Contents lists available at ScienceDirect

Online Social Networks and Media

journal homepage: www.elsevier.com/locate/osnem

Identifying cross-platform user relationships in 2020 U.S. election fraud and protest discussions

Isabel Murdock [a,*], Kathleen M. Carley [b], Osman Yağan [a]

[a] Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, United States of America
[b] Computer Science, Carnegie Mellon University, Pittsburgh, PA, United States of America

ARTICLE INFO

Keywords:
Social media networks
Multi-platform
Election

ABSTRACT

Understanding how social media users interact with each other and spread information across multiple platforms is critical for developing effective methods for promoting truthful information and disrupting misinformation, as well as accurately simulating multi-platform information diffusion. This work explores five approaches for identifying relationships between users involved in cross-platform information spread. We use a combination of user attributes and URL posting behaviors to find users who appear to purposely spread the same information over multiple platforms or transfer information to new platforms. To evaluate the outlined approaches, we apply them to a dataset of over 24M social media posts from Twitter, Facebook, Reddit, and Instagram relating to the 2020 U.S. presidential election. We then characterize and validate our results using null model analysis and the component structure of the user networks returned by each approach. We subsequently examine the political bias, fact ratings, and performance of the content posted by the identified sets of users. We find that the different approaches yield largely distinct sets of users with different biases and content preferences.

1. Introduction

The increasing number and diversity of available social media platforms have given users more ways to connect with others and share information. Social media users now have a wide selection of platforms to choose from, each providing 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 work [2], especially within the growing field of social cybersecurity [3]. Researchers have characterized and compared the activity of 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, some researchers investigated the multi-platform spread of COVID-19 related misinformation [6,7].

To understand how information spreads over multiple social media platforms, it is useful to be able to identify and characterize users who are actively supporting such information diffusion. With this in mind, our paper focuses on three main goals: (i) to identify social media users on different platforms who appeared to purposely spread the same information over multiple platforms or transfer information between platforms, (ii) to characterize the political bias and fact ratings of the users engaged in multi-platform behaviors, and (iii) to compare the performance of URLs spread by the identified users to those in the complete dataset.

To tackle these objectives, we first outline five approaches for identifying cross-platform pairs of users, each involved in the diffusion of URL-based content over multiple platforms. We subsequently collect a dataset of 24M posts relating to election fraud and election-related protests surrounding the 2020 U.S. presidential election from Twitter, Facebook, Reddit, and Instagram. These platforms were selected for this case study since they are widely used by the American public [1], have been used in previous disinformation and influence campaigns [8,9], and were home to discussions about election fraud and election-related protests during the 2020 U.S. election [10]. Furthermore, the platforms provide a mix of different posting mechanisms and social structures, which allows us to compare the results of the five user pair identification approaches over a diverse set of platforms.

Once we identify cross-platform user pairs from the collected dataset, we perform null model analysis and analyze the component structures and overlaps of the identified user to user networks to characterize and validate the results of each identification approach.

[*] Corresponding author.
E-mail addresses: iem@andrew.cmu.edu (I. Murdock), kathleen.carley@cs.cmu.edu (K.M. Carley), oyagan@andrew.cmu.edu (O. Yağan).

https://doi.org/10.1016/j.osnem.2023.100245
Received 30 June 2022; Received in revised form 13 February 2023; Accepted 14 February 2023
Available online 2 March 2023

We then investigate the types of content, political biases, fact ratings, and performances of the URLs promoted by the user pairs identified through each approach to compare the content they each posted in a multi-platform context.

Building on prior work regarding cross-platform user identification and coordination detection, this work explores multiple approaches for identifying cross-platform user relationships within the context of the 2020 U.S. presidential election and provides insight into the types of users identified through different cross-platform behaviors. It contributes to the study of information flow over multiple platforms by identifying and analyzing the users who contribute to this spread, ultimately supporting the development of effective methods for promoting truthful information and limiting the spread of misinformation in our increasingly complex social media environment.

The rest of the paper is organized as follows. We first describe prior findings related to multi-platform information diffusion and previously proposed methods to identify both individuals across multiple platforms and coordinated groups of users. We then explain our research design, including the approaches used to identify cross-platform pairs of users, the application of these approaches to the 2020 U.S. election dataset, and the methods used to evaluate the results. We subsequently present the results of our case study, characterize the users identified, and discuss the limitations of our approach.

2. Related work

2.1. Multi-platform information diffusion

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 multiple social media platforms, there has recently been an increase in studies looking into multi-platform information spread. Concerning the spread of misinformation, work has been done to define impact indicators and found that the origin of highly reposted information can impact its likelihood to spread across multiple platforms [11]. Other research has focused more specifically on the multi-platform spread of misinformation during natural disasters [12], as well as conspiracy theories relating to the COVID-19 pandemic [6].

The postings of COVID-19 conspiracy-related URLs have been used to analyze the effectiveness of content moderation done by Twitter, Facebook, and Reddit. This work 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 [6]. 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 how Twitter and YouTube were used 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 [14]. Another study, which focused on the Russian disinformation campaign during the 2016 U.S. election, investigated Internet Research Agency (IRA) activity on Twitter, Facebook, and Reddit and suggested that the IRA may have used Reddit to test out content before spreading it on Twitter [8]. Meanwhile, an analysis of the IRA’s use of Twitter and YouTube found that the group relied heavily on YouTube for spreading news and other information, particularly from conservative sources [15].

In addition to how multiple platforms can be used to conduct disinformation campaigns, the cross-platform connections between online hate communities can also be relevant for understanding multi-platform social media interactions. In particular, work analyzing the spread of malicious COVID-19 content over multiple mainstream and less-moderated platforms has shown that hate communities effectively funnel blocked content through less-moderated platforms to avoid detection and moderation. These cross-platform connections make it harder for social media platforms to get rid of fake or harmful information and can obscure the actual level of such content on the platforms [7].

2.2. Identifying cross-platform and coordinated users

An obvious way that information can spread between social media platforms is through individuals who have accounts on multiple platforms. Due to its potential use for understanding this, as well as other applications in areas such as marketing and recommendation services, various frameworks have been proposed for identifying individuals across platforms over the past decade. The main approaches for linking users across platforms include using user attributes, social network relationships, posting behaviors, or a combination of these features to match the most similar accounts across different platforms [16]. User attributes include public profile information, such as display names and biographies. Social network relationships depend on the platform being analyzed and the types of relationships supported by the platform. Posting behaviors refer to the content that users post on different platforms.

Some of the earliest approaches for matching users on different social networks involved matching users based on usernames [17]. Multiple studies have found that social media users tend to prefer to use a main username across multiple platforms [17,18]. This allows usernames to serve as a valuable tool for user identification across platforms. However, while users often select similar usernames, they may use different symbols within their names, such as an underscore on one platform and a period on another. Some of the simple metrics used to match usernames or display names of users across different platforms include Levenshtein distance, Jaro Wrinkler distance, cosine similarity, and Jaccard similarity [19]. Supervised learning approaches, using features based on users’ names and profiles, have achieved high rates of accuracy in linking users across platforms but require labeled training data which can be difficult to acquire [20,21].

Additional efforts have combined username matching with other features to achieve higher performance. For example, a study involving matching users on different tagging platforms found that incorporating the tags that users posted with the purely username similarity approach improved the performance by approximately 9% [22]. Similarly, an approach that combined both user profile information and the ego networks of the users achieved 10% better performance in matching users across platforms than existing methods [23]. These findings highlight that while usernames and other profile information can be helpful for identifying users across platforms, incorporating other information, such as social network structures or posting behaviors, can lead to more reliable results.

A separate but related issue is identifying coordinated accounts across multiple platforms. While the anti-White Helmets and 2016 election IRA work discussed earlier demonstrate how coordinated disinformation campaigns have leveraged multiple platforms in their attacks, they do not address the process of identifying coordinated accounts [8,14,15]. In fact, much of the research in this area has been limited to identifying coordination on a single platform. One proposed method for identifying coordinated accounts relies on behavioral traces or common actions, such as reposting the same content or sharing the same URLs [24]. For example, shared retweeting patterns have been used to identify anti-White Helmet accounts [25], and shared URL posting behaviors have been used to differentiate *organic* URL posting relationships from *organized* ones [26]. Furthermore, imposing a time constraint on the shared behavior allows for identifying synchronized actions and was used to find accounts working together in the Reopen America Movement [27]. All three of these examples focused on identifying coordinated accounts on Twitter.

2.3. 2020 election research

Regarding the 2020 U.S. presidential election, an analysis of Twitter data collected before the election, between June and September, examined user engagement with disinformation and information from conspiracy groups such as QAnon and the role bots played on the platform [28,29]. Another pre-election study based on data collected between January and March looked at Reddit and 4chan. It found that while partisan information and "fringe perspectives" were prevalent in political discussions, users of these platforms avoided posting links to sites that produce low-quality, algorithmically generated misinformation. Instead, YouTube, and particularly alternative news and commentary channels on the platform, played a significant role in amplifying misinformation [30].

As for analysis conducted in the election and post-election time frame, the Election Integrity Partnership's 2020 Election Report provided in-depth insight into how a wide range of social media platforms was used to spread election-related misinformation. It also provided details about the cross-platform nature of how different narratives grew and stayed alive [31]. Furthermore, the report documented how social media users leveraged platforms' specific features and moderation policies, or lack thereof, to spread content as effectively as possible. Additionally, research exploring the impact of content moderation on former President Trump's election-related tweets found that when Twitter blocked his messages, they ended up being posted more often and receiving more attention on other platforms [32].

Concerning the January 6th protests, in particular, a recent cross-platform study involving Twitter and Parler found that similar narratives were discussed on both platforms [33]. However, the external content linked to by users on each platform was somewhat dissimilar, and users who posted the same URLs were more likely to share similar names. This research suggested that combining social media artifacts, such as external URLs, with user expressions, such as usernames, could be a useful in studying cross-platform information diffusion and community dynamics.

Our work builds on this prior research by outlining five approaches for identifying users engaged in cross-platform behaviors, leveraging both user attributes and posting behaviors. We then apply these approaches to a set of 24M social media posts collected from four platforms and study the cross-platform user networks produced by each approach. Rather than characterizing how particular stories were discussed or providing an aggregate comparison of information diffusion across the platforms, this work offers an analysis of how certain types of identified users spread information over multiple platforms. We leverage prior multi-platform work by employing some of the existing strategies for identifying users across platforms and using previously proposed URL posting and propagation metrics to characterize the content spread by the identified users.

