Structural Changes in the Web from 2018–2024

Chad Glazier (Student No. 96999370), Egor Bezriadin (78634029), Joy Umejiego (81265373), and Nicholas Chamberlain (13723192)

University of British Columbia, Okanagan Campus, Kelowna, BC, Canada

Abstract. This study provides a comprehensive view of the web's evolving structure and influence from 2018 to 2024, drawing on large-scale hyperlink datasets from Common Crawl. By focusing on the top 10,000 domains in each period, we examine shifts in network centrality, category prevalence, community structure, and top-level domain (TLD) influence. Our analysis reveals that while some domains displayed early indicators of rising prominence, the majority gained unexpected influence over time. Category patterns, once centered on news and knowledge platforms, have given way to more interactive domains such as social media and web services by 2024. Examination of in-degree distributions and PageRank scores across multiple intermediate snapshots from 2018 to 2024 suggests that the dominance of a few major websites may have peaked mid-decade, with influence now distributed more evenly among a broader range of domains. At the TLD level, entrenched leaders like ".com" experienced declines in connectivity, while ".net" surged, and geopolitical factors contributed to the fading prominence of domains associated with specific countries. Complemented by community detection and k-core analyses, these findings highlight a trend toward a more interconnected, balanced, and globally influenced web. Despite limitations due to data reduction and temporal granularity, this work contributes a valuable, multifaceted perspective on how the web's structure and dominant players have evolved, and may continue to evolve over time.

Keywords: Web-graph · Websites · Domain Names.

1 Introduction

In the past few decades, the world-wide web has rapidly evolved from a simple network of static web pages into a dynamic, interconnected system. This constant growth has enabled some online services to grow exponentially, becoming dominant forces. These services are now used by a significant portion of the global population. For example, Facebook had 3.07 billion monthly users in the year 2023 [1]. Given this immense influence, it is pivotal to better understand these web giants and how they have transformed over time. While previous studies have analyzed the structure and relationships between web pages, they have typically relied on single snapshots of the web.

For this reason, our project investigates the evolution of web networks from 2018 to 2024, focusing on changes in connectivity and influence. The problem centers on determining how the structural properties of the web——such as domain centrality, clustering, and connectivity——have evolved over time. In this context, domains are treated as nodes, and hyperlinks as directed edges. To ensure the analysis remains both computationally feasible and focused on significant players, we are using the 10,000 most significant domains from each time period. This allows us to capture the core structure of the web while maintaining a practical scope. By doing so, we aim to uncover key patterns in the web's most influential sectors.

Our analysis addresses five key questions regarding the evolution of web networks between 2018 and 2024:

1. How have the top domain categories evolved between 2018 and 2024, and what trends can be observed?
2. How has the popularity of the most influential websites changed throughout 2018-2024?
3. How has the in-degree distribution changed from 2018 to 2024?
4. How does a domain's top-level extension affect its influence and connectivity in the network?
5. Were there early indications in 2018 that certain websites might become influential by 2024, and how have their influence and connectivity evolved over time?

Using web-graphs from Common Crawl [2], we will process and analyze data to reveal key metrics that highlight fundamental changes in the web's structure. This approach combines quantitative analysis of network properties with qualitative assessments of domain categories, offering a comprehensive view of how the architecture of the web has evolved over time.

2 Related Work

Kumar et. al. [3] have modeled the web in graphs using directed graphs, where nodes stand for static web pages and directed edges symbolize the hyperlinks connecting the web pages. One well-known approach in this research domain involves random graph models to perform simulations of web-like structures [4]. Dependency is introduced between the formations of edges, meaning that links are formed by "copying" from some of the existing nodes, rather than purely randomly selected ones. These models allow for the evolution of the graph, implying that one introduces new vertices over time.

Although they provide great insight into the structure of the web, these models are simulating and not using real-time data. Our research tries to fill in this gap by using real-time data of web domains to model and analyze the web dynamically. Such an approach gives a much better view of how web connectivity changes with emergent technologies, different social spheres of influence, and the proliferation of diverse domains, which helps understanding of the real-world evolution of web networks.

3 Methodology

Four web-graphs from Common Crawl were used to answer our research questions, one from each of 2018 [5], 2020 [6], 2022 [7], and 2024 [8]. Any further mention of an "original web-graph" refers to one of these graphs.

