# Studying the Evolution of Content Providers in the Internet Core

Esteban Carisimo[†§*], Carlos Selmo[‡], J. Ignacio Alvarez-Hamelin[†‡*] and Amogh Dhamdhere[§]
† Universidad de Buenos Aires, ‡ Instituto Tecnológico de Buenos Aires, § CAIDA/UC San Diego, * CONICET
{carisimo,ihameli}@cnet.fi.uba.ar, cselmo@itba.edu.ar, amogh@caida.org

***Abstract*—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. 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.**

## I. 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].

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].

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 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 AS-level internetwork over the last two decades. We first focus on seven large CPs, 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. We then take a broader view, characterizing the set of ASes in the core of the 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/.

## II. 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 [17]–[20], 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 [21], QoE performance [22], 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 [23], network administration [24], and Facebook's SDN [11]. Bottger et al. [25] 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 [26].

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 [27], [28].

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"*.

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 [29]–[31]. 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.

## III. Methodology and dataset

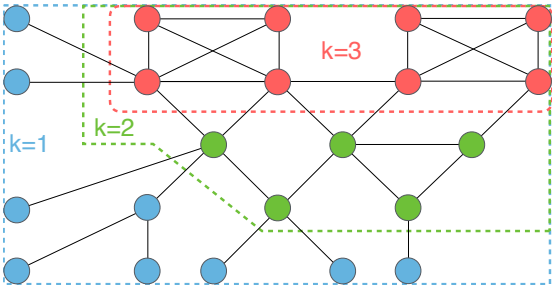


Fig. 1. Example of a k-core decomposition of a given graph.

**k-core decomposition:** 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.

We refer to the *core* of the network as the subset of ASes that are densely connected. To compute the set of such ASes, we use $k$-core decomposition, a naturally applicable tool from the graph-theoretic literature. Although simpler graph metrics, such as node degree, may indicate whether an AS is densely connected or not, these metrics are not as robust as the k-core. For instance, to define the core using node degree, it would be necessary to set a threshold, while the k-core inherently defines a *community* of densely connected nodes.

A $k$-core of a graph $\mathcal{G}$ is the maximum induced subgraph in which all the vertices have at least degree $k$ (see [32]). 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. Figure 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* [33], which means that there are $k$ different paths between two ASes of the same $k$-core.

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 [33], which also provides network visualization, and NetworkX, a python library.

**AS graph datasets:** To apply the above k-core decomposition methodology on the Internet graph longitudinally,

we need periodic historical snapshots of the 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 dataset ([1]) is derived from BGP dumps taken from RouteViews and RIPE RIS collectors [34] from 1998 to present, and contains AS links observed at the BGP collectors along with an inferred business relationships. We use a second dataset which consists of AS links extracted from traceroutes from CAIDA's Archipelago [35] vantage points towards every routed /24 prefix[2]. The two 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 [36]. To get the most complete picture of 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. To view the *k*-core decomposition using only the BGP dataset or traceroute dataset, we refer the reader to a website with these visualizations.[3]

A limitation of our methodology is that CPs also serve content from caches located within ISPs [12], [25], 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; we leave an analysis of cache infrastructure to future work.

## IV. 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], [37]. 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), Microsoft (AS8075) and Netflix (AS2906). Although CPs may have more than one ASN, we study the evolution of their primary ASNs. Publicly available AS sibling datasets can be incomplete and need semi-manual verification; we leave a consideration of sibling ASes for future work.

We chose these CPs based on publicly available information such as PeeringDB [38] and Sandvine reports [39]. According to their PeeringDB records, Akamai, Facebook and Netflix have heavily outbound traffic with levels over 10 Tbps, 1 Tbps and 1 Tbps, respectively. In addition, Sandvine reported in 2016 that Netflix dominated the peak period traffic with 35% of the traffic share, followed by YouTube with 17% and Amazon Video with 4% [39] in North America. The report also mentioned that FaceTime and iCloud (Apple) and Skype and Xbox (Microsoft) are among the top sources of peak period traffic. Cloud computing is also responsible for large data transfers, and this market is led by Amazon with 42% of the share, Microsoft 15% and Google 7% [40].

