A Long Way to the Top:
Significance, Structure, and Stability of Internet Top Lists

Quirin Scheitle
Technical University of Munich

Oliver Hohlfeld
RWTH Aachen University

Julien Gamba
IMDEA Networks Institute/
Universidad Carlos III de Madrid

Jonas Jelten
Technical University of Munich

Torsten Zimmermann
RWTH Aachen University

Stephen D. Strowes
RIPE NCC

Narseo Vallina-Rodriguez
IMDEA Networks Institute/ICSI

ABSTRACT

A broad range of research areas including Internet measurement, privacy, and network security rely on lists of target domains to be analysed; researchers make use of target lists for reasons of necessity or efficiency. The popular Alexa list of one million domains is a widely used example. Despite their prevalence in research papers, the soundness of top lists has seldom been questioned by the community: little is known about the lists' creation, representativity, potential biases, stability, or overlap between lists.

In this study we survey the extent, nature, and evolution of top lists used by research communities. We assess the structure and stability of these lists, and show that rank manipulation is possible for some lists. We also reproduce the results of several scientific studies to assess the impact of using a top list at all, which list specifically, and the date of list creation. We find that *(i)* top lists generally overestimate results compared to the general population by a significant margin, often even an order of magnitude, and *(ii)* some top lists have surprising change characteristics, causing high day-to-day fluctuation and leading to result instability. We conclude our paper with specific recommendations on the use of top lists, and how to interpret results based on top lists with caution.

CCS CONCEPTS

• **Networks → Network measurement**;

1 INTRODUCTION

Scientific studies frequently make use of a sample of DNS domain names for various purposes, be it to conduct lexical analysis, to measure properties of domains, or to test whether a new algorithm

works on real domains. Internet top lists, such as the Alexa or Cisco Umbrella Top 1M lists, serve the purpose of providing a reputedly representative sample of Internet domains in popular use. These top lists can be created with different methods and data sources, resulting in different sets of domains.

The prevalence and opacity of these lists could have introduced an unchecked bias in science—for 10 networking venues in 2017 alone, we count 69 publications that use a top list. This potential bias is based on the fact that curators of such top lists commonly conceal the data sources and ranking mechanism behind those lists, which are typically seen as a proprietary business asset in the search engine optimisation (SEO) space [4]. This leaves researchers using those lists with little to no information about content, stability, biases, evolution and representativity of their contents.

In this work, we analyse three popular top lists—Alexa Global [1], Cisco Umbrella [2], and Majestic Million [3]—and discuss the following characteristics:

Significance: In a survey of 687 networking-related papers published in 2017, we investigate if, to what extent, and for what purpose, these papers make use of Internet top lists. We find that 69 papers (10.0%) make use of at least one top list (*cf.*, §3).

Structure: Domain properties in different top lists, such as the surprising amount of invalid top-level domains (TLDs), low intersections between various lists (<30%), and classifications of disjunct domains, are investigated in §5.

Stability: We conduct in-depth longitudinal analyses of top list stability in §6, revealing daily churn of up to 50% of domains.

Ranking Mechanisms: Through controlled experiments and reverse engineering of the Alexa toolbar, we shed light on the ranking mechanisms of different top lists. In one experiment, we place an unused test domain at a 22k rank in Umbrella (*cf.*, §7).

Research Result Impact: Scientific studies that use top lists for Internet research measure characteristics of the targets contained in each list, or the related infrastructure. To show the bias inherent in any given target list, we run several experiments against top lists and the general population of all *com/net/org* domains. We show that top lists significantly exaggerate results, and that results even depend on the day of week a list was obtained (*cf.*, §8).

We discuss related work and specific recommendations in §9.

Throughout our work, we aim to adhere to the highest ethical standards, and aim for our work to be fully reproducible. We share code, data, and additional insights under

2 DOMAIN TOP LISTS

This section provides an overview of various domain top lists and their creation process. Each of these lists is updated daily.

Alexa: The most popular and widely used top list is the Alexa Global Top 1M list [1]. It is generated based on web activity monitored by the Alexa browser plugin[1], "directly measured sources" [6] and "over 25,000 different browser extensions" [7] over the past three months [8] from "millions of people" [6]. No information exists on the plugin's user base, which opens questions on potential biases in terms of, *e.g.*, geography or age of its user base. Alexa lists are generally offered for sale, with few free offerings. Paid offerings include top lists per country, industry, or region. The Global Top 1M list is the most popular free offering, available with no explicit license, and was briefly suspended in late 2016.

Cisco Umbrella: Another popular top list is the Cisco Umbrella 1M, a top list launched in mid-December 2016. This list contains the Top 1M domains (including subdomains) as seen by Cisco's OpenDNS service [2]. This DNS-based nature is fundamentally different from collecting website visits or links. Hence, the Umbrella list contains Fully Qualified Domain Names (FQDN) for any kind of Internet service, not just web sites as in the case of Alexa or Majestic. Without explicit license, it is provided "free of charge".

Majestic: The third top list is the Majestic Million [3], released in October 2012. This creative commons licensed list is based on Majestic's web crawler. It ranks sites by the number of /24 IPv4-subnets linking to that site [9]. This is yet another data collection methodology and, similar to Alexa, heavily web-focused. While the Majestic list is currently not widely used in research, we still include it in our study for its orthogonal mechanism, its explicitly open license, and its availability for several years.

Other Top Lists: There are few other top lists available, but as those are little used, not consistently available, or fluctuate in size, we will not investigate them in detail in this paper. Quantcast [10] provides a list of the Top 1M most frequently visited websites per country, measured through their web intelligence plugin on sites. Only the US-based list can be downloaded; all other lists can only be viewed online and show ranks only when paid. The Statvoo [11] list provides an API and a download for their Top 1M sites, but has frequently been inaccessible in the months before this publication. Statvoo does not offer insights about the metrics they use in their creation process. The Chrome UX report [12] publishes telemetry data about domains popular with Chrome users. It does not, however, rank domains or provide a static-sized set of domains. We also exclude the SimilarWeb Top Sites ranking [13] as it is not available for free and little used in science.

3 SIGNIFICANCE OF TOP LISTS

Scientific literature often harnesses one or more of the top lists outlined in §2. To better understand how often and to what purpose top lists are used by the literature, we survey 687 recent publications.

3.1 Methodology

We survey papers published at 10 network-related venues in 2017, listed in Table 1. First, we search the 687 papers published at these venues for keywords[2] in an automated manner. Next, we inspect matching papers manually to remove false positives (*e.g.*, Amazon's Alexa home assistant, or an author named Alexander), or papers that mention the lists without actually using them as part of a study.

Finally, we reviewed the remaining 69 papers (10.0%) that made use of a top list, with various aims in mind: to understand the top lists used (§3.2), the nature of the study and the technologies measured (§3.3), whether the study was dependent on the list for its results (§3.4), and whether the study was possibly replicable (§3.5). Table 1 provides an overview of the results.

We find the field of Internet measurement to be most reliant on top lists, used in 22.2% of the surveyed papers. Other fields also use top lists frequently, such as security (8.5%), systems (6.4%) and web technology (7.9%).

3.2 Top Lists Used

We first investigate which lists and what subsets of lists are typically used; Table 1 provides an overview of lists used in the studies we identified. We find 29 studies using the full Alexa Global Top 1M, the most common choice among inspected publications, followed by a surprising variety of Alexa Top 1M subsets (*e.g.*, Top 10k).

All papers except one [69] use a list collated by Alexa. This paper instead uses the Umbrella Top 100 list to assess importance of ASes showing BGP bursts. No paper in our study used the Majestic list.

A study may also use multiple distinct subsets of a list. For example, one study uses the Alexa Global Top 1k, 10K, 500K and Top 1M at different stages of the study [61]. We count these as distinct use-cases in the right section of Table 1.

We also find that 59 studies exclusively use Alexa as a source for domain names. Ten papers use lists from more than one origin; one paper uses the Alexa Global Top 1M, the Umbrella Top 1M, and various DNS zone files as sources [21]. In total, two studies make use of the Cisco Umbrella Top 1M [21, 67].

Category and country-specific lists are also being used: eight studies use country-specific lists from Alexa, usually choosing only one country; one study selected 138 countries [26]. Category-based lists are rarer still: two studies made use of category subsets [17, 71].

3.3 Characterisation of Studies

To show that top lists are used for various types of studies, we look at the range of topics covered and technologies measured in our surveyed papers. For each paper we assigned a broad *purpose*, and the network *layer* in focus.

Purposes: For all papers, we reviewed the broad area of study. The largest category we identified encompasses various aspects of security, across 38 papers in total: this includes phishing attacks [81, 82], session safety during redirections [83], and domain squatting [58], to name a few. Nine more papers study aspects of privacy & censorship, such as the Tor overlay network [61], or user tracking [35]. Network or application performance is also a popular area: ten papers in our survey focus on this, *e.g.*, HTTP/2 server

---

[1]Available for Firefox and Chrome. Internet Explorer discontinued June 2016 [5]
[2]Those being: "alexa", "umbrella", and "majestic"

Table 1: Left: Use of top lists at 2017 venues. The 'dependent' column indicates whether we deemed the results of the study to rely on the list used ('Y'), or that the study relies on a list for verification ('V') of other results, or that a list is used but the outcome doesn't rely on the specific list selected ('N'). The 'date' column indicates how many papers stated the date of list download or measurement. **Right:** Type of lists used in 69 papers from left. Multiple counts for papers using multiple lists.

Venue	Area	Papers	using list #	using list %↓	# dependent Y	# dependent V	# dependent N	# date? List	# date? Study	References
ACM IMC	Measurements	42	11	26.2%	8	2	1	1	3	[14–24]
PAM	Measurements	20	4	20.0%	3	1	0	0	0	[25–28]
TMA	Measurements	19	3	15.8%	1	1	1	0	0	[29–31]
USENIX Security	Security	85	12	14.1%	8	4	0	2	0	[32–43]
IEEE S&P	Security	60	5	8.3%	3	2	0	1	1	[44–49]
ACM CCS	Security	151	11	7.3%	4	5	2	1	1	[50–60]
NDSS	Security	68	3	4.4%	2	0	1	0	0	[61–63]
ACM CoNEXT	Systems	40	4	10.0%	2	1	1	0	1	[64–68]
ACM SIGCOMM	Systems	38	3	7.9%	3	0	0	0	0	[69–71]
WWW	Web Tech.	164	13	7.9%	11	1	1	2	3	[72–84]
Total		687	69	10.0%	45	17	7	7	9	

Alexa Global Top …			
1M	29	5k	2
100k	2	1k	5
75k	1	500	8
50k	2	400	1
25k	2	300	1
20k	1	200	1
16k	1	100	8
10k	11	50	3
8k	1	10	1

Alexa Country:	2
Alexa Category:	2
Umbrella 1M:	3
Umbrella 1k:	1

push [72], mobile web performance [71], and Internet latency [26]. Other studies look at economic aspects such as hosting providers.

Layers: We also reviewed the network layers measured in each study. Many of the papers we surveyed focus on web infrastructure: 22 of the papers are concerned with content, 8 focus on the HTTP(S) protocols, and 7 focus on applications (*e.g.*, browsers [39, 40]).

Studies relating to core network protocols are commonplace: DNS [32, 36, 51, 52, 61] (we identified 3 studies relating to domain names as separate from DNS *protocol* measurements [24, 58, 63]), TCP [19, 31], and IP [14, 15, 18, 30, 64, 69], and TLS/HTTPS [21, 37, 38, 50, 57, 76, 83] layer measurements are common in our survey.

Finally, we identify 12 studies whose experimental design measures more than one specific layer; *e.g.*, cases studying a full connection establishment (from initial DNS query to HTTP request).

We conclude from this that top lists are frequently used to explicitly or implicitly measure DNS, IP, and TLS/HTTPS characteristics, which we investigate in depth in §8.

3.4 Are Results Dependent on Top Lists?

In this section, we discuss how dependent study results are on top lists. For this, we fill the "dependent" columns in Table 1 as follows:

Dependent (Y): Across all papers surveyed, we identify 45 studies whose results may be affected by the list chosen. Such a study would take a list of a certain day, measure some characteristic over the set of domains in that list, and draw conclusions about the measured characteristic. In these cases, we say that the results *depend* on the list being used: a different set of domains in the list may have yielded different results.

Verification (V): We identify 17 studies that use a list only to verify their results. A typical example may be to develop some algorithm to find domains with a certain property, and then use a top list to check whether these domains are popular. In such cases, the algorithm developed is independent of the list's content.