3. Research approach

First, we describe the five approaches used for identifying users engaged in cross-platform behaviors. The approaches differ in either the means by which they identify cross-platform pairs of users or in the types of relationships they aim to identify. Rather than using a combination of these approaches, we explore each one individually to compare the user networks obtained through each method and the types of content each group shared. Additionally, through null model comparisons and network structure analysis, we provide validation of the user pairs returned by each approach.

Once the approaches for cross-platform identification are discussed, we describe the 2020 U.S. election dataset. This data is combined with political bias and credibility ratings to ultimately characterize the types of content posted by the identified users and compare their biases and credibility preferences across approaches. Finally, we measure the performance of the URLs shared by each set of cross-platform users.

3.1. Approaches for identifying potential cross-platform content spreaders

The primary goal of this work is to investigate multiple approaches for identifying users engaging in cross-platform behaviors. These approaches leverage different activities that social media users may engage in to have an impact across multiple platforms or to spread content between platforms. Each approach identifies pairs of users exhibiting certain behaviors and, in doing so, constructs a user to user network where the edges reflect the cross-platform behavior of interest.

By using multiple approaches, we can find users based on behaviors that might have different levels of prevalence on different platforms. For example, the use of hashtags is less common on Reddit than Twitter, so approaches using synchronized actions involving hashtags may have limited use in identifying coordination involving Reddit users. Similarly, while users on Twitter and Facebook often have an interest in using recognizable usernames and display names, anonymity is a central feature of Reddit. Consequently, identifying users across platforms based on similar names may be less effective in this case.

In addition to the shared name and synchronized action approaches, we leverage the URL-posting behaviors of users across the different platforms to identify users who repeatedly introduce the same content to their respective social networks. Additionally, while some of these approaches aim to find the same individual or organization across platforms, others only attempt to identify users acting in multi-platform ways, such as repeatedly cross-posting content from one platform to another (see Table 1). These types of one-way relationships can be beneficial for understanding how content may be intentionally spread to new platforms and therefore is of interest to us. Below, we introduce the five user pair identification approaches.

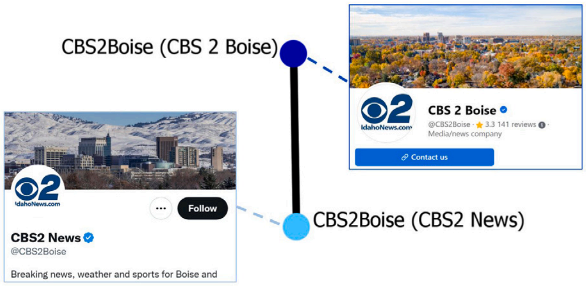
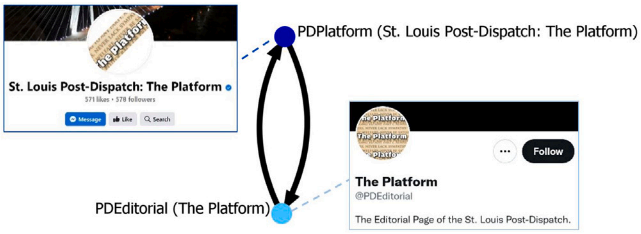
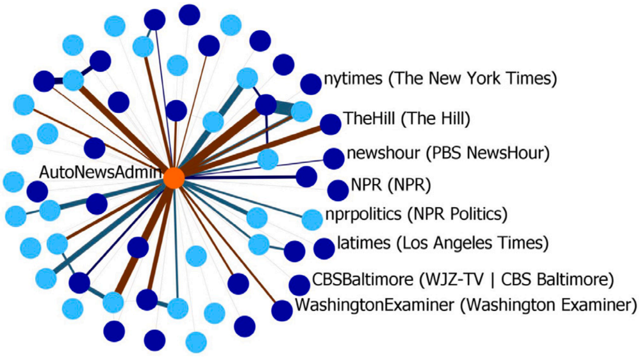
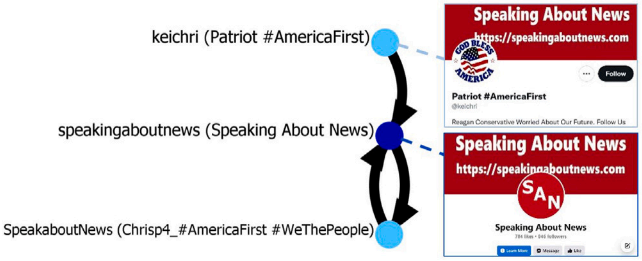
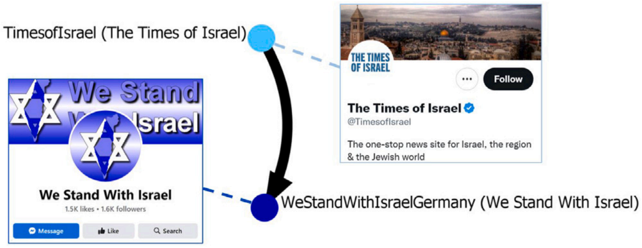
3.1.1. Same name users

As discussed earlier, an established and intuitive approach for identifying users across platforms relies on quantifying the similarity between usernames or display names across platforms. Prior studies have shown that people tend to reuse the same main username across platforms [17,18]. It is also common to use slight variations of their usernames or display names.

To consider the users that share similar names across platforms, we identify pairs of users on different platforms that both posted the same URLs and had a Levenshtein distance of ≤ 1 between their usernames or display names. In other words, these are users who posted the same content on different platforms and have very similar names associated with their accounts. We use Levenshtein distance, which measures the minimum number of insertions, deletions, or substitutions needed to transform one string into another, and disregard any comparisons involving names less than three characters long [33].

Additionally, to account for slight variations in their names across platforms, we calculate the distance of both the users' original usernames and display names, as well as tokenized and standardized versions of their usernames and display names. To standardize the names, we first split the names on commas, underscores, and periods and then recombine the tokens alphabetically. If the smallest Levenshtein distance from the four comparisons is ≤ 1, and the users posted the same URLs, then we conclude that the pair of cross-platform accounts likely represented the same individual or organization. The identified pair of cross-platform users are then added to the same name users network

Table 1
Use cases of user identification approaches.

Type of users	Relevant attributes/behaviors	Types of relationships	Example
Same name users	Have similar usernames and/or display names and posted the same URLs	Same individual or organization across multiple platforms	 CBS2Boise (CBS 2 Boise) CBS 2 Boise @CBS2Boise · 3.3 141 reviews · Media/news company Contact us CBS2 News @CBS2Boise Follow Breaking news, weather and sports for Boise and CBS2Boise (CBS2 News)
Bidirectional introducers	Introduced the same URLs to their respective platforms with each user posting before the other in multiple instances	Same individual or organization across multiple platforms	 PDPlatform (St. Louis Post-Dispatch: The Platform) St. Louis Post-Dispatch: The Platform 571 likes • 578 followers Message Like Search The Platform @PDEditorial Follow The Editorial Page of the St. Louis Post-Dispatch. PDEditorial (The Platform)
Repeat introducers	Introduced the same URLs to their respective platforms multiple times	Users across platforms that are interested in introducing the same content to their networks including in both bot-like and 2-way relationships	 nytimes (The New York Times) TheHill (The Hill) newshour (PBS NewsHour) NPR (NPR) nprpolitics (NPR Politics) latimes (Los Angeles Times) CBSBaltimore (WJZ-TV | CBS Baltimore) WashingtonExaminer (Washington Examiner) AutoNewsAdmin
Synchronized users	Made multiple posts containing the same URLs and hashtags/filter terms within 5 min of each other	Users across platforms who may be coordinating to post the same URLs and use similar language/messaging	 keichri (Patriot #AmericaFirst) Speaking About News https://speakingaboutnews.com Patriot #AmericaFirst Follow Reagan Conservative Worried About Our Future. Follow Us speakingaboutnews (Speaking About News) Speaking About News https://speakingaboutnews.com Speaking About News SpeakaboutNews (Chrisp4_#AmericaFirst #WeThePeople)
Cross-platform linkers	Repeatedly linked to posts made by a specific user on a different platform	Users who want to promote or boost the social media content of a user from a different platform	 TimesofIsrael (The Times of Israel) THE TIMES OF ISRAEL Follow The Times of Israel @TimesofIsrael The one-stop news site for Israel, the region & the Jewish world We Stand With Israel We Stand With Israel 1.5K likes • 1.6K followers Message Like Search WeStandWithIsraelGermany (We Stand With Israel)

Legend ● - Twitter User ● - Facebook User ● - Reddit User

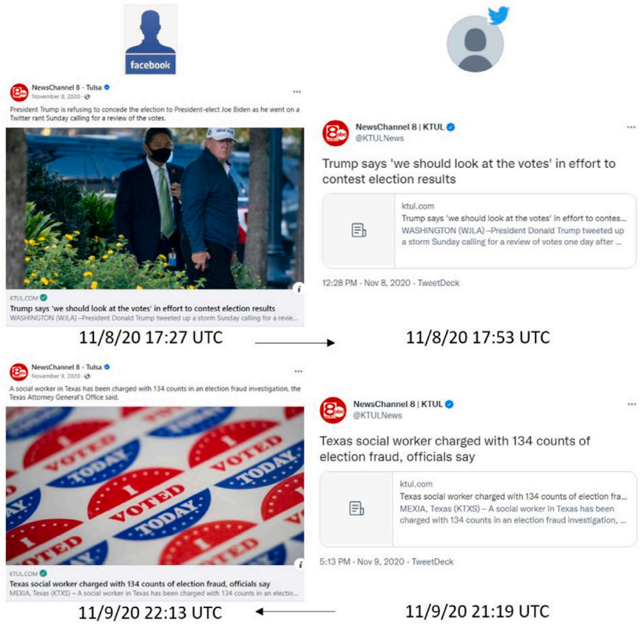
if they were not previously, and an edge is added between the two identified users.

By identifying users who exhibited both name matching and shared URL posting behaviors, this approach leverages prior findings that combining user attributes with posting activities can lead to more accurate and robust results [22,23]. Throughout the remainder of the paper, users identified through this method, and the resulting user to user network, will be referred to as “same name users” and the “same name users network”, respectively.

3.1.2. Bidirectional introducers

Similar to the same name users, this approach aims to find pairs of cross-platform users that are the same individual or organization across different platforms. To do so, it relies on user posting behaviors. More specifically, it focuses on the users that introduce new URLs to their respective platforms.