Due to the size of these web-graphs, which have billions of edges, performing certain analyses and visualization is infeasible. In order to make the data more manageable, each of the original web-graphs was reduced in two steps:

1. The top 30,000 nodes with the highest in-degree in the original web-graph were taken to form a sub-graph. In calculating the in-degree of each node, parallel edges and self-loops were ignored.
2. The PageRank algorithm [9] was run on the $n = 30,000$ sub-graph and the highest-ranking 10,000 nodes were selected to form another sub-graph.

This method was used to reduce each of the original web-graphs to $n = 10,000$ sub-graphs. Any further mention of a "graph" or a "sub-graph" is referring to one of these graphs.

In order to produce the reduced web-graphs, the WebGraph framework [10] was used. Below, there is a list of resources used to produce the reduced web-graphs and the results.

- The code used to produce the sub-graphs, as well as a description of how to reproduce the graphs, is available at the GitHub repository: https://github.com/Chad-Glazier/webgraph-reduction
- The produced graphs are available in CSV format on the Google Drive: https://drive.google.com/drive/folders/1odjh6_URj1K8rUjA6yI06YoV7yLSN79e?dmr=1&ec=wgc-drive-hero-goto
- All code used to produce the results can be found at the GitHub repository: https://github.com/Chad-Glazier/reduced-webgraph-analysis Note that, because the data is large, it is not included inside of this repository. You can download the data from the Google Drive.

3.1 Research Question 1: How have the top domain categories evolved between 2018 and 2024?

To compare the different types of domains, the top 200 domains were categorized carefully by hand. The categories were created to broadly classify websites with the least amount of miscellaneous domains. In total, 13 categories were created.

There were 3 relevant analyses performed on both the 2018 and 2024 data to achieve a top-down view. The first analysis was a pie chart showing the percentages of each category. This was done to provide an easy way to visualize the size comparison of each domain type. The second analysis calculated the top 3 in-degree, closeness, and betweenness centrality for each category. This was performed to examine the general relationships between each type of domain. The third analysis calculated which two categories were the most connected by in-degree. This was to determine which two categories were the most interconnected with each other by hyperlinks. There were two other 2 analyses performed on communities and k-cores, but were deemed less important and moved to Appendix B.

3.2 Research Question 2: How has the popularity of the most influential websites changed?

In order to select the most influential domains from each of the 2018, 2020, 2022, and 2024 sub-graphs, the top 100 domains from each were truncated to their first two domain names. E.g., "com.facebook.m" and "com.facebook" would be combined and their in-degree shares and PageRank scores would be summed. This was to avoid diluting the visualization with several different domains that are part of the same organization. From this collapsed set of domains, the 10 with the highest PageRank scores were selected for each year; since many of the domains appear in the top 10 for multiple years, this led to less than 20 distinct domains. For each of these domains, their relative PageRank scores and in-degree shares were are plotted across time.

It was noticed that many of these top websites were not meant for end-users, but were instead CDNs, web proxies, and other web services. To make the data more relevant, any domain that was not meant for end-users was excluded (e.g., "fonts.googleapis.com").

3.3 Research Question 3: How has the in-degree distribution changed from 2018 to 2024?

The four sub-graphs from 2018, 2020, 2022, and 2024 were used to consider the in-degree distribution of nodes for each year. Two plots were made: the first shows the actual distribution of in-degree share among nodes, and the second shows the cumulative in-degree share for the top x nodes for each year. Notably, although the nodes are selected from the reduced sub-graphs, the in-degree share of each node refers to the proportion of links which point to that domain in the original, full-sized web-graph.

3.4 Research Question 4: How does a domain's top-level extension affect its influence and connectivity in the network?

To explore this, domain-only graph datasets from 2018 and 2024 were analyzed, focusing on metrics like PageRank and connectivity for the top-level domains (TLDs). TLDs from the domain names were extracted in both datasets and used to calculate aggregate metrics for each, including total and average PageRank. Then, metrics for 2018 and 2024 were compared to identify emerging and declining TLDs. Connectivity trends were visualized and cross-connections between emerging and declining TLDs were analyzed.

3.5 Research Question 5: Were there early indications in 2018 that certain websites might become influential by 2024?

To assess early indicators of web pages or domains that might become influential in later years, centrality metrics such as degree centrality, betweenness centrality, and PageRank were calculated. Nodes with moderate degree but high betweenness centrality and high PageRank were of particular interest, as they might represent emerging influencers with growing importance in the network. These metrics enabled the identification of web pages that were not yet dominant in 2018 but exhibited rapid growth or strategic positioning, suggesting their potential to become influential domains in the future.

