

Evaluating the Long-term Effects of Parameters on the Characteristics of the Tranco Top Sites Ranking

Victor Le Pochat

imec-DistriNet, KU Leuven

Tom Van Goethem

imec-DistriNet, KU Leuven

Wouter Joosen

imec-DistriNet, KU Leuven

Abstract

Although researchers often use top websites rankings for web measurements, recent studies have shown that due to the inherent properties and susceptibility to manipulation of these rankings, they potentially have a large and unknown influence on research results and conclusions. As a response, we provide Tranco [8], a research-oriented approach for aggregating these rankings transparently and reproducibly.

We analyze the long-term properties of the Tranco ranking and determine whether it contains a balanced set of domains. We compute how well Tranco captures websites that are responsive, regularly visited and benign. Through one year of rankings, we also examine how the default parameters of Tranco create a stable, robust and comprehensive ranking.

Through our evaluation, we provide an understanding of the characteristics of Tranco that are important for research and of the impact of parameters on the ranking composition. This informs researchers who want to use Tranco in a sound and reproducible manner.

1 Introduction

When measuring the prevalence of (security) practices and issues or evaluating novel tools and approaches across the web, researchers often rely on rankings of the most popular websites to obtain a representative sample of domains for their study, most often the Alexa top 1 million ranking [2]. Even though these rankings are widely used in research, the methods for composing these rankings are opaque, not well-known and rarely questioned by the research community. In fact, several recent unannounced changes in ranking availability [3] and methods [8] highlight how dependent web-related research is on these rankings, and how these may unknowingly have a large impact on research results and conclusions.

Only recently have researchers started to examine the rankings and their potential influences on Internet measurement and web security research, finding flaws in their inherent properties and susceptibility to malicious manipulation, in particular the most widely used Alexa ranking [8, 11, 12]. However,

access to reliable and representative lists of domains remains important, as it enables researchers to study the web in a sound manner.

In prior work, we therefore proposed Tranco [8], a ranking that is oriented towards research by providing improved characteristics, a transparent method and reproducible rankings. This ranking aggregates four existing rankings (Alexa, Majestic, Quantcast and Umbrella) over a customizable period of time, and allows to apply filters that tailor the list to a researcher's needs. Both a daily updated ranking that uses default parameters and a service to generate custom rankings are available to researchers at <https://tranco-list.eu/>.

Top sites rankings should contain a sufficient number of websites that are available for crawling without errors, and that can actually be considered as popular. Ideally, such rankings should also capture changes in popularity over time while still being sufficiently stable as to enable longitudinal studies. While previous work assessed these desirable properties for the four existing rankings, such an extensive analysis of the Tranco ranking has not yet been done.

We now evaluate Tranco similarly to the other rankings, in order to determine whether it is a valid research-oriented alternative to the currently used rankings. Moreover, using the openly available source code, we analyze the parameters available in Tranco, such as the aggregation period or scoring method, and assess their impact on the ranking's composition. Finally, we assess whether the default parameters selected for the daily updated Tranco ranking provide a suitable set of domains that can be used in a broad set of studies.

Our contributions are the following: 1) we generate the Tranco ranking over one year to evaluate its long-term properties, 2) we quantify the unresponsive and malicious sites to inform researchers' assumptions of domain characteristics [8], 3) we compare Tranco with existing popularity rankings, finding a larger overlap with more stable lists and a good overlap with observed web traffic, and 4) we find that the default parameters provide a stable and consistent ranking.

2 Methods of the Tranco ranking

The Tranco ranking aggregates four existing top sites rankings, that all use different vantage points and data collection periods to compute the scores for their ranking [8]:

- Alexa¹ ranks 1 million mainly ‘pay-level domains’² based on web traffic collected either from users of its ‘Alexa Traffic Rank’ browser extension or from website visitors through an analytics script. Ranks are based on 1 day of data.
- Majestic³ ranks 1 million mainly ‘pay-level domains’ based on incoming links collected through a web crawl. Ranks are based on 120 days of data.
- Quantcast⁴ ranks around 500,000 mainly ‘pay-level domains’ based on web traffic collected either from website visitors through an analytics script or from ISP data. Ranks are based on 30 days of data.
- Umbrella⁵ ranks 1 million ‘pay-level domains’ and sub-domains based on DNS traffic collected through its OpenDNS resolvers. Ranks are based on 2 days of data.

The aggregate score of each domain in the Tranco ranking is calculated as the sum of scores across all rankings within the aggregation period. These individual scores are derived from the rank value in a component ranking: either through the Borda method, where the score is the total number of items minus the rank, or the Dowdall method, where it is the inverse of the rank [5].

