The Web Unpacked: A Quantitative Analysis of Global Web Usage

Henrique S. Xavier[a]
NIC.br, Av. das Nações Unidas, 11541, 7º Andar, São Paulo, SP, Brazil

Keywords: World Wide Web, Traffic Share, Website Visits, Web Browsing, Web Usage, Platforms.

Abstract: This paper presents an analysis of global web usage patterns based on data from 250,000 websites monitored by SimilarWeb. We estimate the total web traffic and investigate its distribution among domains and industry sectors. We detail the characteristics of the top 116 domains, which comprise an estimated one-third of all web traffic. Our analysis scrutinizes their content sources, access requirements, offline presence, and ownership features, among others. Our analysis reveals that a diminutive number of top websites captures the majority of visits. Search engines, news and media, social networks, streaming, and adult content emerge as primary attractors of web traffic, which is also highly concentrated on platforms and USA-owned websites. Much of the traffic goes to for-profit but mostly free-of-charge websites, highlighting the dominance of business models not based on paywalls.

1 INTRODUCTION

The World Wide Web (also called the Web) is an online decentralized and owner-free information system released to the public domain in 1993 (CERN, 1993). It is an application of the Internet protocol suite, a computer communication protocol that is also decentralized and owner-free. The Web itself has no moderation or gatekeepers and, for a long time, had almost no governmental regulation. Currently, the Web is the best-known and most pervasive Internet application. Web browsing is a dominant computer activity (Crichton et al., 2021), and Web-based companies figure among the largest and most valuable companies in the World (Evans and Gawer, 2016).

The importance of the Web, its lack of explicit management, and its vast potential have raised questions on what are its realized, actual characteristics in terms of main uses, applications, and content types, how visits are distributed among websites, who are the content publishers, how is the content typically produced, and much more. Answering these questions can provide valuable insights for developing public policies related to the Web, such as regulating new activities, relationships, and business models to promote positive outcomes and reduce negative impacts on society.

In the past, the Web has been thoroughly analyzed in terms of its content and graph properties (e.g., the number of webpages covering a given topic (Chakrabarti et al., 2002; Huizingh, 2000) and the number of hyperlinks connecting them (Adamic and Huberman, 2000b; Huberman, 2001)). However, previous accounts about the characteristics of the Web have primarily been based on metrics other than usage (e.g., link structure, number of websites, or companies' market value) or were built on anecdotal accounts and non-quantitative data. Some exceptions are: (Adamic and Huberman, 2000a; Webster and Lin, 2002) that showed for a restricted subset of users and websites that visitors are distributed among websites in a power-law-like fashion; and (Agarwal and Sastry, 2022), that used Alexa's ranking list to select the 100 most visited domains, Google trends to estimate their popularity and Fortiguard's classification to verify the popularity of different topics, among other things.

This paper aims to provide a data-driven general picture of web usage. Based on data about the number of monthly visits to web domains, we aim to estimate:

1. the number of domains (and which ones) required to form a representative picture of global web usage;
2. the importance of topics explored on the Web; and
3. the prevalence of broad website characteristics (e.g., access barriers, content sources, and ownership) typically encountered in web visits.

Our complete analysis can be accessed at

[a] https://orcid.org/0000-0002-9601-601X

Xavier, H.
The Web Unpacked: A Quantitative Analysis of Global Web Usage.
DOI: 10.5220/0012905900003825
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 20th International Conference on Web Information Systems and Technologies (WEBIST 2024)*, pages 183-190
ISBN: 978-989-758-718-4; ISSN: 2184-3252
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

http://github.com/cewebbr/web-unpacked.

2 DATA SET

The data analyzed in this paper are from SimilarWeb's Starter Plan, covering August to October 2023. SimilarWeb is a company that measures visits to numerous websites by combining different user monitoring methods. Other companies that provide similar data include Semrush and Ahrefs. The Starter Plan provides data about the most visited domains in each of the 210 industry segments specified by SimilarWeb, limited to 10,000 domains per industry. During the covered period, data was available for 1,336,963 domains. For each domain, SimilarWeb estimates the average monthly visits, month-over-month visit variation, and average visit duration, among other metrics. Due to limited accuracy, data is not provided for domains with less than 5,000 monthly visits. In general, subdomains are not distinguished from their parent domains (e.g., visits to docs.google.com are accounted for as part of visits to google.com along with visits to other Google subdomains), and transitions from one subdomain to another do not count as an extra domain visit, just as a single one.