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.

We also investigate whether CPs belong to the TOPcore in each geographical region, defined as the Regional Internet Registries (RIR) regions. We repeat the analysis of speed and date of arrival 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.

### A. Tracking the evolution of the Big Seven

Figure 2 shows the monthly evolution of the normalized CP-core on the Ark+BGP dataset. A first observation is that as of the end of 2017, all the studied CPs have already joined the TOPcore, which is 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 Microsoft 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 [41] as multimedia content began to dominate the Internet traffic share [42]. 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], [43].

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.

*a) 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

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

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

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

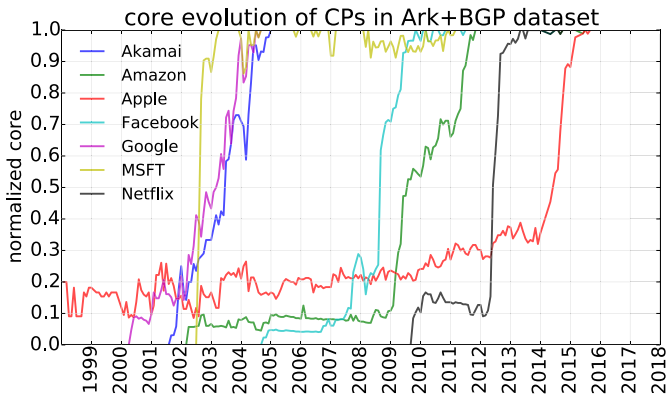


Fig. 2. k-core evolution of the *Big Seven*. All of these CPs have reached the TOPcore.

rather than generating content, they have always aimed to have a large number of peers. Moreover, Akamai acquired Speedera networks, a rival third-party CDN, in 2005 to consolidate its market position as well as to enlarge its platform. According to Figure 2, Akamai had already reached the TOPcore when it purchased Speedera networks.

*b) Amazon:* Amazon's infrastructure deployment appears to have occurred in two steps, according to Figure 2. This is further corroborated by information provided on Amazon's website [44]. In 2009 Amazon established its datacenter in Northern California, which matches with the first growth. Between 2010 and 2012, Amazon established datacenters in several parts of the world, which would explain the second growth spurt from 2010 to 2012. In addition to the datacenter deployment, Amazon established dozens of PoPs all over the world to boost expansion, which correlates with its rise to the TOPcore. Finally, we find that the WHOIS record for Amazon's DNS nameservers zone (e.g. `awsdns-39.net`) was created in late 2010, which coincides with the last spurt in its core growth. DNS nameservers are essential elements of Amazon's cloud infrastructure, required to load balance traffic among locations. For instance, Slack, which is hosted on Amazon, has `slack.com` NS records pointing to `ns-606.awsdns-11.net` among other AWS nameservers.

*c) Apple:* We find that Apple's AS reached the TOPcore in 2015 after a fairly quick growth. According to publicly documented information, Apple has been steadily off-loading its content from Akamai onto its own CDN since 2013 [45]. Apple's traffic share has been growing rapidly in recent years fueled by large data transfers due to software updates, such a new OS releases [46] or security patches. This is one of the motivating reasons for Apple to build its own CDN. 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 [47].

*d) 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 [48] which coincides with Facebook's expansion period and rise to the TOPcore. Although Facebook has kept on growing exponentially since then, the massive growth during that period encouraged Facebook to establish multiple peering agreements that enabled it to reach the TOPcore. In addition, the WHOIS record for `fbcdn.net`, which stands for Facebook CDN, was created in 2007 when Facebook's expansion was happening.

*e) Google:* Google was launched in September 1997 and in just a couple of years became the most popular search engine [49]. Over time, as Google started serving large volumes of video traffic via the acquisition of YouTube in 2006 [50], it expanded its CDN to get as close as possible to "eyeball" networks and achieve high QoE for users. However, looking carefully into Google's peers in the early days, between 1999 and 2003, even before the CDN was deployed, it had several agreements with tier-1 transit providers. Before December 2002, Google already peered with 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.