To be considered an “introducer”, the user must be the first user in the dataset to post a given URL on their respective platform. From this



NewsChannel 8 - Tulsa
November 8, 2020 ·

President Trump is refusing to concede the election to President-elect Joe Biden as he went on a Twitter rant Sunday calling for a review of the votes.

KTUL.COM
Trump says 'we should look at the votes' in effort to contest election results
WASHINGTON (WJLA) –President Donald Trump tweeted up a storm Sunday calling for a revie...

11/8/20 17:27 UTC

NewsChannel 8 | KTUL
@KTULNews

Trump says 'we should look at the votes' in effort to contest election results

ktul.com
Trump says 'we should look at the votes' in effort to contes...
WASHINGTON (WJLA) –President Donald Trump tweeted up a storm Sunday calling for a review of votes one day after ...

12:28 PM · Nov 8, 2020 · TweetDeck

11/8/20 17:53 UTC

NewsChannel 8 - Tulsa
November 9, 2020 ·

A social worker in Texas has been charged with 134 counts in an election fraud investigation, the Texas Attorney General's Office said.

KTUL.COM
Texas social worker charged with 134 counts of election fraud, officials say
MEXIA, Texas (KTXS) – A social worker in Texas has been charged with 134 counts in an electio...

11/9/20 22:13 UTC

NewsChannel 8 | KTUL
@KTULNews

Texas social worker charged with 134 counts of election fraud, officials say

ktul.com
Texas social worker charged with 134 counts of election fra...
MEXIA, Texas (KTXS) – A social worker in Texas has been charged with 134 counts in an election fraud investigation, ...

5:13 PM · Nov 9, 2020 · TweetDeck

11/9/20 21:19 UTC

Fig. 1. Example of the bidirectional property of the bidirectional introducers. Here, the Facebook account `NewsChannel 8 - Tulsa` introduced a URL on Facebook on Nov. 8 before the Twitter account `NewsChannel8|KTUL` introduced it on Twitter. Then, on Nov. 9, the Twitter account introduced a different URL before the Facebook account. This behavior would need to occur in both directions multiple times in order for the pair of users to be classified as bidirectional introducers, while only one direction is required for them to be repeat introducers.

set of introducers, we identify pairs of cross-platform users who introduced the same URLs as each other multiple times. We then consider the temporal order in which the users of the identified cross-platform pairs introduced each of the URLs. We filter out any relationships in which one user always introduced the URLs before the other. This bidirectional condition is made more robust by requiring each user in the relationship to have introduced URLs before the other user multiple times (see Fig. 1). If these conditions are satisfied, the cross-platform pair of users is added to the bidirectional introducers' user to user network, and the weight of the edge connecting them is equivalent to the number of shared URLs they introduced.

This bidirectional condition is enforced as it helps filter out bot-like or unidirectional cases, where one user just repeatedly posts what someone from a different platform posts. The resulting "bidirectional introducers" consist of users who introduced the same content to their respective platforms multiple times in both orders. The intuition behind this approach can be understood through the perspective of news organizations who may write articles and, soon after publishing, post their articles' URLs on their social media accounts across different platforms. However, this type of behavior is not limited to news organizations. One such example is the anti-White Helmets campaign mentioned earlier. The group would create YouTube videos and then post tweets containing URLs linking to the videos, thereby introducing the YouTube content to their Twitter networks through the URLs [14].

The bidirectional condition helps ensure that the identified user pairs are, in fact, the same individual or organization across platforms, thus limiting the chance of having false positives. However, such entities may always introduce content on one platform before the other, whether through automated systems or intentional policies. Due to this, the number of identified bidirectional introducers acts as a lower bound for the full set of users intentionally spreading the same content across multiple platforms. This is part of the motivation for the next approach in our analysis.

3.1.3. Repeat introducers

Using the same concept of introducers as described in the previous approach, this "repeat introducers" approach relaxes the bidirectional condition. This approach only requires the identified users to have repeatedly introduced the same URLs to their respective platforms, regardless of the order in which they introduced the content. Similar to the bidirectional introducers approach, the requirement to have *repeatedly* introduced the same content helps filter out relationships that may have occurred by chance.

Broadening our scope compared to the bidirectional introducers, this approach allows us to also identify users who may have engaged in bot-like cross-platform spreading processes, as well as those who may use automation services or follow posting policies which result in posts consistently appearing on one platform before another. Rather than aiming to identify pairs of cross-platform accounts that are run by the same organization or individual, this approach attempts to identify a wider set of users who are involved in multi-platform content spreading.

3.1.4. Synchronized users

Rather than focusing only on users who *introduced* the same content, this approach considers users who just posted the same content regardless of whether it had already been posted on the given platforms. However, to filter out relationships that are likely to occur out of coincidence from users posting popular URLs, this approach also requires that the users included the same hashtags or key terms within their posts. In doing so, this approach draws on the synchronized action framework proposed for identifying potentially coordinated groups of users [27].

To implement this type of coordinated user detection, we consider pairs of users from different platforms who repeatedly posted the same URLs and hashtags within a 5-min window of each other. More specifically, the identified users must have made posts within 5 min of each other that include both the same URL and the same hashtags or filter terms, and they must have done this multiple times. If this is the case, then the cross-platform user pair is added to the synchronized users network, and an edge is added between the users with a weight equivalent to the number of posts the coordination was detected.

3.1.5. Cross-platform linkers

Unlike the previous four approaches that rely on shared URL-posting behaviors, this approach uses direct linking between the cross-platform users. The "cross-platform linkers" are identified by building a directed network of cross-platform user pairs who repeatedly linked to social media posts made on different platforms by the same user. Therefore, these users repeatedly make posts on their platform that link to a different platform user's content.

Similar to the cross-platform user pairs returned by the repeat introducers approach, we expect that the pairs returned by this approach will involve both one-way and two-way relationships. For example, if there is a user, u_1, on platform A who follows someone, u_2, on platform B and repeatedly makes posts on platform A that contain links to u_2's posts on platform B, then u_1 and u_2 would be returned as a cross-platform user pair by this approach. However, this would not indicate that u_1 and u_2 were the same person or organization, but instead that there was a cross-platform content spreading relationship between the two users. Furthermore, we expect that popular users with larger audiences are more likely to be included in the pairs identified by this approach since their posts are more likely to be cross-posted to other platforms.

An important note about these five approaches for identifying cross-platform pairs of users is that they do not yield mutually exclusive sets of user pairs. Instead, they identify users who appear to intentionally spread information to new platforms or be working together to spread information over multiple platforms simultaneously. Furthermore, multi-platform user pairs that fall under multiple approaches may be more likely to be users acting intentionally to spread information over multiple platforms.

3.2. Validation through null models

A significant challenge in evaluating the success of the five approaches above for identifying users with cross-platform behaviors is the lack of ground truth to validate the observed results. Additionally, three of the approaches rely solely on shared posting behaviors to identify users, meaning they may identify users based on coincidences in posting orders and the temporal nature of URL content.

To evaluate how the number of users and relationships identified by the bidirectional introducers, repeat introducers, and synchronized users approaches compares to those we may expect to arise from chance alone, we perform a null model analysis. Drawing on null model theories used for hypothesis testing in the context of social networks, we perform pre-network data resampling to generate our random values for each approach [34,35].

Three different null models are constructed, with increasing levels of similarity to our observed dataset, yet they all aim to preserve the temporal characteristics of URLs and content creation. For this reason, the post timing and content used in the models remain consistent with our collected dataset. However, the authors of the posts are randomly assigned. Once the assignment process is complete, we can identify the bidirectional introducers, repeat introducers, and synchronized users in the randomized datasets. This then allows us to measure the number of such users and relationships we may expect to occur if the users had posted independently from each other across the platforms and the approaches only identified relationships that arose by chance.

3.2.1. Uniform null model

The first null model we construct is the simplest. It assumes that all users from a given platform in our dataset are equally likely to have made a post. The following process is performed to simulate this situation and produce the associated random values:

(1) An empty pool of users is created for each platform.
(2) We loop through every URL-containing post in the original dataset and add the post's author to the user pool associated with the post's platform if the author is not already in the user pool.
(3) The resampling is performed by looping through each post in the original dataset, noting the platform the post was made on, and assigning a new author to the post by selecting uniformly at random (with replacement) a user from the respective platform's user pool.
(4) The bidirectional introducers, repeat introducers, and synchronized users are identified in the randomized dataset.
(5) The number of nodes (users) and edges (cross-platform user pairs) identified by each approach are recorded.

Steps 3–5 in the above process are repeated 1000 times to produce a null model distribution to compare with our observed numbers of users and cross-platform user pairs. While simple, this null model allows us to simulate the situation in which the users of a given platform are equally likely to have been the author of a given post and act independently across the platforms. Consequently, we can measure how likely it would be to identify the number of cross-platform user pairs we observed in the real-world dataset in this setting where no actual cross-platform activity exists.

3.2.2. Proportional null model

Rather than assuming that all users are equally likely to be the author of a given post, the proportional null model considers the number of posts made by each user in the dataset. In effect, the users are assigned randomly to posts with probabilities proportional to the number of times they made posts in the observed dataset. To implement this model, we use a similar process as the uniform null model:

(1) An empty pool of users is created for each platform.
(2) Users are added to the pools such that a user i is added n_i times to the pool of the platform it is a member of, where n_i is the number of URL-containing posts that user i posted.
(3) The resampling is performed by looping through each post in the original dataset, noting the platform the post was made on, and assigning it a new author by randomly selecting, with replacement, a user from the pool associated with the post's platform.
(4) The bidirectional introducers, repeat introducers, and synchronized users are identified in the randomized dataset.
(5) The number of nodes (users) and edges (cross-platform user pairs) identified by each approach are recorded.

Again, steps 3–5 are repeated 1000 times to collect the null model distribution. By resampling the post authors with probabilities proportional to the number of posts made by each user, this method allows us to preserve the activity levels of each user and simulate the setting where some users are more likely than others to make posts. However, this process still results in a dataset in which the authors of the posts have been randomized, and the users' assignment to specific URLs is independent across the platforms.