In describing our results, we will use the following terms.

- **Moderate degree**: This describes a node with a number of connections (degree centrality) that falls between the 25th and 75th percentiles. These nodes are not the most connected but still have a reasonable number of links in the network.
- **High betweenness**: This refers to a node that is in the top 10% for betweenness centrality. These nodes act as important bridges that connect different parts of the network.
- **High PageRank**: This refers to a node that is in the top 10% for PageRank. These are influential nodes based on how many important nodes link to them.
- **Emerging Domain**: This refers to a domain that was rapidly gaining attention in 2018 and went on to have significant influence in 2024. Conversely, a **non-emerging domain** is one that was *not* identified as gaining popularity in 2018, but went on to become influential by 2024 anyway.

To understand what non-emerging nodes had in common, we analyzed their centrality metrics from 2018.

4 Results

The categories with the most connections between them in the $n = 200$ categorized sub-graphs were as follows.

- In 2018, *News* and *Knowledge Base* had the most connections with 646 edges between them.
- In 2024, *Social Media* and *Web Services* had the most connections with 349 edges between them.

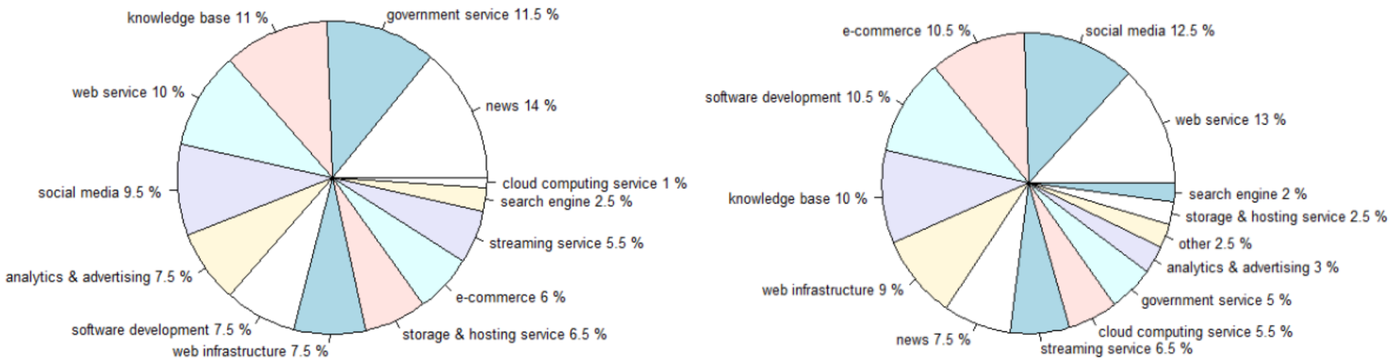


Fig. 1: Proportion of website categories from the top 200 domains from 2018 (left) and 2024 (right).

Table 1: Top three categories from 2018 and 2024 based on in-degree, betweenness, and closeness.

Centrality Measure	**Category (2018)**	**Value**	**Category (2024)**	**Value**
In-degree	Cloud computing	203.0000	Knowledge base	76.3500
	News	191.8214	Social media	74.8400
	Search engine	128.8000	Web service	72.1923
Betweenness	Cloud computing	160.3454	Software development	138.5736
	News	110.9136	Knowledge base	110.1038
	Search engine	110.8893	Social media	98.2797
Closeness	Cloud computing	0.0040	Social media	0.0033
	News	0.0039	Knowledge base	0.0033
	Knowledge base	0.0034	Web service	0.0032