*f) Microsoft:* Similar to Google, Microsoft has been serving large volumes of online content since the mid-1990s, such as hotmail and MSN. Figure 2 shows that Microsoft entered the TOPcore in late 2002. An analysis of Microsoft's peers shows that by December 2002, Microsoft had a large number of connections with tier-1 Transit Providers, e.g., Level3 (AS3549), TATA (AS6453), Telia (AS1299), Telefonica (AS12956), Sprint (AS1239) and NTT (AS2914). The presence of peering links with multiple tier-1 ASes which were in the TOPcore resulted in Microsoft also entering the TOPcore.

*g) Netflix:* In 2012, it took Netflix less than a year to move from core $k^* = 0.1$ to the TOPcore. Netflix started to offer 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 [51] platform in 2012. Netflix's strategy to switch from third-party CDNs to its own infrastructure manifests itself as a sharp increase in its normalized core value between 01/2012 and 09/2012 as shown in Figure 2. The transition of Netflix from using third-party CDNs to using its OpenConnect platform also led to a number of peering disputes with large access providers over interconnection fees, e.g., with Comcast in 2013 [52].

In summary, all of the studied CPs moved from third-party CDNs to private CDNs and entered the TOPcore. In particular, Apple, Facebook, Microsoft 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 [41]. Despite losing major clients, Figure 2 shows that Akamai is still in the TOPcore, which means that Akamai's peering agreements do not depend exclusively on these large clients.

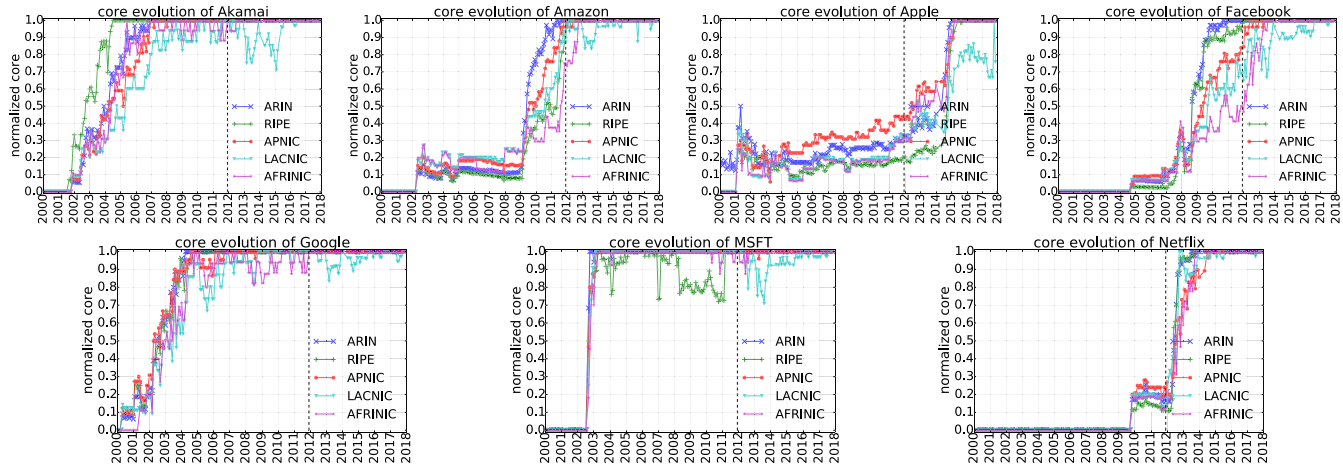


Fig. 3. k-core evolution of the *Big Seven* in each RIR. The dashed line displays the beginning of NetAcuity geolocation database.