3.2.3. Introducers proportional null model

Taking the proportional model one step further, we now distinguish between cases where users *introduced* the URLs and where they posted a URL that had already appeared in the dataset. To do this, we use the following procedure:

(1) Create two sets of empty user pools for each platform, one for URL-introducing posts (i.e., "introducers user pool") and the other for non-introducing URL posts (i.e., "non-introducers user pool").
(2) For each URL-containing post in the dataset, we determine if it is the first time the URL appeared in our dataset on the given platform. If it is the first time, we add the post's author to the platform's introducers user pool. If it is not, we add the post's author to the platform's non-introducers user pool. Similar to the user pool created for the proportional null model, the number of times a user is in a respective pool is proportional to the number of times they made posts introducing a URL or the number of times they made a post containing a URL that had already been introduced.
(3) The resampling is performed by looping through each post in the original dataset and noting the platform the post was made on and whether it contained a URL that had not appeared on the platform before. If it contained the first instance of the URL, a user was randomly selected, with replacement, from the platform's introducers user pool and assigned to be the author of the post. Otherwise, a user was randomly selected from the platform's non-introducers user pool and assigned to the post.
(4) The bidirectional introducers, repeat introducers, and synchronized users are identified in the randomized dataset.
(5) The number of nodes (users) and edges (cross-platform user pairs) identified by each approach are recorded.

Similar to the other two models, steps 3–5 are repeated 1000 times to get the null model distribution which we compare to our observed number of users and user pairs. By separating out instances where URLs were first posted and the users who introduced the content, this model reflects the setting where introducers are unique from other users and are more likely to introduce content than users who never introduced content in our dataset.

We anticipate that out of the three null models, this model's results will be the closest to the observed values. Yet, due to the independence that this model maintains between the assignment of users across different platforms, it allows us to determine if the observed values are significant compared to what would be expected to happen by chance under this setting.

Table 2
Data filter terms.

Election fraud	Election protests
Corrupt election	Do not certify
Dead voters	Maga civil war
Deceased voters	March for trump
Dominion voting systems	March to save america
Election fraud	Million maga march
Election integrity	Saveamerica
Fake election	Save america rally
Fake votes	Stop the fraud
Fraudulent election	Stop the steal
Legal votes only	Wild protest
Legitimate votes only	
Massive corruption	
Rigged election	
Stolen election	
Voter fraud	

Both space-separated and hashtag forms of the filter terms (e.g., "corrupt election" and "corruptelection") were used to collect the data.

3.3. Validation through network component analysis

Two of the approaches for identifying potential cross-platform spreaders aim to find relationships between users who are the same individual or organization across multiple platforms. Due to this, we would expect that the resulting same name users network and bidirectional introducers network would be primarily composed of many small components. While some organizations operate a few different pages on a given platform (e.g., for different offices or sectors of their business), finding large components in the resulting networks would suggest that the same name users and bidirectional introducers approaches were not successful in only identifying individual organizations or entities.

Therefore, for both the same name users network and the bidirectional introducers network, we plot the size distribution of the components in each network and evaluate the presence of small components. Furthermore, we perform this analysis on the networks produced by the other three approaches to provide additional context to our findings and compare the size distributions across all five approaches.

Since the same name users and bidirectional introducers approaches aim to find the same type of user relationships, we also plot the size distribution of the intersection of their networks. We hypothesize that cross-platform user pairs identified by both approaches will be more likely to be the same individual or organization, and therefore we will find a higher percentage of dyads and triads in the intersection network. This analysis of the distribution of component sizes in each network allows us to study whether the structures of the cross-platform user networks align with the types of relationships we expect them to return.

3.4. Overlaps of user nodesets and cross-platform pairs

After considering the structure of the user networks, we measure the overlap between the edges (i.e., cross-platform user relationships) identified in each approach. We do this for two reasons. First, if any of the approaches return extremely similar edge sets, we may be able to conclude that such approaches are redundant and that only one of them needs to be considered. Second, the same name user network can provide validation and context to the other user networks. For example, a large proportion of the bidirectional introducers sharing similar names would help indicate that the cross-platform user pairs returned by this approach are, in fact, the same entities across platforms.

We perform this overlap analysis by finding the percentage of edges from each approach that appears in the other approaches' user networks. Since the approaches return networks with potentially different node sets and sizes, this metric allows us to gauge whether the networks returned subsets of each other. For instance, we know this to be the case between the bidirectional introducers and repeat introducers, as the bidirectional introducers must fulfill the repeat requirement of the repeat introducers.

3.5. Data collection

With the approaches for identifying cross-platform user pairs and the strategies for validating the results outlined, we now describe the dataset collected to explore these approaches through a case study. Since understanding the cross-platform ways in which information spread regarding election fraud and protests during the 2020 U.S. election could be helpful for limiting such spread in the future, we collected posts made regarding those topics on Twitter, Facebook, Reddit, and Instagram between Oct. 1, 2020, and Jan. 19, 2021.

3.5.1. Collection process

We first compiled a list of hashtags and phrases pertaining to election fraud and election-related protests (see Table 2). We included both the hashtags and the space-separated phrase versions of them in our data collection because hashtags are less prevalent on Reddit than on the other platforms. Additionally, since we are not focused on analyzing how these hashtags were discussed but on using them as a means to collect posts that potentially contain multi-platform URLs, using both the hashtags and phrases allows us to collect a larger set of posts across all four platforms.

The election fraud-related filter 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 filter terms are related, they focus more on calls to action, such as "do not certify", and references to marches and rallies. Since our work focuses on discussions about fraud and protests, we do not presume to know the accuracy or credibility of the posts containing the filter terms, just that they were related to fraud claims or protests.

We collected social media posts made between Oct. 1, 2020 and Jan. 19, 2021 that contained case-insensitive versions of the filter terms. The time range includes two months of pre-election discussions before the Nov. 3, 2020 election. It also contains posts made following the election as major fraud narratives developed and protests occurred, including the Jan. 6 Capitol attack.

Tweets were gathered using Twitter's full-archive search and included public posts that were not deleted or removed before the data was collected in Mar. 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. Also collected through CrowdTangle, the Instagram posts come from public accounts with more than 50k followers and verified accounts. Reddit posts and comments were collected using the Pushshift API and include posts and comments made in public subreddits.

Overall, over 23M tweets, 726k Facebook posts, 23k Instagram posts, and 262k Reddit posts and comments containing the filter terms in their titles, text, or URLs were collected.

3.5.2. URL standardization and selection

From these posts, we collected and cleaned the URLs found within them so that URLs linking to the same content, despite being shortened or having slightly different formats, mapped to the same representative URL. This involved first standardizing the URLs' prefixes and then using SMaPP's urlExpander tool [36] to identify shortened ones and expand them when possible. From this set of standardized URLs, we extracted two sets of URLs: multi-platform URLs and cross-platform linking URLs (see Fig. 2).

The multi-platform URLs are URLs that were posted to multiple platforms in our dataset and contained the filter terms within the URLs themselves. To identify these URLs, we first removed the query parameters at the end of the URLs, excluding those linking to YouTube, similar to prior work [33]. We then filtered for only the URLs that contained the filter terms within the URL itself. This requirement of

Table 3
Breakdown of unique multi-platform URLs, postings of those URLs, and users who posted them from each platform.

	Number of unique URLs	Number of postings	Number of users
Twitter	19,661	2,490,056	674,594
Facebook	18,277	111,703	23,830
Reddit	5454	22,178	8852
Instagram	86	129	81

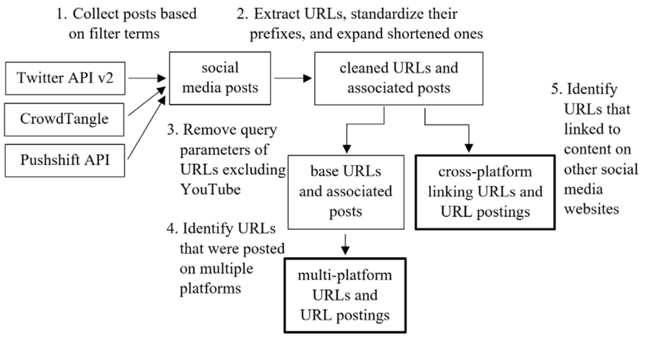

1. Collect posts based on filter terms
2. Extract URLs, standardize their prefixes, and expand shortened ones
Twitter API v2
CrowdTangle
Pushshift API
social media posts
cleaned URLs and associated posts
3. Remove query parameters of URLs excluding YouTube
base URLs and associated posts
cross-platform linking URLs and URL postings
5. Identify URLs that linked to content on other social media websites
4. Identify URLs that were posted on multiple platforms
multi-platform URLs and URL postings

Fig. 2. Process to collect the post data, extract URL postings, and filter for multi-platform and cross-platform linking URLs.

the filter terms appearing in the URLs helped us focus on the URLs that were more likely to be related to the topics of fraud and protests. It also meant that our collected data was more likely to contain all of the instances in which the URLs were posted, assuming we had access to the posts. This is particularly important for our user pair identification approaches which rely on finding the introducers of URLs.

The final set of multi-platform URLs and their associated social media posts contained approximately 19.8k URLs and 2.6M posts. Out of the four platforms, the Twitter dataset contained the most posts and consequently the largest number of posts involving the multi-platform URLs (see Table 3). Additionally, over 90% of the multi-platform URLs appeared on Twitter and Facebook, while approximately 28% appeared on Reddit. Less than 1% appeared on Instagram. This was expected since URLs are harder to post on Instagram and, therefore, less prevalent across the platform.

The second type of URL, cross-platform linking URLs, are URLs that linked to social media platforms other than the one that they were posted on, e.g., a URL in a Reddit post that linked to a tweet. As we are interested in studying social media users who repeatedly linked to content posted by a specific user on a different platform, we only consider URLs that linked to one of our platforms of focus. Overall, there were 25.7k cross-platform linking URLs posted in 72k social media posts (see Fig. 3).

Out of the two collected datasets, the multi-platform URL dataset was used to identify the same name users, bidirectional introducers, repeat introducers, and synchronized users, as those relationships all rely on URL content that was posted on multiple platforms. On the other hand, the cross-platform post linkers were found using the cross-platform linking URL dataset.

3.6. Content classifications

While the cross-platform URLs clearly linked to social media websites, we wanted to explore other types of content linked to by the identified user pairs. This information could give us insight into the types of users or organizations identified by the different approaches. Therefore, the most-posted website domains from the multi-platform URLs were classified into one of the following categories:

(1) **News:** Recognized and reputable news organizations, including international, national, and local outlets, e.g., nytimes.com.