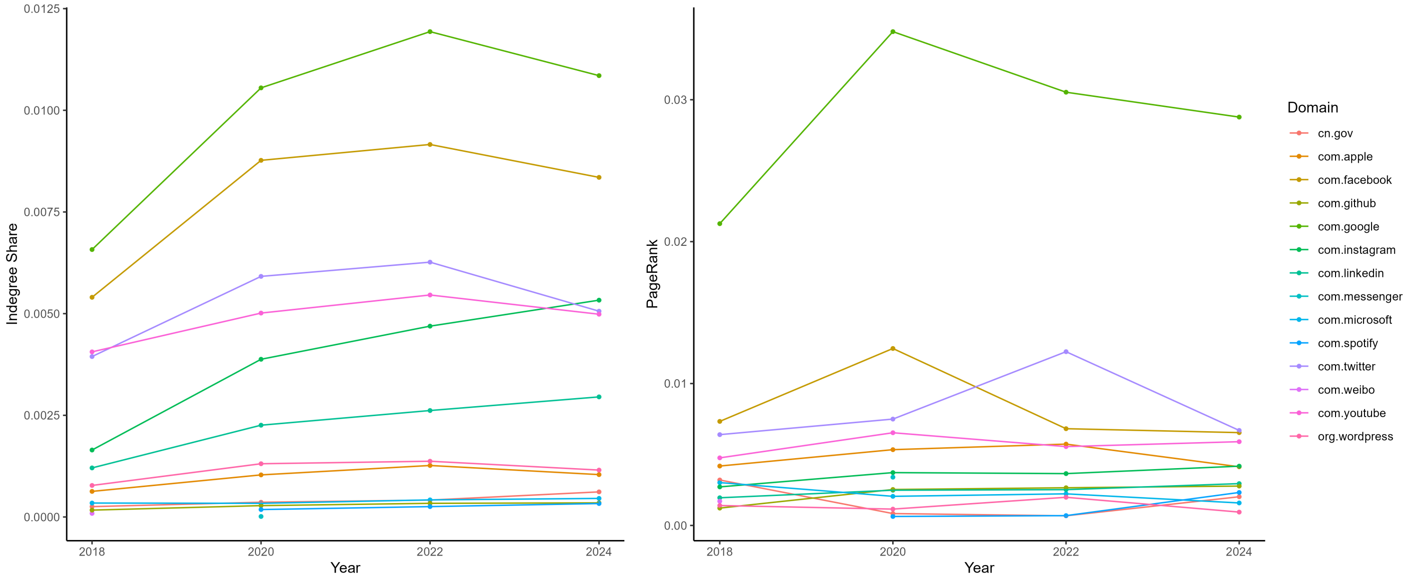


Fig. 2: Changes in in-degree share (left) and PageRank scores (right) for the top 10 end-user domains from each year.

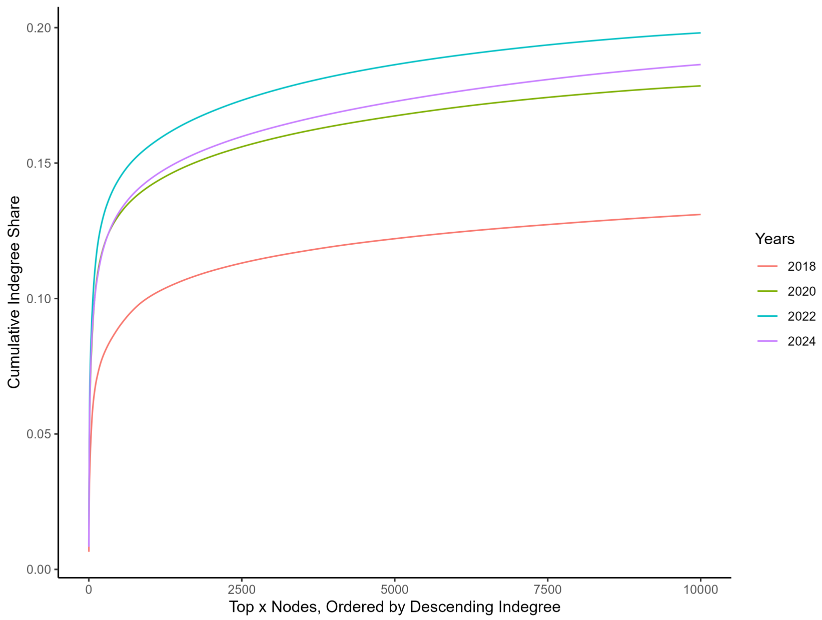


Fig. 3: Cumulative in-degree share for each year.

Table 2: Connectivity changes in the top TLDs.

TLD	Connections (2018)	Connections (2024)	% Change
com	2,061,037	1,449,484	−29.6%
ru	798,322	25,959	−96.7%
org	480,236	543,344	+13.1%
edu	120,098	134,841	+12.2%
uk	90,116	86,187	−4.3%
su	60,449	246	−99.5%
gov	59,493	80,596	+35.4%
net	51,284	97,778	+90.6%
de	47,931	41,754	−12.8%
jp	41,795	34,242	−18.0%