### B. Evolution by geographical region

Next, we compare the evolution of the *Big Seven* by geographical region. To determine which regions an AS is present in, we use the NetAcuity [53] geolocation database to geolocate each prefix advertised by an AS in a given snapshot. The (in)accuracy of geolocation databases has been studied extensively [54]. However, previous work has found that the NetAcuity database is mostly reliable for country-level geolocation [55]. 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 [4] 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.

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.

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

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.

*1) Geographical evolution of the Big Seven:* Figure 3 shows the evolution of each CP by RIR. We find that all CPs have reached the 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 [44]. Figure 3 shows that Amazon's core trends follow its documented infrastructure deployment. Facebook, which has been part of the worldwide 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 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 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 [56].

Since the *Big Seven* are all U.S. based companies, one might expect that they first reached the TOPcore in ARIN, and later expanded to developing regions such as LACNIC and AFRINIC. Figure 3 shows, however, that Akamai, Google and Microsoft 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 [57]. The reason for this discrepancy is that Akamai, Google and Microsoft had peering links with tier-1 transit providers present in those regions,

TABLE I
PERCENTAGE OF LOCAL PEERS IN EACH REGION

| | % | Akamai | Amazon | Apple | Facebook | Google | MSFT | Netflix |
|---|---|---|---|---|---|---|---|---|
| ARIN | 2007 | 0.33 | 1.0 | 0.73 | 0.51 | 1.0 | 0.68 | 0 |
| | 2012 | 0.45 | 0.49 | 0.6 | 0.49 | 0.43 | 0.52 | **0.86** |
| | 2017 | 0.41 | 0.4 | 0.42 | 0.44 | 0.39 | 0.43 | 0.39 |
| RIPE | 2007 | 0.75 | 0.0 | 0.0 | 0.68 | 0.0 | 0.42 | 0 |
| | 2012 | 0.74 | **0.75** | 0.15 | 0.75 | **0.81** | **0.76** | 0.14 |
| | 2017 | 0.71 | 0.68 | **0.67** | 0.73 | 0.7 | 0.7 | **0.77** |
| APNIC | 2007 | 0.21 | 0.0 | 0.4 | 0.21 | 0.0 | 0.13 | 0 |
| | 2012 | **0.45** | **0.24** | 0.29 | 0.37 | 0.27 | 0.24 | 0.0 |
| | 2017 | 0.47 | 0.37 | 0.4 | 0.37 | **0.38** | **0.37** | **0.39** |
| LACNIC | 2007 | 0.0 | 0.0 | 0.0 | 0.11 | 0.0 | 0.0 | 0 |
| | 2012 | 0.11 | **0.35** | 0.0 | 0.36 | 0.07 | 0.08 | 0.0 |
| | 2017 | **0.56** | 0.53 | **0.17** | **0.51** | **0.56** | **0.5** | **0.57** |
| AFRINIC | 2007 | 0.05 | 0.0 | 0.0 | 0.11 | 0.0 | 0.04 | 0 |
| | 2012 | 0.0 | 0.0 | 0.0 | 0.05 | 0.0 | 0.0 | 0.0 |
| | 2017 | **0.23** | 0.03 | 0.07 | **0.14** | **0.07** | **0.1** | **0.1** |

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.

Netflix and Apple were the latest to enter the worldwide TOPcore as well as the 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 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 in the TOPcore of every RIR except LACNIC where it reached the TOPcore in Jan 2017.

*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 I 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 I shows that Akamai, Google and Microsoft 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 Figure 3 where all of the CPs belong to every TOPcore, Table I 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 $0.10$.

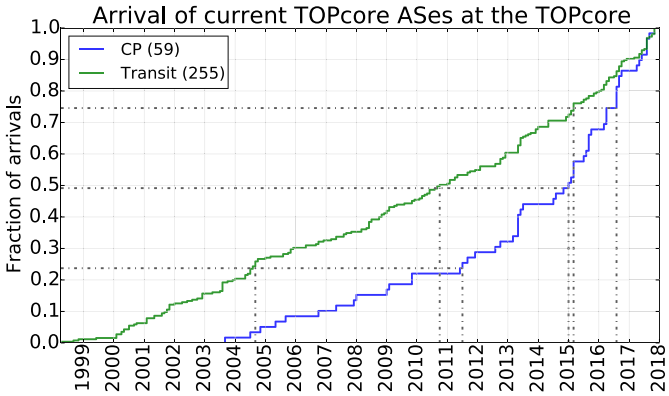


Fig. 4. Date of first arrival at the TOPcore for ASes which currently compose the TOPcore.