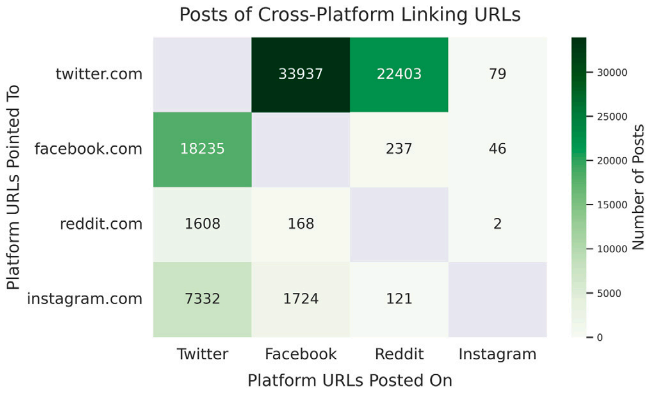


Fig. 3. Number of social media posts on each platform that contained URLs linking to one of the three other social media platforms. This shows that Twitter was the most linked to platform, with the strongest linking relationship occurring between Facebook and Twitter. Reddit was linked to the least but had a significant amount of linking to Twitter.

(2) **Political opinion:** Political blogs and news commentary websites, including news-focused websites with strong partisan perspectives, e.g., bongino.com.
(3) **Entertainment/culture:** News organizations with a focus on culture or entertainment or tabloid journalism outlets, e.g., tmz.com.
(4) **Social media:** Social media websites that connect users and provide platforms for sharing information, e.g., youtube.com.
(5) **Political:** Websites created to support a particular political candidate or advocate for political interests, e.g., donaldjtrump.com.
(6) **Investigative/government:** Websites run by the government, academic institutions, or investigative organizations, e.g., cisa.gov.
(7) **Political Satire:** Outlets dedicated to political satire that publicly disclose their satirical nature, e.g., babylonbee.com.
(8) **Other:** Websites that do not fall into any of the above categories.

To determine which domains to classify, we first calculated the number of times each domain was posted in the full multi-platform dataset. We then manually classified the domains, one by one, in order of decreasing popularity until at least 90% of the URL postings had an associated classification. At that point, we wanted to ensure that a reasonable amount of the content posted by the identified user pairs was classified. To that end, we ranked the remaining unclassified domains in terms of the number of times they were posted by the identified users. We proceeded to classify additional domains from this list in order of decreasing popularity until at least 85% of the URLs posted by the identified user pairs were classified. In total, 1161 website domains were assigned content classifications.

This classification process was intended to provide a rough approximation of the types of content linked to by the multi-platform URLs and the identified cross-platform user pairs. Therefore, only a single coder classified the domains, and consequently, the labels have limited reliability. In the future, additional coders could perform the classification process, and reliability tests could validate the associated results. However, given that these classifications are only used to generally characterize the types of content linked to by multi-platform URLs and that there are over 1k website domains classified, we accept the limitations of this approach for this work.

3.7. Bias and credibility classifications

To further classify the multi-platform URLs according to the political bias and credibility of the websites they linked to and characterize the bias and credibility of the identified user pairs, we used Media

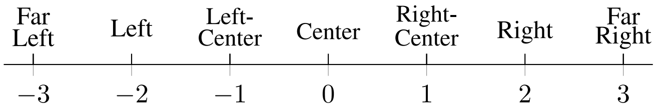


Fig. 4. Mapping of bias labels to bias values.

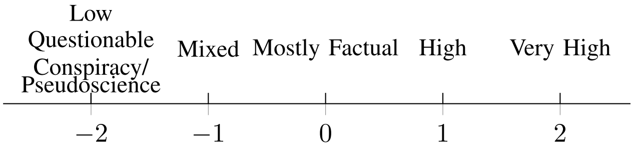


Fig. 5. Mapping of factual rating labels to factual rating values.

Bias/Fact Check (MBFC) website's bias and factual reporting ratings [37]. MBFC is an independent website that provides bias and credibility ratings on hundreds of media sources and organizations. In particular, MBFC's dataset is valuable as it allows us to separate bias from factual accuracy and analyze both the spread of highly biased news and that of low factual accuracy. MBFC's ratings have been employed in previous studies, including ones regarding news sharing in online communities [38], media bias detection [39], and the spread of COVID-19 misinformation [40].

Approximately 62% of the multi-platform URLs in our final dataset linked to websites classified by MBFC. This meant we could label these URLs based on the bias and factual ratings of the websites they linked to, while URLs linking to unclassified websites were left unassigned. The resulting bias categories that appeared in the dataset were: Left, Left-Center, Center, Right-Center, Right, and Far Right. Although the MBFC dataset includes Far Left-classified sources, they did not appear within the multi-platform URL dataset. The fact ratings in the dataset were: Very High, High, Mostly Factual, Mixed, and Low. Since MBFC also provides lists of Questionable and Conspiracy/Pseudoscience websites, we labeled URLs linking to those sources as questionable or conspiracy/pseudoscience-related.

3.7.1. User bias and factual information scores

Using the now-labeled URLs, users could be assigned average bias scores based on the bias of the content they posted. To quantify the political bias of each user, we first took the MBFC-classified URLs that a given user posted and mapped the URLs' bias labels to numerical values. We used the same mapping as defined by Weld et al. [38], where negative values indicate left-leaning biases, and positive values indicate right-leaning biases (see Fig. 4). Once the bias labels were mapped to bias values for the given user, we took the simple average of these values to get the user's **average bias score**.

We also assigned the users factual information scores according to the factual ratings of the content they posted. This involved a similar process of taking the MBFC-classified URLs a given user posted and mapping their factual ratings to factual values. For this mapping, negative values indicate weaker factual ratings, and positive values indicate stronger factual ratings (see Fig. 5). Of note, URLs linking to Questionable or Conspiracy/Pseudoscience sources were given the lowest factual values, along with those linking to Low-rated websites. Once the factual values were collected for the user, we again calculated the simple average to get the user's **factual information score**.

From these scores, we compare the identified user pairs' distributions of political bias and factual ratings. This allows us to determine if the different cross-platform user identification approaches identify user pairs with different content preferences.

3.8. Content performance measures

Finally, to compare how content introduced or posted by the identified cross-platform user pairs performed relative to each other, we used the following URL propagation metrics: number of posts, posting life span, posting speed, and number of lives. Together these metrics help us understand how much the URLs were posted, how long they remained active on the platforms, and how quickly they spread.

(a) Number of Posts. The number of posts is simply the number of times each URL was posted in the dataset.

(b) Posting Life Span. The posting life span is the total time, in days, between the first and last time a given URL was posted in our dataset.

(c) Posting Speed. The posting speed, which is a modified version of the retweet propagation speed defined by Shahi et al. [41], is measured as the total number of times a URL was posted divided by its total life span in hours. It provides the average number of times a given URL was posted per hour during its life span. If a URL was only posted once, it is assigned a posting speed of zero.

(d) Number of Lives. The number of lives of a URL is defined as the number of active posting periods it has. An active posting period is any time interval when a URL is posted without a break longer than 24 h. Once a 24-h break in a posted URL occurs, the next time the URL is posted, a new life begins, and the number of lives of the URL increases by one.

Not only are these metrics used to quantify the performance of URLs posted by the identified user pairs, but they are also used to compare the performance of their content to the rest of the multi-platform URLs in the dataset. This helps us to evaluate whether the user pairs tended to be involved in the spread of high performing, well-posted content, or rather if they appeared to share less popular and fewer re-posted URLs. Note that while URLs collected at the beginning and end of our data collection period may not be fully represented in the dataset and therefore have inaccurate metric values, this impact is minimized by the fact that they make up a relatively small percentage of the full dataset. Only 5% of the posts collected occurred in the first or last week of the data collection period. Additionally, the metrics are only used in a comparative sense between content posted by the users in each approach, rather than to accurately describe the multi-platform content diffusion itself.

4. Results

4.1. Identified potential cross-platform relationships

We present the results from the five cross-platform user pair identification approaches outlined in Section 3.1 in Table 4. The cross-platform linkers approach produced the most user pairs, more than double returned by any other approach. Meanwhile, the bidirectional introducers approach identified only 225 cross-platform user pairs, and the synchronized users approach returned 59.

It is not surprising that the cross-platform linkers approach returned the largest number of cross-platform relationships. As discussed previously, we expected this approach to return user pairs involving popular and well-followed accounts whose content was more likely to be shared by others, including on alternative platforms from where they were originally posted. Additionally, this approach does not aim to identify only 2-way or bidirectional relationships.

Conversely, since the bidirectional introducers approach is constrained to identifying 2-way relationships in which the users involved are the same entity across multiple platforms, we would expect it to produce fewer user pairs. As expected, it only returned 225 user pairs, significantly less than the 2816 pairs returned by the same name users approach. This suggests that the bidirectional introduction of content across different platforms may be a less common behavior exhibited by

Table 4
Breakdown of users and user pairs identified by each approach.

Type of users	Twitter users	Facebook users	Reddit users	Instagram users	Total # of users	# of user pairs	# of URLs involved
Same name users	2682	2512	85	32	5311	2816	6649
Bidirectional introducers	201	204	12	0	417	225	2081
Repeat introducers	1342	1273	113	6	2738	1807	7218
Synchronized users	50	38	4	5	97	59	365
Cross-platform linkers	3307	2967	960	187	7421	7656	6638

these types of users, as opposed to using similar names across platforms. By examining the overlap in users returned by each approach, we can further evaluate whether the bidirectional introducers approach primarily identifies a subset of the same name users or produces additional user pairs. This overlap analysis is discussed in Section 4.2.3.

There were also differences in the number of unique URLs posted by the user pairs identified through each approach. Notably, the bidirectional introducer user pairs introduced an average of 9.2 URLs per pair, while the repeat introducers introduced an average of 4 URLs per pair. This means that the repeat introducer pairs who also exhibited the bidirectional behavior were more prolific in introducing new content to their respective platforms. Also, further suggesting that the cross-platform linkers were linking to popular accounts, we find that there was only an average of 0.9 unique URLs per user pair. This was the lowest out of all of the approaches, meaning there was more overlap in the content posted by these user pairs than the rest.

Finally, the breakdown of the platforms that the identified users belonged to illustrates how some of the approaches may be more successful on certain platforms. For example, only 3% of the same name user pairs involved a user from Reddit, whereas 12.5% of the cross-platform linking user pairs involved a Reddit user. This is consistent with the fact that anonymity is a prominent feature of Reddit, and we found that tweets would commonly be cross-posted on Reddit. Additionally, the lack of Instagram users present in user pairs from approaches involving repeated postings of URLs makes sense since URLs are not commonly posted on the platform.

