Multi-Platform Analysis of 2020 U.S. Election Fraud and Protest Related Posts [*]

Isabel Murdock[1], Kathleen M. Carley[1], and Osman Yağan[1]

Carnegie Mellon University, Pittsburgh PA 15213, USA
{iem, oyagan}@andrew.cmu.edu, kathleen.carley@cs.cmu.edu

Abstract. Understanding how social media users interact and spread information both within and across different platforms is critical for accurately simulating and developing effective methods for promoting truthful information and disrupting misinformation. In order to understand how information was shared regarding election fraud and protests during the 2020 U.S. Presidential Election, we collect relevant Twitter, Facebook, and Reddit posts made between Oct. 1, 2020 and Jan. 19, 2021. The posts are then analyzed to compare communication within each platform, as well as to characterize cross-platform referencing and URL-based content spread. We find that social media posts relating to election fraud and protests differed in both content and timing across the platforms, and that there were significant cross-platform interactions.

Keywords: Social media networks · Multi-platform · Election.

1 Introduction

The increasing number and diversity of available social media platforms have not only given users more ways to connect with others and share information, but also complicated the ways in which misinformation and disinformation can spread. Social media users have a wide selection of platforms to choose from, each of which provide unique social structures, posting mechanisms, and recommendation systems. On top of that, many users engage with multiple social media websites daily and, in doing so, provide new pathways through which information can spread between different communities on different platforms [1].

This has led multi-platform-based research to become a critical area of research [2], especially within the field of social cybersecurity [3]. Researchers have characterized and compared the activity and users on different social media platforms, ranging from studying different methods of political campaigning across platforms [4] to identifying linguistic differences in posts made on Facebook and Twitter [5]. More recently, work has been done involving the multi-platform spread of COVID-19 related misinformation [6, 7]. In particular, postings of COVID-19 conspiracy-related URLs have been used to analyze the effectiveness of content moderation done by Twitter, Facebook, and Reddit [6].

The goal of this work is to provide a multi-platform analysis of how information was shared and spread between platforms regarding election fraud and election protests surrounding the 2020 U.S. presidential election. The Election Integrity Partnership's 2020 Election Report provides in-depth insight into the ways in which a wide range of social media platforms were used to spread misinformation, as well as details about the cross-platform nature of how different narratives grew and stayed alive [8]. This work furthers the discussion by collecting a set of over 24M social media posts and using them to study how, in an aggregate sense, users discussed election fraud and protests. Rather than characterizing how particular stories were discussed, this work focuses on comparing the main ways that posts regarding fraud and protests differed across the platforms, and how cross-platform URLs can reveal multi-platform trends.

2   Data

In order to analyze how social media users on Twitter, Facebook, and Reddit discussed election fraud and protests surrounding the 2020 U.S. presidential election, we first compiled a list of hashtags and phrases pertaining to election fraud and election-related protests (see Table 1). The election fraud key terms include both general mentions of fraud and fake votes, as well as phrases relating to more specific narratives such as Dominion Voting Systems and dead voters. While the election protest terms are related, they focus more on calls to action such as "do not certify" and references to marches and rallies.

Table 1. Data filter terms.

Election Fraud	Election Protests
corrupt election (corruptelection)	do not certify (donotcertify)
dead voters (deadvoters)	maga civil war (magacivilwar)
deceased voters (deceasedvoters)	march for trump (marchfortrump)
dominion voting systems (dominionvotingsystems)	march to save america (marchtosaveamerica)
election fraud (electionfraud)	million maga march (millionmagamarch)
election integrity (electionintegrity)	
fake election (fakeelection)	saveamerica
fake votes (fakevotes)	save america rally (saveamericarally)
fraudulent election (fraudulentelection)	stop the fraud (stopthefraud)
legal votes only (legalvotesonly)	stop the steal (stopthesteal)
legitimate votes only (legitimatevotesonly)	wild protest (wildprotest)
massive corruption (massivecorruption)	
rigged election (riggedelection)	
stolen election (stolenelection)	
voter fraud (voterfraud)	

Using these key terms, we collected social media posts made between Oct. 1, 2020 and Jan. 19, 2021 that contained case insensitive versions of them. This time range includes both pre-election discussions and, more interestingly, posts made about major fraud narratives and protest events following the Nov. 3, 2020 election.