Using this scoring method, any subset of the four providers can be aggregated over any aggregation period. However, in order to offer an readily available, easy to use and consistent ranking to researchers, a standard ranking with default settings is generated daily and published online⁶. In this default ranking, all four rankings are aggregated over a period of 30 days, with domains being scored with the Dowdall method.

In order to support studies that require domains with specific properties, certain filters can be applied to obtain a set of appropriate domains. Domains can be filtered on their components: only keeping ‘pay-level domains’, certain TLDs, certain subdomains (e.g. only `login`) and/or one domain with a certain second-level label (e.g. only one `google.*` domain). Moreover, a researcher can require domains to appear in the rankings of a certain number of providers or for a certain number of days. Finally, domains can be checked against

¹<https://www.alexa.com/topsites>

²A pay-level domain is a domain name that a consumer or business can directly register, and consists of a subdomain of a public suffix or effective top-level domain [9] (e.g. `.com` but also `.co.uk`).

³<https://majestic.com/reports/majestic-million>

⁴<https://www.quantcast.com/top-sites/>

⁵<https://umbrella-static.s3-us-west-1.amazonaws.com/index.html>

⁶<https://tranco-list.eu/>

other lists, such as the Chrome User Experience Report [7] or Google Safe Browsing [6], to retain only regularly visited websites (Section 3.2) or domains that have not been flagged as malicious (Section 3.5) respectively. In the default Tranco ranking, only pay-level domains are retained.

All domains in the union of the aggregated rankings receive a score. As the component rankings contain some domains unique to them, this union is usually larger than one million domains. The default ranking is truncated to one million domains, in line with existing rankings. The larger union also means that after filtering, there usually still remain at least one million domains that satisfy the applied filters.

3 Analysis of Tranco’s properties

To evaluate the properties of the Tranco ranking, we generate rankings following the method described in our prior work [8], using the publicly available source code⁷ and an archive of Tranco’s component rankings. Even though (custom) rankings can be generated online, we generate them separately to reduce the burden on this service and to reduce processing time by parallelizing and optimizing the generation (e.g. generating 14-day lists from pairs of 7-day lists).

Our analyses are based on those that we previously conducted on the four rankings that constitute Tranco [8]. When we describe a ranking for a date D with an aggregation period of N days, this ranking aggregates the component rankings from $N - 1$ days before D until and including D . Unless otherwise mentioned, we use the default parameters of the Tranco ranking: all four component rankings, an aggregation period of 30 days, the Dowdall method for scoring domains, only retaining pay-level domains and truncating the ranking at one million domains. We calculate aggregate scores over 1 day starting from April 1, 2018; as we construct rankings aggregated over longer periods from those calculated for shorter periods, we generate the default 30-day rankings starting from April 30, 2018 until April 30, 2019.

3.1 Similarity with component rankings

The four component rankings of Tranco have different properties (e.g. stability over time) due to the differences in data sourcing and processing (e.g. the data collection period). Even though Tranco considers its component rankings equally when calculating global domain scores, these differences can cause the rankings to have a different similarity with and influence on Tranco. Moreover, the resulting differences in ranking composition may also affect other properties, such as the responsiveness of domains in Tranco. In this section, we analyze the similarity of the component rankings with Tranco; throughout the rest of this paper, we analyze the contribution of each component ranking to the other properties of Tranco.

⁷<https://github.com/Distrinet/tranco-list>

RBO p	Subset with weight		Rank-biased overlap of Tranco and			
	85.2%	99.9996%	Alexa	Majestic	Quantcast	Umbrella
0.9	10	100	67.18%	70.32%	54.41%	25.92%
0.99	100	1K	60.65%	54.73%	40.61%	23.21%
0.999	1K	10K	56.40%	54.01%	36.17%	17.36%
0.9999	10K	100K	56.02%	56.66%	40.42%	14.46%
0.99999	100K	1M	54.95%	55.76%	43.55%	14.42%

Table 1: Average rank-biased overlap (RBO) in April 2019 between Tranco and its component rankings for five values of the parameter p , i.e. weightings of subsets.

Table 1 shows the rank-biased overlap [13] (a similarity measure where better ranks receive a higher weight, configured through a parameter p) between Tranco and its component rankings, averaged over April 2019. Alexa and Majestic have a similar overlap with Tranco, at 55.0% and 55.8% respectively when the top 100,000 is weighted at 85.2%. Quantcast is third at 43.6%: its shorter list means that it can contribute fewer domains to Tranco. Finally, Umbrella has a low overlap at 14.4%: this can be attributed to the subdomains in Umbrella that are not retained in the default Tranco ranking. In general, overlap improves when a smaller subset of the rankings is more heavily weighted (lower p), showing that rankings tend to agree more on the head of the ranking but less on the long tail.