4.2. *Validation of the identified user relationships*

Having applied the five approaches to the real-world dataset and found that all of them produced cross-platform user pairs, we now present the results of the null model analysis, the network component distribution plots, and the overlaps of the user networks to validate the relationships identified.

4.2.1. *Null model analysis*

We perform the null model analysis to measure the significance of the cross-platform user pairs identified against those we might expect from chance alone. For this analysis, we relied on the direct approach of determining significance with null models [34,42]. With this approach, we derived a p value directly from each of the null model distributions, as the proportion of the model's random trials that produced more user pairs than our observed numbers from the social media data. We find that, for all three null models described earlier, the number of bidirectional introducers, repeat introducers, and synchronized users identified in the real-world dataset was significant compared to the null models' distributions, all with p values less than 0.01.

With the uniform null model, under which each URL-posting user on a given platform was equally likely to make one of that platform's posts, no bidirectional introducer user pairs were identified across all of the 1000 simulation runs. Across the repeat introducers simulations, a maximum of 1 cross-platform relationship was ever identified in a single run. The synchronized users were the most prevalent in this null model, with a maximum of 16 users and 14 edges being identified in a single run. However, this was still less than the 97 synchronized users found in the election dataset, and the average across the simulations was 0.4 synchronized users and 0.3 relationships.

Table 5
Proportional model simulations.

	Bidirectional introducers		Repeat introducers		Synchronized users	
	Users	Pairs	Users	Pairs	Users	Pairs
Avg	0	0	12.3	7.9	3.0	2.0
Max	2	1	28	18	17	14
Observed	**417**	**225**	**2738**	**1807**	**97**	**59**

Table 6
Proportional introducers model simulations.

	Bidirectional introducers		Repeat introducers		Synchronized users	
	Users	Pairs	Users	Pairs	Users	Pairs
Avg	3.8	2.5	119.1	119.4	2.2	1.4
Max	10	8	153	160	13	11
Observed	**417**	**225**	**2738**	**1807**	**97**	**59**

Considering the proportional null model, which accounted for each user's URL posting activity levels on their respective platform, we again find that the observed values are significant. Across the 1000 simulation runs, we found a maximum of only 1 bidirectional introducer relationship and 18 repeat introducer relationships (see Table 5). Given that 225 bidirectional introducer relationships and 1807 repeat introducer relationships were identified in the real-world data, the maximum values of the null model distribution remain significantly lower than the real-world findings.

For the synchronized users, the maximum number of synchronized users and relationships across the simulations remains close to the uniform null model, with 17 users and 14 relationships being the largest values found across the runs (see Table 5). Again, these remain smaller than the observed values from the election dataset.

Finally, we compare the observed values to the distributions produced by the proportional introducers null model, which considers cases where URLs were introduced to each platform separately from the rest. It consequently preserves the introducing activity levels of the users on each platform during the author reassignment process.

As this model had the least abstraction from the collected dataset, it makes sense that it produced results closest to the observed values. However, the maximum values across the 1000 runs of this model still remained significantly less than the observed ones, providing p values less than 0.01. The maximum number of bidirectional introducers found during a single run was 10, and the maximum number of relationships was 8, far less than the 417 and 225 found in the election dataset (see Table 6). For the repeat introducers, the maximum values across the simulation runs were 153 users and 160 relationships, whereas the observed values from the election data were 2738 and 1807, respectively. Similarly, for the synchronized users, the maximum values were 13 users and 11 relationships, while the real-world dataset contained 97 synchronized users and 59 relationships between them.

These differences between the simulation results with the null models and the observed values from the dataset suggest that the social media users in our dataset were not operating independently across the platforms. This aligns with our expectations since news organizations, political figures, and similar entities often operate accounts across platforms. While our evaluation of the effectiveness of our approaches for identifying cross-platform user behaviors is limited by the fact that we do not have ground truth regarding user identities across the

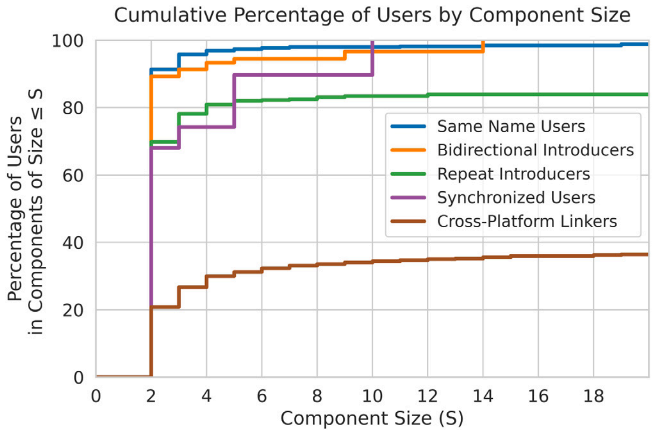


Fig. 6. Cumulative percentage of nodes in each network that belong to components less than or equal to the given sizes. We see that the same name users and bidirectional introducers networks had more than 90% of users in components of size ≤3. Unlike the other networks, the repeat introducers and cross-platform linkers networks had components of sizes ≥100 nodes which result in their cumulative percentages remaining below 100% in this plot, which is cut off at components of size 19 users.

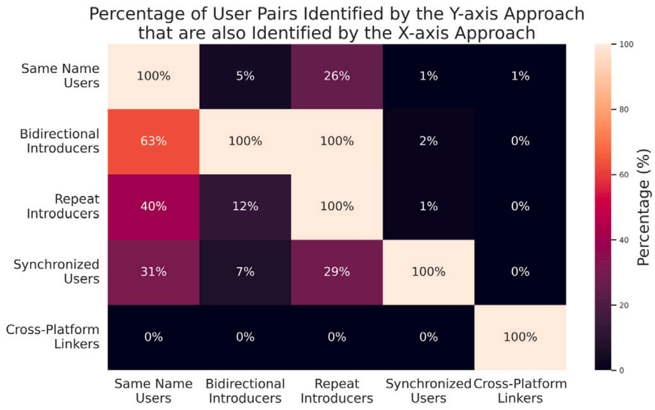


Fig. 7. Heatmap displaying the percentage of user pairs in the y-axis approach that are also identified by the x-axis approach. 100% of the user pairs in the bidirectional introducers network are in the repeat introducers network as they are guaranteed to fulfill the repeat introducers requirement. The next highest percentage is 63% of bidirectional introducers also appearing in the same name users network.

platforms, these results indicate that the approaches do not simply identify patterns arising from chance alone.

4.2.2. *Networks component size distributions*

The secondary analysis we perform to evaluate the results of the user pair identification approaches involves considering the sizes of the components returned by each approach. In particular, we are interested in whether the same name users and bidirectional introducers networks are mostly composed of dyads and triads. This is because these approaches aim to find the same type of user relationships, namely the same individual or organization across platforms. Since we expect most individuals and organizations to have a limited number of accounts across platforms, these approaches should produce networks mainly composed of small components.

For both the same name users network and the bidirectional introducers network, most of the users in the networks belong to dyads. In fact, about 91% of users in the same name users network belong to dyads, and 4% belong to triads (see Fig. 6). Similarly, 89% of users in the bidirectional introducers network are a part of dyads, and 2% are part of triads. Considering slightly larger components, we find that almost all of the users, 98% of same name users and 97% of bidirectional introducers, are in components of size 10 or smaller. While these results do not guarantee that the individuals identified by either approach are the same users across platforms, the absence of large components in these networks concurs with our expectations.

As expected, the repeat introducers, synchronized users, and cross-platform linkers networks have a lower percentage of users in small components than the other two approaches. Only 78% of the repeat introducers, 74% of the synchronized users, and 27% of the cross-platform linkers belong to components of size 3 or smaller. Of note, the size of the largest component was 442 users in the repeat introducers network and 4645 in the cross-platform linkers network. Since these approaches intended to find cross-platform user relationships that included bots and one-way cross-platform behaviors, it makes sense that such components would arise in the resulting networks. As discussed, popular users on individual platforms are more likely to be linked to or have their content cross-posted to new platforms by many different users and, therefore, could cause large components to form. Similarly, bots on platforms like Reddit may track multiple users on other platforms and cross-post their content to Reddit, further increasing the likelihood of large components appearing in these networks.

As described earlier, we are also interested in the structure of the intersection of the same name users and bidirectional introducers networks. Since they aim to identify the same type of users, it makes sense that cross-platform pairs of users identified by both approaches would be more likely to be the same individual or organization. In the intersection network, which only contains cross-platform relationships identified by both approaches, all users are in components of size 2. Therefore, by combining approaches, the relationships that had resulted in components larger than 2 were filtered out, and the network was decomposed into separate dyads. This supports the hypothesis that cross-platform user pairs identified by both of these approaches may yield a more accurate set of cross-platform users.

4.2.3. *Overlaps across the identified cross-platform pairs*

Rather than focusing on the structure of the resulting user networks, we now explore the amount of overlap between user relationships identified in each approach. To do this, we consider the percentage of edges (i.e., user pairs) returned by each approach that were also identified by each of the other approaches (see Fig. 7). This is because the node sets (i.e., users) returned by each approach might be different. By considering the percentage of user relationships in each network that were also identified by other approaches, we can get a sense of whether or not the networks returned subsets of each other.

Starting with the same name users, we find that 26% of the user pairs returned by this approach were also a part of the repeat introducers network. This makes sense since we anticipated that most of the accounts that share the same name are the same entity across platforms and, consequently, could be interested in introducing the same content across their accounts. While it may initially be surprising that only 5% of the same name user pairs also appear within the bidirectional introducers network, this is primarily because the bidirectional introducers network is much smaller than the same name users network. The maximum percentage of same name user pairs that could have been found in the bidirectional introducers network is only 8%.

When we flip the comparison to consider the share of bidirectional introducer pairs that appear in the same name users network, we find that 63% are in the same name users network. Since the goal of the bidirectional introducers approach was to find the same individuals and organizations across platforms, the fact that most of them shared extremely similar names provides additional validation of this approach. It supports the conclusion that the pairs identified are the same user or individual acting on multiple platforms. Meanwhile, the fact that it was not a complete overlap suggests that the bidirectional introducers approach identified additional accounts that did not share similar names but may still likely be the same individual or organization. Also, all of the bidirectional introducer pairs were in the repeat introducers

Table 7
Top cross-platform user pairs from each approach based on number of shared URLs.

