

# Multi-Platform URL-Based Information Diffusion from Three U.S. Political Events

Isabel Murdock<sup>1</sup>[0000-0002-7932-5216], Kathleen M. Carley<sup>1</sup>[0000-0002-6356-0238],  
and Osman Yag an<sup>1</sup>[0000-0002-7057-2966]

Carnegie Mellon University, Pittsburgh PA 15213, USA  
`iem@andrew.cmu.edu`

**Abstract.** Using datasets from three political events, this paper analyzes the diffusion of URL-based content across Twitter, Facebook, and Reddit. Posts regarding the 2020 U.S. election, the Dobbs v. Jackson Supreme Court decision, and the 2022 U.S. election are examined to identify multi-platform diffusion trends. We show that the datasets contain content with varied political biases and factual ratings. Thus, the presented findings regarding the performance of multi-platform URLs and the order in which they spread across the platforms are more likely to generalize to additional political situations and be relevant for informing interventions against misinformation diffusion.

**Keywords:** social media · multi-platform diffusion · misinformation.

## 1 Introduction

Despite increasing attention to the spread of harmful content online, social media platforms have continued to serve as mediums for spreading misinformation. Recent examples include the multi-platform spread of COVID-19 related misinformation [9] and fraud claims related to the 2020 U.S. election [3]. Combined with the increase in AI tools that can assist in generating misleading content, which can ultimately be used to influence electoral processes, sow societal divisions, and cause distrust of democratic institutions, misinformation and disinformation have become central threats to global stability [12].

By identifying trends that arise in datasets of varied political contexts, the findings presented in this paper supplement prior work on multi-platform information diffusion [5, 8]. We focus specifically on Twitter, Facebook, and Reddit and analyze data from three U.S.-based political events. Collecting social media data from the 2020 and 2022 U.S. elections and the Dobbs v. Jackson Supreme Court decision, we measure and characterize how different types of URLs were posted and spread between the three platforms. We first identify URL-based content that appeared on multiple platforms. We then compare the performance and diffusion of the URLs based on the platforms on which they appeared. This allows us to evaluate whether multi-platform URLs were posted more and had greater longevity than their single-platform counterparts. We conclude by analyzing the order in which the URLs appeared on the platforms, which provides insights for potential countermeasures to limit the spread of misinformation.

## 2 Related Work

Much of the prior work on information diffusion over social media has focused on information spread within a single platform. However, as it has become increasingly apparent that misinformation spreads across social media platforms, there has recently been an increase in studies looking into multi-platform information diffusion. Concerning the spread of fake information, work has found that the origin of highly reposted information can impact its likelihood of spreading across multiple platforms [2]. Another line of research that analyzed postings of COVID-19 conspiracy-related URLs found that the information paths between platforms are complex and content dependent, and that fringe social media sites are not the sole contributors to the spread of conspiracy theories [9]. While much of this work on information diffusion has been data-driven, theoretical research has also examined how information can spread faster and further when users are connected to additional, conjoining networks [13].

Another relevant research topic is the role that multiple platforms can play in the execution of disinformation campaigns. One study analyzed the use of Twitter and YouTube in a campaign against the White Helmets during the Syrian civil war. It found that those targeting the White Helmets used YouTube, as well as “alternative” news websites, in a complementary and effective way to direct users from Twitter to large sets of videos on YouTube [11]. Another study, which focused on the Russian disinformation campaign during the 2016 U.S. election, investigated Internet Research Agency activity on Twitter, Facebook, and Reddit and suggested that it may have used Reddit to test out content before spreading it on Twitter [6].

## 3 Approach

We examine the spread of information over multiple platforms by collecting data from Twitter, Facebook, and Reddit as outlined in section 4. We focus on these platforms as they are widely used by the American public [1] and have been used in previous disinformation and influence campaigns [6]. Furthermore, users on the platforms often share content in the form of URLs, providing a convenient mechanism to study information spread. Due to this, as well as their prevalence in the collected datasets, URLs serve as the basis for studying multi-platform information diffusion in this work.

We identify and characterize two categories of URLs from our collected datasets: “single-platform URLs” and “multi-platform URLs.” To do this, the URLs that appear in the collected posts are first extracted and standardized so those linking to the same content map to the same representative URL. We then separate the URLs that only appear on a single platform in the given dataset, i.e., single-platform URLs, and those that appear on multiple ones, i.e., multi-platform URLs. The final step is filtering for the URLs that contain their respective dataset’s collection terms. This helps us focus on the URLs that are more likely to be related to the topics of interest and have all of their posts represented in the datasets, assuming we had access to the posts.

To understand the types of content linked to by the multi-platform URLs, their domains are classified into one of eight categories ranging from mainstream news to investigative/government websites. To further classify the bias and credibility of the collected multi-platform URLs, we use ratings from Media Bias/Fact Check (MBFC) [7], an independent website that provides bias and credibility ratings on hundreds of media sources. Since our collected datasets relate to news and politics, we anticipate that much of the single- and multi-platform URLs link to websites classified by MBFC. This allows us to label these URLs with the bias and factual ratings of the websites they link to. The resulting bias categories in the datasets are Left, Left-Center, Center, Right-Center, and Right. The fact ratings in the dataset are: Very High, High, Mostly Factual, Mixed, and Low. Since MBFC also provides lists of Questionable and Conspiracy/Pseudoscience websites, URLs linking to those sources are labeled as questionable or conspiracy/pseudoscience-related.