Independent (N): Eight studies cite and use a list, but we determine that their results are not necessarily reliant on the list. These papers typically use a top list as one source among many, such that changes in the top list would likely not affect the overall results.

3.5 Are Studies Replicable?

Repeatability, replicability, and reproducibility are ongoing concerns in Computer Networks [85, 86] and Internet Measurement [87]. While specifying the date of when a top list was downloaded, and the date when measurements where conducted, are not necessarily sufficient to reproduce studies, they are important first steps.

Table 1 lists two "date" columns that indicate whether the list download date or the measurement dates were given[3]. Across all 69 papers using top lists, only 7 stated the date the list was retrieved, and 9 stated the measurement date. Unfortunately, only 2 papers give both the list and measurement data and hence fulfil these basic criteria for reproducibility. This does not necessarily mean that the other papers are not reproducible, authors may publish the specific top list used as part of data, or authors might be able to provide the dates or specific list copies upon inquiry. However, recent investigations of reproducibility in networking hints that this may be an unlikely expectation [87, 88]. We find two papers that explicitly discuss instability and bias of top lists, and use aggregation or enrichment to stabilise results [45, 67].

3.6 Summary

Though our survey has a certain level of subjectivity, we consider its broad findings meaningful: *(i)* that top lists are frequently used, *(ii)* that many papers' results depend on list content, and *(iii)* that few papers indicate precise list download or measurement dates.

---

[3]We require a specific day to be given to count a paper, the few papers just citing a year or month were counted as no date given

We also find that the use of top lists to measure network and security characteristics (DNS, IP, HTTPS/TLS) is common. We further investigate how top list use impacts result quality and stability in studies by measuring these layers in §8.

4 TOP LISTS DATASET

For the three lists we focus on in this study, we source daily snapshots as far back as possible. Many snapshots come from our own archives, and others were shared with us by other members of the research community, such as [89–91]. Table 2 gives an overview of our datasets along with some metrics discussed in §5. For the Alexa list, we have a dataset with daily snapshots from January 2009 to March 2012, named *AL0912*, and another from April 2013 to April 2018, named *AL1318*. The Alexa list underwent a significant change in January 2018; for this we created a partial dataset named *AL18* after this change. For the Umbrella list, we have a dataset spanning 2016 to 2018, named *UM1618*. For the Majestic Million list, we cover June 2017 to April 2018.

As many of our analyses are comparative between lists, we create a JOINT dataset, spanning the overlapping period from June 2017 to the end of April 2018. We also sourced individual daily snapshots from the community and the Internet Archive [92], but only used periods with continuous daily data for our study.

5 STRUCTURE OF TOP LISTS

In this section, we analyse the structure and nature of the three top lists in our study. This includes questions such as top level domain (TLD) coverage, subdomain depth, and list intersection.

DNS Terms used in this paper, for clarity, are the following: for *www.net.in.tum.de*, *.de* is the public suffix[4] (and top level domain), *tum.de* is the base domain, *in.tum.de* is the first subdomain, and *net.in.tum.de* is the second subdomain. Hence, *www.net.in.tum.de* counts as a third-level subdomain.

5.1 Domain Name Depth and Breadth

A first characteristic to understand about top lists is the scope of their coverage: how many of the active TLDs do they cover, and how many do they miss? How deep are they going into specific subdomains, choosing trade-offs between breadth and depth?

TLD Coverage is a first indicator of list breadth. Per IANA [94, 95], 1,543 TLDs exist as of May 20th, 2018. Based on this list, we count valid and invalid TLDs per list. The average coverage of valid TLDs in the JOINT period is ≈700 TLDs, covering only about 50% of active TLDs. This implies that measurements based on top lists may miss up to 50% of TLDs in the Internet.

At the Top 1k level we find quite different behaviour with 105 valid TLDs for Alexa, 50 for Majestic, but only 13 (*com/net/org* and few other TLDs) for Umbrella. We speculate that this is rooted in DNS administrators from highly queried DNS names preferring the smaller set of professionally managed and well-established top level domains over the sometimes problematic new gTLDs [96–98].

Invalid TLDs occur neither in any Top 1k domains nor in the Alexa Top 1M domains, but as a minor count in the Majestic Top 1M (7 invalid TLDs, resulting in 35 domain names), and significant

count in the Umbrella Top 1M: there, we can find 1,347 invalid TLDs[5], in a total of 23k domain names (2.3% of the list). This is an early indicator of a specific characteristic in the Umbrella list: invalid domain names queried by misconfigured hosts or outdated software can easily get included into the list.

Comparing valid and invalid TLDs also reveals another structural change in the Alexa list on July 20th, 2014: before that date, Alexa had a fairly static count of 206 invalid and 248 valid TLDs. Perhaps driven by the introduction of new gTLDs from 2013 [99], Alexa changed its filtering: After that date, invalid TLDs have been reduced to ≈0, and valid TLDs have shown continued growth from 248 to ≈800. This confirms again that top lists can undergo rapid and unannounced changes in their characteristics, which may significantly influence measurement results.

Subdomain Depth is an important property of top lists. Base domains offer more breadth and variety in setups, while subdomains may offer interesting targets besides a domain's main web presence. The ratio of base to subdomains is hence a breadth/depth trade-off, which we explore for the three lists used. Table 2 shows the average number of base domains (μ_{BD}) per top list. We note that Alexa and Majestic contain almost exclusively base domains with few exceptions (*e.g.*, for `blogspot`). In contrast, 28% of the names in the Umbrella list are base domains, *i.e.*, Umbrella emphasises depth of domains. Table 2 also details the subdomain depth for a single-day snapshot (April 30, 2018) of all lists. As the Umbrella list is based on DNS lookups, such deep DNS labels can easily become part of the Umbrella list, regardless of the origin of the request. In fact, Umbrella holds subdomains up to level 33 (*e.g.*, domains with extensive *www* prefixes or '.'-separated OIDs).

We also note that the base domain is usually part of the list when its subdomains are listed. On average, each list contains only few hundred subdomains whose base domain is not part of the list.

Domain Aliases are domains with the same second-level domain, but different top-level domains, *e.g.*, *google.com* and *google.de*. Table 2 shows the number of domain aliases as DUP_{SLD}. We find a moderate level of ≈5% of domain aliases within various top lists, with only 1.5% for Majestic. Analysis reveals a very flat distribution, with the top entry *google* at ≈200 occurrences.

5.2 Intersection between Lists

We next study intersection between lists—all 3 lists in our study promise a view on the most popular domains (or websites) in the Internet, hence measuring how much these lists agree[6] is a strong indicator of bias in list creation. Figure 1a shows the intersection between top lists over time during the JOINT period. We see that the intersection is quite small: for the Top1M domains, Alexa and Majestic share 285k domains on average during the JOINT duration. Alexa and Umbrella agree on 150k, Umbrella and Majestic on 113k, and all three only on 99k out of 1M domains.

For the Top1k lists, the picture is more pronounced. On average during the JOINT period, Alexa and Majestic agree on 295 domains, Alexa and Umbrella on 56, Majestic and Umbrella on 65, and all three only on 47 domains.

---

[4]per Public Suffix List [93], a browser-maintained list aware of cases such as *co.uk*.

[5]Examples for invalid TLDs: *instagram*, *localdomain*, *server*, *cpe*, *0*, *big*, *cs*
[6]To control for varying subdomain length, we first normalise all lists to unique base domains (cf. μ_{BD} in Table 2, reducing *e.g.*, Umbrella to 273k base domains)

Table 2: Datasets: mean of valid TLDs covered (μ_{TLD}), mean of base domains (μ_{BD}), mean of sub-domain level spread (SD_n for share of n-th level subdomains, SD_M for maximum sub-domain level), mean of domain aliases (DUP_{SLD}), mean of daily change (μ_Δ) and mean of new (*i.e.*, not included before) domains per day (μ_{NEW}). Footnote 4: Average after Alexa's change in January 18.

List	Top	Dataset	Dates	$\mu_{TLD} \pm \sigma$	$\mu_{BD} \pm \sigma$	SD_1	SD_2	SD_3	SD_M	DUP_{SLD}	μ_Δ	μ_{NEW}
Alexa	1M	AL0912	29.1.09–16.3.12	248 ± 2	$973k \pm 2k$	1.6%	0.4%	≈0%	4	$47k \pm 2k$	23k	n/a
Alexa	1M	AL1318	30.4.13–28.1.18	545 ± 180	$972k \pm 6k$	2.2%	0.1%	≈0%	4	$49k \pm 3k$	21k	5k
Alexa	1M	AL18	29.1.18–30.4.18	771 ± 8	$962k \pm 4k$	3.7%	≈0%	≈0%	4	$45k \pm 1k$	483k	121k
Alexa	1M	JOINT	6.6.17–30.4.18	760 ± 11	$972k \pm 7k$	2.6%	≈0%	≈0%	4	$51k \pm 4k$	147k	38k
Umbrella	1M	JOINT	6.6.17–30.4.18	580 ± 13	$273k \pm 13k$	49.9%	14.7%	5.9%	33	$15k \pm 1k$	100k	22k
Majestic	1M	JOINT	6.6.17–30.4.18	698 ± 14	$994k \pm 617$	0.4%	≈0%	≈0%	4	$49k \pm 1k$	6k	2k
Alexa	1k	JOINT	6.6.17- 30.4.18	105 ± 3	990 ± 2	1.3%	0.0%	0.0%	1	22 ± 2	9 (78^4)	4 (8^4)
Umbrella	1k	JOINT	6.6.17–30.4.18	13 ± 1	317 ± 6	52.0%	14%	≈0%	6	11 ± 2	44	2
Majestic	1k	JOINT	6.6.17–30.4.18	50 ± 1	939 ± 3	5.9%	0.1%	0.1%	4	32 ± 1	5	.8
Umbrella	1M	UM1618	15.12.16–30.4.18	591 ± 45	$281k \pm 16k$	49.4%	14.5%	5.7%	33	$15k \pm 1k$	118k	n/a

This disparity between top domains suggests a high bias in the list creation. We note that even both web-based lists, Alexa and Majestic, only share an average of 29% of domains.

Standing out from Figure 1a is the fact that the Alexa list has changed its nature in January 2018, reducing the average intersection with Majestic from 285k to 240k. This change also introduced a weekly pattern, which we discuss further in §6.2. We speculate that Alexa might have reduced its 3-month sliding window [8], making the list more volatile and susceptible to weekly patterns. We contacted Alexa about this change, but received no response.

5.3 Studying Top List Discrepancies

The low intersection between Umbrella and the other lists could be rooted in the DNS vs. web-based creation. Our hypothesis is that the web-based creation of Alexa and Majestic lists tends to miss domains providing embedded content as well as domains popular on mobile applications [67, 100]. In this section, we explore the origin of discrepancies across domain lists.

We aggregate the Alexa, Umbrella, and Majestic Top 1k domains from the last week of April 2018, and analyse the set of 3,005 disjunct domains across these lists, *i.e.*, those found only in a single list. 40.7% of these domains originate from Alexa, 37.1% from Umbrella, and 22.1% from Majestic. Subsequently, we identify whether the disjunct domains are associated with mobile traffic or third-party advertising and tracking services not actively visited by users, but included through their DNS lookups. We opt against utilizing domain classifiers such as the OpenDNS Domain Tagging service [101], as it has been reported that categories are vague and coverage is low [100].

Instead, we use the data captured by the Lumen Privacy Monitor [102] to associate domains with mobile traffic for more than 60,000 Android apps, and use popular anti-tracking blacklists such as MalwareBytes' hpHosts ATS file [103]. We also check if the domains from a given top list can be found in the aggregated Top 1M of the other two top lists during the same period of time. Table 3 summarises the results. As we suspected, Umbrella has significantly more domains flagged as "mobile traffic" and third-party advertising and tracking services than the other lists. It also has the lowest proportion of domains shared with other Top 1M lists.

Table 3: Share of one-week Top 1k disjunct domains present in hpHosts (blacklist), Lumen (mobile), and Top 1M of other top lists.

List	# Disjunct	% hpHosts	% Lumen	% Top 1M
Alexa	1,224	3.10%	1.55%	99.10%
Umbrella	1,116	20.16%	39.43%	25.63%
Majestic	665	1.95%	3.76%	93.63%

This confirms that Umbrella is capable of capturing domains from any device using OpenDNS, such as mobile and IoT devices, and also include domains users are not aware of visiting, such as embedded third-party trackers in websites. Alexa and Majestic provide a web-specific picture of popular Internet domains.

6 STABILITY OF TOP LISTS

Armed with a good understanding of the *structure* of top lists, we now focus on their stability over time. Research has revealed hourly, daily and weekly patterns on ISP traffic and service load, as well as significant regional and demographic differences in accessed content due to user habits [104–107]. We assess whether such patterns also manifest in top lists, as a first step towards understanding the impact of studies selecting a top list at a given time.