The quality of SimilarWeb data has been evaluated in previous academic works and is considered good enough for relative ballpark measurements between domains. Jansen *et al.* used Google Analytics for 86 websites as a truth table and identified a systematic bias of 20% in SimilarWeb's monthly visits estimates (Jansen et al., 2022). Prantl compared monthly visits from SimilarWeb and NetMonitor[1] for 485 Czech websites and found an average absolute difference of about 42% between these two sets of measurements (Prantl and Prantl, 2018). Enterprise reports have indicated similar scenarios (Hardwick, 2018; Diachuk et al., 2021). However, given that our data on monthly visits cover over seven orders of magnitude (from 5 thousand to 86 billion), the reported deviations do not significantly affect the overall distribution of visits among websites. The data still provides precise ranks, with a Pearson correlation of 95% between monthly visit estimates from SimilarWeb and Google Analytics (Jansen et al., 2022) and a Spearman correlation of 96% between SimilarWeb's and NetMonitor's ranks (Prantl and Prantl, 2018). Also, the present work relies only on relative measurements, so any systematic biases are irrelevant.

It is important to note that web traffic exhibits annual patterns that could impact the representativeness of our three-month data sample (Liu et al., 2017; Liu et al., 2018). By analyzing five years of Google Trends data for the most visited domain names across various industries, we found that Western e-commerce sites like amazon.com and walmart.com see increased searches before Christmas, while searches for weather-related websites like weather.com and accuweather.com rise during school vacations. Assuming these annual patterns in Google Trends reflect similar patterns in actual domain visits, we estimate that our data may deviate from its annual average by up to 24%, as seen in the case of espn.com. However, major domains like youtube.com, instagram.com, and xvideos.com show only minor variations (i.e., less than 5%). These estimates suggest that, while seasonality is a factor, it is unlikely to affect our findings significantly.

For each industry, we assumed that SimilarWeb is complete up to a certain threshold (the domain with the lowest average monthly visits). In other words, we assumed that domains missing from an industry's list must have average monthly visits lower than the least visited domain in the list. However, it is essential to note that being complete in a given industry sector does not imply the entire dataset is complete, as SimilarWeb's Starter Plan limits the number of domains per industry to 10,000. To ensure completeness in the entire dataset, we must enforce a threshold on average monthly visits of 140,484, the largest individual industry threshold. This complete dataset, considered in our analysis, contains 254,661 domains and will be ranked by average monthly visits.

3 ANALYSIS

3.1 Traffic Distribution

Fig. 1 shows a log-log plot of the monthly visits V vs. rank position p for the domains in our dataset. A power-law $V(p) = V_0 p^{\beta}$ appears as a straight line in such a plot. We obtain a best-fit model to our data with $V_0 = 4.1 \times 10^{11}$ and $\beta = -1.19$. The pink bands represent a 2σ-equivalent interval (i.e., from the 2nd to the 98th percentile) for the month-over-month visit variation measured by SimilarWeb. The narrowness of the bands demonstrates that the properties of the traffic distribution should be reasonably stable over time (at least for a few months) and that the overall ranking of the domains is not dramatically altered from on month to the next.

To calculate the share of the total web traffic captured by each domain in our dataset, we first need an estimate of this total traffic. For that, we extrapo-

[1] https://www.spir.cz/projekty/netmonitor/

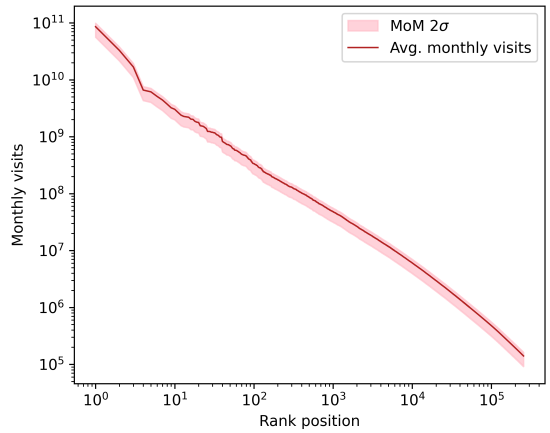

Figure 1: Average monthly visits as a function of the domain's position in the rank. Light bands represent a 2σ variation in the visits from month to month.