Overall, over 23M tweets, 726K Facebook posts, and 262K Reddit posts and comments containing the key terms in their titles, text, or URLs were collected. Tweets were gathered using Twitter's full-archive search and include public posts that were not deleted or removed by the time the data was collected in March 2021. Facebook posts were collected using CrowdTangle and come from public Facebook groups with more than 95K members (or US-based groups with more than 2K members), pages with more than 50K likes, and verified profiles and pages with at least 100K followers. Reddit posts and comments were collected using the Pushshift API and include posts and comments made in public subreddits.

Each collected post was classified as relating to fraud, protests, or both, according to the key terms that it contained. For all of the platforms, there were more election fraud-related posts than protest-related ones (see Table 2). Relatively few posts contained both fraud and protest key terms.

Table 2. Posts and URLs collected from each platform.

	# of Posts			# of URL Postings		
	Fraud	Protests	Both	Fraud	Protests	Both
Twitter	19,593,141	3,874,322	318,960	11,872,690	3,298,517	306,188
Facebook	563,747	140,317	22,448	533,065	133,577	21,153
Reddit	228,503	19,171	14,883	110,530	10,817	86,603

Additionally, we collected the URLs occurring in each post. Overall, URLs were posted about 16M times (see Table 2). We cleaned the URLs and used SMaPP's urlExpander tool [9] to identify shortened ones. The shortened URLs were then expanded, when possible, using URL X-ray [10]. This process was used to create a URL dictionary so that URLs linking to the same content, despite being shortened or having slightly different formats, would map to the same representative URL. At the end of this process, there was a final set of approximately 2M unique URLs.

3   Analysis

Our analysis focuses on two main goals: (i) comparing the differences in communication within each of the three platforms, and (ii) understanding how information was spread across them. In order to tackle the first objective, we examine the temporal patterns of fraud and protest-related posts for each platform, the

post length and net positive reactions those posts received, and the types of URL content and political leaning of news sources contained within them. For the cross-platform analysis, we look into the level and general motivations of cross-platform linking, as well as the amount, type, and timing of URL-based content that appeared on multiple platforms.

3.1   Platform Comparison

Posts Over Time. In terms of fraud-related posts over time, the first significant increase in activity occurred on the day after the election. Twitter and Facebook reached their peak number of fraud posts a couple of days later on Nov. 6 (see Fig. 1), the day that both Pennsylvania and Georgia switched from Trump to Biden leading in the ballot counts [11, 12]. As such, it makes sense fraud was looked to as an explanation for those developments. Interestingly, Reddit's number of fraud-related posts and comments peaked a day afterward on Nov. 7, the day that most major news networks declared Biden the winner of the election [13]. For all three platforms, the election fraud posts continued to be made at a relatively high level for the week following the election.
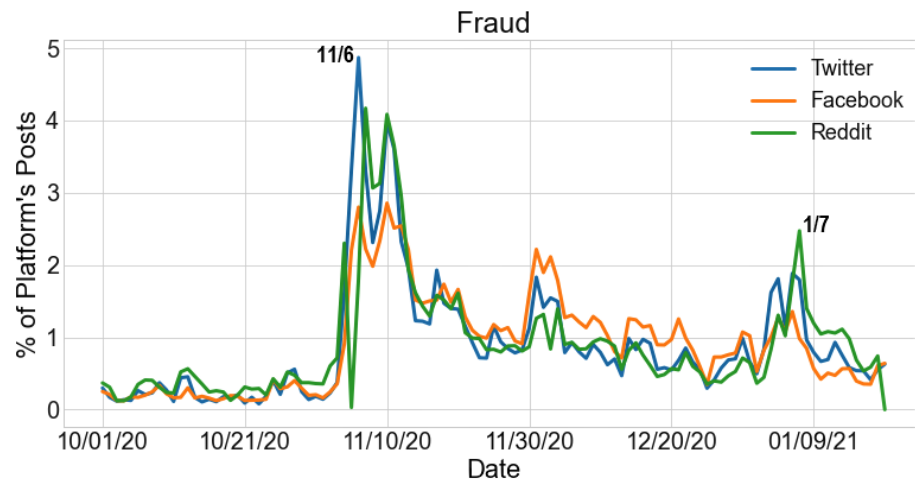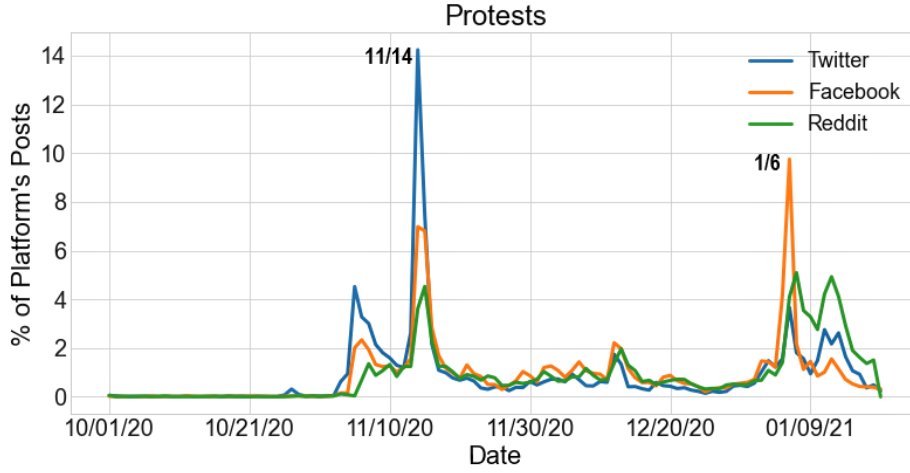