6.1 Daily Changes

We start our analysis by understanding the composition and evolution of top lists on a daily basis. As all top lists have the same size, we use the raw count of daily changing domains for comparison.

Figure 1b shows the count of domains that were removed daily, specifically the count of domains present in a list on day *n* but not on day *n*+1. The Majestic list is very stable (6k daily change), the Umbrella list offers significant churn (118k), and the Alexa list used to be stable (21k), but drastically changed its characteristic in January 2018 (483k), becoming the most unstable list.

The fluctuations in the Umbrella list, and in the Alexa list after January 2018, are weekly patterns, which we investigate closer in §6.2. The average daily changes are given in column μ_Δ of Table 2.

Which Ranks Change? Previous studies of Internet traffic revealed that the distribution of accessed domains and services follows a power-law distribution [68, 104–107]. Therefore, the ranking
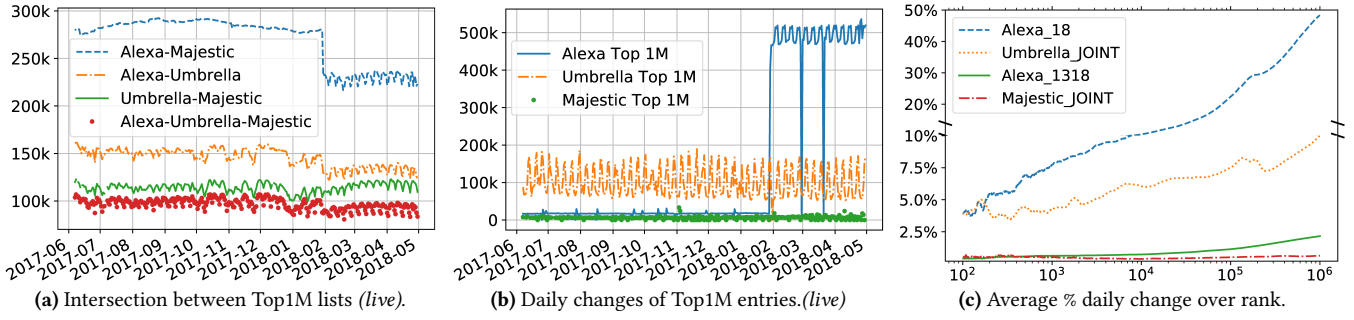
(a) Intersection between Top1M lists *(live)*.

(b) Daily changes of Top1M entries.*(live)*

(c) Average % daily change over rank.

Figure 1: Intersection, daily changes and average stability of top lists (y-axis re-scaled at 10% in right plot). *Click for live version/source code*

of domains in the long tail should be based on significantly smaller and hence less reliable numbers.

Figure 1c displays the stability of lists depending on subset size. The y-axis shows the mean number of daily changing domains in the top X domains, where X is depicted on the x-axis. For example, an x-value of 1000 means that the lines at this point show the average daily change per list for the Top 1k domains. The figure shows instability increasing with higher ranks for Alexa and Umbrella, but not for Majestic. We plot Alexa before and after its January 2018 change, highlighting the significance of the change across all its ranks–even its Top 1k domains have increased their instability from 0.62% to 7.7% of daily change.

New or In-and-out Domains? Daily changes in top lists may stem from new domains joining, or from previously contained domains re-joining. To evaluate this, we cumulatively sum all the unique domains ever seen in a list in Figure 2a, *i.e.*, a list with only permutations of the same set of domains would be a flat line. Majestic exhibits linear growth: every day, about 2k previously not included domains are added to it — approximately a third of the 6k total changing domains per day (*i.e.*, 4k domains have rejoined). Over the course of a year, the total count of domains included in the Majestic list is 1.7M. Umbrella adds about 20k new domains per day (out of 118k daily changing domains), resulting in 7.3M domains after one year. Alexa grows by 5k (of 21k) and 121k (of 483k) domains per day, before and after its structural change in January 2018. Mainly driven from the strong growth after Alexa's change, its cumulative number of domains after one year is 13.5M. This means that a long-term study of the Alexa Top 1M will, over the course of this year, have measured 13.5M distinct domains.

Across all lists, we find an average of 20% to 33% of daily changing domains to be new domains, *i.e.*, entering the list for the first time. This also implies that 66% to 80% of daily changing domains are domains that are repeatedly removed from and inserted into a list. We also show these and the equivalent Top 1k numbers in column μ_{NEW} of Table 2.

This behaviour is further confirmed in Figure 2b. In this figure, we compute the intersection between a fixed starting day and the upcoming days. We compute it seven times, with each day of the first week of the JOINT dataset as the starting day. Figure 2b shows the daily median value between these seven intersections.

This shows several interesting aspects: *(i)* the long-term trend in temporal decay per list, confirming much of what we have seen before (high stability for Majestic, weekly patterns and high instability for Umbrella and the late Alexa list) *(ii)* the fact that for Alexa

and Umbrella, the decay is non-monotonic, *i.e.*, a set of domains is leaving and rejoining at weekly intervals.

For How Long are Domains Part of a Top List? We investigate the average number of days a domain remains in both the Top 1M and Top 1k lists in Figure 2c. This figure displays a CDF with the number of days from the JOINT dataset in the x-axis, and the normalised cumulative probability that a domain is included on the list for X or fewer days. Our analysis reveals significant differences across lists. While about 90% of domains in the Alexa Top 1M list are in the list for 50 or fewer days, 40% of domains in the Majestic Top 1M list remain in the list across the full year. With this reading, lines closer to the lower right corner are "better" in the sense that more domains have stayed in the list for longer periods, while lines closer to the upper left indicate that domains get removed more rapidly. The lists show quite different behaviour, with Majestic Top 1k being the most stable by far (only ≈ 26% of domains present on < 100% of days), and being followed by Majestic Top 1M, Umbrella Top 1k, Alexa Top 1k, Umbrella Top 1M, and Alexa Top 1M. The Majestic Top 1M list offers stability similar to the Alexa and Umbrella Top 1k lists.

6.2 Weekly Patterns

We now investigate the weekly[7] pattern in the Alexa and Umbrella lists as observed in Figure 1b. We generally do not include Majestic as it does not display a weekly pattern. In this section, we resort to various statistical methods to investigate those weekend patterns. We will describe each one of them in their relevant subsection.

How Do Domain Ranks Change over the Weekends? The weekly periodical patterns shown in Figure 1b show that list content depends on the day of the week. To investigate this pattern statistically, we calculate a weekday and weekend distribution of the rank position of a given domain and compute the distance between those two distribution using the Kolmogorov-Smirnov (KS) test. This method allows us to statistically determine to what degree the distribution of a domain's ranks on weekdays and weekends overlap, and is shown in Figure 3a. We include Majestic as a base line without a weekly pattern. For Alexa Top 1M, we can see that ≈35% of domains have a KS distance of one, meaning that their weekend and weekday distributions have no data point in common. This feature is also present in Umbrella's rank, where over 15% of domains have a KS distance of 1. The changes are less pronounced for the Top 1k Alexa and Umbrella lists, suggesting that the top

---

[7]It is unclear what cut-off times list providers use, and how they offset time zones. For our analysis, we map files to days using our download timestamp
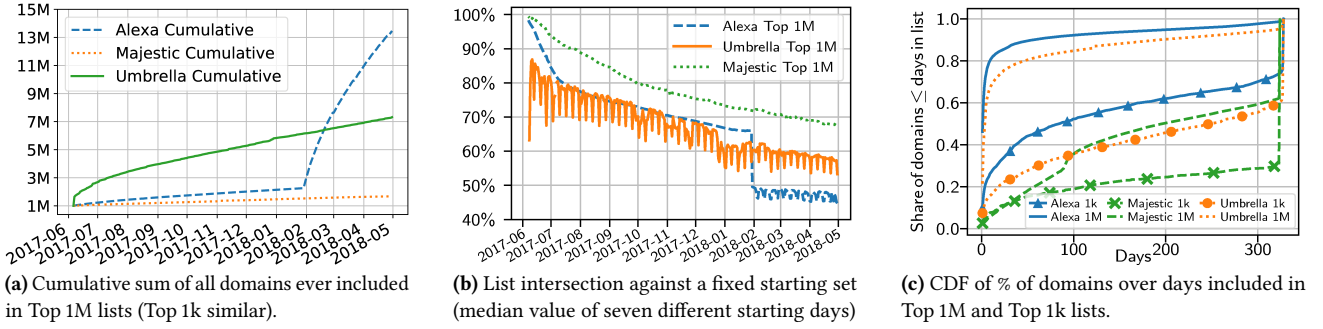
(a) Cumulative sum of all domains ever included in Top 1M lists (Top 1k similar).

(b) List intersection against a fixed starting set (median value of seven different starting days)

(c) CDF of % of domains over days included in Top 1M and Top 1k lists.

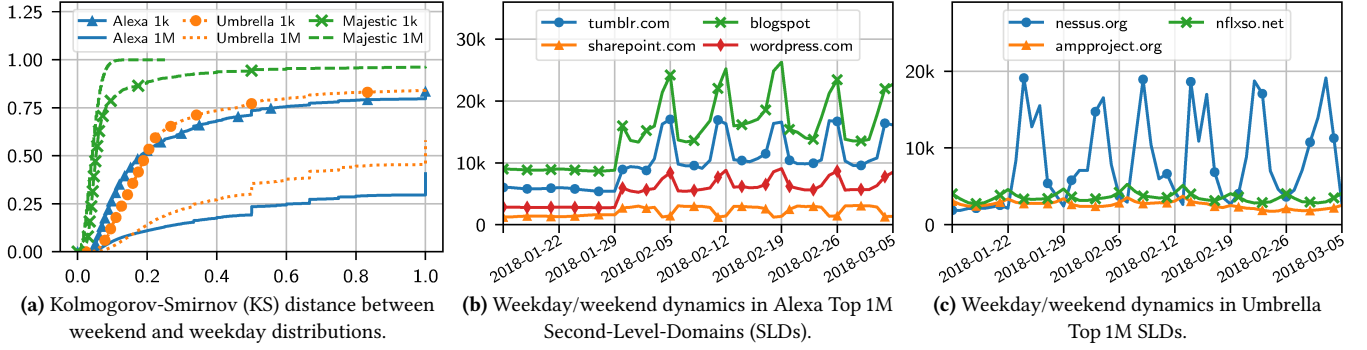**Figure 2:** Run-up and run-down of domains; share of days that a domains spend in a top list for the JOINT dataset.

(a) Kolmogorov-Smirnov (KS) distance between weekend and weekday distributions.

(b) Weekday/weekend dynamics in Alexa Top 1M Second-Level-Domains (SLDs).

(c) Weekday/weekend dynamics in Umbrella Top 1M SLDs.

Figure 3: Comparison of weekday vs. weekend distributions and dynamics in Second-Level-Domains (SLDs).

domains are more stable. As a reference, the KS distance when comparing weekdays to weekdays and weekends to weekends is much lower. For 90% of domains in Alexa or Umbrella (Top 1k or Top 1M) the distance is lower than 0.05. The KS distance is lower than 0.02 for all of the domains in Majestic rankings (Top 1k or Top 1M). This demonstrates that a certain set of domains, the majority of them localised in the long-tail, present disjunct rankings between weekends and weekdays.

What Domains are More Popular on Weekends? This leads to the question about the nature of domains changing in popularity with a weekly pattern. To investigate this, we group domains by "second-level-domain" (SLD), which we define as the label left of a public suffix per the Public Suffix list [93]. Figures 3b and 3c display the time dynamics of SLD groups for which the number of domains varies by more than 40% between weekdays and weekends. For Alexa, we can see stable behaviour before its February 2018 change. We see that some groups such as blogspot.*[8] or tumblr.com are significantly more popular on weekends than on weekdays. The opposite is true for domains under sharepoint.com (a web-based Microsoft Office platform). Umbrella shows the same behaviour, with nessus.org (a threat intelligence tool) more popular during the week, and ampproject.org (a dominant website performance optimisation framework), and nflxso.net (a Netflix domain) more popular on weekends. These examples confirm that different Internet usage on weekends[9] is a cause for the weekly patterns.

---

[8]We include all blogspot.* domains in the same group

[9]Our data indicates prevailing Saturday and Sunday weekends

6.3 Order of Domains in Top Lists

As top lists are sorted, a statistical analysis of order variation completes our view of top lists' stability. We use the Kendall rank correlation coefficient [108], commonly known as Kendall's τ coefficient, to measure rank correlation, *i.e.*, the similarity in the *order* of lists. Kendall's correlation between two variables will be high when observations have a similar order between the two variables, and low when observations have a dissimilar (or fully different for a correlation of -1) rank between the two variables.