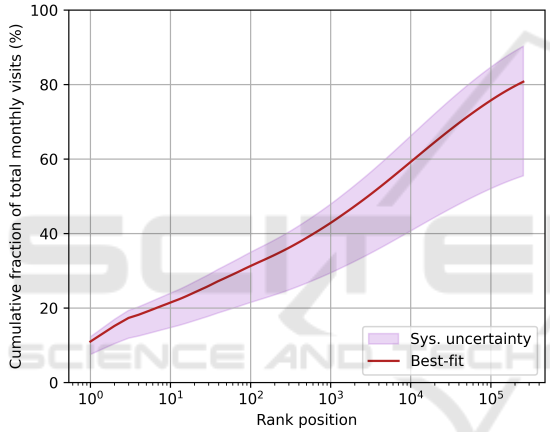

Figure 2: Best estimate of the cumulative traffic share of domains as a function of position in the rank (red line). The violet band represents the systematic uncertainty.

lated the best-fit power law to position 354 million, the number of registered domains as of June 2023 (Verisign, 2023), and aggregated the measured traffic along with the estimated traffic beyond position 254,661, resulting in a total of 781 billion visits per month. With 5.35 billion internet users as of January 2024 (Petrosyan, 2024), this amounts to 4.9 domain visits per person per day, a reasonable ballpark figure: previous work has shown that users from a metropolitan area in the United States visited an average of 20.1 websites per day on their computers, while the least active user (out of 257) visited an average of 2.9 (Crichton et al., 2021). Including rural regions and other countries should lead to a lower average. Fig. 2 shows the estimated traffic share for the 254,661 domains in our data.

To estimate a systematic uncertainty on the shares of total monthly web traffic, we fitted a power law up to position $p_i = 10(p_{\text{last}}/10)^{i/99}$, with $i = 0, \ldots, 99$ and $p_{\text{last}} = 254,661$, representing several hypothetical cases where we have data only for the first p_i domains. The smallest and largest β obtained, along with the respective power-law scaling factor $V_0 = V(p_{\text{last}})/p_{\text{last}}^{\beta}$, were used as alternative extrapolations to estimate a systematic uncertainty interval on the total monthly web traffic. The resulting interval for the traffic shares is shown as a violet band in Fig. 2.

Our analysis reveals a significant concentration of web traffic in a few domains. Our best estimate suggests that 50% of all web traffic is directed to the top 3,000 domains, and 80% is allocated to domains in our dataset. Table 1 presents values for our best traffic extrapolation and for the systematic uncertainty boundaries with the smallest and largest traffic concentration, denoted as Sys. − and Sys. +. These values are: the power-law exponent, β; the total monthly web traffic, $\sum V$; the average number of domain visits per person per day, $\bar{u}$; the extrapolated traffic for the last domain in the rank (among 354 million), V_{min}; the number of top domains that accumulate 50% of total web traffic, $p_{50\%}$; and the Gini inequality coefficient.

The V_{min} value for the Sys. − boundary is implausibly large, indicating that the traffic share must decline faster at the end of the rank, making the actual total traffic likely closer to our best estimate. The fact that the domains in our dataset likely represent the destination of 80% of all web traffic makes them an excellent sample of the current state of the Web, at least in terms of content access rate.

3.2 Popular Web Industries

In order to estimate the most popular industries on the Web, we aggregated the monthly visits into 187 non-overlapping categories defined by SimilarWeb. Our analysis shows that traffic coalesces in a few industries, with 50% directed to just five segments and 80% directed to 26 segments. In most cases, despite the power-law distribution of traffic within industries, monthly visits are not monopolized by a single domain or company, and the most popular domains accumulate at most 15% of their industry's traffic. However, there are exceptions, as noted below.

Figure 3 demonstrates the prevalence of "Search Engines" as the most visited industry. This prevalence is expected due to the lack of a built-in index on the Web, making search engines entry points for most users. This trend is further amplified by the integration of address and search bars in web browsers, leading users to search even for known websites instead of directly typing their URLs (Cannon, 2008).

Table 1: Main monthly traffic share attributes under different traffic extrapolations.