Same name users		Bidirectional introducers		Repeat introducers		Synchronized users		Cross-platform linkers	

Newsweek(T)
Newsweek(F)
	208	TruthSeeker___(T)
		thetruthseeker1(F)
			102	TruthSeeker___(T)
				thetruthseeker1(F)
					102	SpeakaboutNews(T)
						speakingaboutnews(F)
							42	politicalscrap1(T)
								politicalscrapbooknet(F)
									129

EpochTimes(T)
epochtimes(F)
	200	ConservNewsDly(T)
		ConservativeNewsDly(F)
			67	keichri(T)
				speakingaboutnews(F)
					96	keichri(T)
						speakingaboutnews(F)
							21	barnes_law(T)
								reddit_feed_bot(R)
									97

IndyUSA(T)
IndependentUS(F)
	150	OANN(T)
		OneAmericaNewsNetwork(F)
			53	ConservNewsDly(T)
				ConservativeNewsDly(F)
					67	FreedomWireNews(T)
						Freedomwirenewz(F)
							16	Thomas1774Paine(T)
								reddit_feed_bot(R)
									97

ConservNewsDly(T)
ConservativeNewsDly(F)
	134	Newsweek(F)
		Newsweek(F)
			40	RandyMBell(T)
				TruthForTheTimes(F)
					65	NcsVentures(T)
						ncsnewstoday(F)
							11	JackPosobiec(T)
								reddit_feed_bot(R)
									84

wbradleyjr1(T)
wbradleyjr1(R)
	112	realTuckFrumper(T)
		hillreporter(F)
			28	OANN(T)
				OneAmericaNewsNetwork(F)
					53	PatriotPlanet(T)
						PatriotPlanetOfficial(F)
							7	TBUNEWS(F)
								TBUNEWS(T)
									75

(T) indicates Twitter user, (F) indicates Facebook user, and (R) indicates Reddit user.

network since they are repeat introducers who exhibited a bidirectional relationship in posting order.

Moving on to the repeat introducers user pairs, we find that 40% of them were also identified by the same name users approach. This is a smaller percentage than the 63% of bidirectional introducers who shared similar names and is consistent with the assumptions about the types of user relationships identified by the repeat introducers approach, namely that it contains both links between the same entities across platforms, as well as one-way relationships such as bot-like reposting relationships. Additionally, some organizations or individuals who use similar names over multiple platforms might always post information in a particular order or use a post-sharing service, which results in this one-way behavior.

Regarding the synchronized users approach, 31% of the user pairs in this network were also in the same name users network, and 29% were in the repeat introducers network. Since the synchronized users posted the same URLs as each other, and the number of synchronized users identified in the dataset was significantly smaller than the rest of the approaches, it makes sense that they would have some overlap with the repeat introducers and same name users.

As for the cross-platform linkers, only minimal percentages of the user pairs identified by this approach appeared within the other user networks. Even accounting for the large number of user pairs in the cross-platform linkers network, these findings suggest that the cross-platform linking approach found different users and relationships than the other approaches.

Together, these findings are consistent with the fact that the approaches take different behaviors into account and identify different types of relationships between users. They also suggest that the different approaches identified unique sets of users that would not have been found using only a single approach. For this reason, using a combination of approaches that identify organizations and individuals acting on multiple platforms based on different behaviors could be helpful for finding cross-platform pairs of users engaged in the spread of multi-platform and cross-platform content. Additionally, the types of accounts that researchers are interested in analyzing should impact the multi-platform behaviors considered.

4.3. Analysis of content posted by the identified user pairs

Having determined that the different approaches returned different sets of users and relationships, we now explore the types of content each set of users posted. This includes both the content categories of the multi-platform URLs and the bias and factual ratings of the news content posted by the user pairs identified by each approach.

4.3.1. Content comparison

In terms of the content that the user pairs of each group linked to, the same name users, bidirectional introducers, and repeat introducers shared similar categories of content (see Fig. 8). They all posted news the most, followed by political opinion and blog-type websites. This is consistent with the top user pairs identified by each of the three approaches being largely news organizations or news-related accounts (see Table 7). Out of the three approaches, the same name users had the smallest percentage of political opinion content. This could indicate that the same name users were more likely to be news organizations or otherwise less likely to share content linking to mostly opinion-related websites.

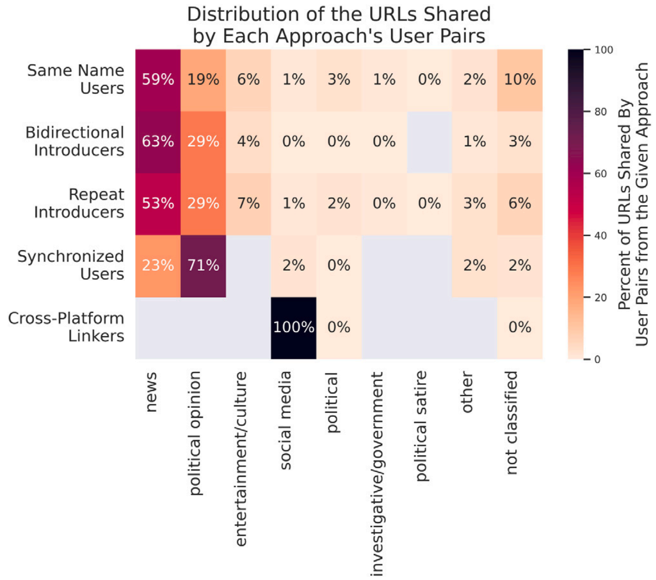


Fig. 8. Heatmap displaying the percentage of URL postings made by the user pairs identified by the various approaches. Content categories with less than 2% of the URLs from all of approaches are excluded. The most posted categories involved news and political blog type websites, as well as social media websites for the case of the cross-platform linkers.

Conversely, the synchronized users linked to political blogs and commentary websites far more than news organizations, 71% to 23%. This suggests that the synchronized users approach may have identified users with different goals and content preferences. Additionally, we find that for all identified groups of users except the cross-platform post linkers, none of the content categories besides news and political opinion received more than 8% of the group's URL postings.

4.3.2. Bias and fact comparison

As for the political biases of the users who exhibited the multi-platform behaviors, we find that the synchronized and cross-platform linking users were more right-leaning than the rest (see Fig. 9). More than a third of the users identified by each approach had average bias

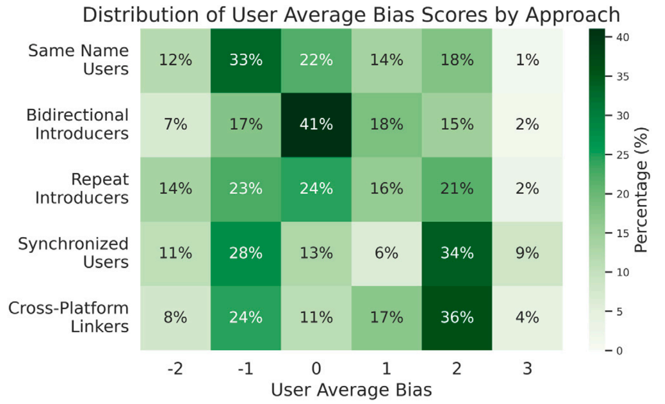


Fig. 9. Heatmap displaying the distributions of user average bias scores of the users in each *y*-axis user group. User average bias scores were rounded to the nearest integer. Negative bias scores indicate left-leaning bias and positive scores indicate right-leaning bias. There were no sources with bias values of −3 (i.e., far left) in the dataset so it was excluded from this figure.

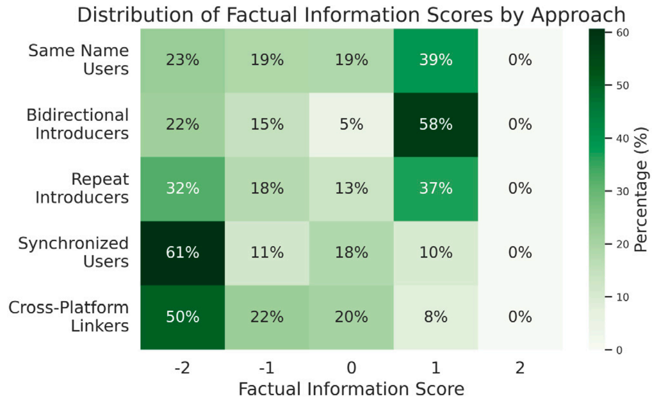


Fig. 10. Heatmap displaying the distributions of factual information scores of the users identified by each approach. User factual information scores were rounded to the nearest integer. Higher scores indicate more factual sources.

scores between 1.5 and 2. Combined with the content finding above, we can conclude that the synchronized users were mostly engaged in promoting content linking to right-leaning political blogs and commentary websites.

Meanwhile, the bidirectional repeat introducers were the least biased out of all the groups, and the same name users exhibited the most substantial left-leaning bias. Together with the previous content results, these findings suggest that the same name users primarily linked to left-leaning mainstream news sources, while the bidirectional repeat introducers tended to introduce content from less biased mainstream news sources.

The distributions of factual information scores for the users in each group also highlight the synchronized users for posting low factual information (see Fig. 10). Of the synchronized users, 68% had factual information scores less than −1.5, which was 19% more than any other user group. Furthermore, we see that the same name users and bidirectional repeat introducers had the biggest shares of users with high factual information scores, i.e., scores greater than 0.5. This again indicates that these pairs of users were sharing URLs linking to higher credibility websites.

4.4. Performance comparison

Now, we consider whether the content shared by the different sets of user pairs had different performance characteristics. One main finding is that the URLs introduced by the bidirectional and repeat introducers underperformed relative to the full set of multi-platform URLs. They also performed worse than the URLs posted by other identified user pairs. The content introduced by bidirectional introducers and repeat introducers tended to be posted less and have shorter posting life spans (see Figs. 11, 12). Although they had faster posting speeds than the multi-platform URLs and therefore may have appeared to spread quickly across the platforms, this more likely resulted from the URLs being posted fewer times in a shorter period rather than being more viral (see Fig. 13). Additionally, this content was the least likely to be posted again in the dataset once it had a 24-h break in being posted (see Fig. 14).