In Figure 4, we show the CDF of Kendall's τ rank correlation coefficient for the Alexa, Umbrella and Majestic Top 1k domains in two cases: *(i)* for day to day comparisons; *(ii)* for a static comparison to the first day in the JOINT dataset. For analysis, we can compare the percentage of very strongly correlated ranks, *i.e.*, the ranks for which Kendall's τ is higher than 0.95. For day to day comparisons, Majestic is clearly most similar at 99%, with Alexa (72%) and Umbrella (40%) both showing considerably dissimilarities.

When compared for a reference day, very strong correlation drops below 5% for all lists. This suggests that the order variations are not perceived in the short term, but may arise when considering longer temporal windows.

Investigating the Long Tail: To compare higher and lower ranked domains, we take three exemplary domains from the Top 100 and the lower ranks as examples. Table 4 summarises the results. For each of the six domains, we compute the highest, median, and lowest rank over the duration of the JOINT dataset. The difference of variability between top and bottom domains is striking and in line with our previous findings: the ranks of top domains are fairly stable, while the ranks of bottom domains vary drastically.

Table 4: Rank variation for some more and less popular websites in the Top 1M lists.

Domain	Highest rank			Median rank			Lowest rank		
	Alexa	Umbrella	Majestic	Alexa	Umbrella	Majestic	Alexa	Umbrella	Majestic
google.com	1	1	1	1	1	1	2	4	8
facebook.com	3	1	2	3	6	2	3	8	19
netflix.com	21	1	455	32	2	515	34	487	572
jetblue.com	2,284	14,291	4,810	3,133	29,637	4,960	5,000	56,964	5,150
mdc.edu	25,619	177,571	24,720	35,405	275,579	26,122	88,093	449,309	30,914
puresight.com	183,088	593,773	687,838	511,800	885,269	749,819	998,407	999,694	869,872
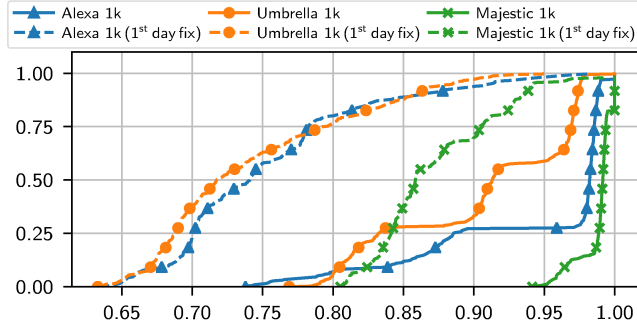
Figure 4: CDF of Kendall's τ between top lists.

6.4 Summary

We investigate the stability of top lists, and find abrupt changes, weekly patterns, and significant churn for some lists. Lower ranked domains fluctuate more, but the effect heavily depends on the list and the subset (Top 1k or Top 1M). We can confirm that the weekly pattern stems from leisure-oriented domains being more popular on weekends, and give examples for domain rank variations.

7 UNDERSTANDING AND INFLUENCING TOP LISTS RANKING MECHANISMS

We have seen that top lists can be rather unstable from day to day, and hence we investigate what traffic levels are required and at what effort it is possible to manipulate the ranking of a certain domain. As discussed previously, the Alexa list is based on its browser toolbar and "various other sources", Umbrella is based on OpenDNS queries, and Majestic is based on the count of subnets with inbound links to the domain in question. In this section, we investigate the ranking mechanisms of these top lists more closely.

7.1 Alexa

Alexa obtains visited URLs through "over 25,000 different browser extensions" to calculate site ranks through visitor and page view statistics [7, 109]. There is no further information about these tool-bars besides Alexa's own toolbar. Alexa also provides data to The Internet Archive to add new sites [92]. It has been speculated that Alexa provides tracking information to feed the Amazon recommen-dation and profiling engine since Amazon's purchase of Alexa in 1999 [110]. To better understand the ranking mechanism behind the Alexa list, we reverse engineer the Alexa toolbar[10] and investigate what data it gathers. Upon installation, the toolbar fetches a unique

---

[10]We detail the reverse engineering process in our dataset

identifier which is stored in the browser's local storage, called the *Alexa ID* (*aid*). This identifier is used for distinctly tracking the device. During installation, Alexa requests information about age, (binary) gender, household income, ethnicity, education, children, and the toolbar installation location (home/work). All of these are linked to the *aid*. After installation, the toolbar transfers for each visited site: the page URL, screen/page sizes, referer, window IDs, tab IDs, and loading time metrics. For a scarce set of 8 search en-gine and shopping URLs[11], referer and URL are anonymised to their host name. For all other domains, the entire URL, including all GET parameters, is transmitted to Alexa's servers under *data.alexa.com*. Because of the injected JavaScript, the visit is only transmitted if the site actually exists and was loaded. In April 2018, Alexa's API DNS name had a rank of ≈30k in the Umbrella list, indicating at least 10k unique source IP addresses querying that DNS domain name through OpenDNS per day (cf §7.2).

Due to its dominance, the Alexa rank of a domain is an impor-tant criterion in domain trading and search engine optimisation. Unsurprisingly, there is a gray area industry of sites promising to "optimise" the Alexa rank of a site for money [111–113]. Although sending synthetic data to Alexa's backend API should be possible at reasonable effort, we refrain from doing so for two reasons: *(i)* in April 2018, the backend API has changed, breaking communication with the toolbar, and *(ii)* unclear ethical implications of actively injecting values into this API. We refer the interested reader to le Pochat *et al.* [114], who have recently succeeded in manipulating Alexa ranks through the toolbar API.

7.2 Umbrella

As the Umbrella list is solely based on DNS queries through the OpenDNS public resolver, it mainly reflects domains frequently resolved, not necessarily domains visited by humans, as confirmed in §5.3. Examples are the Internet scanning machines of various research institutions, which likely show up in the Umbrella ranking through automated forward-confirmed reverse-DNS at scanned hosts, and not from humans entering the URL into their browser. Building a top list based on DNS queries has various trade-offs and parameters, which we aim to explore here. One specifically is the TTL value of a DNS domain name. As the DNS highly relies on caching, TTL values could introduce a bias in determining pop-ularity based on DNS query volume: domain names with higher Time-To-Live values can be cached longer and may cause fewer DNS queries at upstream resolvers. To better understand Umbrella's

---

[11]As of 2018-05-17, these are google.com, instacart.com, shop.rewe.de, youtube.com, search.yahoo.com, jet.com and ocado.com
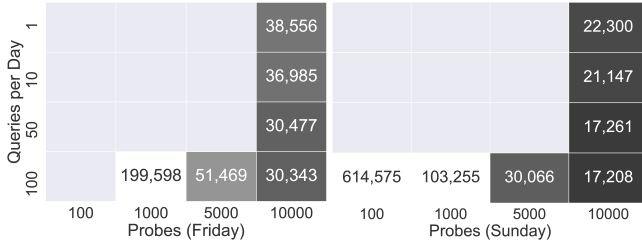
Figure 5: Umbrella rank depending on probe count, query frequency, and weekday (Friday left, Sunday right). Empty fields indicate the settings did not result in a Top 1M ranking.

ranking mechanism and query volume required, we set up 7 RIPE Atlas measurements [115], which query the OpenDNS resolvers for DNS names under our control.

Probe Count versus Query Volume: We set up measurements with 100, 1k, 5k, and 10k RIPE Atlas probes, and at frequencies of 1, 10, 50, and 100 DNS queries per RIPE Atlas probe per day [115]. The resulting ranks, stabilised after several days of measurement, are depicted in Figure 5. A main insight is that the number of probes has a much stronger influence than the query volume per probe: 10k probes at 1 query per day (a total of 10k queries) achieve a rank of 38k, while 1000 probes at 100 queries per day (a total of 100k queries) only achieve rank 199k.

It is a reasonable and considerate choice to base the ranking mechanism mainly on the number of unique sources, as it makes the ranking less susceptible to individual heavy hitters.

Upon stopping our measurements, our test domains quickly (within 1-2 days) disappeared from the list.

TTL Influence: To test whether the Umbrella list normalises the potential effects of TTL values, we query DNS names with 5 different TTL values from 1000 probes at a 900s interval [116]. We could not determine any significant effect of the TTL values: all 5 domains maintain a distance of less than 1k list places over time.

This is coherent with our previous observation that the Umbrella rank is mainly determined from the number of clients and not the query volume per client: as the TTL volume would mainly impact the query volume per client, its effect should be marginal.

7.3 Majestic

The Majestic Million top list is based on a custom web crawler mainly used for commercial link intelligence [117]. Initially, Majestic ranked sites by the raw number of referring domains. As this had an undesired outcome, the link count was normalised by the count of referring /24-IPv4-subnets to limit the influence of single IP addresses [118]. The list is calculated using 90 days of data [119]. As this approach is similar to PageRank [120], except that Majestic does not weigh incoming links by the originating domain, it is to be expected that referral services can increase a domain's popularity. We can, however, not see an *efficient* way to influence a domain's rank in the Majestic list without using such referral services. Le Pochat *et al.* [114] recently influenced a domain's rank in the Majestic link through such purchasing of back links.

8 IMPACT ON RESEARCH RESULTS

§3 highlighted that top lists are broadly used in networking, security and systems research. Their use is especially prevalent in Internet measurement research, where top lists are used to study aspects across all layers. This motivates us to understand the impact of top list usage on the outcome of these studies. As the replication of all studies covered in our survey is not possible, we evaluate the impact of the lists' structure on research results in the Internet measurement field by investigating *(i)* common layers, such as DNS and IP, that played a role in many studies, and *(ii)* a sample of specific studies across a variety of layers, aiming for one specific study per layer.

We evaluate those scientific results with 3 questions in mind: *(i)* what is the bias when using a top list as compared to a general population of all *com/net/org* domains[12] *(ii)* what is the difference in result when using a different top list? *(iii)* what is the difference in result when using a top list from a different day?

8.1 Domain Name System (DNS)

A typical first step in list usage is DNS resolution, which is also a popular research focus (cf. §3). We split this view into a record type perspective (*e.g.*, IPv6 adoption) and a hosting infrastructure perspective (*e.g.*, CDN prevalence and AS mapping). For both, we download lists and run measurements daily over the course of one year.

8.1.1 Record Type Perspective. We investigate the share of NX-DOMAIN domains and IPv6-enabled domains, and the share of CAA-enabled domains as an example of a DNS-based measurement study [122]. Results are shown in Table 5 and Figure 6.

Assessing list quality via NXDOMAIN: We begin by using NXDOMAIN as a proxy measure for assessing the quality of entries in the top lists. An NXDOMAIN error code in return to a DNS query means that the queried DNS name does not exist at the respective authoritative nameserver. This error code is unexpected for allegedly popular domains. Ideally, a top list would only provide existing domains. Surprisingly, we find the amount of NX-DOMAIN responses in both the Umbrella (11.5%) and the Majestic (2.7%) top lists higher than in the general population of *com/net/org* domains (0.8%). This is in alignment with the fact that already ≈23k domains in the Umbrella list belong to non-existent top-level domains (*cf.*, §5.1). Figure 6a shows that the NXDOMAIN share is, except for Umbrella, stable over time. We found almost no NX-DOMAINs among Top 1k ranked domains. One notable exception is `teredo.ipv6.microsoft.com`, a service discontinued in 2013 and unreachable, but still commonly appearing at high ranks in Umbrella, probably through requests from legacy clients.

This also highlights a challenge in Majestic's ranking mechanism: while counting the number of links to a certain website is quite stable over time, it also reacts slowly to domain closure.