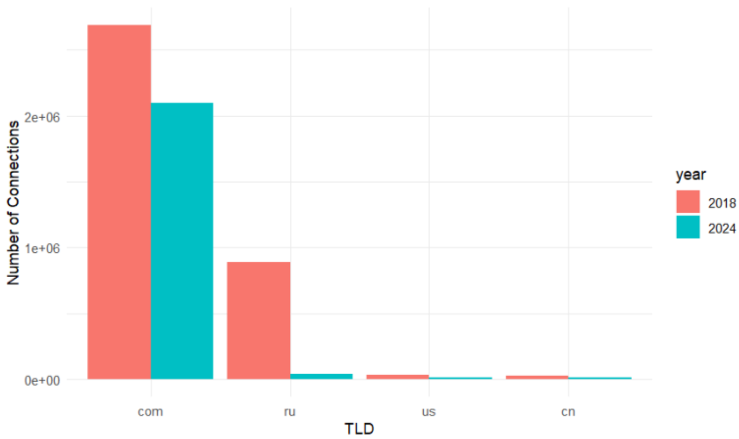


Fig. 4: Connectivity trends for top declining TLDs (2018 vs. 2024)

35 nodes (only 2.25% of all nodes in 2018) were identified as emerging and later became popular in 2024.

Table 3: Top 10 emerging nodes with high centrality in each year.

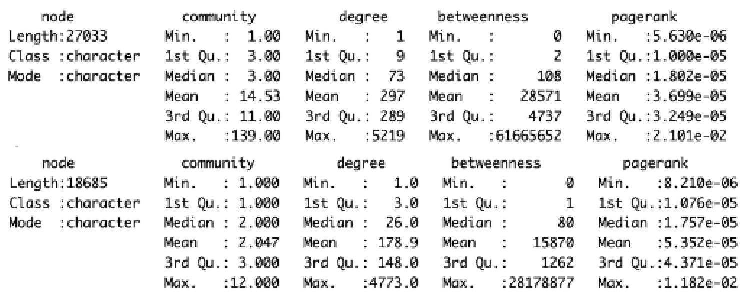
2018	**2024**
cn.west	com.google.sites
xyz.eco-corner	com.medium
com.toocle.china	ph.telegra
com.monfr	com.google.groups
com.examw	org.bitbucket
la.51	ly.bit
cn.gov.beian	org.wikipedia.en
cn.gov	com.github
com.sohu	co.rentry
com.knoji	com.google.play

Table 4: Distribution of emerging nodes in 2024.

Category	**Count**	**Percentage**
Top 10% Popular Nodes in 2024	2,998	—
Popular in 2024 but Not Emerging Nodes	2,963	99%
Popular in 2024 but Not Emerging Nodes Present in 2018	2,182	74%
Popular in 2024 Present in 2018	2,217	74%
New Nodes (2019–2024)	781	26%

Table 5: Centrality scores of non-emerging popular nodes in 2024 based on 2018 metrics.

Metric (in 2018)	**Percentage of Non-Emerging but Popular Nodes in 2024**
Top 10% Degree Centrality	42.3%
Top 10% Betweenness Centrality	34.3%
Top 10% PageRank	33.4%



```
      node            community          degree       betweenness          pagerank
 Length:27033       Min.   :  1.00   Min.   :   1   Min.   :        0   Min.   :5.630e-06
 Class :character   1st Qu.:  3.00   1st Qu.:   9   1st Qu.:        2   1st Qu.:1.000e-05
 Mode  :character   Median :  3.00   Median :  73   Median :      108   Median :1.802e-05
                    Mean   : 14.53   Mean   : 297   Mean   :    28571   Mean   :3.699e-05
                    3rd Qu.: 11.00   3rd Qu.: 289   3rd Qu.:     4737   3rd Qu.:3.249e-05
                    Max.   :139.00   Max.   :5219   Max.   :61665652   Max.   :2.101e-02
      node            community          degree          betweenness          pagerank
 Length:18685       Min.   : 1.000   Min.   :   1.0   Min.   :       0   Min.   :8.210e-06
 Class :character   1st Qu.: 1.000   1st Qu.:   3.0   1st Qu.:       1   1st Qu.:1.076e-05
 Mode  :character   Median : 2.000   Median :  26.0   Median :      80   Median :1.757e-05
                    Mean   : 2.047   Mean   : 178.9   Mean   :   15870   Mean   :5.352e-05
                    3rd Qu.: 3.000   3rd Qu.: 148.0   3rd Qu.:    1262   3rd Qu.:4.371e-05
                    Max.   :12.000   Max.   :4773.0   Max.   :28178877   Max.   :1.182e-02
```