Extrap.	β	$\sum V$	$\bar{u}$	$V_{\min}$	$p_{50\%}$	Gini
Sys. −	-0.83	$1.14 \cdot 10^{12}$	7.1	341.1	62,621	82%
Best	-1.19	$7.81 \cdot 10^{11}$	4.9	27.3	2,938	97%
Sys. +	-1.52	$6.98 \cdot 10^{11}$	4.3	2.4	1,342	99%

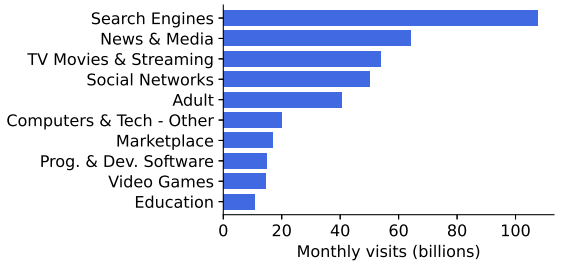

Figure 3: Monthly visits to the top 10 industries. We shortened some industry names to save space on the plot.

In this industry, google.com dominates with 79% of the traffic (considering the top 10,000 domains in the segment). The following competitors are the Chinese search engine baidu.com, with 4.7%, and the Russian search engine yandex.ru, with 3.0%.

In the "TV Movies and Streaming" segment, youtube.com (owned by Alphabet, like google.com) captures 61% of the industry's traffic. This segment is succeeded by "Social Networks and Online Communities," another case where the industry's traffic predominantly flows to a single company. When combining facebook.com, instagram.com, and whatsapp.com, Meta receives 52% of the segment's traffic. Among the remaining top 26 industries, only "Dictionaries and Encyclopedias" has a domain (wikipedia.org) with more than half (56%) of the total industry's traffic.

It is essential to observe that domains categorized as "Search Engines," such as google.com, yandex.ru, and baidu.com, offer more services than just web search. Consequently, a portion of the traffic attributed to the "Search Engines" segment pertains to other industries listed in SimilarWeb's classification, such as "Email" and "File Sharing and Hosting." For google.com, ignoring its subdomains unrelated to search results in a 15% reduction in traffic, maintaining the integrity of our rankings and keeping search as Google's main service. Conversely, the traffic directed to these subdomains would place Google as the main player in several industries had SimilarWeb classified them properly. This is the case of "Email", with mail.google.com; "Programming and Developer Software', with docs.google.com; and "File Sharing and Hosting", with drive.google.com.

3.3 Manual Inspection of Top Domains

SimilarWeb data provides no other information about the domains besides visitation metrics and the domain's industry. To address this gap, the author manually inspected and researched the 116 most visited domains, which collectively capture between 22% and 36% of the total web traffic (with a best estimate of 32%). The objective was to answer the following questions about them:

1. Does the website primarily offer Software as a Service (SaaS)?
2. Does the website produce its content? For marketplaces, does the domain owner sell its own products on the site?
3. Does the website function as a platform for user-generated content? For e-commerce platforms, does the site allow third-party sellers to list their products? Comments on original content (such as those found on some news websites) were insufficient to classify the site as a platform. We did not consider news aggregators or similar sites that curate content from the Web as platforms since the content is not user-provided.
4. Does the website require users to log in before accessing its main content?
5. Does the website charge users for access to its main content and features? We ignored charges for additional features and, for marketplaces, the charges imposed on sellers.
6. Is the domain owner's primary business or activity related to the Web, or is a significant portion of their business conducted offline? We considered news websites primarily web-related, but not e-commerce sites that sell their own products.
7. According to Wikipedia, who is the ultimate owner of the domain? In the case of subsidiaries, we identified the ultimate parent company as the domain's final owner. For partial ownership, we only considered parent companies that owned at least 50% of the subsidiaries.
8. Is the ultimate owner of the domain a for-profit organization?
9. According to Wikipedia, in which country is the domain's ultimate owner located?

We have made our annotations regarding these questions available online[2]. While the analysis of this subset of domains does not guarantee a comprehensive representation of total Web traffic, the trends observed within it are likely to extend to the following several thousand domains, thus reflecting the characteristics of a significant portion of Web usage.

Among the top 116 listed domains, Microsoft is the largest owner with 11 domains. These include linkedin.com, openai.com[3], bing.com, github.com, and several others associated with Microsoft itself and its SaaS products like office365.com. Following is Amazon, the owner of twitch.tv, imdb.com, and five Amazon-related domains. Alphabet holds ownership of five domains, including youtube.com and various Google domains, while Meta possesses four (facebook.com, instagram.com, whatsapp.com, and messenger.com). The first non-US company in terms of top domains is the Russian VK, which owns dzen.ru, vk.com, mail.ru, and ok.ru, followed by the Japanese SoftBank Group Corp., which owns three domains.