Tracking IPv6 adoption has been the subject of several scientific studies such as [126, 127]. We compare IPv6 adoption across top lists and the general population, for which we count the number of domains that return at least one routed IPv6 address as an

---

[12] *com/net/org* is still only a 45% sample of the general population (156.7M of 332M domains as per [121]), but more complete and still unbiased samples are difficult to obtain due to ccTLDs' restrictive zone file access policies. [21, 122–125]

Table 5: Internet measurement characteristics compared across top lists and general population, usually given as $\mu \pm \sigma$. For each cell, we highlight if it significantly (50%[6]) exceeds ▲ or falls behind ▼ the base value (1k / 1M, 1M / *com/net/org*), or not ■ . **In almost all cases (▲ and ▼), top lists significantly distort the characteristics of the general population.**

Study	Alexa 1K	Umbrella 1K	Majestic 1K	Alexa 1M	Umbrella 1M	Majestic 1M	com/net/org 157.24M ± 172K
NXDOMAIN[1]	▼~0.0% ± 0.0%	▼~0.0% ± 0.0%	▼~0.0% ± 0.0%	▼0.13% ± 0.02	▲11.51% ± 0.9	▲2.66% ± 0.09	0.8% ± 0.02
IPv6-enabled[2]	▲22.7% ± 0.6	▲22.6% ± 1.0	▲20.7% ± 0.4	▲12.9% ± 0.9	▲14.8% ± 0.8	▲10.8% ± 0.2	4.1% ± 0.2
CAA-enabled[1]	▲15.3% ± 0.9	▲5.6% ± 0.3	▲27.9% ± 0.3	▲1.7% ± 0.1	▲1.0% ± 0.0	▲1.5% ± 0.0	0.1% ± 0.0
CNAMEs[3]	■53.1% ± 1.1	▲44.46% ± 0.43	▲64.8% ± 0.34	■44.1% ± 1	▼27.86% ± 1	▲39.81% ± 0.15	51.4% ± 1.7
CDNs (via CNAME)[3]	▲27.5% ± 0.89	▲29.9% ± 0.37	▲36.1% ± 0.22	▲6% ± 0.6	▲10.14% ± 0.63	▲2.6% ± 0.01	1.3% ± 0.004
Unique AS IPv4 (avg.)[3,4]	256 ± 5	132 ± 4	250 ± 3	19511 ± 597	16922 ± 584	17418 ± 61	34876 ± 53
Unique AS IPv6 (avg.)[3,4]	44 ± 5	26 ± 2	48 ± 30	1856 ± 56	2591 ± 157	1236 ± 793	3025 ± 9
Top 5 AS (Share)[3]	▲52.68% ± 1.74	▲53.33% ± 1.75	▲51.74% ± 1.73	▲25.68% ± 0.67	■33.95% ± 1.06	▲22.29% ± 0.17	40.22 ± 0.09
TLS-capable[5]	▲89.6%	▲66.2%	▲84.7%	▲74.65%	■43.05%	▲62.89%	36.69%
HSTS-enabled HTTPS[5]	▲22.9%	▲13.0%	▲27.4%	▲12.17%	▲11.65%	■8.44%	7.63%
HTTP2[3]	▲47.5% ± 0.75	▲36.3% ± 2.4	▲36.6% ± 0.72	▲26.6% ± 0.88	▲19.11% ± 0.63	▲19.8% ± 0.15	7.84% ± 0.08

1: μ Apr, 2018 2: μ of JOINT period (6.6.17−30.4.18) 3: μ Apr, 2018 - 8. May, 2018 4: no share, thus no ▼ , ■ , or ▲ 5: Single day/list in May, 2018 6: For base values over 40%, the test for significant deviation is 25% and 5σ.

(a) % of NXDOMAIN responses. (b) % of IPv6 Adoption. (c) % of CAA-enabled domains.
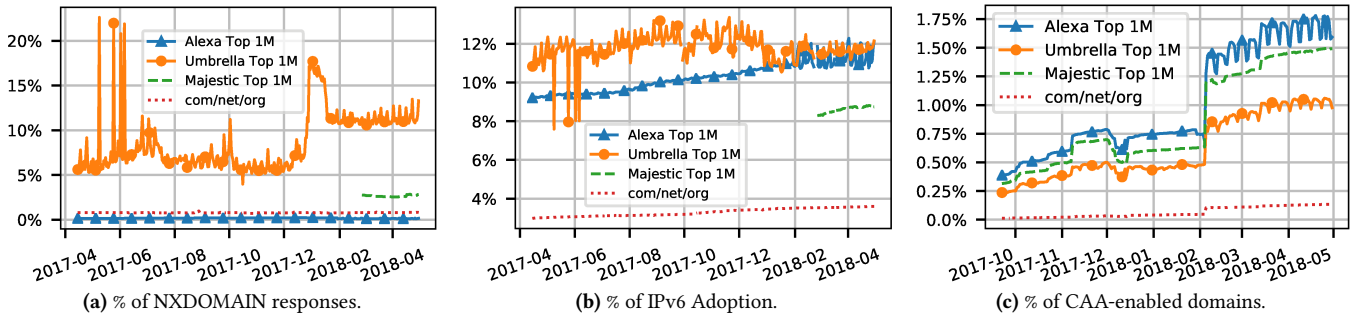
Figure 6: DNS characteristics in the Top 1M lists and general population of about 158M domains.

AAAA record or within a chain of up to 10 CNAMEs. At 11–13%, we find IPv6 enablement across top lists to significantly exceed the general population of domains at 4%. Also, the highest adoption lies with Umbrella, a good indication for IPv6 adoption: when the most frequently resolved DNS names support IPv6, many subsequent content requests are enabled to use IPv6.

CAA Adoption: Exemplary for other record types, we also investigate the adoption of Certification Authority Authorization (CAA) records in top lists and the general population. CAA is a rather new record type, and has become mandatory for CAs to check before certificate issuance, *cf.*, [122, 128]. We measure CAA adoption as described in [122], *i.e.*, the count of base domains with an *issue* or *issuewild* set. Similar to IPv6 adoption, we find CAA adoption among top lists (1–2%) to significantly exceed adoption among the general population at 0.1%. Even more stunning, the Top 1k lists feature a CAA adoption of up to 28%, distorting the 0.1% in the general population by two magnitudes.

Takeaway: The DNS-focused results above highlight that top lists may introduce a picture where results significantly differ from the general population, a popularity bias to be kept in mind. Figure 6 also shows that Umbrella, and recently Alexa, can have different results when using a different day. The daily differences, ranging, *e.g.*, from 1.5–1.8% of CAA adoption around a mean of 1.7% for Alexa, are not extreme, but should be accounted for.

8.1.2 Hosting Infrastructure Perspective. Domains can be hosted by users themselves, by hosting companies, or a variety of CDNs. The hosting landscape is subject to a body of research that is using top lists to obtain target domains. Here, we study the share of hosting infrastructures in different top lists.

CDN Prevalence: We start by studying the prevalence of CDNs in top lists and the general population of all *com/net/org* domains. Since many CDNs use DNS CNAME records, we perform daily DNS resolutions in April 2018, querying all domains both raw *www*-prefixed. We match the observed CNAME records against a list of CNAME patterns for 77 CDNs [129] to identify CDN use.

We first observe that the prevalence of CDNs differs by list and domain rank (see Table 5), with all Top 1M lists exceeding the general population by at least a factor of 2, and all Top 1k lists exceeding the general population by at least a factor of 20. When grouping the CDN ratio per list by weekdays (see Figure 7a), we observe minor influences of weekends vs. weekdays due to the top list dynamics described in §6.2.

After adoption of CDNs in general, we study the structure of CDN adoption. We analyse the top 5 CDNs and show their distribution in Figure 7 to study if the relative share is stable over different lists. We thus show the fraction of domains using one of the top 5 CDNs for both a subset of the Top 1k and the entire list of Top 1M domains per list. We first observe that the relative share of the top 5 CDNs differs by list and rank (see Figure 7b), but is generally very
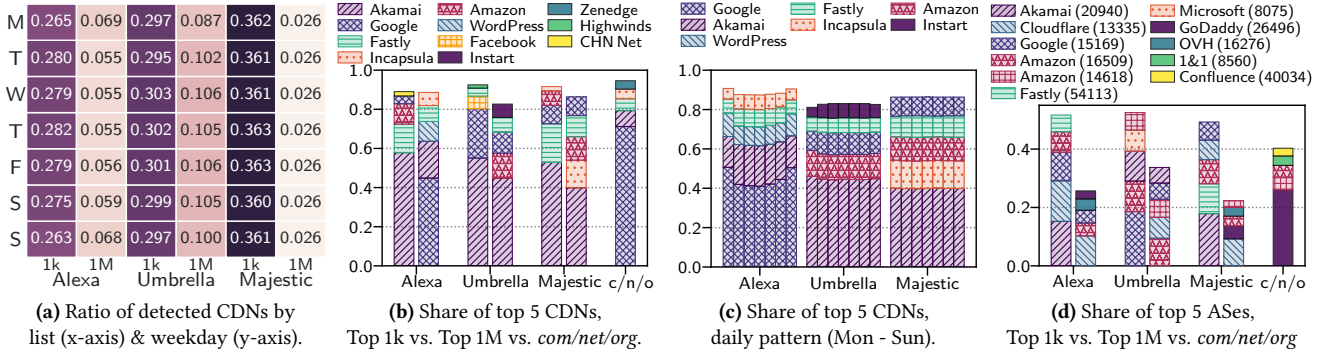
(a) Ratio of detected CDNs by list (x-axis) & weekday (y-axis).

(b) Share of top 5 CDNs, Top 1k vs. Top 1M vs. *com/net/org*.

(c) Share of top 5 CDNs, daily pattern (Mon - Sun).

(d) Share of top 5 ASes, Top 1k vs. Top 1M vs. *com/net/org*

Figure 7: Overall CDN ratio, ratio of top 5 CDNs, and ratio of top 5 ASes, dependent on list, list type, and weekday.

high at >80%. The biggest discrepancy is between using a top list and focusing on the general population of *com/net/org* domains. Google dominates the general population with a share of 71.17% due to many (private) Google-hosted sites. Domains in top lists are more frequently hosted by typical CDNs (*e.g.*, Akamai). Grouping the CDN share per list by weekday in Figure 7c shows a strong weekend/weekday pattern for Alexa, due to the rank dynamics discussed in §6.2. Interestingly, the weekend days have a higher share of Google DNS, indicating that more privately-hosted domains are visited on the weekend.

These observations highlight that using a top list or not has significant influence on the top 5 CDNs observed, and, if using Alexa, the day of list creation as well.

ASes: We next analyse the distribution of Autonomous Systems (AS) that announce a DNS name's A record in BGP, as per Route Views data from the day of the measurement, obtained from [130]. First, we study the AS diversity by counting the number of different ASes hit by the different lists. We observe lists to experience large differences in the number of unique ASes (*cf.*, Table 5); while Alexa Top 1M hits the most ASes, *i.e.*, 19511 on average, Umbrella Top 1M hits the fewest, *i.e.*, 16922 on average. To better understand which ASes contribute the most IPs, we next focus on studying the top ASes. Figure 7d shows the top 5 ASes for the Top 1k and Top 1M domains of each list, as well as the set of *com/net/org* domains. We observe that both the set and share of involved ASes differ by list.

We note that the general share of the top 5 ASes is 40% in the general population, compared to an average of 53% in the Top 1k and an average of 27% in the Top 1M lists.

In terms of structure, we further observe that GoDaddy (AS26496) clearly dominates the general population with a share of 25.99%, while it only accounts for 2.74% on the Alexa Top 1M and for 4.45% on the Majestic Top 1M.

While Alexa and Majestic share a somewhat similar distribution for both the Top 1M and Top 1k lists, Umbrella offers a quite different view, with a high share of Google/AWS hosted domains, which also relates to the CDN analysis above.

This view is also eye-opening for other measurement studies: with a significant share of a population hosted by different 5 ASes, it is of no surprise that certain higher layer characteristics differ.

8.2 TLS

In line with the prevalence of TLS studies amongst the surveyed top list papers in §3, we next investigate TLS adoption among lists and the general population. To probe for TLS support, we instruct zgrab to visit each domain via HTTPS for one day per list in May 2018. As in the previous section, we measure all domains with and without *www* prefix (except for Umbrella that contains subdomains), as we found greater coverage for these domains. We were able to successfully establish TLS connections with 74.65% of the Alexa, 62.89% of the Majestic, 43.05% of the Umbrella, and 36.69% of the *com/net/org* domains (*cf.*, Table 5). For Top 1k domains, TLS support further increases by 15–30% per list.

These results show TLS support to be most pronounced among Alexa-listed domains, and that support in top lists generally exceeds the general population.

HSTS: As one current research topic [21], we study the prevalence of HTTP Strict Transport Security (HSTS) among TLS enabled domains. We define a domain to be HSTS–enabled if the domain provides a valid HSTS header with a *max-age* setting >0. Out of the TLS-enabled domains, 12.17% of the Alexa, 11.65% of the Umbrella, 8.44% of the Majestic, and 7.63% of the com/net/org domains provide HSTS support (see Table 5). Only inspecting Top 1k domains again increases support significantly to 22.9% for Alexa, 13.0% for Umbrella, and 27.4% for Majestic. HSTS support is, again, over-represented in top lists.