### 3.1 URL Diffusion Measures

To address how the URLs spread differently, based on the platforms that they were posted on and the types of content that they linked to, we use the following URL propagation metrics for each dataset:

- **Number of Posts:** The number of times each URL was posted.
- **Posting Life Span:** The total time, in days, between the first and last time a given URL was posted.
- **Posting Speed:** The average number of times a URL was posted per hour. This is a modified version of the retweet propagation speed defined by Shahi et al. [10]. If a URL is only posted once, it has a posting speed of zero.
- **Number of Lives:** The number of active posting periods a URL has. An active posting period is any time interval when a URL is posted without a break longer than 24 hours. It is similar to the concept of active sessions that has been used to study user behavior [4].

Together, these metrics help us understand how much the URLs were posted, how long they lasted on the platforms, and how quickly they spread across each platform. We also analyze the timing of the multi-platform URLs in diffusing across the different platforms. We consider the order in which the URLs first appeared on each platform, along with the time delay it took for them to appear on the second and third platforms. We also study the delays in peak posting of the multi-platform URLs on each platform by taking the 24-hour simple moving average of the URLs’ hourly posting signal on each platform.

## 4 Data

This paper uses three multi-platform datasets related to U.S. political events. The tweets were gathered using Twitter’s API v2 full-archive search. The Facebook posts were collected using CrowdTangle and come from public groups

**Table 1.** Data collection parameters and collected multi-platform URLs and postings.

Dataset	Time frame	Topic(s)	Twitter	Facebook	Reddit
2020 election	1 Oct 2020 - 19 Jan 2021	election fraud and protests	11K URLs 2.2M posts	10K URLs 39K posts	3K URLs 15K posts
Dobbs decision	1 May 2022 - 31 Jul 2022	abortion bans/access and pro-life terms	14K URLs 857K posts	13K URLs 32K posts	3K URLs 11K posts
2022 election	1 Oct 2022 - 30 Nov 2022	close senate, house, and gubernatorial races	11K URLs 922K posts	9K URLs 18K posts	2K URLs 6K posts

and pages. Meanwhile, the Reddit posts and comments were collected using the Pushshift dataset and include posts and comments from public subreddits.

The collection parameters for each dataset are displayed in Table 1. The first dataset relates to the 2020 U.S. election and was collected based on a list of phrases related to election fraud and election-related protests. The second dataset focuses on discussions about the Dobbs v. Jackson Women’s Health Organization Supreme Court decision. The final dataset relates to the 2022 U.S. midterm election. Its posts were collected based on phrases related to seven states with close gubernatorial, U.S. Senate, or U.S. House races: Arizona, Georgia, North Carolina, Nevada, Ohio, Pennsylvania, and Wisconsin.

## 5 Analysis of Multi-Platform URLs

As shown in Table 1, all of the datasets contain a significant number of multi-platform URLs. Most of the content linked to by the URLs relates to news, with more than 86% of the multi-platform URL postings in each dataset having an MBFC rating. From these URLs, we find that the political bias and fact ratings vary across the datasets. The 2020 election dataset was more balanced than the other two in terms of political bias but contained more low fact content. In contrast, the Dobbs dataset was the most moderate but the least balanced. Although its collection terms included phrases associated with pro-life and conservative positions, left-leaning content dominated this dataset. It also tended to contain more higher fact URLs than the other two datasets. The prevalence of news content in the datasets and the diversity of the websites they tended to link to enables us to identify more generalizable multi-platform trends.

One such trend relates to the general performance of the multi-platform URLs compared to those appearing on a single platform. We find that the URLs appearing on multiple platforms tended to be posted more. This held for all three datasets and was especially true for URLs that appeared on all three platforms. We also find that the multi-platform URLs had longer posting life spans, with the URLs appearing on all three platforms tending to remain in the social media ecosystem the longest. They similarly had the most lives.

Since the URLs that appeared on multiple platforms had the opportunity to spread in parallel across three different social networks, we expect their levels of posting to be greater. However, when we compare the performance of the multi-platform URLs on each platform individually to the given platform’s

single-platform URLs, we find that the multi-platform URLs still outperform the single-platform ones. This was consistent across the three datasets and all of the platforms. We used two-sample Kolmogorov-Smirnov and two-sample Anderson-Darling tests to determine whether the differences in the within-platform performance were statistically significant. In all but one case, the differences in the performance of the multi- and single-platform URLs were significant ( $\alpha = 0.01$ ).

