool that helps to perform complex systems analysis, New J. Phys. 10 (12) (2008) 125003.
[46] CAIDA, AS Rank, http://as-rank.caida.org/ (07 2018).
[47] T. Böttger, F. Cuadrado, S. Uhlig, Looking for hypergiants in peeringdb, SIGCOMM Comput. Commun. Rev. 48 (3) (2018) 13–19, http://dx.doi.org/10.1145/3276799.3276801, URL http://doi.acm.org/10.1145/3276799.3276801.
[48] R. Oliveira, D. Pei, W. Willinger, B. Zhang, L. Zhang, The (in)completeness of the observed internet as-level structure, IEEE/ACM Trans. Netw. 18 (1) (2010) 109–122, http://dx.doi.org/10.1109/TNET.2009.2020798.
[49] Y.-C. Chiu, B. Schlinker, A.B. Radhakrishnan, E. Katz-Bassett, R. Govindan, Are we one hop away from a better internet? in: Proceedings of the 2015 Internet Measurement Conference, in: IMC '15, ACM, New York, NY, USA, 2015, pp. 523–529, http://dx.doi.org/10.1145/2815675.2815719, URL http://doi.acm.org/10.1145/2815675.2815719.
[50] V. Giotsas, M. Luckie, B. Huffaker, K. Claffy, Ipv6 as relationships, cliques, and congruence, in: J. Mirkovic, Y. Liu (Eds.), Passive and Active Measurement, Springer International Publishing, Cham, 2015, pp. 111–122.
[51] CAIDA, Archipelago (Ark) Measurement Infrastructure, https://www.caida.org/projects/ark/.
[52] Y. Hyun, A. Broido, k. claffy, Traceroute and BGP AS Path Incongruities, Tech. rep., Cooperative Association for Internet Data Analysis (CAIDA) (Mar 2003).
[53] Stefan Meinders, The New Internet, ENOG11 (2016).
[54] CAIDA, Mapping Autonomous Systems to Organizations: CAIDAs Inference Methodology, https://www.caida.org/research/topology/as2org/.
[55] Sandvine, Global internet phenomena report Spring 2011 (2011).
[56] V. Stocker, G. Smaragdakis, W. Lehr, S. Bauer, The growing complexity of content delivery networks: Challenges and implications for the internet ecosystem, Telecommun. Policy 41 (10) (2017) 1003–1016.
[57] Amazon, AWS global infrastructure, https://aws.amazon.com/es/about-aws/global-infrastructure/ (2017).
[58] Apple Insider, Apples in-house CDN efforts spell trouble for Akamai as infrastructure biz warns of losses, https://bit.ly/2I4mbBl (2016).
[59] ARS Technica, Apples multi-terabit, $100M CDN is live - with paid connection to Comcast, https://bit.ly/2KgVADT (2014).
[60] LA Times, Apple's original TV production to begin small: We are just starting out, https://bit.ly/2VXUIGf (2017).
[61] Yahoo Finance, Number of active users at Facebook over the years, https://finance.yahoo.com/news/number-active-users-facebook-over-years-214600186--finance.html (2012).
[62] Tom Hormby, The Rise of Google: Beating Yahoo at Its Own Game, http: //lowendmac.com/2013/the-rise-of-google-beating-yahoo-at-its-own-game/.
[63] New York Times, Google to Acquire YouTube for $1.65 Billion, http://www.nytimes.com/2006/10/09/business/09cnd-deal.html.
[64] Netflix Media Center, "Announcing the Netflix Open Connect Network", https://media.netflix.com/en/company-blog/announcing-the-netflix-open-connect-network (2012).
[65] Seeking Alpha, Apple, Microsoft And Facebook Bring More Traffic To In-House CDNs, Impacting Akamais Media Business, https://seekingalpha.com/article/3613736-apple-microsoft-facebook-bring-traffic-house-cdns-impacting-akamais-media-business (2015).
[66] Geoff Huston, Measuring IPv6 Deployment, https://meetings.ripe.net/ripe-56/presentations/Huston-Measuring_IPv6_Deployment.pdf (2008).
[67] C. Kaufmann, Akamai's V6 Rollout Plan and Experience from a CDN Point of View, MENOG9, 2011.
[68] Amazon Web Services, Elastic Load Balancing Announces Support for IPv6, Zone Apex Support and Security Group Integration, https://aws.amazon.com/es/about-aws/whats-new/2011/05/24/elb-ipv6-zoneapex-securitygroups/ (2011).
[69] Amazon, AWS IPv6 Update ? Global Support Spanning 15 Regions & Multiple AWS Services, https://aws.amazon.com/es/blogs/aws/aws-ipv6-update-global-support-spanning-15-regions-multiple-aws-services/ (2017).
[70] Dyn Blog, IPv6: One Operating System at a Time, https://dyn.com/blog/ipv6-one-operating-system-at-a-time-2/ (2015).
[71] Lorenzo Colitti, IPv6 at Google, https://www.ripe.net/participate/meetings/roundtable/february-2009/LorenzoIPv6atGoogle.pdf (2009).
[72] Facebook Engineering, Adding :face: to every IP: Celebrating IPv6's one-year anniversary, https://www.facebook.com/notes/facebook-engineering/adding-face-to-every-ip-celebrating-ipv6s-one-year-anniversary/10151492544578920/ (2009).

[73] Netflix Tech Blog, Enabling Support for IPv6, https://medium.com/netflix-techblog/enabling-support-for-ipv6-48a495d5196f (2012).
[74] Netflix Tech Blog, Building fast.com, https://medium.com/netflix-techblog/building-fast-com-4857fe0f8adb (2016).
[75] Digital Element, NetAcuity, https://www.digitalelement.com/solutions/.
[76] I. Poese, S. Uhlig, M.A. Kaafar, B. Donnet, B. Gueye, Ip geolocation databases: Unreliable? SIGCOMM Comput. Commun. Rev. 41 (2) (2011) 53–56, http://dx.doi.org/10.1145/1971162.1971171, URL http://doi.acm.org/10.1145/1971162.1971171.
[77] M. Gharaibeh, A. Shah, B. Huffaker, H. Zhang, R. Ensafi, C. Papadopoulos, A look at router geolocation in public and commercial databases, in: Proceedings of the 2017 Internet Measurement Conference, in: IMC '17, ACM, New York, NY, USA, 2017, pp. 463–469, http://dx.doi.org/10.1145/3131365.3131380, URL http://doi.acm.org/10.1145/3131365.3131380.
[78] Quinn, James, Being Open: How Facebook Got Its Edge, NANOG68, 2016.
[79] Galperin, Hernan, Connectivity in Latin America and the Caribbean: The role of Internet Exchange Points, Internet Society, 2013.
[80] CAIDA, CAIDAs AS classification list, http://data.caida.org/datasets/as-classification/.
[81] Twitch blog, Twitch is 4th in Peak US Internet Traffic, https://blog.twitch.tv/twitch-is-4th-in-peak-us-internet-traffic-90b1295af358 (2014).
[82] Business insider, Amazons streaming service Twitch is pulling in as many viewers as CNN and MSNBC, https://www.businessinsider.com/twitch-is-bigger-than-cnn-msnbc-2018-2 (2018).
[83] J. Deng, G. Tyson, F. Cuadrado, S. Uhlig, Internet scale user-generated live video streaming: The twitch case, in: M.A. Kaafar, S. Uhlig, J. Amann (Eds.), Passive and Active Measurement, Springer International Publishing, Cham, 2017, pp. 60–71.
[84] Peering DB, TWITCH (AS46489) entry at Peering DB, https://www.peeringdb.com/net/1956 (2018).