8.3 HTTP/2 Adoption

One academic use of top lists is to study the adoption of upcoming protocols, *e.g.*, HTTP/2 [125, 131]. The motivation for probing top listed domains can be based on the assumption that popular domains are more likely to adopt new protocols and are thus promising targets to study. We thus exemplify this effect and the influence of different top lists by probing domains in top lists and the general population for their HTTP/2 adoption.

We try to fetch the domains' landing page via HTTP/2 by using the nghttp2 library. We again *www*-prefix all domains in Alexa and Majestic. In case of a successfully established HTTP/2 connection, we issue a GET request for the / page of the domain. We follow up to 10 redirects and if actual data for the landing page is transferred via HTTP/2, we count the domain as HTTP/2-enabled. We probe top lists on a daily basis and the larger zone file on a weekly basis.
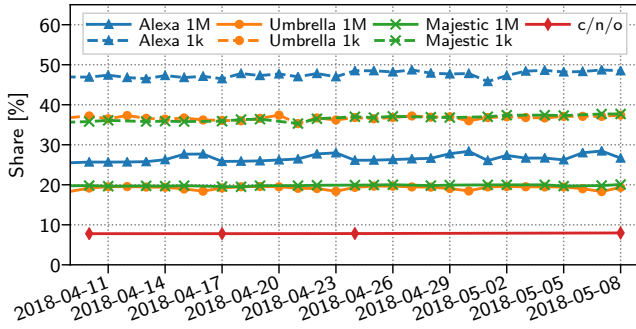
Figure 8: HTTP/2 adoption over time for the Top 1k and Top 1M lists and *com/net/org* domains.

We show HTTP/2 adoption in Figure 8. First, we observe that the HTTP/2 adoption of all *com/net/org* domains is 7.84% on average and thus significantly lower than for domains listed in Top 1M lists, (up to 26.6% for Alexa) and even more so for Top 1k lists, which show adoption around 35% or more.

One explanation is that, as shown above, popular domains are more likely hosted on progressive infrastructures (*e.g.*, CDNs) than the general population.

We next investigate HTTP/2 adoption between top lists based on Figure 8. Unsurprisingly, we observe HTTP/2 adoption differs by list and by weekday for those lists with a weekday pattern (*cf.*, §6.2). We also note the extremely different result when querying the Top 1k lists as compared to the general population.

8.4 Takeaway

We have analysed the properties of top lists and the general population across many layers, and found that top lists *(i)* generally show significantly more extreme measurement results, *e.g.*, protocol adoption. This effect is pronounced to up to 2 orders of magnitude for the Top 1k domains. Results can *(ii)* be affected by a weekly pattern, *e.g.*, the % of protocol adoption may yield a different result when using a list generated on a weekend as compared to a weekday. This is a significant limitation to be kept in mind when using top lists for measurement studies.

9 DISCUSSION

We have shown in §3 that top lists are being frequently used in scientific studies. We acknowledge that using top lists has distinct advantages—they provide a set of relevant domains at a small and stable size that can be compared over time. However, the use of top lists also comes with certain disadvantages, which we have explored in this paper.

First, while it is the stated purpose of a top list to provide a sample biased towards the list's specific measure of popularity, these samples do not represent the general state in the Internet well: we have observed in §8 that almost all conceivable measurements suffer significant bias when using a Top 1M list, and excessive bias in terms of magnitudes when using a Top 1k list. This indicates that domains in top lists exhibit behaviour significantly different from the general population—quantitative insights based on top list domains likely will not generalise.

Second, we have shown that top lists can significantly change from day to day, rendering results of one-off measurements unstable. A similar effect is that lists may be structurally different on weekends and weekdays, yielding differences in results purely based on the day of week when a list was downloaded.

Third, the choice of a certain top list can significantly influence measurement results as well, *e.g.*, for CDN or AS structure (*cf.*, §8.1.2), which stems from different lists having different sampling biases. While these effects can be desired, *e.g.*, to find many domains that adopt a certain new technology, it leads to bad generalisation of results to "the Internet", and results obtained from measuring top lists must be interpreted very carefully.

9.1 Recommendation for Top List Use

Based on our observations, we develop specific recommendations for the use of top lists. §3 has revealed that top lists are used for different purposes in diverse fields of research. The impact of the specific problems we have discussed will differ by study purpose, which is why we consider the following a set of core questions to be considered by study authors—and not a definite guideline.

Match Choice of List to Study Purpose: Based on a precise understanding of what the domains in a list represent, an appropriate list type should be carefully chosen for a study. For example, the Umbrella list represents DNS names queried by many individual clients using OpenDNS (not only PCs, but also mobile devices and IoT devices), some bogus, some non-existent, but overall a representation of typical DNS traffic, and may form a good base for DNS analyses. The Alexa list gives a solid choice of functional websites frequently visited by users, and may be a good choice for a human web-centered study. Through its link-counting, the Majestic list also includes "hidden" links, and may include domains frequently loaded, but not necessarily knowingly requested by humans. To obtain a reasonably general picture of the Internet, we recommend to scan a large sample, such as the "general population" used in §8, *i.e.*, the set of all *com/net/org* domains.

Consider Stability: With lists changing up to 50% per day, insights from measurement results might not even generalise to the next day. For most measurement studies, stability should be increased by conducting repeated, longitudinal measurements. This also helps to avoid bias from weekday vs. weekend lists.

Document List and Measurement Details: Studies should note the precise list (*e.g.*, Alexa Global Top 1M), its download date, and the measurements date to enable basic replicability. Ideally, the list used should be shared in a paper's dataset.

9.2 Desired Properties for Top Lists

Based on the challenges discussed in this work, we derive various properties that top lists should offer:

Consistency: The characteristic, mainly structure and stability, of top lists should be kept static over time. Where changes are required due to the evolving nature of the Internet, these should be announced and documented.

Transparency: Top list providers should be transparent about their ranking process and biases to help researchers understand and potentially control those biases. This may, of course, contradict the business interests of commercial list providers.

Stability: List stability faces a difficult trade-off: While capturing the ever-evolving trends in the Internet requires recent data, many typical top list uses are not stable to changes of up to 50% per day. We hence suggest that lists should be offered as long-term (*e.g.*, a 90-day sliding window) and short-term (*e.g.*, only the most recent data) versions.

9.3 Ethical Considerations

We aim to minimise harm to all stakeholders possibly affected by our work. For active scans, we minimise interference by following best scanning practices [132], such as maintaining a blacklist, using dedicated servers with meaningful rDNS records, websites, and abuse contacts. We assess whether data collection can harm individuals and follow the beneficence principle as proposed by [133, 134].

Regarding list influencing in §7, the ethical implications of inserting a test domain into the Top 1M domains is small and unlikely to cause any harm. In order to influence Umbrella ranks, we generated DNS traffic. For this, we selected query volumes unlikely to cause problems with the OpenDNS infrastructure or the RIPE Atlas platform. Regarding the RIPE Atlas platform, we spread probes across the measurements as carefully as possible: 10k probes queried specific domains 100, 50, 10, and 1 times per day. In addition, 100, 1000, and 5000 probes performed an additional 100 queries per day. Per probe, that means 6,100 probes generated 261 queries per day (fewer than 11 queries per hour), and another 3,900 probes generated 161 queries per day. Refer to Figure 5 to visualise the query volume. That implies a total workload of around 2,220,000 queries per day. As the OpenDNS service is anycasted across multiple locations, it seems unlikely that our workload was a problem for the service.

10 RELATED WORK

We consider our work to be related to three fields:

Sound Internet Measurements: There exists a canon of work with guidelines on sound Internet measurements, such as [132, 135–137]. These set out useful guidelines for measurements in general, but do not specifically tackle the issue of top lists.

Measuring Web Popularity: Understanding web popularity is important for marketing as well as for business performance analyses. A book authored by Croll and Power [138] warns site owners about the potential instrumentation biases present in Alexa ranks, specially with low-traffic sites. Besides that, there is a set of blog posts and articles from the SEO space about anecdotal problems with certain top lists, but none of these conduct systematic analyses [4, 139].

Limitations of Using Top Lists in Research: Despite the fact that top lists are widely used by research papers, we are not aware of any study focusing on the content of popular lists. However, a number of research papers mentioned the limitations of relying on those ranks for their specific research efforts [45, 67]. Wählisch *et al.* [90] discuss the challenges of using top lists for web measurements. They demonstrate that results vary when including *www* subdomains, and investigate root causes such as routing failures. The aforementioned recent work by le Pochat *et al.* [114] focuses on manipulating top lists.

11 CONCLUSION

To the best of our knowledge, this is the first comprehensive study of the structure, stability, and significance of popular Internet top lists. We have shown that use of top lists is significant among networking papers, and found distinctive structural characteristics per list. List stability has revealed interesting highlights, such as up to 50% churn per day for some lists. We have closely investigated ranking mechanisms of lists and manipulated a test domain's Umbrella rank in a controlled experiment. Systematic measurement of top list domain characteristics and reproduction of studies has revealed that top lists in general significantly distort results from the general population, and that results can depend on the day of week. We closed our work with a discussion on desirable properties of top lists and recommendations for top list use in science. We share code, data, and additional insights under

https://toplists.github.io

For long-term access, we provide an archival mirror at the TUM University Library: https://mediatum.ub.tum.de/1452290.

REFERENCES

[1] Alexa. Top 1M sites. https://www.alexa.com/topsites, May 24, 2018. http://s3.dualstack.us-east-1.amazonaws.com/alexa-static/top-1m.csv.zip.

[2] Cisco. Umbrella Top 1M List. https://umbrella.cisco.com/blog/blog/2016/12/14/cisco-umbrella-1-million/.

[3] Majestic. https://majestic.com/reports/majestic-million/, May 17, 2018.

[4] Matthew Woodward. Ahrefs vs Majestic SEO – 1 Million Reasons Why Ahrefs Is Better. https://www.matthewwoodward.co.uk/experiments/ahrefs-majestic-seo-1-million-domain-showdown/, May 23, 2018.

[5] Alexa. The Alexa Extension. https://web.archive.org/web/20160604100555/http://www.alexa.com/toolbar, June 04, 2016.

[6] Alexa. Alexa Increases its Global Traffic Panel. https://blog.alexa.com/alexa-panel-increase/, May 17, 2018.

[7] Alexa. Top 6 Myths about the Alexa Traffic Rank. https://blog.alexa.com/top-6-myths-about-the-alexa-traffic-rank/, May 22, 2018.

[8] Alexa. What's going on with my Alexa Rank? https://support.alexa.com/hc/en-us/articles/200449614, May 17, 2018.

[9] Majestic. Majestic Million CSV now free for all, daily. https://blog.majestic.com/development/majestic-million-csv-daily/, May 17, 2018.

[10] Quantcast. https://www.quantcast.com/top-sites/US/1.

[11] Statvoo. https://statvoo.com/top/sites, May 17, 2018.

[12] Google. Chrome User Experience Report. https://developers.google.com/web/tools/chrome-user-experience-report/, May 15, 2018.

[13] SimilarWeb Top Websites Ranking. https://www.similarweb.com/top-websites.

[14] Vasileios Giotsas, Philipp Richter, Georgios Smaragdakis, Anja Feldmann, Christoph Dietzel, and Arthur Berger. Inferring BGP Blackholing Activity in the Internet. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, November 2017.

[15] Srikanth Sundaresan, Xiaohong Deng, Yun Feng, Danny Lee, and Amogh Dhamdhere. Challenges in Inferring Internet Congestion Using Throughput Measurements. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, November 2017.

[16] Zhongjie Wang, Yue Cao, Zhiyun Qian, Chengyu Song, and Srikanth V. Krishnamurthy. Your State is Not Mine: A Closer Look at Evading Stateful Internet Censorship. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, November 2017.

[17] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Nicolas Kourtellis, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. The Web Centipede: Understanding How Web Communities Influence Each Other Through the Lens of Mainstream and Alternative News Sources. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, November 2017.

[18] Austin Murdock, Frank Li, Paul Bramsen, Zakir Durumeric, and Vern Paxson. Target Generation for Internet-wide IPv6 Scanning. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, November 2017.

[19] Jan Rüth, Christian Bormann, and Oliver Hohlfeld. Large-scale Scanning of TCP's Initial Window. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, November 2017.

[20] Umar Iqbal, Zubair Shafiq, and Zhiyun Qian. The Ad Wars: Retrospective Measurement and Analysis of Anti-adblock Filter Lists. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, November 2017.

[21] Johanna Amann, Oliver Gasser, Quirin Scheitle, Lexi Brent, Georg Carle, and Ralph Holz. Mission Accomplished?: HTTPS Security After Diginotar. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, November 2017.