Figure 1 shows in detail which subsets of the four component rankings contribute to the Tranco ranking on April 30, 2019⁸. The Majestic ranking has the highest intersection of 627,341 domains, with its top 500,000 being almost evenly represented throughout the full Tranco ranking. Quantcast shares 338,588 domains (70.32% of its domains), with a high influence on the very top of the Tranco ranking. Alexa shares 421,916 domains, which are mostly well ranked; even though it shares about one third of domains less than Majestic, its contribution to the better ranked domains in Tranco translates into a similar rank-biased overlap. Finally, Umbrella shares only 165,244 domains (all pay-level) across all its one million domains, almost exclusively with the Tranco top 200,000.

We see that more stable rankings, i.e. Majestic and Quantcast, have a higher overlap with Tranco. Given that the four component rankings do not tend to agree on which domains are the most popular [8], we estimate that this higher overlap is not due to one ranking containing more domains that are also included in any of the other rankings. Instead, this is due to domains in more stable lists being ranked repeatedly (i.e. over a longer period) and therefore receiving a higher aggregate score. A more evenly distributed contribution of each ranking can therefore be achieved by reducing the aggregation period. Overall, while some rankings may have a higher contribution due to their stability, each component ranking contributes to some extent to the Tranco ranking.

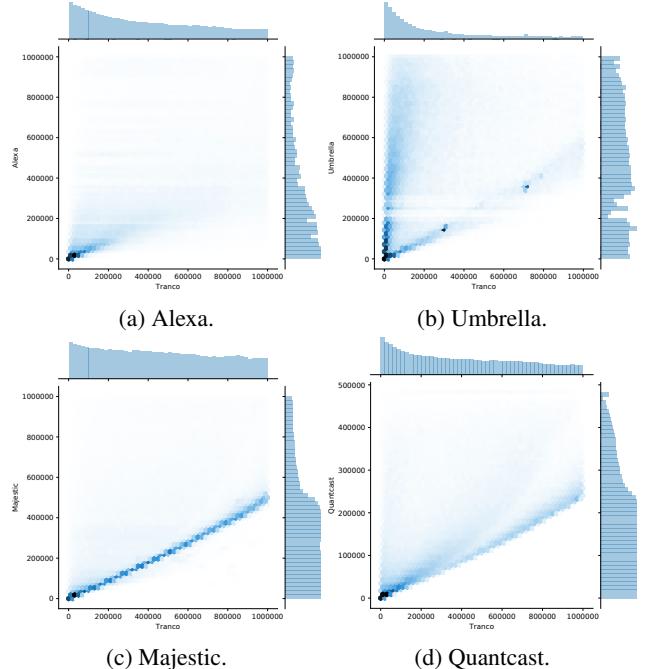


Figure 1: Contribution of component rankings to the Tranco ranking of April 30, 2019. The right-hand histogram shows which domains from the component ranking appear in Tranco, while the top histogram shows at which ranks they appear. The heat map shows whether these contributions are evenly distributed or focused on specific parts of the rankings.

3.2 Comparison with web traffic

The Chrome User Experience Report is a data set released by Google that contains website performance metrics for over 4.3 million distinct domains (May 2019) [7]. These domains are said to be ‘popular destinations on the web’, having been observed sufficiently regularly in Chrome user traffic. While the report is not designed to be a ranking, we can still use it to assess whether Tranco contains domains that are regularly visited in the currently most popular browser [10]. This would mean that Tranco contains a representative sample of actual websites, and can therefore reliably be used to comprehensively study ecosystems on the web.

Table 2 shows that in April 2019, on average 62.24% of domains in the Tranco ranking were included in the Chrome User Experience Report of April 2019. This places Tranco after Quantcast and Alexa but before Majestic and Umbrella. Table 2 shows that overlap improves for smaller subsets of the ranking, with over 90% of the top 10,000 domains appearing in the Chrome User Experience Report.

We find that almost two thirds of domains in the Tranco ranking are sufficiently visited by Chrome users to be included in the Chrome User Experience Report. This indicates that the majority of Tranco’s domains reflect real web traffic. Moreover, the domains in the Tranco ranking can be filtered

⁸ Available at <https://tranco-list.eu/list/7PJX/1000000>.

Subset	Alexa	Majestic	Quantcast	Umbrella	Tranco
10	100%	90%	100%	33%	100%
100	94.7%	76.4%	100.0%	21.1%	91.6%
1K	93.33%	85.51%	98.54%	14.36%	90.73%
10K	92.54%	85.60%	96.19%	11.04%	90.60%
100K	87.13%	57.05%	94.71%	11.63%	82.47%
1M	68.12%	40.45%	76.90%	13.96%	62.24%

Table 2: Average overlap in April 2019 between the Chrome User Experience Report and top sites rankings.

upfront on whether they appear in the Chrome User Experience Report, guaranteeing that the resulting list consists only of actually visited websites.