Fig. 1. Percent of each platform's election fraud posts by day they were posted.

There was also a relatively small increase in fraud-related posts made on Twitter and Facebook in the lead up to Jan 6. Again, we find that Reddit lagged the other two platforms and had more posts made on Jan. 7, the day after the Capitol riot, rather than before or during the event.

As for the posts related to election protests, while there was a small uptick in posts following the election, the most significant spikes were on days of major marches and protests (see Fig. 2). The first spike in posts occurred on the day

of the Million MAGA March in Washington D.C., Nov. 14 [14]. Despite having had a significant number of posts on that date, Reddit had even more posts on Nov. 15.

Fig. 2. Percent of each platform's election protests posts by day they were posted.

The next significant spike in protest-related posts occurred on Jan. 6. In particular, Facebook saw the largest relative increase in posts, whereas Reddit again had more posts on the day after the event. Reddit's uptick in posts discussing protests was sustained for the week following the event. Surprisingly, Twitter did not have as large of a relative increase in tweets in relation to Jan. 6 as the other two platforms. This could be due to users deleting relevant tweets or Twitter's own moderation efforts.

Overall, we found that all of the platforms had increases in fraud and protest-related posts around similar dates and that they coincided with major election events and protests. That being said, Reddit appeared to lag about a day behind Twitter and Facebook in terms of peaks in relevant posts, which may indicate its use as a platform for discussing major events that had already occurred, rather than planning or energizing such events beforehand.

Post Length and Reactions. To measure and compare the length of the tweets, Facebook posts, and Reddit posts and comments in our dataset, post length is defined as the number of words in each post. The reaction that each post received is quantified based on the net positive reactions for each platform. For Twitter, the net positive reaction is the number of likes a tweet received. For Facebook, it is the number of likes, loves, wows, and cares a post had, minus its sad and angry reacts. For Reddit, it is simply the score of the post or comment, which is calculated by Reddit as its upvotes minus downvotes.

In general, tweets were the shortest posts and Facebook posts were the longest (see Table 3). That being said, our dataset contained some exceptionally long Reddit posts. Also, Reddit posts tended to be shorter than Reddit comments. Across all of the platforms, fraud-related posts tended to be longer than protest-related ones. Additionally, for both Facebook and Reddit, posts containing both fraud and protest keywords were the longest. Such posts may have tied multiple narratives together which resulted in longer posts.

Table 3. Posts length (# of words in post) and net positive reaction.