Fig. 5: Number of communities of non-emerging nodes in each year

5 Discussion

5.1 Research Question 1: How have the top domain categories evolved between 2018 and 2024?

The results for this question have a couple implications. In 2018, the most popular websites were news sites. Figure 1 shows that, among the top 200 categorized websites, the news category had the highest representation at 14%. News also had the second-highest in-degree, closeness, and betweenness centrality (Table 1). Additionally, news sites had the most connections with the knowledge base category, making these two categories the most tightly integrated. This makes sense, as both act as major information centers. In 2024, the landscape shifted significantly, with the most popular domains being web services and social media, accounting for for 13% and 12.5% of the top 200 domains, respectively. Furthermore, these two categories had the highest number of connections between them. Social media also ranked second in in-degree, third in betweenness centrality, and first in closeness centrality (Table 1). These results suggest that social media has overtaken news as the most prominent domain category.

5.2 Research Question 2: How has the popularity of the most influential websites changed?

The most popular websites seem to be retaining much of their popularity in terms of PageRank, but their share of in-degree has slightly declined from previous years—–notice the upside-down cup shape of the highest-ranked domains in Figure 2. This might suggest that the most popular websites have peaked around 2020-2022 and the distribution of in-degree is being more evenly spread out. However, this is not the only possible explanation. Another possibility is that ultra-popular websites are diversifying their domains (for example, Google now has several domains for their wide range of products and web services), and the actual organizations behind them are still growing in popularity.

5.3 Research Question 3: How has the in-degree distribution changed from 2018 to 2024?

In order to draw any conclusions about changes in in-degree distribution, we need a broader view. To this end, the cumulative in-degree share of each year's webgraph is visualized in Figure 3. From these plots, it appears as though the greatest concentration of in-degree in the top nodes was in 2022, and the distribution now is more similar to that of 2020. In other words, from the data we have, it appears as though the degree inequality of the web has uncharacteristically decreased from 2022 to 2024.

5.4 Research Question 4: How does a domain's top-level extension affect its influence and connectivity in the network?

The results of the TLD analysis offer several important insights into the structure of the web between 2018 and 2024. First, ".com" obviously remains dominant, confirming its central role in the network. However, its connectivity has notably declined, dropping from 2.06 million connections in 2018 to 1.45 million in 2024 (Table 2). This decrease suggests that while ".com" is still significant, the web is slowly diversifying, and influence is spreading to other TLDs, particularly ".net".

The rise of ".net" is a clear example of such a shift. With its connections almost doubling from 2018 to 2024, ".net" appears to be absorbing links from other TLDs, especially those in decline like ".ru" and ".cn". This change might reflect the growing importance of infrastructure-focused TLDs, as ".net" is increasingly used for hosting, content delivery, and other technical services. The rise of ".net" shows how TLDs that cater to versatile and global use cases can gain influence as the web evolves.

Conversely, the sharp decline of ".ru" (which is the country-code TLD for Russia) highlights how geopolitical factors can significantly impact the role of a TLD in the network. The reduction in ".ru" connections, from over 800,000 in 2018 to fewer than 40,000 in 2024 (Table 4), aligns with the global isolation of Russia following the start of the war in Ukraine in 2022 [11]. Many global companies and providers cut ties with Russian domains, and the internal web space also became more restricted, leading to ".ru" losing both external and internal connectivity. Interestingly, ".us" (the ccTLD for the United States) also shows a decline. This suggests that ".us", once more actively used for regional government and education-related content, may now be less prominent as other TLDs like ".gov" and ".edu" serve these purposes more directly. This trend could also reflect shifts in how regional domains are used, with many activities consolidated under more general-purpose or global TLDs like ".com".

5.5 Research Question 5: Were there early indications in 2018 that certain websites might become influential by 2024?

The results of our analysis suggests that only a small fraction of the emerging domains from 2018 were influential by 2024, which demonstrates that early potential influential success is not strongly predictive of future prominence. The data in Table 3 shows that none of the top 10 domains in 2018 remained among the 10 most central in 2024[1]. This shows a transition toward centralized, authoritative platforms that can reach a wider audience, such as Google, Medium, and YouTube. The authors developed the analysis of emerging centralities with evidence that some emerging domains from 2018 remain important, although emerging social media and knowledge-sharing platforms have reshaped the network

How did the remaining domains become prominent? Table 5 highlights that in 2024, other "non-emerging but popular" domains had relatively high centrality scores in 2018, meaning they were already positioned well within the network. These domains may have used their strategic positioning in terms of connections and authority to become more influential over time, emphasizing the importance of network position for sustainable success.