[22] Joe DeBlasio, Stefan Savage, Geoffrey M. Voelker, and Alex C. Snoeren. Tripwire: Inferring Internet Site Compromise. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, November 2017.

[23] Shehroze Farooqi, Fareed Zaffar, Nektarios Leontiadis, and Zubair Shafiq. Measuring and Mitigating Oauth Access Token Abuse by Collusion Networks. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, November 2017.

[24] Janos Szurdi and Nicolas Christin. Email Typosquatting. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, November 2017.

[25] Enrico Bocchi, Luca De Cicco, Marco Mellia, and Dario Rossi. The Web, the Users, and the MOS: Influence of HTTP/2 on User Experience. In *International Conference on Passive and Active Network Measurement*, pages 47–59. Springer, 2017.

[26] Ilker Nadi Bozkurt, Anthony Aguirre, Balakrishnan Chandrasekaran, P Brighten Godfrey, Gregory Laughlin, Bruce Maggs, and Ankit Singla. Why is the Internet so Slow?! In *International Conference on Passive and Active Network Measurement*, pages 173–187. Springer, 2017.

[27] Stephen Ludin. Measuring What is Not Ours: A Tale of 3rd Party Performance. In *Passive and Active Measurement: 18th International Conference, PAM 2017, Sydney, NSW, Australia, March 30-31, 2017, Proceedings*, volume 10176, page 142. Springer, 2017.

[28] Kittipat Apicharttrisorn, Ahmed Osama Fathy Atya, Jiasi Chen, Karthikeyan Sundaresan, and Srikanth V Krishnamurthy. Enhancing WiFi Throughput with PLC Extenders: A Measurement Study. In *International Conference on Passive and Active Network Measurement*, pages 257–269. Springer, 2017.

[29] Alexander Darer, Oliver Farnan, and Joss Wright. FilteredWeb: A framework for the Automated Search-based Discovery of Blocked URLs. In *Network Traffic Measurement and Analysis Conference (TMA), 2017*, pages 1–9. IEEE, 2017.

[30] Jelena Mirkovic, Genevieve Bartlett, John Heidemann, Hao Shi, and Xiyue Deng. Do You See Me Now? Sparsity in Passive Observations of Address Liveness. In *Network Traffic Measurement and Analysis Conference (TMA), 2017*, pages 1–9. IEEE, 2017.

[31] Quirin Scheitle, Oliver Gasser, Minoo Rouhi, and Georg Carle. Large-scale Classification of IPv6-IPv4 Siblings with Variable Clock Skew. In *Network Traffic Measurement and Analysis Conference (TMA), 2017*, pages 1–9. IEEE, 2017.

[32] Paul Pearce, Ben Jones, Frank Li, Roya Ensafi, Nick Feamster, Nick Weaver, and Vern Paxson. Global Measurement of DNS Manipulation. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security '17)*, August 2017.

[33] Rachee Singh, Rishab Nithyanand, Sadia Afroz, Paul Pearce, Michael Carl Tschantz, Phillipa Gill, and Vern Paxson. Characterizing the Nature and Dynamics of Tor Exit Blocking. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security '17)*, August 2017.

[34] Tao Wang and Ian Goldberg. Walkie-Talkie: An Efficient Defense Against Passive Website Fingerprinting Attacks. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security '17)*, August 2017.

[35] Sebastian Zimmeck, Jie S Li, Hyungtae Kim, Steven M Bellovin, and Tony Jebara. A Privacy Analysis of Cross-device Tracking. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security '17)*, August 2017.

[36] Taejoong Chung, Roland van Rijswijk-Deij, Balakrishnan Chandrasekaran, David Choffnes, Dave Levin, Bruce M Maggs, Alan Mislove, and Christo Wilson. A Longitudinal, End-to-End View of the DNSSEC Ecosystem. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security '17)*, August 2017.

[37] Katharina Krombholz, Wilfried Mayer, Martin Schmiedecker, and Edgar Weippl. "I Have No Idea What I'm Doing" – On the Usability of Deploying HTTPS. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security '17)*, August 2017.

[38] Adrienne Porter Felt, Richard Barnes, April King, Chris Palmer, Chris Bentzel, and Parisa Tabriz. Measuring HTTPS Adoption on the Web. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security '17)*, August 2017.

[39] Ben Stock, Martin Johns, Marius Steffens, and Michael Backes. How the Web Tangled Itself:Uncovering the History of Client-Side Web (In)Security. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security '17)*, August 2017.

[40] Pepe Vila and Boris Köpf. Loophole: Timing Attacks on Shared Event Loops in Chrome. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security '17)*, August 2017.

[41] Jörg Schwenk, Marcus Niemietz, and Christian Mainka. Same-Origin Policy: Evaluation in Modern Browser. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security '17)*, August 2017.

[42] Stefano Calzavara, Alvise Rabitti, and Michele Bugliesi. CCSP: Controlled Relaxation of Content Security Policies by Runtime Policy Composition. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security '17)*, August 2017.

[43] Fang Liu, Chun Wang, Andres Pico, Danfeng Yao, and Gang Wang. Measuring the Insecurity of Mobile Deep Links of Android. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security '17)*, August 2017.

[44] Paul Pearce, Roya Ensafi, Frank Li, Nick Feamster, and Vern Paxson. Augur: Internet-Wide Detection of Connectivity Disruptions. In *IEEE Symposium on Security and Privacy*, 2017.

[45] Sumayah Alrwais, Xiaojing Liao, Xianghang Mi, Peng Wang, Xiaofeng Wang, Feng Qian, Raheem Beyah, and Damon McCoy. Under the Shadow of Sunshine: Understanding and Detecting Bulletproof Hosting on Legitimate Service Provider Networks. In *IEEE Symposium on Security and Privacy*, 2017.

[46] Oleksii Starov and Nick Nikiforakis. XHOUND: Quantifying the Fingerprintability of Browser Extensions. In *IEEE Symposium on Security and Privacy*, 2017.

[47] Chaz Lever, Platon Kotzias, Davide Balzarotti, Juan Caballero, and Manos Antonakakis. A Lustrum of Malware Network Communication: Evolution and Insights. In *IEEE Symposium on Security and Privacy*, 2017.

[48] James Larisch, David Choffnes, Dave Levin, Bruce M. Maggs, Alan Mislove, and Christo Wilson. CRLite: A Scalable System for Pushing All TLS Revocations to All Browsers. In *IEEE Symposium on Security and Privacy*, 2017.

[49] Nethanel Gelernter, Senia Kalma, Bar Magnezi, and Hen Porcilan. The Password Reset MitM Attack. In *IEEE Symposium on Security and Privacy*, 2017.

[50] Milad Nasr, Amir Houmansadr, and Arya Mazumdar. Compressive Traffic Analysis: A New Paradigm for Scalable Traffic Analysis. In *CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, November 2017.

[51] Daiping Liu, Zhou Li, Kun Du, Haining Wang, Baojun Liu, and Haixin Duan. Don't Let One Rotten Apple Spoil the Whole Barrel: Towards Automated Detection of Shadowed Domains. In *CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, November 2017.

[52] Thomas Vissers, Timothy Barron, Tom Van Goethem, Wouter Joosen, and Nick Nikiforakis. The Wolf of Name Street: Hijacking Domains Through Their Nameservers. In *CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, November 2017.

[53] Ada Lerner, Tadayoshi Kohno, and Franziska Roesner. Rewriting History: Changing the Archived Web from the Present. In *CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, November 2017.

[54] Yinzhi Cao, Zhanhao Chen, Song Li, and Shujiang Wu. Deterministic Browser. In *CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, November 2017.

[55] Yizheng Chen, Yacin Nadji, Athanasios Kountouras, Fabian Monrose, Roberto Perdisci, Manos Antonakakis, and Nikolaos Vasiloglou. Practical Attacks Against Graph-based Clustering. In *CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, November 2017.

[56] Sebastian Lekies, Krzysztof Kotowicz, Samuel Groß, Eduardo A. Vela Nava, and Martin Johns. Code-Reuse Attacks for the Web: Breaking Cross-Site Scripting Mitigations via Script Gadgets. In *CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, November 2017.

[57] Milad Nasr, Hadi Zolfaghari, and Amir Houmansadr. The Waterfall of Liberty: Decoy Routing Circumvention That Resists Routing Attacks. In *CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, November 2017.

[58] Panagiotis Kintis, Najmeh Miramirkhani, Charles Lever, Yizheng Chen, Rosa Romero-Gómez, Nikolaos Pitropakis, Nick Nikiforakis, and Manos Antonakakis. Hiding in Plain Sight: A Longitudinal Study of Combosquatting Abuse. In *CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, November 2017.

[59] Doowon Kim, Bum Jun Kwon, and Tudor Dumitraş. Certified Malware: Measuring Breaches of Trust in the Windows Code-Signing PKI. In *CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, November 2017.

[60] Peter Snyder, Cynthia Taylor, and Chris Kanich. Most Websites Don't Need to Vibrate: A Cost-Benefit Approach to Improving Browser Security. In *CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, November 2017.

[61] Benjamin Greschbach, Tobias Pulls, Laura M. Roberts, Phillip Winter, and Nick Feamster. The Effect of DNS on Tor's Anonymity. In *24th Annual Network and Distributed System Security Symposium, NDSS 2017*, NDSS '17, February 2017.

[62] Tobias Lauinger, Abdelberi Chaabane, Sajjad Arshad, William Robertson, Christo Wilson, and Engin Kirda. Thou Shalt Not Depend on Me: Analysing the Use of Outdated JavaScript Libraries on the Web. In *24th Annual Network and Distributed System Security Symposium, NDSS 2017*, NDSS '17, February 2017.

[63] Najmeh Miramirkhani, Oleksii Starov, and Nick Nikiforakis. Dial One for Scam: A Large-Scale Analysis of Technical Support Scams. In *24th Annual Network and Distributed System Security Symposium, NDSS 2017*, NDSS '17', February 2017.

[64] Marc Anthony Warrior, Uri Klarman, Marcel Flores, and Aleksandar Kuzmanovic. Drongo: Speeding Up CDNs with Subnet Assimilation from the Client. In *CoNEXT '17: Proceedings of the 13th International Conference on Emerging Networking EXperiments and Technologies*. ACM, December 2017.

[65] Shinyoung Cho, Rishab Nithyanand, Abbas Razaghpanah, and Phillipa Gill. A Churn for the Better. In *CoNEXT '17: Proceedings of the 13th International Conference on Emerging Networking EXperiments and Technologies*. ACM, December 2017.

[66] Wai Kay Leong, Zixiao Wang, and Ben Leong. TCP Congestion Control Beyond Bandwidth-Delay Product for Mobile Cellular Networks. In *CoNEXT '17: Proceedings of the 13th International Conference on Emerging Networking EXperiments and Technologies*. ACM, December 2017.

[67] Mario Almeida, Alessandro Finamore, Diego Perino, Narseo Vallina-Rodriguez, and Matteo Varvello. Dissecting DNS Stakeholders in Mobile Networks. In *CoNEXT '17: Proceedings of the 13th International Conference on Emerging Networking EXperiments and Technologies*. ACM, December 2017.

[68] David Naylor, Richard Li, Christos Gkantsidis, Thomas Karagiannis, and Peter Steenkiste. And Then There Were More: Secure Communication for More Than Two Parties. In *CoNEXT '17: Proceedings of the 13th International Conference on Emerging Networking EXperiments and Technologies*. ACM, December 2017.

[69] Thomas Holterbach, Stefano Vissicchio, Alberto Dainotti, and Laurent Vanbever. Swift: Predictive fast reroute. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, SIGCOMM '17. ACM, August 2017.

[70] Costas Iordanou, Claudio Soriente, Michael Sirivianos, and Nikolaos Laoutaris. Who is Fiddling with Prices?: Building and Deploying a Watchdog Service for E-commerce. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, SIGCOMM '17. ACM, August 2017.

[71] Vaspol Ruamviboonsuk, Ravi Netravali, Muhammed Uluyol, and Harsha V. Madhyastha. Vroom: Accelerating the Mobile Web with Server-Aided Dependency Resolution. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, SIGCOMM '17. ACM, August 2017.

[72] Sanae Rosen, Bo Han, Shuai Hao, Z. Morley Mao, and Feng Qian. Push or Request: An Investigation of HTTP/2 Server Push for Improving Mobile Web Performance. In *Proceedings of the 26th International Conference on World Wide Web*, 2017.

[73] Elias P. Papadopoulos, Michalis Diamantaris, Panagiotis Papadopoulos, Thanasis Petsas, Sotiris Ioannidis, and Evangelos P. Markatos. The Long-Standing Privacy Debate: Mobile Websites vs Mobile Apps. In *Proceedings of the 26th International Conference on World Wide Web*, 2017.