	Post Length			Net Positive Reaction		
	Twitter	Facebook	Reddit	Twitter	Facebook	Reddit
Avg	21	122	85	4	143	5
Min	1	1	1	0	-59,589	-323
Max	108	2,173	6,254	562,707	979,633	60,541
Median	21	57	47	0	1	1

Similar to the overall post lengths, we also found that Facebook posts contained, on average, more URLs than Reddit posts or comments did which, in turn, contained more URLs than Twitter posts did. While URLs were ubiquitous in posts from all platforms, the longer post lengths supported by Facebook and Reddit allowed social media users to include more content and URLs within individual posts.

As for net positive reactions on those posts, Facebook posts received, on average, much larger reactions. This, however, is due to the fact that the Facebook data only reflected reactions to posts in highly followed public pages and groups. We also found that most posts received minimal reactions, although some garnered hundreds of thousands of positive reactions. On average, fraud-related posts on Facebook received about 40 more net positive reactions than protest-related posts did. Whereas on Reddit, the score for protest-related posts averaged about 7 points higher than fraud-related posts. Across all platforms, posts containing both fraud and protest keywords received the lowest average net positive reaction.

Types of URL Content. To compare the types of content most prevalent on each platform, roughly 800 of the most posted website domains were classified into one of the following categories: social media, news, political, media sharing, research/government, blog, and other. This resulted in approximately 88% of URLs posted being classified. All other non-classified domains and URLs were categorized as unclassified.

Social media was the most common type of URL posted, making up 57% of Twitter URL postings, 40% of Facebook URL postings, and 26% of Reddit URL postings. News was the next most posted category, accounting for 19-23%

of each platform's URL postings. On Twitter and Facebook, news links were more likely, and social media links were less likely, to be posted multiple times. Conversely, news URLs were less likely than social media ones to be reposted on Reddit.

In terms of the fraud and protest categories, links to social media made up 10% more of the URLs from protest posts than fraud posts across all platforms. News, on the other hand, was 10% more common in the URLs from fraud posts than protest ones.

This analysis suggests that, regardless of the platform, the topic of discussion may impact the type of URL content referenced by users. It also indicates that users may be inclined to post different types of content depending on the platform they are using.

Political Leaning of News Sources. News URLs were further classified according to their website's media bias ranking from AllSides [15]. Those included in the AllSides database were assigned one of the following categories: Left, Left-Center, Center, Right-Center, Right. Less mainstream, local, or foreign news sources not included in the database were given a bias label of None.

While news postings in the Twitter and Reddit datasets have a left-leaning bias, we find that the Facebook dataset has a right-leaning bias (see Fig. 3). That said, all platforms have more left-center postings than left ones, and right-center postings than right ones. Additionally, on Twitter, URLs from left-center sources were more likely to be posted multiple times, whereas, on Facebook, URLs from right-center sources were more likely to be posted repeatedly.
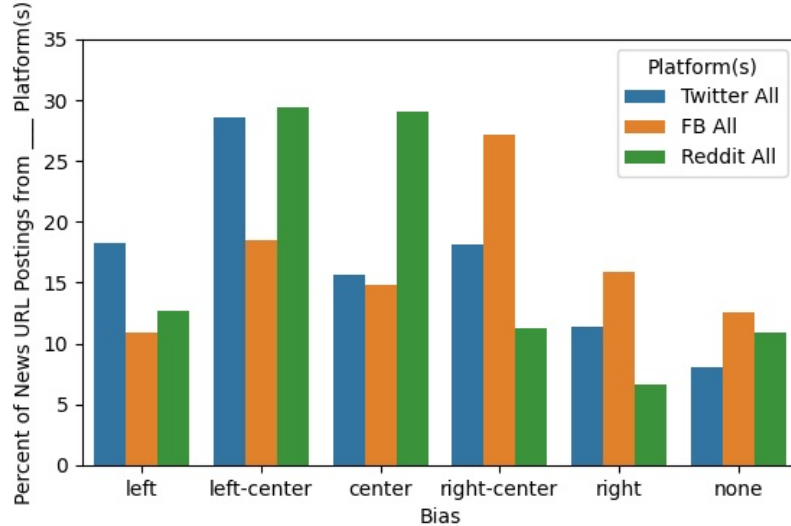
Fig. 3. Breakdown of political bias of news URL postings for each platform.