Suppose we rank the final owners by their aggregated traffic. In that case, Alphabet receives 120 billion visits monthly, accounting for almost 50% of the total traffic of the top 116 domains (see Fig. 4). This ranking also roughly follows a power-law distribution, but its best-fit exponent, β, is more negative than the best-fit for the disaggregated top 116 domains: -1.22 versus -0.98. It is important to note that this decrease in β is not an artifact caused by the aggregation itself. Randomly aggregating domains would typically blend famous and less known domains, evening out the traffic and resulting in a less negative β. Thus, there must be a socio-economic explanation for why the same companies own highly visited domains. Although advancing such an explanation is beyond the scope of this paper, we conjecture that this phenomenon partially stems from large companies acquiring popular websites (e.g., youtube.com, linkedin.com, dzen.ru, twitch.tv, imdb.com, and github.com were independent websites later purchased by their current owners).

When we aggregate the top domains' visits by the ultimate owner's country, the concentration becomes even more pronounced. The best-fit power-law exponent reaches $\beta = -2.00$, with the United States capturing 80% of the total traffic of the top 116 domains.

[2]https://github.com/cewebbr/web-unpacked/blob/main/data/cleaned/domains-annotated_v03.csv

[3]While we attribute ownership of OpenAI to Microsoft, the precise nature of their relationship remains a topic of debate.

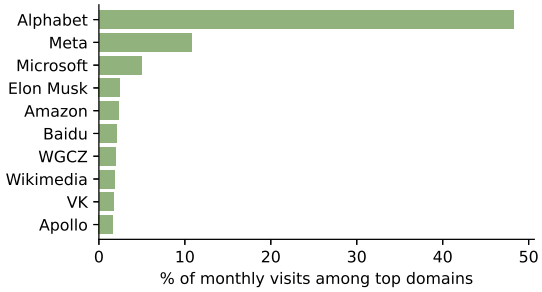

Figure 4: Fraction of the top 116 domains' total traffic aggregated by final owner. The plot shows only the top 10 final owners.

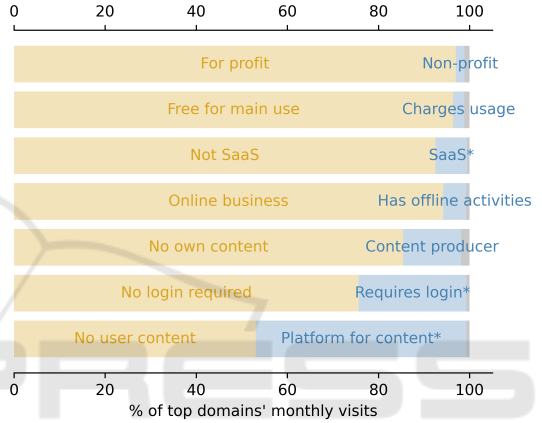

Figure 5: Fraction of the top 116 domains' total traffic directed to domains classified under seven binary properties. The gray segments denote traffic directed to unclassified domains. Segments marked with * would be larger depending on how we deal with large search engines (see text).

This dominance is followed by China (4.5%), Russia (3.7%), Japan (2.6%), the Czech Republic (2.1%), and South Korea (0.97%).

Fig. 5 summarizes the traffic shares associated with the binary answers to the remaining questions (1 to 6 and 8). Each bar in the plot illustrates how the traffic is distributed among domains subjected to a specific binary classification. Notably, the vast majority (97%) of the traffic to the top 116 domains flows to domains owned by profit-seeking companies. Interestingly, among these 116 domains, only two are not-for-profit: wikipedia.org and archiveofourown.org. This observation is intriguing, especially considering that 96% of the top traffic goes to websites that do not charge for access or their main functionalities. This seeming contradiction underscores emerging business models that rely on other forms of making a profit (Lee et al., 2006; Georgieva et al., 2015; Hermes et al., 2020b).

Figure 5 emphasizes that 85% of the total traffic

to top domains flows to platforms that do not generate their content, such as social media platforms, marketplaces, news aggregators, and search engines. Conversely, 46% of the traffic goes to websites where users themselves create content. Excluding traffic to search engines, which neither produce content nor allow users to post content, the proportion of traffic allocated to user-generated content platforms rises to 77%, with minimal impact on the distribution across other domain categorizations. This large percentage underscores the central role of user-content platforms in the Web's contemporary landscape. Finally, it is essential to highlight that while large search engines were annotated as not SaaS and as not requiring a login, they may provide other services that would be classified differently.