3.3 Stability over time

The stability of a ranking determines how much the retrieval date impacts the obtained set of domains, which is particularly important for longitudinal studies. Stable rankings yield a very similar set of domains, but may however not capture sudden increases or decreases in popularity.

Figure 2 shows that the default Tranco ranking has a much higher stability than the Alexa and Umbrella ranking, at around 0.6%, on par with Majestic and Quantcast. Figure 3 shows that this stability extends to smaller subsets of the ranking, with usually less than 1% daily change. Changes in the top 10 or 100 do occur, but are normally limited to one domain changing between consecutive days.

Figure 4 shows that aggregating over 1 day already removes a large part of the volatility introduced by the Alexa ranking. Moreover, for longer periods than the default 30 days, the improvement in stability is relatively small. Finally, the ‘weekend effect’, where the set of domains in rankings based on weekend traffic differs significantly from those in weekday rankings [11], is largely subdued when aggregating over longer periods, including the default 30 days.

When a ranking is calculated for an aggregation period longer than 1 day, the set of component rankings will only differ in the first and last day of the period, meaning that only these two days affect the stability between the rankings of two consecutive days. To assess whether Tranco maintains stability over a longer period of time, we analyze the difference between rankings that are spaced apart with the length of their aggregation period. Figure 5 shows that rankings aggregated over 30 days or less produce a relatively stable set of domains, changing around 10% over time when no anomalies are present. Rankings aggregated over longer periods are less stable; moreover, anomalies (such as those discussed in Section 3.6) have a longer-lasting negative effect on stability.

Finally, Figure 6 shows the global change of the set of domains in Tranco over one year. For the full top million, 33.13% of domains are new when comparing the rankings of April 30, 2018 and April 29, 2019. For smaller subsets of the

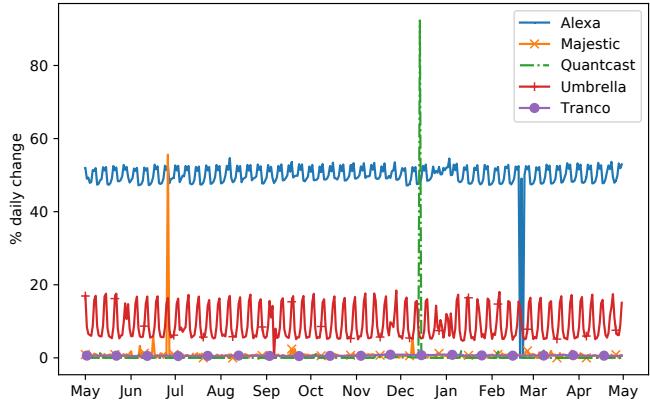


Figure 2: Stability over time from May 2018 until April 2019 of Tranco and its four component rankings, measured as the difference between rankings of two consecutive days.

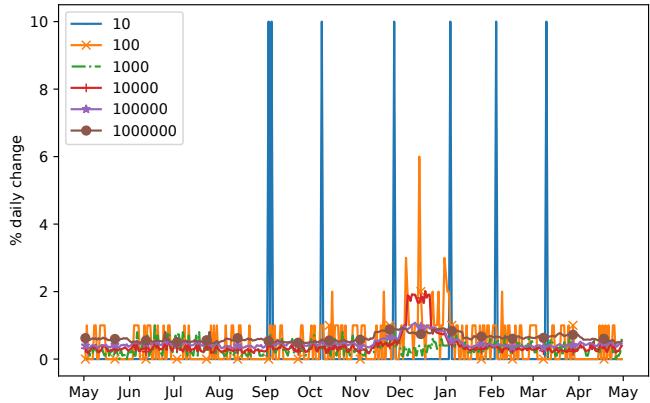


Figure 3: Stability over time from May 2018 until April 2019 of Tranco for different subsets, measured as the difference between rankings of two consecutive days.

ranking, this global change is even lower: the top 10,000 sees only a 16.66% change. The observed level of global change is comparable to that of Majestic and lower than that of Alexa and Umbrella [12], and shows that Tranco provides stability even in the long term while still capturing genuine changes in popularity over time.

Existing rankings suffer from a high volatility, sometimes unknown to researchers [8, 12]. By aggregating over longer periods, Tranco provides a set of domains that does not change significantly over time, reducing the influence of the exact date on which the ranking is downloaded and therefore improving the soundness and reproducibility of (longitudinal) studies.