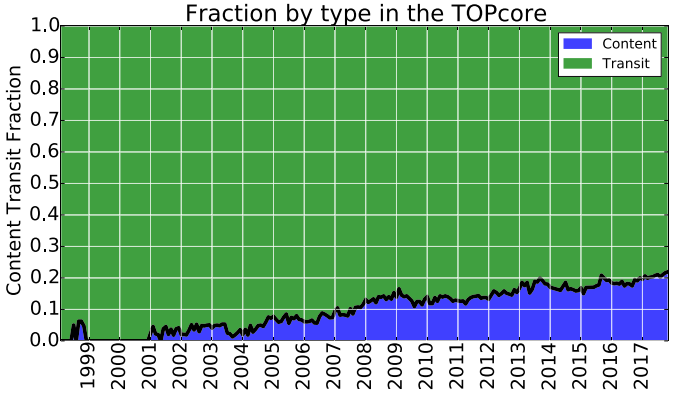


Fig. 5. Monthly evolution of the fraction of CPs and Transit in the TOPcore.

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.

## V. THE TOPCORE BEYOND THE BIG SEVEN

We conclude our analysis by looking at other networks in the TOPcore. Specifically, we investigate how many networks are in the TOPcore, what type of networks they are (transit or content), what fraction of the TOPcore networks are accounted for by content networks, and *how quickly* those networks reached the TOPcore. 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 $k^* = 1$. By this broader definition, we had 314 ASes in the TOPcore — 59 Content Providers and 255 Transit/Access Providers according to CAIDA's AS classification [58].

TABLE II
ORIGIN ACCORDING TO WHOIS FOR TOPCORE ASES

| | ARIN | RIPE | APNIC | LACNIC | AFRINIC | Unknown |
|---|---|---|---|---|---|---|
| Content | 36 | 20 | 3 | 0 | 0 | - |
| Transit | 35 | 165 | 38 | 3 | 8 | 6 |

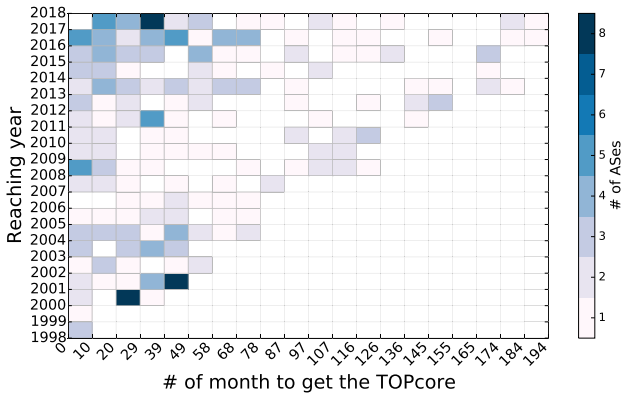


Fig. 6. Correlation between speed of growth and date of arrival at the TOPcore.

Figure 4 shows the fraction of these 314 ASes (separated into Content and Transit) that first reached the TOPcore over time. This plot clearly shows that over time, more CPs have been peering aggressively and 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 arrive in the TOPcore has increased since 2011. The arrival of Transit Providers, on the other hand, appears steady over the years.

Table II shows the geographical distribution of ASes in the TOPcore. We see that CPs in the TOPcore are mostly from ARIN and RIPE (with the exception of 3 from APNIC). However, among Transit Providers, RIPE has significantly more ASes than other regions. AFRINIC and LACNIC have negligible or no presence in either category. APNIC has a considerable number of Transit Providers but few CPs.