[1] Table 6 in Appendix A includes 20 nodes and has the same property.

Finally, the network structure underwent the greatest change in the number of domain communities, which has become fewer by almost 120 in the years from 2018 to 2024 (see the metrics in Figure 5). This implies a shift towards greater cohesion of communities, with small communities merging into bigger, more connected ones. This trend and decline in domain community count, degree, betweenness, and PageRank suggests an increasingly distributed and balanced network, which decreases the dominance of certain domains, while favoring a more even distribution of influence.

5.6 Limitations

Challenges related to the size of data and data processing arose and thus the decision was made to reduce the dataset into $n = 10,000$ sub-graphs. This brought about sampling bias since the sub-graphs may not fully represent the overall web-graph and this could affect result generalizability. Limited temporal coverage, with only four web-graphs only covered in a period of six years, could have failed to capture prolonged trends or anomalies. For example, it's possible that the degree inequality of the web has not actually decreased or that it will continue to increase and 2024 is just an outlier.

Future research on web-graphs might utilize several enhancements. Expanding datasets to include the whole web-graphs over longer periods of time will give a complete image of web evolution. Better computational tools, such as GPUs will allow for the investigation of larger datasets without losing detail in insights. In addition to that, close examination of emerging node life cycles stands to help identify factors that influence their rise and decline. Finally, integrating the geopolitical and socioeconomic data in detail would provide another layer of understanding of external influences on web connectivity.

6 Conclusion

The internet has undergone many changes since its inception, but it is now clear just how much it has shifted in six years. By using a large amount of data from across the internet and distilling it to its most influential domains, we uncovered a few changes and trends. The period from 2018 to 2024 saw the web transition from news and information-sharing platforms to social media and online services as dominant categories, with Google and YouTube strengthening their hold. For TLDs, ".com" declined in use while ".net" rose with an increase in diversification, and ".ru" notably declined in influence partly due to geopolitical events. Community structures became consolidated on the web, where smaller communities merged into larger connected clusters. Broadly, it appears as though the web is gradually reducing degree inequality and becoming more evenly-distributed. However, this study only includes data from 2018 to 2024. By 2018, the internet was already well established, and many of the trends observed in this study were likely ongoing beforehand. A future study with a dataset spanning a longer time frame would likely produce even more pronounced results.

References

1. Meta Platforms, Inc.: Meta Reports Fourth Quarter and Full Year 2023 Results; Initiates Quarterly Dividend. Press Release, January 31, 2024. https://investor.fb.com/investor-news/press-release-details/2024/Meta-Reports-Fourth-Quarter-and-Full-Year-2023-Results-Initiates-Quarterly-Dividend/default.aspx (accessed November 20, 2024).
2. Common Crawl: Common Crawl Homepage. https://commoncrawl.org (accessed November 20, 2024).
3. Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., Upfal, E.: Stochastic models for the web graph. In: Proceedings of the 41st Annual Symposium on Foundations of Computer Science, Redondo Beach, CA, USA, pp. 57–65 (2000). https://doi.org/10.1109/SFCS.2000.892065
4. J. Wu and B. Xu: A Method to Support Web Evolution by Modeling Static Structure and Dynamic Behavior. In: 2009 International Conference on Computer Engineering and Technology, Singapore, 2009, pp. 458-462 (2009). doi: 10.1109/ICCET.2009.165.
5. Common Crawl: Web Graph Release cc-main-2018-19-nov-dec-jan. https://data.commoncrawl.org/projects/hyperlinkgraph/cc-main-2018-19-nov-dec-jan/index.html (accessed November 14, 2024).
6. Common Crawl: Web Graph Release cc-main-2020-21-oct-nov-jan. https://data.commoncrawl.org/projects/hyperlinkgraph/cc-main-2020-21-oct-nov-jan/index.html (Accessed November 14, 2024).
7. Common Crawl: Web Graph Release cc-main-2021-22-oct-nov-jan. https://data.commoncrawl.org/projects/hyperlinkgraph/cc-main-2021-22-oct-nov-jan/index.html (Accessed November 14, 2024).
8. Common Crawl: Web Graph Release cc-main-2024-jul-aug-sep https://data.commoncrawl.org/projects/hyperlinkgraph/cc-main-2024-jul-aug-sep/index.html (Accessed November 14, 2024).
9. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank citation ranking: Bringing order to the web. Technical Report, Stanford InfoLab (1999). Available at: http://ilpubs.stanford.edu:8090/422/.
10. Boldi, P., Vigna, S.: The WebGraph framework I: Compression techniques. In: Proceedings of the Thirteenth International World-Wide Web Conference, pp. 595–601. ACM Press (2004).
11. Human Rights Watch. Russia: With Tech Firms Pulling Out, Internet Spiraling into Isolation (2022). https://www.hrw.org/news/2022/03/14/russia-tech-firms-pulling-out-internet-spiraling-isolation (Accessed December 1st, 2024).
12. Blondel, V.D., Guillaume, J.-L., Lambiotted, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