3.4 Responsiveness

If the websites of the domains in a ranking are unavailable, the size of the studied sample shrinks, which makes a study less comprehensive. Moreover, websites that produce an error when visited may not be representative of ‘normal’ websites

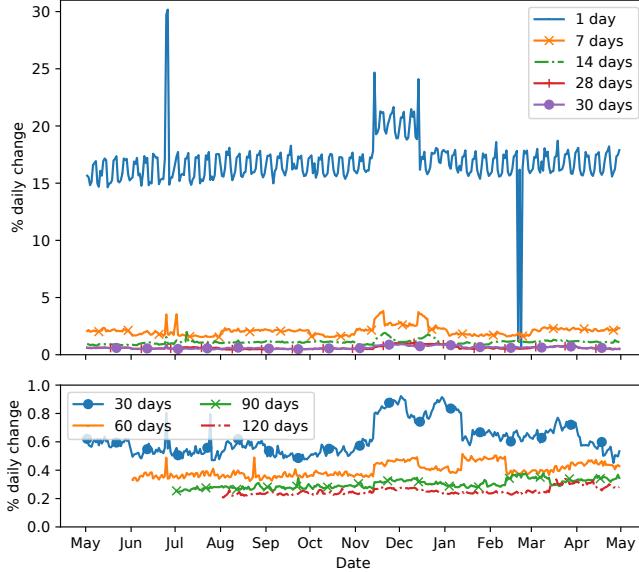


Figure 4: Stability over time from May 2018 until April 2019 of Tranco for eight aggregation periods, measured as the difference between rankings of two consecutive days.

and could therefore skew measurements.

Table 3 shows the distribution of HTTP status codes for a crawl of the root web pages for the Tranco ranking of May 14, 2019⁹ conducted from May 14 to 16, 2019. For the full ranking, we find that 85.17% of domains respond with HTTP status code 200, indicating an available website. 4.14% report another status code, indicating that the server is responsive but that no content is provided on the root page of the domain. Finally, 10.68% of domains could not be crawled. Smaller subsets of the ranking see an increased success rate.

In order to understand the source of crawl failures, we manually analyze the 109 domains in the top 1,000 that do not have status code 200. 32 domains are used for analytics or advertising, serving tracking scripts and advertisements or forming part of a redirection chain. 27 domains are used in a content delivery network, usually serving content that is embedded on other (popular) sites. 8 domains are part of some other type of infrastructure, e.g. serving Windows updates (`windowsupdate.com`) or as a portal for mobile users (`metropcs.mobi`). 22 domains blocked our crawler with e.g. a 403 Forbidden status code. Finally, 20 domains failed to be crawled for some other reason, e.g. only serving content on a subdomain. This analysis shows that Tranco also covers domains that experience high traffic volumes but through other means than website visits. This complements e.g. Alexa's focus on website traffic and therefore makes Tranco appropriate for a broad set of studies (e.g. into the security of popular non-website domains).

By filtering out domains that do not appear in the Chrome

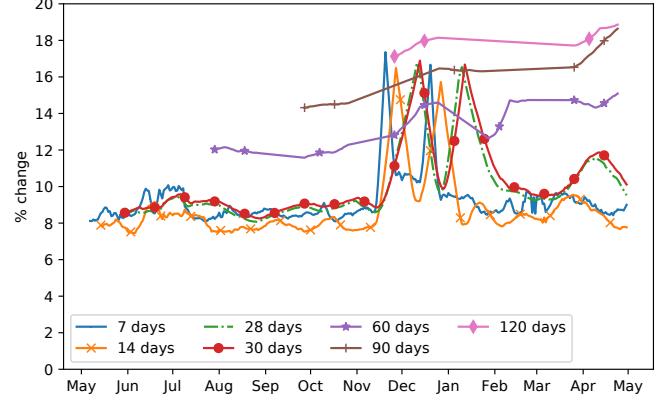


Figure 5: Stability over time from May 2018 until April 2019 of Tranco for eight aggregation periods of N days, measured as the difference between the rankings of dates $D - N$ and D .

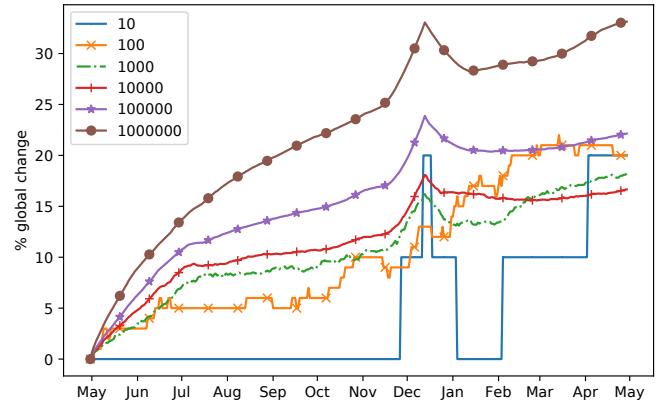


Figure 6: Difference between the Tranco ranking of April 30, 2018 and rankings from May 2018 until April 2019 for different subsets.