As for which news sources garnered more net positive user reactions on each platform, Reddit saw the clearest bias with 62% of net positive reactions going to left-leaning URLs, and only 2% of the reactions going to right-leaning URLs. One explanation for this, in addition to the fact that fewer right-leaning than left-leaning URLs were posted in general, is that URLs from left news sources were posted in communities with, on average, five times as many subscribers as were ones from right sources.

Facebook, on the other hand, saw a slightly smaller, but still outsized response to content from right-leaning sources. In particular, 58% of net positive reactions went to right-leaning URLs and 23% went to left-leaning ones. Twitter, however, had user reactions proportional to the number of URLs shared in each bias category.

3.2   Cross-Platform Analysis

Cross-Platform Linking. Within our dataset, there was significant cross-platform linking. In particular, Twitter and Facebook saw the largest amount of linking, with Facebook users posting 10,877 URLs linking to Twitter content and Twitter users posting 5,441 URLs to Facebook content. The cross-platform linking to Reddit was somewhat weaker, with Twitter users posting 822 URLs and Facebook users posting 59 URLs to Reddit content. Reddit users posted 5,196 URLs linking to Twitter content and 136 linking to Facebook content.

Users also posted URLs linking to content on other mainstream social media platforms, such as Instagram 9,168 times and TikTok 3,947 times. Lower moderation platforms had a strong presence in this dataset as well with 14,843 URLs linking to Parler, 3,045 linking to Gab, and 1,042 linking to CloutHub.

Some of the main reasons why users linked to other social media platforms within our dataset were to share links to protest organizing and planning information, to provide alternative ways to view censored content, to cite news or evidence of fraud from other platforms, and to promote accounts or membership on alternative platforms. These results are consistent with the Election Integrity Partnership's findings [8]. We also found posts that discussed suspicious behavior or censorship occurring on other platforms, such as Facebook's banning of a "Stop the Steal" group in Nov. 2020.

Cross-Platform Content. Content that appeared on multiple platforms also gives insight into how information spread between platforms. Overall, we found over 59K URLs that appeared on at least two of the platforms. Again, the strongest relationship was between Twitter and Facebook, which share 51,194 URLs. Twitter and Reddit share 15,412 URLs and Facebook and Reddit share 9,359 URLs. Out of those links, 7,888 appeared on all three platforms.

These URLs are mostly news, followed by social media and media sharing. However, the social media URLs were more likely, on average, to be posted multiple times compared to the news ones. In terms of the political bias of the news URLs, those that were posted on Twitter and Facebook were fairly balanced but all others had a left-leaning bias.