Figure 6 shows 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 Figure 2. Figure 6 shows that 172 of the 314 ASes joined the TOPcore between 2011 and 2018 and most of them moved from lower cores in just a few months, where the average speed was 61 months. This fast evolution of the 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.

Next, we investigate the composition of ASes in the TOPcore over time. In Figure 5, 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 the TOPcore has been steadily increasing; as of the October 2017 snaspshot,

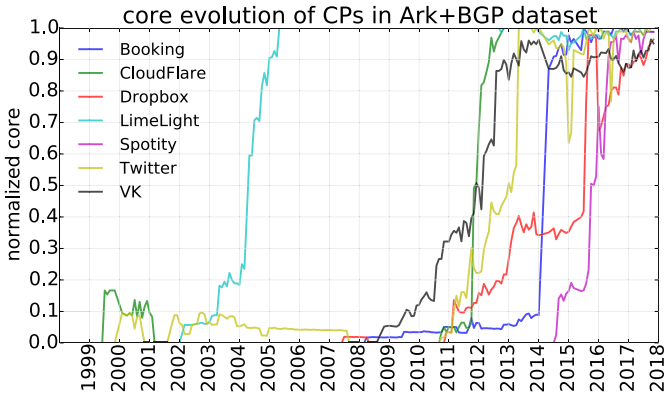


Fig. 7. k-core evolution of CPs other than the *Big Seven*.

22% of ASes in the TOPcore were CPs. Note that the absolute number of ASes in the TOPcore has been increasing as well, which implies that the TOPcore has been incorporating more CPs than Transit ASes over time.

Finally, Figure 7 shows the evolution of seven CPs that have joined the TOPcore in recent years (different from the Big Seven). Interestingly, there are ASes in this set that are not normally considered among the top CPs, such as Booking.com or Spotify. 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/

## VI. 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.

## VII. ACKNOWLEDGMENTS

This work was partially founded by UBACyT 2014 (20020130200122BA) and NSF grant CNS-1513847. EC acknowledges CONICET Argentina for a PhD fellowship.

## REFERENCES

[1] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, "Internet inter-domain traffic," in *ACM SIGCOMM*. ACM, 2010.
[2] K. C. Claffy, G. C. Polyzos, and H.-W. Braun, "Traffic characteristics of the t1 nsfnet backbone," in *INFOCOM'93*. IEEE, 1993.
[3] N. Chatzis, G. Smaragdakis, A. Feldmann, and W. Willinger, "There is more to IXPs than meets the eye," *ACM SIGCOMM CCR*, 2013.
[4] G. Huston, "The death of Transit,," 2016.
[5] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "The flattening internet topology: Natural evolution, unsightly barnacles or contrived collapse?" in *Proceedings of PAM*. Springer, 2008.

[6] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, and W. Willinger, "Anatomy of a large european ixp," in *Proceedings of ACM SIGCOMM*, 2012.

[7] T. Leighton, "Improving performance on the internet," *Comm. of the ACM*, vol. 52, no. 2, pp. 44–51, 2009.

[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 *et al.*, "B4: Experience with a globally-deployed software defined wan," *ACM SIGCOMM CCR*, vol. 43, no. 4, pp. 3–14, 2013.

[10] D. E. Eisenbud, C. Yi, C. Contavalli, C. Smith, R. Kononov, E. Mann-Hielscher, A. Cilingiroglu, B. Cheyney, W. Shang, and J. D. Hosein, "Maglev: A fast and reliable software network load balancer." in *NSDI*, 2016.

[11] B. Schlinker, H. Kim, T. Cui, E. Katz-Bassett, H. V. Madhyastha, I. Cunha, J. Quinn, S. Hasan, P. Lapukhov, and H. Zeng, "Engineering egress with edge fabric: Steering oceans of content to the world," in *Proceedings of ACM SIGCOMM*, 2017.

[12] A.-J. Su, D. R. Choffnes, A. Kuzmanovic, and F. E. Bustamante, "Drafting behind akamai: Inferring network conditions based on cdn redirections," *IEEE/ACM ToN*, vol. 17, no. 6, pp. 1752–1765, 2009.