User Experience Report (Section 3.2), i.e. domains that are not regularly seen in a major browser, the share of responsive domains increases: 93.20% of domains from Tranco seen in the Chrome User Experience Report respond with status code 200, 1.89% respond with another status code, and 4.91% could not be crawled.

For the full Tranco ranking, Majestic shares the most sites that do not respond with status code 200, at 96,839 domains; Alexa shares 37,343 domains, Umbrella 34,486 and Quantcast 24,823. However, the full Majestic ranking does not have the most unresponsive domains of all four rankings [8]. The higher share of unresponsive domains contributed by the Majestic ranking may therefore rather be explained by the higher contribution of Majestic to Tranco overall.

Overall, the large majority of Tranco domains hosts a responsive website. This means that crawling those domains provides a sufficiently large sample, allowing for a broader and more representative overview of the web.

⁹ Available at <https://tranco-list.eu/list/666X/1000000>.

Subset	Success (200)	Client error (4xx)	Server error (5xx)	Other status code	Failure
10	100.0%	0%	0%	0%	0%
100	92%	2%	0%	0%	6%
1K	89.1%	3.4%	0.2%	0.2%	7.1%
10K	86.24%	3.97%	0.57%	0.08%	9.14%
100K	84.99%	3.32%	0.80%	0.09%	10.79%
1M	85.17%	2.77%	1.30%	0.07%	10.68%

Table 3: Responsiveness of (subsets of) the sites in the Tranco ranking of May 14, 2019.

	10K	100K	1M
Malware	1	24	187
Social engineering	1	21	1,486
Unwanted software	2	34	189
Potentially harmful application	0	0	8
Total (unique domains)			1,851

Table 4: Number of sites in the Tranco ranking of May 14, 2019 that were flagged by Google Safe Browsing.

3.5 Benignness

Even though researchers and security companies often assume that very popular sites are by definition benign, and therefore whitelist them, malicious domains are present in top sites rankings [8]. Conversely, the rank of popular malicious domains could indicate how many victims they affect.

Table 4 shows that 1,851 unique domains on the Tranco ranking of May 14, 2019⁷ were flagged by Google Safe Browsing [6]. 80.3% of these perform ‘social engineering’ (e.g. phishing). Within the top 10,000, four sites were flagged. Note that this represents a lower bound of malicious domains: unflagged domains can be benign but their malicious character could also not yet have been discovered.

Most malicious sites can be attributed to the Majestic ranking, at 1,168 domains; 703 malicious domains appear in the Alexa ranking, 594 in Quantcast and 210 in Umbrella. Of the four top 10,000 sites that are flagged, all appear in the Majestic ranking, 3 in Alexa and Umbrella, and 1 in Quantcast. Although the higher share of malicious domains from the Majestic ranking may again reflect the higher contribution of Majestic to Tranco overall, Majestic does tend to include more malicious domains in and of itself [8].

Although popular websites are often assumed to be benign, we find that the four rankings that are aggregated into Tranco, and therefore also the Tranco ranking itself, still contain some malicious domains. Given that Google Safe Browsing can be reliably queried for one million domains, the Tranco ranking can be prefiltered to exclude the malicious domains.

3.6 Anomalies

Throughout the one year of rankings that we evaluate, we observe three major anomalies in the four component rankings. These have (temporary) effects on the composition of the Tranco ranking: they explain the peaks and sudden jumps observed in our results throughout this paper.

On June 25, 2018, the Majestic ranking was truncated, containing only 445,000 instead of the usual million domains. Between November 14 and December 13, 2018, Quantcast’s ranking contained only around 38,000 instead of the usual 500,000 domains, discarding all domains for which traffic was estimated [8]. Finally, on February 20 and 22, 2019, the Alexa ranking was a duplicate of the previous day’s list.

These anomalies mainly cause a higher or lower than average daily change in composition (as can be seen in Figure 4), due to changing contributions of the four component rankings. However, this effect is more outspoken for shorter aggregation periods, so larger periods including the default 30 days can smooth out these anomalies.

Moreover, we see that anomalies are not limited to one particular ranking, showing that real-world data collection and processing is susceptible to error. Tranco reduces the impact of these anomalies by aggregating data from multiple providers, such that sudden changes in one ranking’s composition do not immediately result in a widely varying set of domains.

3.7 Combination method

The Tranco ranking supports two methods of calculating domain scores: Borda (total number of items minus rank) and Dowdall (inverse of rank) [5]. The latter reflects previous observations of a Zipf-like distribution in domain popularity [1, 4]. Figure 7 shows that these two methods produce moderately similar rankings, with a rank-biased overlap [13] of between 50% and 80% depending on its parameter p , which indicates how heavily a smaller subset of the ranking is weighted. In terms of stability, Figure 8 shows that the Borda method produces a slightly more stable ranking, but overall stability is comparably high.