We employed clustering techniques to categorize the top 116 domains into four groups. We aimed to minimize the average Hamming distances within each group across a 7-dimensional binary feature space constructed from our questionnaire responses. The domains within each cluster exhibit similar characteristics, allowing us to delineate the clusters using the following archetypes:

1. *SaaS:* This group comprises domains primarily characterized by SaaS offerings, devoid of offline activities and requiring user authentication. While some domains charge users for their core services, many do not. Examples include openai.com, zoom.us, office365.com, and canva.com. Our clustering process assigned 12 domains to this group.
2. *Open Content Providers:* Domains in this cluster serve as content producers, offering access to their content without imposing a paywall or requiring user authentication. Notably, this is the only group to include businesses with offline operations. Examples encompass samsung.com, walmart.com, bbc.co.uk, weather.com, and yahoo.com. Amazon.com was categorized within this group by our clustering algorithm, along with other 31 domains.
3. *Platforms:* This cluster encompasses domains whose activities are confined to the online realm, where users contribute the entirety of the content, and access to the websites is free. Prominent examples comprise youtube.com, roblox.com, chaturbate.com, tiktok.com, wikipedia.org, and booking.com. Interestingly, search engines and messaging web services such as telegram.org and discord.com were assigned to this group by our clustering method. In total, 67 domains were classified under this archetype.

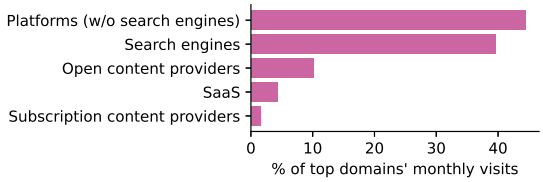

Figure 6: Fraction of the top 116 domains' total traffic directed to each of the 4 clusters. Search engines were manually removed from cluster group "Platforms".

4. *Subscription Content Providers:* Domains in this category are online content producers that mandate user authentication and levy charges for accessing their content. Examples include nytimes.com, espn.com, netflix.com, and disneyplus.com. Our clustering process identified five domains within this group.

Figure 6 shows that the cluster designated as "Platforms" is the primary destination for traffic among the top 116 domains, even when excluding search engines from consideration. While some traffic directed toward search engines could be labeled as SaaS due to their supplementary services, the preeminence of the Platforms cluster remains unassailable.

4 CONCLUSIONS

To the best of our knowledge, our estimation of web usage concentration is the only published one to date. Its implications for web studies are profound, as it underscores, through quantitative data, that the characteristics of the Web, particularly in terms of usage, can be reasonably inferred from a minimal fraction of all registered domains. While our dataset covers a mere 0.07% of all domains, it is likely to encapsulate around 80% of total web traffic.

Our in-depth analysis of the 116 most visited domains, collectively responsible for approximately one-third of all web traffic, reveals that nearly all of this traffic goes towards websites owned by for-profit digital tech companies. This observation highlights the Web primarily as a vehicle for commercial enterprise. Notably, this enterprise predominantly manifests as purely online ventures rather than online extensions of offline activities. Originally conceived as an open and uncharted domain, the Web has evolved into a realm ripe for commercial exploitation.

Furthermore, it is noteworthy that almost all of the 116 domains analyzed do not charge users for access or the core functionalities offered by them. This lack of paywalls highlights prevalent business models on the Web, which rely on alternative revenue streams.

Such revenue generation methods may include charging for supplementary features, as observed in domains like google.com, youtube.com, roblox.com, amazon.com, and github.com. Additionally, some companies secure venture capital funding to pursue a "growth at all costs" strategy, as exemplified by platforms such as whatsapp.com and quora.com (Kutcher et al., 2014; Ernst & Young LLP, 2014; Constine, 2013). Moreover, certain companies engage in two-sided markets, where the user base serves more as an asset than traditional customers, providing valuable data, customers, and audiences for the other side of the market. Prominent examples of this approach include facebook.com, booking.com, ebay.com, and youtube.com (Zuboff, 2019).