[13] N. Economides and J. Tåg, "Network neutrality on the internet: A two-sided market analysis," *Information Economics and Policy*, vol. 24, no. 2, pp. 91–104, 2012.

[14] A. Dhamdhere and C. Dovrolis, "The internet is flat: Modeling the transition from a transit hierarchy to a peering mesh," in *Proceedings of ACM CoNEXT*. ACM, 2010.

[15] ——, "Ten years in the evolution of the internet ecosystem," in *Proceedings of ACM IMC*, 2008.

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

[17] J. Dilley, B. Maggs, J. Parikh, H. Prokop, R. Sitaraman, and B. Weihl, "Globally distributed content delivery," *IEEE Internet Computing*, vol. 6, no. 5, pp. 50–58, 2002.

[18] G. Pallis and A. Vakali, "Insight and perspectives for content delivery networks," *Comm. of the ACM*, vol. 49, no. 1, pp. 101–106, 2006.

[19] C. Huang, A. Wang, J. Li, and K. W. Ross, "Understanding hybrid cdn-p2p: why limelight needs its own red swoosh," in *Proceedings of NOSSDAV*, 2008.

[20] M. Pathan, R. Buyya, and A. Vakali, "Content delivery networks: State of the art, insights, and imperatives," *Content Delivery Networks*, pp. 3–32, 2008.

[21] M. Calder, X. Fan, Z. Hu, E. Katz-Bassett, J. Heidemann, and R. Govindan, "Mapping the expansion of google's serving infrastructure," in *Proceedings of ACM IMC*, 2013.

[22] P. Casas, A. D'Alconzo, P. Fiadino, A. Bär, A. Finamore, and T. Zseby, "When youtube does not work?analysis of qoe-relevant degradation in google cdn traffic," *IEEE TNSM*, vol. 11, no. 4, 2014.

[23] Q. Huang, K. Birman, R. van Renesse, W. Lloyd, S. Kumar, and H. C. Li, "An analysis of facebook photo caching," in *Proceedings of ACM SOSP*, 2013.

[24] Y.-W. E. Sung, X. Tie, S. H. Wong, and H. Zeng, "Robotron: Top-down network management at facebook scale," in *Proceedings of ACM SIGCOMM*, 2016.

[25] T. Böttger, F. Cuadrado, G. Tyson, I. Castro, and S. Uhlig, "A hyper-giant's view of the internet," *ACM SIGCOMM CCR*, 2017.

[26] M. Calder, A. Flavel, E. Katz-Bassett, R. Mahajan, and J. Padhye, "Analyzing the performance of an anycast cdn," in *Proceedings of ACM IMC*, 2015.

[27] A. Gupta, M. Calder, N. Feamster, M. Chetty, E. Calandro, and E. Katz-Bassett, "Peering at the internet's frontier: A first look at isp interconnectivity in africa." in *PAM*, 2014.

[28] R. Fanou, P. Francois, and E. Aben, "On the diversity of interdomain routing in africa," in *PAM*, 2015.

[29] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, "K-core decomposition of Internet graphs: hierarchies, self-similarity and measurement biases," *Networks and Heterogeneous Media*, vol. 3, no. 2, pp. 371–293, 2008.

[30] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "K-core organization of complex networks," *Phys. Rev. Lett.*, vol. 96, no. 4, p. 040601, 2006.

[31] C. Orsini, E. Gregori, L. Lenzini, and D. Krioukov, "Evolution of the internet k-dense structure," *IEEE/ACM ToN*, vol. 22, no. 6, pp. 1769–1780, Dec. 2014.

[32] V. Batagelj and M. Zaveršnik, "Fast algorithms for determining (generalized) core groups in social networks," *Advances in Data Analysis and Classification*, vol. 5, no. 2, pp. 129–145, 2011.

[33] M. G. Beiró, J. I. Alvarez-Hamelin, and J. R. Busch, "A low complexity visualization tool that helps to perform complex systems analysis," *New J. Phys*, vol. 10, no. 12, p. 125003, 2008.