However, Figure 9 shows that for small subsets (e.g. the top 1,000) the Dowdall method is more robust against anomalies in the data. This is due to the rescaling of ranks, which gave the anomalous Quantcast rankings in November and December 2018 a disproportionately high influence on the Tranco ranking. While the combination method produces similar sets of domains, the default Dowdall method results in a more stable list if anomalies are present.

3.8 Structure

The default Tranco ranking only includes only ‘pay-level domains’, referring to domain names that a customer can directly

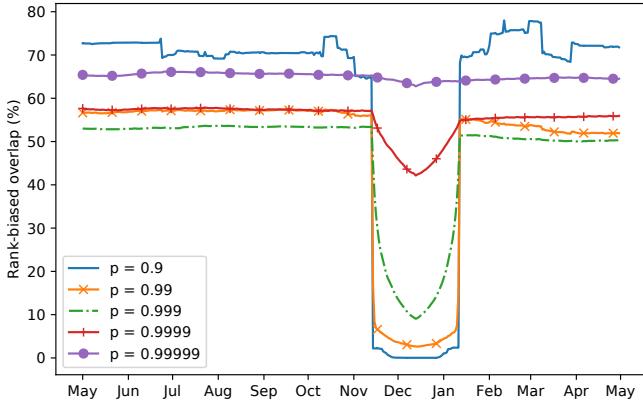


Figure 7: Similarity over time from May 2018 until April 2019 between rankings using the two scoring methods available in Tranco, measured as the rank-biased overlap between rankings of two consecutive days.

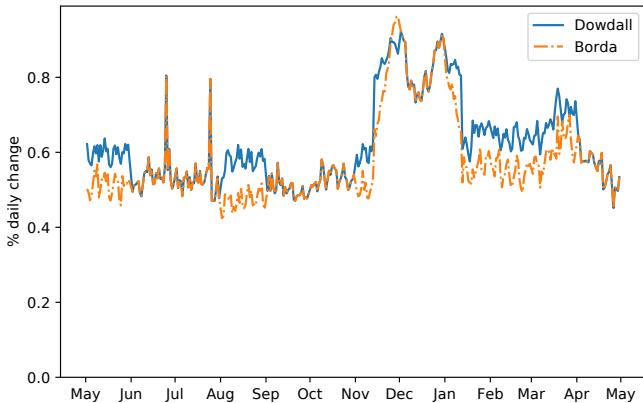


Figure 8: Stability over time from May 2018 until April 2019 between the two combination methods available in Tranco for the full ranking, measured as the difference between rankings of two consecutive days.

register, as certain top-level domains do not allow direct registrations under the TLD (e.g. `.uk`, requiring to register under e.g. `.co.uk`). Due to the way in which pay-level domains are determined, by checking domains against the Public Suffix List [9] to extract the TLD and then taking the next label as the pay-level domain, no subdomains nor invalid domains remain in the default Tranco ranking. This means that the final set of domains captures a higher variety of valid hosts, content and ownership, allowing for more comprehensive studies.

4 Related work

Scheitle et al. [12] study the stability and similarity of the Alexa, Majestic and Umbrella lists, measure the potential impact on (Internet measurement) research and list guidelines for using the rankings in a sound and reproducible manner.

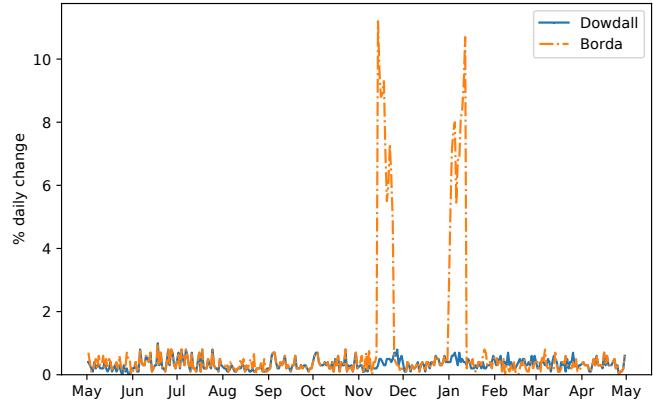


Figure 9: Stability over time from May 2018 until April 2019 between the two combination methods available in Tranco for the top 1,000, measured as the difference between rankings of two consecutive days.

Rweyemamu et al. [11] conduct a more detailed analysis of three effects of the rankings' methods on their composition, using these findings to extend the ranking usage guidelines.