Having established that the multi-platform URLs outperformed the single-platform ones, we now consider the order in which the multi-platform URLs appeared on the different platforms. We find that they tended to appear on Twitter and Facebook before Reddit across all of the datasets. Within the Dobbs dataset, the close relationship between Twitter and Facebook was especially present, with almost a quarter of the URLs being introduced to the platforms within a minute of each other and more than half being introduced within an hour. In general, more URLs were posted on Twitter before Facebook. As for the multi-platform URLs appearing on Reddit, they tended to appear on Twitter or Facebook first and then spread to Reddit. Almost half of the URLs that appeared on Twitter and Reddit took longer than a day to appear on Reddit after being posted on Twitter. For URLs posted to both Facebook and Reddit, we find that they tended to appear on Facebook first but spread to Reddit with a generally shorter lag than was found between Twitter and Reddit.

Compared to the delays in the first appearance on the platforms, the delays in the peak posting of the multi-platform URLs across the different platforms were considerably smaller. This was true for all combinations of the platforms but was particularly noticeable for the URLs on Twitter and Reddit. With this in mind, the results suggest that while there may be time to enact interventions on Reddit before content from Twitter or Facebook spreads to the platform, waiting until such content goes viral or reaches its peak posting on Twitter or Facebook may be too late to be effective.

## 6 Conclusion

This paper uses datasets from three political events to explore the multi-platform diffusion of URL-based content. The different contexts of each dataset translated to them having varied political biases and fact ratings. Due to this, we argue that the multi-platform behaviors identified across all three datasets are more likely to apply to additional contexts. Among our findings from this work is that the multi-platform URLs outperform those appearing on a single platform. This further motivates the importance of understanding how misinformation spreads in a multi-platform environment. Additionally, as there was considerable delay in content spreading to Reddit, there could be opportunities to conduct interventions before potentially harmful content appears on the platform.

**Acknowledgments.** This work was supported in part by the National Science Foundation through Grants #1813637 and #2225513, Office of Naval Research through Grant N000141712675, and the Army Research Office through Grant #W911NF-22-1-0181. I. Murdock was supported by a Knight Fellowship through the CMU IdeaS

Center. The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Knight Foundation, the National Science Foundation, the Office of Naval Research, the Army Research Office, or the US Government.

## References

1. Auxier, B., Anderson, M.: Social media use in 2021. Pew Research Center (April 2021), <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/>
2. Chen, Z., Aslett, K., Rosiere, J., Reynolds, J., Nagler, J., Tucker, J., Bonneau, R.: An automatic framework to continuously monitor multi-platform information spread. *CEUR Workshop Proceedings* **2890** (2021)
3. Fischer, S.: Reddit bans subreddit group "r/donaldtrump". *Axios* (Jan 2021), <https://www.axios.com/2021/01/08/reddit-bans-rdonaldtrump-subreddit>
4. Kooti, F., Moro, E., Lerman, K.: Twitter session analytics: Profiling users' short-term behavioral changes. In: Spiro, E., Ahn, Y.Y. (eds.) *Social Informatics*, vol. 10047, pp. 71–86. Springer International Publishing, Cham (2016). [https://doi.org/10.1007/978-3-319-47874-6\\_6](https://doi.org/10.1007/978-3-319-47874-6_6)
5. Lepird, C.S., Ng, L.H.X., Wu, A., Carley, K.M.: What news is shared where and how: A multi-platform analysis of news shared during the 2022 U.S. midterm elections. *Social Media + Society* **10**(2), 20563051241245950 (2024). <https://doi.org/10.1177/20563051241245950>
6. Lukito, J.: Coordinating a multi-platform disinformation campaign: Internet research agency activity on three U.S. social media platforms, 2015 to 2017. *Political Communication* **37**(2), 238–255 (2020). <https://doi.org/10.1080/10584609.2019.1661889>
7. Media Bias / Fact Check: Bias and factual reporting ratings, <https://mediabiasfactcheck.com>, last accessed Jan. 2022
8. Murdock, I., Carley, K.M., Yağan, O.: Identifying cross-platform user relationships in 2020 U.S. election fraud and protest discussions. *Online Social Networks and Media* **33**, 100245 (2023). <https://doi.org/https://doi.org/10.1016/j.osnem.2023.100245>
9. Papakyriakopoulos, O., Medina Serrano, J., Hegelich, S.: The spread of covid-19 conspiracy theories on social media and the effect of content moderation. *Harvard Kennedy School (HKS) Misinformation Review* (2020). <https://doi.org/10.37016/mr-2020-034>
10. Shahi, G.K., Dirkson, A., Majchrzak, T.A.: An exploratory study of covid-19 misinformation on twitter. *Online Social Networks and Media* **22**, 100104 (2021). <https://doi.org/10.1016/j.osnem.2020.100104>
11. Starbird, K., Wilson, T.: Cross-platform disinformation campaigns: Lessons learned and next steps. *Harvard Kennedy School (HKS) Misinformation Review* (2020). <https://doi.org/10.37016/mr-2020-002>
12. World Economic Forum: The global risks report 2024 (January 2024), <https://www.weforum.org/publications/global-risks-report-2024/>
13. Yağan, O., Qian, D., Zhang, J., Cochran, D.: Conjoining speeds up information diffusion in overlaying social-physical networks. *IEEE Journal on Selected Areas in Communications* **31**(6), 1038–1048 (2013). <https://doi.org/10.1109/JSAC.2013.130606>