[74] Sanae Rosen, Bo Han, Shuai Hao, Z. Morley Mao, and Feng Qian. Extended Tracking Powers: Measuring the Privacy Diffusion Enabled by Browser Extensions. In *Proceedings of the 26th International Conference on World Wide Web*, 2017.

[75] Deepak Kumar, Zane Ma, Zakir Durumeric, Ariana Mirian, Joshua Mason, J. Alex Halderman, and Michael Bailey. Security Challenges in an Increasingly Tangled Web. In *Proceedings of the 26th International Conference on World Wide Web*, 2017.

[76] Gareth Tyson, Shan Huang, Felix Cuadrado, Ignacio Castro, Vasile C. Perta, Arjuna Sathiaseelan, and Steve Uhlig. Exploring HTTP Header Manipulation In-The-Wild. In *Proceedings of the 26th International Conference on World Wide Web*, 2017.

[77] Luca Soldaini and Elad Yom-Tov. Inferring Individual Attributes from Search Engine Queries and Auxiliary Information. In *Proceedings of the 26th International Conference on World Wide Web*, 2017.

[78] Kyungtae Kim, I Luk Kim, Chung Hwan Kim, Yonghwi Kwon, Yunhui Zheng, Xiangyu Zhang, and Dongyan Xu. J-Force: Forced Execution on JavaScript. In *Proceedings of the 26th International Conference on World Wide Web*, 2017.

[79] Dolière Francis Some, Nataliia Bielova, and Tamara Rezk. On the Content Security Policy Violations due to the Same-Origin Policy. In *Proceedings of the 26th International Conference on World Wide Web*, 2017.

[80] Milivoj Simeonovski, Giancarlo Pellegrino, Christian Rossow, and Michael Backes. Who Controls the Internet? Analyzing Global Threats using Property Graph Traversals. In *Proceedings of the 26th International Conference on World Wide Web*, 2017.

[81] Ajaya Neupane, Nitesh Saxena, and Leanne Hirshfield. Neural Underpinnings of Website Legitimacy and Familiarity Detection: An fNIRS Study. In *Proceedings*

[82] of the 26th International Conference on World Wide Web, 2017.

[83] Qian Cui, Guy-Vincent Jourdan, Gregor Bochmann, Russell Couturier, and Vio Onut. Tracking Phishing Attacks Over Time. In *Proceedings of the 26th International Conference on World Wide Web*, 2017.

[84] Li Chang, Hsu-Chun Hsiao, Wei Jeng, Tiffany Hyun-Jin Kim, and Wei-Hsi Lin. Security Implications of Redirection Trail in Popular Websites Worldwide. In *Proceedings of the 26th International Conference on World Wide Web*, 2017.

[85] Enrico Mariconti, Jeremiah Onaolapo, Sharique Ahmad, Nicolas Nikiforou, Manuel Egele, Nick Nikiforakis, and Gianluca Stringhini. What's in a Name? Understanding Profile Name Reuse on Twitter. In *Proceedings of the 26th International Conference on World Wide Web*, 2017.

[86] ACM. Result and Artifact Review and Badging. https://www.acm.org/publications/policies/artifact-review-badging, Acc. Jan 18 2017.

[87] Quirin Scheitle, Matthias Wählisch, Oliver Gasser, Thomas C. Schmidt, and Georg Carle. Towards an Ecosystem for Reproducible Research in Computer Networking. In *ACM SIGCOMM 2017 Reproducibility Workshop*, 2017.

[88] Matthias Flittner, Mohamed Naoufal Mahfoudi, Damien Saucez, Matthias Wählisch, Luigi Iannone, Vaibhav Bajpai, and Alex Afanasyev. A Survey on Artifacts from CoNEXT, ICN, IMC, and SIGCOMM Conferences in 2017. *SIGCOMM Comput. Commun. Rev.*, 48(1):75–80, April 2018.

[89] Damien Saucez and Luigi Iannone. Thoughts and Recommendations from the ACM SIGCOMM 2017 Reproducibility Workshop. *ACM SIGCOMM Computer Communication Review*, 48(1):70–74, 2018.

[90] Mark Allman. Comments On DNS Robustness. *IMC*, 2018.

[91] Matthias Wählisch, Robert Schmidt, Thomas C Schmidt, Olaf Maennel, Steve Uhlig, and Gareth Tyson. RiPKI: The tragic story of RPKI deployment in the Web ecosystem. In *Proceedings of the 14th ACM Workshop on Hot Topics in Networks*, page 11. ACM, 2015.

[92] Ralph Holz, Lothar Braun, Nils Kammenhuber, and Georg Carle. The SSL Landscape - A Thorough Analysis of the X.509 PKI Using Active and Passive Measurements. In *IMC*, Nov. 2011.

[93] The Internet Archive. Alexa Crawls. https://archive.org/details/alexacrawls, May 22, 2018.

[94] Mozilla. Public Suffix List: commit 2f9350. https://github.com/publicsuffix/list/commit/85fa8fbdf, Apr. 20, 2018.

[95] IANA. TLD Directory. http://data.iana.org/TLD/tlds-alpha-by-domain.txt, May 20, 2018.

[96] ICANN. Notices of Termination and Status of gTLD. https://www.icann.org/resources/pages/gtld-registry-agreement-termination-2015-10-09-en, Apr. 20, 2018.

[97] Nick Parsons. Stop using .IO Domain Names for Production Traffic. https://hackernoon.com/stop-using-io-domain-names-for-production-traffic-b6aa17eeac20, May 21, 2018.

[98] Matthew Bryant. The .io Error – Taking Control of All .io Domains With a Targeted Registration. https://thehackerblog.com/the-io-error-taking-control-of-all-io-domains-with-a-targeted-registration/, May 21, 2018.

[99] Tomislav Lombarovic. Be aware: How domain registrar can kill your business. https://www.uptimechecker.io/blog/how-domain-registrar-can-kill-your-business, May 21, 2018.

[100] ICANN. new gTLD Program Timeline. https://newgtlds.icann.org/en/program-status/timelinesen, Apr. 20, 2018.

[101] Abbas Razaghpanah, Rishab Nithyanand, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Mark Allman, Christian Kreibich, and Phillipa Gill. Apps, Trackers, Privacy, and Regulators: A Global Study of the Mobile Tracking Ecosystem. In *NDSS*, 2018.

[102] OpenDNS. Domain Tagging. https://domain.opendns.com.

[103] Abbas Razaghpanah, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Christian Kreibich, Phillipa Gill, Mark Allman, and Vern Paxson. Haystack: A multi-purpose mobile vantage point in user space. *arXiv preprint arXiv:1510.01419*, 2015.

[104] hpHosts. hpHosts Domain Blacklist, May 21, 2018. https://hosts-file.net/.

[105] Anukool Lakhina, Konstantina Papagiannaki, Mark Crovella, Christophe Diot, Eric D Kolaczyk, and Nina Taft. Structural Analysis of Network Traffic Flows. In *ACM SIGMETRICS Performance Evaluation Review*, volume 32, pages 61–72. ACM, 2004.

[106] Konstantina Papagiannaki, Nina Taft, Z-L Zhang, and Christophe Diot. Long-term Forecasting of Internet Backbone Traffic: Observations and Initial Models. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 2, pages 1178–1188. IEEE, 2003.

[107] Phillipa Gill, Martin Arlitt, Zongpeng Li, and Anirban Mahanti. Youtube Traffic Characterization: A View from the Edge. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 15–28. ACM, 2007.

[108] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 1–14. ACM, 2007.

[109] Maurice G Kendall. A New Measure of Rank Correlation. *Biometrika*, 30(1/2):81–93, 1938.

[109] Alexa. How are Alexa's traffic rankings determined? https://support.alexa.com/hc/en-us/articles/200449744, May 17, 2018.

[110] Adam Feuerstein. E-commerce loves Street: Critical Path plans encore. San Francisco Business Times, https://www.bizjournals.com/sanfrancisco/stories/1999/05/24/newscolumn4.html, May 1999.

[111] Alexa Specialist. http://www.improvealexaranking.com/, May 22, 2018.

[112] Rankboostup. https://rankboostup.com/, May 22, 2018.

[113] UpMyRank. http://www.upmyrank.com/, May 22, 2018.

[114] Victor Le Pochat, Tom Van Goethem, and Wouter Joosen. Rigging Research Results by Manipulating Top Websites Rankings. arXiv preprint arXiv:1806.01156, June 4, 2018.

[115] RIPE Atlas. Measurement IDs 124307{26,28-33}.

[116] RIPE Atlas. Measurement IDs 124674{03-10}.

[117] Majestic. About Majestic. https://blog.majestic.com/about/, May 22, 2018.

[118] Majestic. Majestic Million – Reloaded! https://blog.majestic.com/company/majestic-million-reloaded/, May 22, 2018.

[119] Majestic. A Million here… A Million there…. https://blog.majestic.com/case-studies/a-million-here-a-million-there/, May 22, 2018.

[120] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[121] Verisign. The Domain Name Industry Brief 2017Q4, 2018.

[122] Quirin Scheitle, Taejoong Chung, Jens Hiller, Oliver Gasser, Johannes Naab, Roland van Rijswijk-Deij, Oliver Hohlfeld, Ralph Holz, Dave Choffnes, Alan Mislove, and Georg Carle. A First Look at Certification Authority Authorization (CAA). ACM SIGCOMM CCR, April 2018.

[123] Roland van Rijswijk-Deij, Mattijs Jonker, Anna Sperotto, and Aiko Pras. A High-Performance, Scalable Infrastructure for Large-Scale Active DNS Measurements. IEEE JSAC, 2016.

[124] Oliver Gasser, Quirin Scheitle, Sebastian Gebhard, and Georg Carle. Scanning the IPv6 Internet: Towards a Comprehensive Hitlist. In TMA, 2016.

[125] T. Zimmermann, J. Rüth, B. Wolters, and O. Hohlfeld. How HTTP/2 pushes the web: An empirical study of HTTP/2 Server Push. In 2017 IFIP Networking Conference (IFIP Networking) and Workshops, pages 1–9, June 2017.

[126] Jakub Czyz, Mark Allman, Jing Zhang, Scott Iekel-Johnson, Eric Osterweil, and Michael Bailey. Measuring IPv6 Adoption. In ACM SIGCOMM, 2014.

[127] Steffie Jacob Eravuchira, Vaibhav Bajpai, Jürgen Schönwälder, and Sam Crawford. Measuring web similarity from dual-stacked hosts. In Network and Service Management (CNSM), 2016 12th International Conference on, pages 181–187. IEEE, 2016.

[128] Jukka Ruohonen. An Empirical Survey on the Early Adoption of DNS Certification Authority Authorization. arXiv preprint arXiv:1804.07604, 2018.

[129] Google. WebPagetest CDN domain list, cdn.h. https://github.com/WPO-Foundation/webpagetest/blob/master/agent/wpthook/cdn.h.

[130] University of Oregon. Route Views Project. http://www.routeviews.org.

[131] Matteo Varvello, Kyle Schomp, David Naylor, Jeremy Blackburn, Alessandro Finamore, and Konstantina Papagiannaki. Is the Web HTTP/2 Yet? In Thomas Karagiannis and Xenofontas Dimitropoulos, editors, Passive and Active Measurement, pages 218–232, Cham, 2016. Springer International Publishing.

[132] Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. ZMap: Fast Internet-wide Scanning and Its Security Applications. In USENIX Security, 2013.

[133] David Dittrich, Erin Kenneally, et al. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. US Department of Homeland Security, 2012.

[134] Craig Partridge and Mark Allman. Ethical Considerations in Network Measurement Papers. Communications of the ACM, 2016.

[135] Vern Paxson. Strategies for Sound Internet Measurement. In Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, pages 263–271. ACM, 2004.

[136] Mark Allman. On Changing the Culture of Empirical Internet Assessment. ACM Computer Communication Review, 43(3), July 2013. Editorial Contribution.

[137] Mark Allman, Robert Beverly, and Brian Trammell. Principles for measurability in protocol design. ACM SIGCOMM Computer Communication Review, 47(2):2–12, 2017.

[138] Alistair Croll and Sean Power. Complete web monitoring: watching your visitors, performance, communities, and competitors. " O'Reilly Media, Inc.", 2009.

[139] Michael Arrington. Alexa's Make Believe Internet; Alexa Says YouTube Is Now Bigger Than Google. Alexa Is Useless. https://techcrunch.com/2007/11/25/alexas-make-believe-internet/, 2007.