A Full Tables

Table 6: Top 20 emerging nodes with high centrality in each year.

2018	**2024**
cn.west	com.google.sites
xyz.eco-corner	com.medium
com.toocle.china	ph.telegra
com.monfr	com.google.groups
com.examw	org.bitbucket
la.51	ly.bit
cn.gov.beian	org.wikipedia.en
cn.gov	com.github
com.sohu	co.rentry
com.knoji	com.google.play
com.smartcode	com.printwhatyoulike
com.baidu.zhidao	com.youtube
com.123-free-download	com.rapidapi
ca.blogtv	com.google.docs
us.opendi	com.apple.apps
com.bing.cn	me.t
com.szshjsj	cc.vocus
nl.online-winkelstart	com.forbes
com.nihuili	com.pinterest
com.jianshu	jp.ameblo

B Further Category Analysis

B.1 Methodology and Results

To analyze whether the communities formed by the $n = 200$ category sub-graph using the Louvain community detection algorithm [12]. Figure 6 displays the number of domains from each category that are in each of the 4 communities.

k-core analysis was also conducted to find that in the sub-graphs, the maximum k-core in 2018 included 13 distinct categories, while the maximum k-core in 2024 had 12.

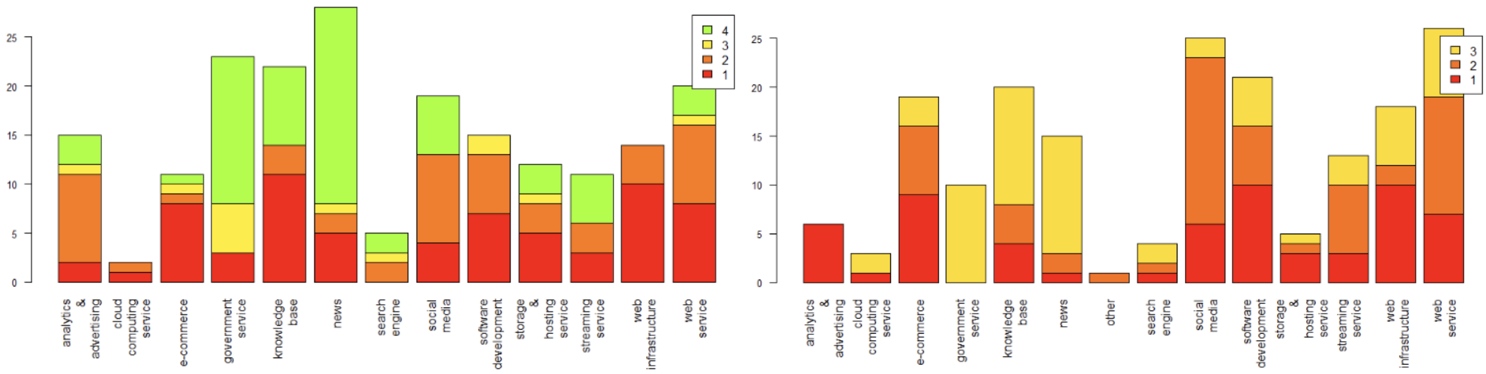


Fig. 6: Number of categories in each identified community for 2018 (left) and 2024 (right) from the $n = 200$ category sub-graphs.

B.2 Discussion

When analyzing the communities (Figure 6), it is evident that no single category forms its own isolated community. However, in 2024, all government service domains were found in community 3. Furthermore, several other categories showed high percentages within specific communities. This indicates that certain categories are more tightly connected, but typically in conjunction with other categories rather than in isolation. This finding is further supported by the k-core analysis. The maximum k-core contains a variety of domains from nearly every category. This demonstrates that there is no single category at the center of the graph, emphasizing the interconnectedness of different domain types.