[34] M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas, and k. claffy, "As relationships, customer cones, and validation," in *Proceedings of ACM SIGCOMM IMC*, 2013.

[35] CAIDA, "Archipelago (Ark) Measurement Infrastructure," https://www.caida.org/projects/ark/.

[36] Y. Hyun, A. Broido *et al.*, "Traceroute and bgp as path incongruities," 2003.

[37] Stefan Meinders, "The New Internet," ENOG11, 2016.

[38] PeeringDB, "https://www.peeringdb.com."

[39] Sandvine, "Global internet phenomena report 2016," 2016.

[40] USA Today, "Does Amazon control the Internet, or does it just feel that way?" https://www.usatoday.com/story/tech/talkingtech/2017/03/01/amazon-control-internet-aws-cloud-services-outage/98548762/, 2017.

[41] Seeking Alpha, "Apple, Microsoft And Facebook Bring More Traffic To In-House CDNs, Impacting Akamai's Media Business," https://seekingalpha.com/article/3613736-apple-microsoft-facebook-bring-traffic-house-cdns-impacting-akamais-media-business, 2015.

[42] Sandvine, "Global internet phenomena report Spring 2011," 2011.

[43] V. Stocker, G. Smaragdakis, W. Lehr, and S. Bauer, "The growing complexity of content delivery networks: Challenges and implications for the internet ecosystem," *Telecommunications Policy*, vol. 41, no. 10, pp. 1003–1016, 2017.

[44] Amazon, "AWS global infrastructure," https://aws.amazon.com/es/about-aws/global-infrastructure/, 2017.

[45] Apple Insider, "Apple's in-house CDN efforts spell trouble for Akamai as infrastructure biz warns of losses," http://appleinsider.com/articles/16/02/10/apples-in-house-cdn-efforts-spell-trouble-for-akamai-as-infrastructure-biz-warns-of-losses, 2016.

[46] ARS Technica, "Apple?s multi-terabit, $100M CDN is live - with paid connection to Comcast," https://arstechnica.com/information-technology/2014/07/apples-multi-terabit-100m-cdn-is-live-with-paid-connection-to-comcast/, 2014.

[47] LA Times, "Apple's original TV production to begin small: 'We are just starting out'," http://beta.latimes.com/business/hollywood/la-fi-ct-apple-television-strategy-planet-apps-20170214-story.html, 2017.

[48] 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.

[49] 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/.

[50] New York Times, "Google to Acquire YouTube for $1.65 Billion," http://www.nytimes.com/2006/10/09/business/09cnd-deal.html.

[51] Netflix Media Center, ""Announcing the Netflix Open Connect Network"," https://media.netflix.com/en/company-blog/announcing-the-netflix-open-connect-network, 2012.

[52] Quartz Media, ""The inside story of how Netflix came to pay Comcast for internet traffic"," https://qz.com/256586/the-inside-story-of-how-netflix-came-to-pay-comcast-for-internet-traffic/, 2014.

[53] Digital Element, "NetAcuity," https://www.digitalelement.com/solutions/.

[54] Poese, Ingmar and Uhlig, Steve and Kaafar, Mohamed Ali and Donnet, Benoit and Gueye, Bamba, "IP geolocation databases: Unreliable?" *ACM SIGCOMM CCR*, vol. 41, no. 2, 2011.

[55] Gharaibeh, Manaf and Shah, Anant and Huffaker, Bradley and Zhang, Han and Ensafi, Roya and Papadopoulos, Christos, "A look at router geolocation in public and commercial databases," in *Proceedings of ACM SIGCOMM IMC*, 2017.

[56] Quinn, James, "Being Open: How Facebook Got It's Edge," *NANOG68*, 2016.

[57] Galperin, Hernan, "Connectivity in Latin America and the Caribbean: The role of internet exchange points," *Internet Society, November*, 2013.

[58] CAIDA, "CAIDA's AS classification list," http://data.caida.org/datasets/as-classification/.