Most traffic to top domains goes towards websites that do not generate their content. Instead, this content originates from various sources, including users (e.g., instagram.com, xvideos.com, github.com, and messenger.com), sellers (ebay.com, booking.com, and aliexpress.com), or other external websites (such as search engines and news aggregators). We named websites that exhibit these predominant characteristics as "Platforms." Platforms compose most of the top domains (67 out of 116) and collectively capture over 80% of its traffic (refer to Fig. 6), underscoring their significance in scholarly discourse (Hein et al., 2020; Hermes et al., 2020b; Zuboff, 2019).

While our platform identification efforts focused on approximately one-third of the web traffic, our data suggests that platform hegemony could extend to more significant portions of web usage. Several prominent industries outlined in Fig. 3 primarily consist of such websites. For instance, "Search Engines," "Social Networks and Online Communities," and "Adult" industries are mostly comprised of platforms. Additionally, "TV Movies and Streaming," although primarily constituted by subscription content providers, sees its traffic largely dominated by youtube.com, a platform. While platform dominance has been previously acknowledged in the financial realm (Evans and Gawer, 2016) and may be readily observed through personal experience, we are unaware of former quantitative evidence of its prevalence over web traffic.

The concentration of web visits on platforms can be seen as a form of privatization of the Web, particularly concerning its usage dynamics. Despite the abundance of domain names and websites – publishing an independent website remains straightforward and accessible to many – visits are a scarce resource that platforms have enclosed. Within the web context, visits represent more than mere audience numbers, as visitors actively contribute to the content hosted on platforms. Web usage, encompassing activities such as viewing, posting, and interacting, predominantly occurs within these platforms, which dictate the rules governing these activities and determine the fate of the content and data generated through these visits.

Additionally, we observe that a staggering 80% of the traffic to the top domains flows to companies in the United States. This substantial concentration, which could potentially characterize the Web as predominantly an American enterprise, has been previously highlighted in terms of companies' market value (Evans and Gawer, 2016), but not in terms of traffic distribution. As discussed in Section 3.3, the accumulation of traffic on specific countries necessitates a socio-economic and political explanation. Hermes *et al.* took a step in this direction by interviewing European experts and top managers to glean insights into the reasons behind the dominance of American platforms in terms of market value (Hermes et al., 2020a). Their findings suggest that factors such as a results-oriented mindset, willingness to take risks, a sizable domestic market, state investment, early-mover advantage, the establishment of technology hubs (such as Silicon Valley), close collaborations with universities, and access to venture capital collectively contribute to American dominance in the web landscape.

The utilization of the Web exhibits, of course, significant diversity, varying among different individuals, social groups, and countries. However, our analysis indicates that, currently, on a global scale, the Web is predominantly characterized by the overwhelming presence of for-profit American platforms with business models not reliant on subscription fees. This observation provides quantitative support for common descriptions of the digital realm, such as "Platform capitalism" (Srnicek, 2016), "Surveillance capitalism" (Zuboff, 2019), and "Technofeudalism" (Varoufakis, 2024). While these descriptions often focus on corporate giants like Alphabet, Meta, Microsoft, and Amazon, potentially conveying that their findings are specific to these companies, our analysis suggests otherwise. These companies not only command a significant portion of web usage but also exhibit characteristics mirrored by numerous smaller entities, indicating that their descriptions may reflect broader trends across the web ecosystem.

ACKNOWLEDGEMENTS

The artificial intelligence systems ChatGPT and Grammarly were used to revise the text for orthography, grammar, and style.

REFERENCES

Adamic, L. A. and Huberman, B. A. (2000a). The nature of markets in the world wide web. https://ssrn.com/abstract=166108.

Adamic, L. A. and Huberman, B. A. (2000b). Power-law distribution of the world wide web. *Science*, 287(5461):2115–2115.

Agarwal, V. and Sastry, N. (2022). "way back then": A data-driven view of 25+ years of web evolution. In *Proceedings of the ACM Web Conference 2022*, pages 3471–3479.

Cannon, R. (2008). More people are using google to navigate, not just search [data]. https://www.searchenginejournal.com/is-google-trumping-the-url/6820. Acessed on 2024-02-29.

CERN (1993). Software release of www into public domain. https://cds.cern.ch/record/1164399.

Chakrabarti, S., Joshi, M. M., Punera, K., and Pennock, D. M. (2002). The structure of broad topics on the web. In *Proceedings of the 11th International Conference on World Wide Web*, pages 251–262.