Timing of Cross-Platform Content. To understand how the cross-platform content spread between platforms, we measured the time difference between first appearances of the URLs for each pair of platforms. For URLs appearing on both Twitter and Facebook, we found that 12% of them first appeared on both platforms within a minute of each other and 81% appeared on the second platform within 24 hours of appearing on the first platform (see Fig. 4). In general, more URLs were first posted on Twitter and then Facebook with a median time difference of about 14 minutes. Additionally, 21% of URLs first appeared on Facebook at least an hour before appearing on Twitter, and 41% of URLs first appeared on Twitter at least an hour before appearing on Facebook.
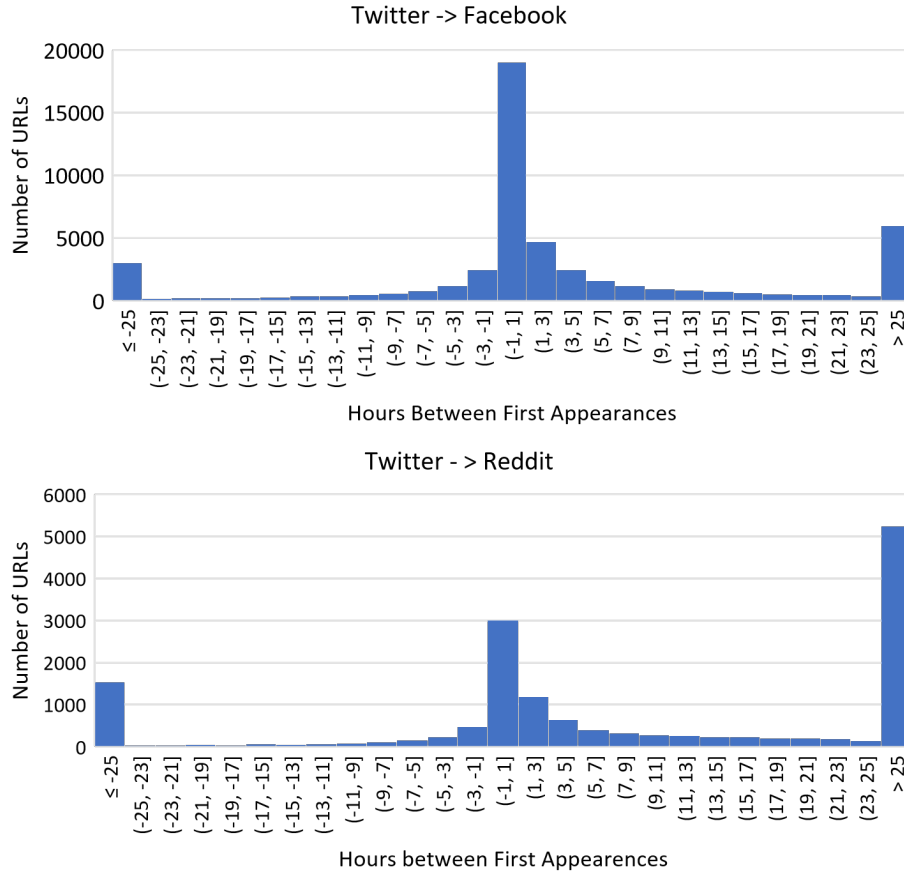
Fig. 4. Histogram of times (in hours) between when URLs first appeared on Twitter until they appeared on Facebook or Reddit. Negative time differences mean that the URLs appeared first on Facebook or Reddit while positive time differences indicate that the URLs appeared first on Twitter.

In order to understand how the approximately 6,000 URLs that appeared on Twitter and Facebook within a minute of each other were spread, we identified the pairs of Twitter and Facebook users responsible for those occurrences. Overall, there were 2,242 unique pairs of users. Out of those pairs, 800 were responsible for multiple instances of the quickly appearing URLs and in total, accounted for 75% of the 6,000 instances. Many of these pairs of accounts were news organization or opinion related accounts. The top pairs are shown in Table 4. Additionally, 33 of the multiple URLs pairs of accounts had at least one account linking to the other in their user descriptions. This suggests that much of the cross-platform spread that occurred in quick succession between Twitter and Facebook was intentional and/or coordinated.

Table 4. Top Pairs of Facebook and Twitter Accounts Responsible for URL Appearances within 1 minute.

Twitter User	Facebook User	# of URLs
TruthSeeker____	thetruthseeker1	102
JohnstonShow	JohnstonShow	100
ConservNewsDly	ConservativeNewsDly	95
RandyMBell	TruthForTheTimes	66
OANN	OneAmericaNewsNetwork	54
ARedPillRep0rt	ARedPillReport	45
EagleEdMartin	EdMartinLive	44
CitizensJourn	citizensjournalVC	39
DFL3CD	DFL3CD	38
KTULNews	NewsChannel8Tulsa	33

Focusing on the cross-platform URLs appearing on Twitter and Reddit instead, we found that there was a stronger tendency for URLs to appear on Twitter and then spread to Reddit (see Fig. 4). The median time difference of 4.9 hours reflects a generally longer lag between first appearances than was found with Twitter and Facebook. We found that only 2% of URLs appeared on both platforms within one minute of each other and only 55% were within the same day. Additionally, 34% of URLs took longer than 24 hours to appear on Reddit after appearing on Twitter. This supports what we found earlier in our temporal analysis of the fraud and protest-related posts.

Finally, for URLs on both Facebook and Reddit, we found that they tended to appear on Facebook first, but that the lag was not generally as long as between Twitter and Reddit, as the median lag was about 30 minutes. 65% of URLs made first appearances on Facebook and Reddit within 24 hours of each other, 30% of URLs were posted on Reddit more than an hour before Facebook, and 47% appeared on Facebook at least an hour before Reddit.