In prior work [8], we analyze the Alexa, Majestic, Umbrella and Quantcast rankings within the context of security research: their analysis includes the representativeness, responsiveness and benignness of rankings. Moreover, we demonstrate that the rankings are susceptible to (malicious) large-scale manipulation. Finally, we provide Tranco, the ranking that we analyze in this paper, as a research-oriented, reproducible alternative to existing top websites rankings.

5 Conclusion

Many studies in web security and Internet measurement research depend on rankings of popular domains. We presented the Tranco ranking as a more research-oriented alternative to existing rankings, emphasizing a transparent method and providing a publicly available archive of reproducible rankings. However, the influence of its parameters on the composition of the ranking had not yet been analyzed. We evaluate Tranco over one year, and find that it has the following research-related properties:

Agreement with existing rankings The more stable Majestic and Quantcast rankings share the most domains with Tranco. Alexa shares fewer domains due to its volatility, but these domains are highly ranked. Umbrella contributes the least to Tranco, as only ‘pay-level domains’ are listed in Tranco by default.

62% of domains in Tranco were observed to be regularly visited in the Chrome browser, indicating that these are genuinely popular websites.

Stability The default aggregation period yields a very stable

ranking; with a smaller period, some volatility is reintroduced, while a larger period only causes minor gains in stability.

Responsiveness When crawling their root page, Tranco contains around 15% unresponsive domains. However, many of these domains do play an important infrastructural role, showing how Tranco captures a balanced set of popular domains beyond regular website traffic.

Benignness Tranco contains around 0.2% domains flagged as malicious.

Anomalies The default Dowdall scoring method is more robust against observed anomalies in Tranco's component rankings.

Structure Tranco contains only valid pay-level domains.

Subsets In general, overlap with existing domain lists, stability and responsiveness improve slightly for smaller subsets of Tranco.

Moreover, these properties can be further improved by creating a customized ranking where appropriate filters are applied.

Our analysis informs researchers who need to use a top websites ranking on those characteristics of Tranco that might be important to their study, and serves as a guide for those who want to customize the ranking to their requirements. In particular, we find that the default parameters selected for the daily updated Tranco ranking, available at <https://tranco-list.eu/>, are overall appropriate and recommended for research such as large-scale security measurements that needs a representative set of domains.

Acknowledgments

This research is partially funded by the Research Fund KU Leuven. Victor Le Pochat holds a PhD Fellowship of the Research Foundation - Flanders (FWO).

References

- [1] Lada A. Adamic and Bernardo A. Huberman. Zipf's law and the Internet. *Glottometrics*, 3:143–150, 2002.
- [2] Alexa Internet, Inc. Does Alexa have a list of its top-ranked websites? <https://web.archive.org/web/20160404003433/https://support.alexa.com/hc/en-us/articles/200449834-Does-Alexa-have-a-list-of-its-top-ranked-websites->, January 2016. Archived on April 4, 2016.
- [3] Alexa Support (@AlexaSupport). Yes, the top 1m sites file has been retired. https://twitter.com/Alexa_Support/status/800755671784308736, November 2016.
- [4] Aaron Clauset, Cosma Rohilla Shalizi, and Mark E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.
- [5] Jon Fraenkel and Bernard Grofman. The Borda count and its real-world alternatives: Comparing scoring rules in Nauru and Slovenia. *Australian Journal of Political Science*, 49(2):186–205, 2014.
- [6] Google Inc. Safe browsing. <https://safebrowsing.google.com/>.
- [7] Google, Inc. Chrome User Experience Report. <https://developers.google.com/web/tools/chrome-user-experience-report/>, January 2018.
- [8] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. Tranco: A research-oriented top sites ranking hardened against manipulation. In *26th Annual Network and Distributed System Security Symposium*, 2019.
- [9] Mozilla Foundation. Public suffix list. <https://publicsuffix.org/>, 2019.
- [10] NetMarketShare. Market share statistics for Internet technologies. <https://netmarketshare.com/>, April 2019.
- [11] Walter Rweyemamu, Tobias Lauinger, Christo Wilson, William Robertson, and Engin Kirda. Clustering and the weekend effect: Recommendations for the use of top domain lists in security research. In *20th International Conference on Passive and Active Measurement*, pages 161–177, 2019.
- [12] Quirin Scheitle, Oliver Hohlfeld, Julien Gamba, Jonas Jelten, Torsten Zimmermann, Stephen D. Strowes, and Narseo Vallina-Rodriguez. A long way to the top: Significance, structure, and stability of Internet top lists. In *Internet Measurement Conference*, pages 478–493, 2018.
- [13] William Webber, Alistair Moffat, and Justin Zobel. A similarity measure for indefinite rankings. *ACM Transactions on Information Systems*, 28(4):1–38, November 2010.