Constine, J. (2013). Quora signals it's favoring search ads for eventual monetization, launches author stats tool. http://techcrunch.com/2013/11/12/quora-confirms-its-favoring-search-ads-for-eventual-monetization-launches-author-stats-tool. Acessed on 2024-04-10.

Crichton, K., Christin, N., and Cranor, L. F. (2021). How do home computer users browse the web? *ACM Trans. Web*, 16(1).

Diachuk, O., Loba, P., and Mirgorodskaya, O. (2021). Comparing accuracy: Semrush vs similarweb. http://www.owox.com/blog/articles/semrush-vs-similarweb. Acessed on 2024-04-18.

Ernst & Young LLP (2014). Whatsapp inc. financial statements. http://www.sec.gov/Archives/edgar/data/1326801/000132680114000047/exhibit991auditedwhatsappi.htm. Acessed on 2024-04-10.

Evans, P. and Gawer, A. (2016). The rise of the platform enterprise: A global survey. http://www.thecge.net/app/uploads/2016/01/PDF-WEB-Platform-Survey_01_12.pdf. Acessed on 2024-04-18.

Georgieva, G., Arnab, S., Romero, M., and de Freitas, S. (2015). Transposing freemium business model from casual games to serious games. *Entertainment Computing*, 9-10:29–41.

Hardwick, J. (2018). Find out how much traffic a website gets: 3 ways compared. http://ahrefs.com/blog/website-traffic. Acessed on 2024-04-18.

Hein, A., Schreieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., and Krcmar, H. (2020). Digital platform ecosystems. *Electron Markets*, 30:87–98.

Hermes, S., Clemons, E., Schreieck, M., Pfab, S., Mitre, M., Böhm, M., Wiesche, M., and Krcmar, H. (2020a). Breeding grounds of digital platforms: Exploring the sources of american platform domination, china's platform self-sufficiency, and europe's platform gap. In *Proceedings of the 28th European Conference on Information Systems (ECIS)*.

Hermes, S., Hein, A., Böhm, M., Pfab, S., Weking, J., and Krcmar, H. (2020b). Digital platforms and market dominance: Insights from a systematic literature review and avenues for future research. In *Proceedings of the 24th Pacific Asia Conference on Information Systems: Information Systems (IS) for the Future*.

Huberman, B. A. (2001). *The Laws of the Web: Patterns in the Ecology of Information*. The MIT Press.

Huizingh, E. K. (2000). The content and design of web sites: an empirical study. *Information & Management*, 37(3):123–134.

Jansen, B. J., Jung, S.-g., and Salminen, J. (2022). Measuring user interactions with websites: A comparison of two industry standard analytics approaches using data of 86 websites. *PLOS ONE*, 17(5):1–27.

Kutcher, E., Nottebohm, O., and Sprague, K. (2014). Grow fast or die slow. http://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/grow-fast-or-die-slow. Acessed on 2024-04-10.

Lee, E., Lee, J., and Lee, J. (2006). Reconsideration of the winner-take-all hypothesis: Complex networks and local bias. *Management Science*, 52(12):1838–1848.

Liu, J., Li, X., and Guo, Y. (2017). Periodicity analysis and a model structure for consumer behavior on hotel online search interest in the us. *International Journal of Contemporary Hospitality Management*, 29.

Liu, Z., Yan, Y., and Hauskrecht, M. (2018). A flexible forecasting framework for hierarchical time series with seasonal patterns: A case study of web traffic. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 889–892.

Petrosyan, A. (2024). Number of internet and social media users worldwide as of january 2024. http://www.statista.com/statistics/617136/digital-population-worldwide. Acessed on 2024-02-27.

Prantl, D. and Prantl, M. (2018). Website traffic measurement and rankings: competitive intelligence tools examination. *International Journal of Web Information Systems*, 14(4):423–437.

Srnicek, N. (2016). *Platform Capitalism*. John Wiley & Sons, 1 edition.

Varoufakis, Y. (2024). *Technofeudalism: What Killed Capitalism*. Melville House, 1 edition.

Verisign (2023). Verisign domain name industry brief: 354 million domain name registrations in the first quarter of 2023. http://blog.verisign.com/domain-names/verisign-q1-2023-the-domain-name-industry-brief. Acessed on 2024-02-21.

Webster, J. G. and Lin, S.-F. (2002). The internet audience: Web use as mass behavior. *Journal of Broadcasting & Electronic Media*, 46(1):1–12.

Zuboff, S. (2019). *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. Public Affairs, 1 edition.