In order to look into the time differences between when the cross-platform URLs had their peak usage in each platform, we identified the most popular

posting period for each URL and platform. To do so, for each URL, we found its peak 24 hours of being posted on each platform. If the peak 24 hours did not consist of at least 2 posts, the URL/platform combination was removed from this analysis. For the remaining URLs, we found the hour within the peak 24 hours with the most posts and used that hour as the representative most popular time. Then, for all URLs with representative times for multiple platforms, we calculated the difference in most popular posting times for each platform pair.

In general, the difference in most popular posting times tended to fall within 24 hours of each other (81% of URLs on Twitter and Facebook, 70% of URLs on Twitter and Reddit, and 74% of URLs on Facebook and Reddit). Additionally, between 14-20% of URLs for each pair of platforms had peaks at the same hour. Interestingly, while slightly more URLs peaked on Twitter before Reddit, more URLs peaked on Reddit before Facebook. More URLs also peaked on Twitter before Facebook, though these variations were largely only a few hours apart.

4   Conclusion

Based on the presented results, we found that Twitter, Facebook, and Reddit not only differ in their post lengths and user reactions, but also their temporal posting behaviors, types of URL content, and the political bias of their users to news content. Additionally, there was a significant level of cross-platform referencing and URL posting, with Twitter and Facebook having the most reciprocal relationship. In terms of the spread of cross-platform URLs, they tended to appear first on Twitter and then spread to the other platforms, but they often peaked on the same day.

Although these findings are limited insofar as the dataset only contains public data and, in the case of Facebook, primarily reflects popular and well-followed accounts, they suggest general trends for further investigation and insights for developing multi-platform models.

References

1. Auxier, B., Anderson, M.: Social Media Use in 2021. Pew Research Center (2021)
2. Tucker, J. A., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., Nyhan, B.: Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. (2018) http://dx.doi.org/10.2139/ssrn.3144139
3. Carley, K. M., Cervone, G., Agarwal, N., Liu, H.: Social Cyber-Security. In: Bisgin, H., Hyder, A., Dancy, C., and Thomson, R. (eds.) International Conference SBP-BRiMS 2018, LNCS, vol. 10899. pp. 389–394. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93372-6_42
4. Bossetta, M.: The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election. Journalism & Mass Communication Quarterly. **95**(2), 471–496 (2018)
5. Lin, H., Qiu, L.: Two Sites, Two Voices: Linguistic Differences between Facebook Status Updates and Tweets. In: Rau, P.L.P. (eds.) Cross-Cultural Design, Cultural Differences in Everyday Life, CCD 2013, LNCS, vol. 8024. Springer, Berlin, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39137-8_48

6. Papakyriakopoulos, O., Medina Serrano, J. C., 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

7. Velásquez, N., Leahy, R., Johnson Restrepo, N., Lupu, Y., Sear, R., Gabriel, N., Jha, O., Goldberg, B., Johnson, N.F.: Hate multiverse spreads malicious COVID-19 content online beyond individual platform control. arXiv e-prints (2020)

8. Center for an Informed Public, Digital Forensic Research Lab, Graphika, Stanford Internet Observatory: The Long Fuse: Misinformation and the 2020 Election. Stanford Digital Repository: Election Integrity Partnership. v1.2.0 (2021)

9. Yin, L.: SMAPPNYU/urlExpander. (2018). https://doi.org/10.5281/zenodo.1345144

10. URL X-ray, http://urlxray.com/. Last accessed 27 May 2021

11. Otterbein, H.: Biden overtakes Trump in Pennsylvania. Politico (2020)

12. Speare-Cole, R.: Joe Biden Taking Georgia Lead Sparks Wave of Memes as Democrats on Course to Flip State Blue. Newsweek (2020)

13. Inskeep, S.: Timeline: What Trump Told Supporters For Months Before They Attacked. NPR (2021)

14. Khalil, A., Freking, K.: Thousands rally behind Trump, insisting he won race he lost. Associated Press (2020)

15. AllSides Media Bias Ratings, https://www.allsides.com/media-bias/media-bias-ratings. Last accessed 24 May 2021