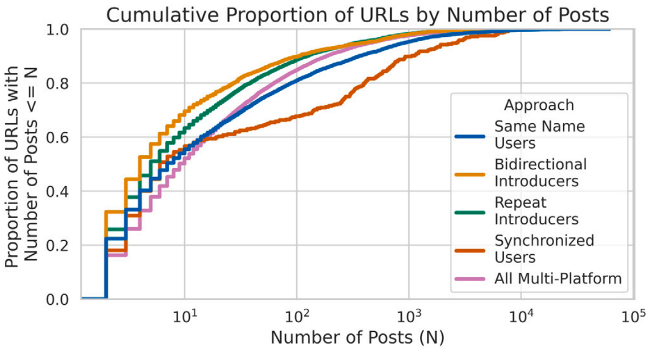


Fig. 11. Cumulative proportion of URLs shared by the user pairs from each approach based on their number of posts. We see that the URLs introduced by the bidirectional introducers and repeat introducers tended to be posted less, while a large share of the URLs posted by the synchronized users was highly posted. Compared to the performance of the full set of multi-platform URLs, the same name users' URLs had a similar performance but a slightly larger share of highly posted content.

Considering the URLs posted by these bidirectional and repeat introducer pairs, instead of focusing only on the URLs they introduced, we find that their performance is similar to the same name users. The same name user pairs posted URLs with numbers of postings close to the full set of multi-platform URLs, except for having slightly larger tails (see Fig. 11). They did not remain active in the dataset for as long, though, with an average life span of 9.2 days as opposed to 11.3 days for the multi-platform URLs (see Fig. 13).

The synchronized users' URLs had the most unique performance characteristics of all the URL subsets. This is likely due, in part, to the relatively small number of URLs posted by the synchronized users' user pairs. Unlike the rest of the user pairs, a large share of the synchronized users' URLs, more than 30%, received over 100 posts (see Fig. 11). They also tended to have the fastest postings speeds and were the most likely to be posted again after multiple 24-h breaks in being posted.

Altogether, these findings suggest that while the URLs introduced by the repeat and bidirectional introducers tended to perform worse than the rest, the user pairs identified by those approaches tended to generally post content with similar performance to those posted by the same name user pairs. It also indicates that the synchronized users were mostly involved in spreading well-posted and likely popular content.

4.5. Limitations

In terms of the user pair identification approaches explored in this work, we make no claims about their ability to identify all cross-platform user relationships. In fact, by using both the same name users and bidirectional introducers approaches, we recognize that neither approach is comprehensive in finding all of the same individuals or organizations across platforms. Therefore, a user not being identified by any of the approaches does not mean they do not have accounts across multiple platforms. Rather, in this work, we focus on identifying users who appeared to exhibit *multi-platform information spreading behaviors*.

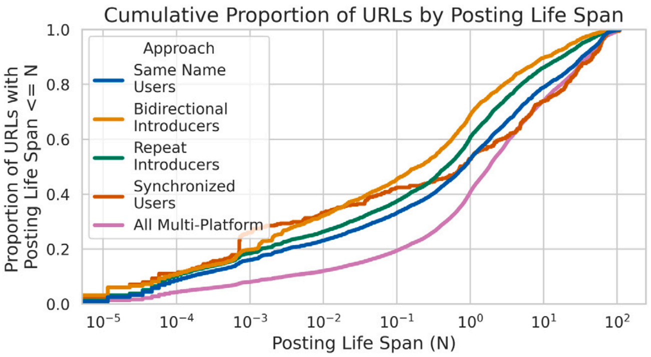


Fig. 12. Cumulative proportion of URLs shared by the user pairs from each approach based on posting life span (i.e., the number of days between the first and last time a given URL was posted in the dataset). URLs introduced by the bidirectional introducers tended have the shortest posting life spans, followed by the repeat introducers and same name users. Overall, the full set of multi-platform URLs tended to be active on the platforms the longest, although around 20% of the URLs posted by the synchronized users had long posting life spans.

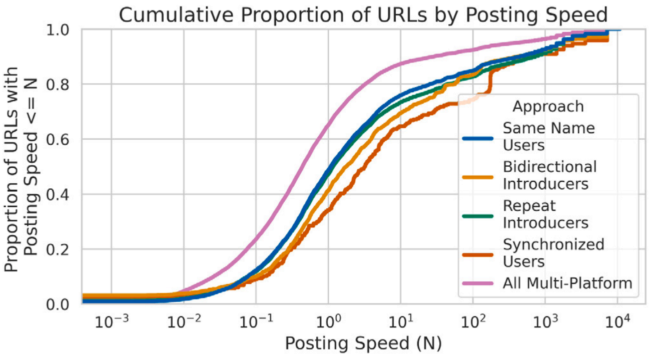


Fig. 13. Cumulative proportion of URLs shared by the user pairs from each approach based on posting speed (i.e., average number of times a given URL was posted per hour). The full set of multi-platform URLs tended to be posted at slower speeds than those posted by the identified user pairs, while the URLs posted by the synchronized users were posted the fastest.

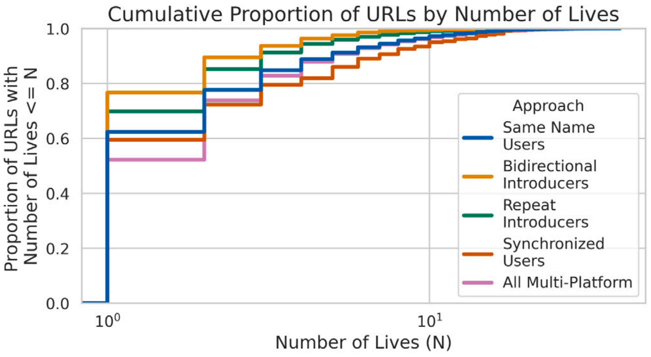


Fig. 14. Cumulative proportion of URLs shared by the user pairs from each approach based on number of lives. The URLs introduced by the bidirectional introducers and the repeat introducers were the least likely to be posted again after a 24-h break in being posted, while the URLs posted by the synchronized users were the most likely to have three or more lives.

Additionally, the lack of ground truth, both in terms of users who actually coordinated online or offline to spread the same information over multiple platforms and those who have active accounts across multiple platforms, limits our ability to evaluate the false positive rate of our identified cross-platform user pairs. This limitation is further complicated by the privacy concerns of releasing the full lists of the identified user pairs. As anonymous accounts are an important and prominent feature of Reddit, we are sensitive to the fact that linking such accounts to users on Twitter or Facebook may deanonymize them.

Though substantial, we attempted to mitigate these limitations by performing null model analysis and an analysis of the component size distributions of the returned user networks. This allowed us to verify that the cross-platform user pair identification approaches relied on behaviors that did not arise from chance alone and that the identified user relationships reflected network structures we would expect from each approach. In future work, we plan to use tools such as Maltego [43] to investigate further the users identified by our approaches and validate our results.

As for the empirical results of our work, a limitation is that the collected multi-platform dataset only contains public data and, in the case of Facebook and Instagram, primarily reflects popular and well-followed accounts. Consequently, our identified user pairs only reflect relationships involving a subset of Facebook and Instagram users. This means that overall number of users pairs identified involving each platform should not be taken as a reflection of the magnitude of related users across platforms, rather we intend to provide a comparison of the number of user pairs identified by each approach within the platform. A second limitation related to data access is that the collected dataset does not contain posts that were deleted or censored prior to our data collection, which also likely resulted in some cross-platform relationships going undetected.

Lastly, the presence and behavior of the users identified within this 2020 U.S. election case study are limited to this context. We do not claim that our results apply to all discussions across Twitter, Facebook, Reddit, and Instagram. We plan to pursue future work exploring these approaches in additional contexts.

5. Conclusion

In this work, we draw on prior cross-platform user identification and user coordination research to explore five approaches for identifying pairs of cross-platform users engaged in multi-platform content-spreading behaviors. These approaches differ in either the attributes and behaviors they leverage to identify the user pairs, or in the types of relationships they aim to uncover. In doing so, they do not rely on the social media platforms of study having similar social structures or public users, but rather that URL-posting is a common practice on the platform.

Within the context of fraud and protest discussions surrounding the 2020 U.S. election, we use the outlined approaches to perform a case study of four social media platforms. Using null models, size distributions of the user network components, and the overlaps in users returned by each approach, we come to three conclusions: (i) the approaches returned relationships that would not be expected if they relied on posting behaviors arising from chance alone; (ii) the sizes of the components returned by each approach align with the types of relationships they aimed to identify; and (iii) the overlap in user pairs between the approaches not only helps support the bidirectional introducers as an approach for identifying the same entities across platforms but also shows that each approach returned a unique set of user relationships. These findings help validate the results of applying the user identification approaches to this dataset and support the benefits of using multiple strategies for identifying users with multi-platform information-spreading behaviors.

While these methods have only been explored in a limited context, the evaluation of each method independently reveals that the users identified by the different approaches shared different types of content, with different political biases and factual ratings, and posted URLs with varying degrees of spreading performance. This suggests that if practitioners are interested in identifying a particular type of multi-platform actor (e.g., news organizations), or comparing behaviors across different types of actors, it may be advantageous and more efficient to use individual identification approaches. On the other hand, if researchers are interested in collecting a set of possible cross-platform actors, they may use the approaches in combination. Finally, if they want to minimize false positives when identifying pairs of cross-platform users, one effective approach may be to require the users to display behaviors from multiple approaches.

Ultimately, this work can be used to help identify and characterize users who facilitate multi-platform content diffusion and highlights the importance of studying the spread of information across multiple social media platforms. Further investigation is needed within additional contexts and platforms to evaluate the types of users identified by the approaches presented here and to measure the role that such users play in facilitating multi-platform content spread. This research is vital for developing realistic multi-platform models of information diffusion and devising effective strategies to mitigate misinformation spread in our complex social media environment.

CRediT authorship contribution statement

Isabel Murdock: Conceptualization, Methodology, Software, Investigation, Data curation, Writing – original draft, Visualization. **Kathleen M. Carley:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Funding acquisition. **Osman Yağan:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported in part by the National Science Foundation through Grants #1813637 and #2225513, the Office of Naval Research through Grant N000141712675, and the Army Research Office through Grant #W911NF-22-1-0181. I. Murdock was supported by the Knight Fellowship and the Boeing Scholarship. 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] B. Auxier, M. Anderson, Social Media Use in 2021, Tech. Rep., Pew Research Center, 2021.
[2] J. Tucker, A. Guess, P. Barbera, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, B. Nyhan, Social media, political polarization, and political disinformation: A review of the scientific literature, SSRN Electron. J. (2018).
[3] K. Carley, G. Cervone, N. Agarwal, H. Liu, Social cyber-security, in: Social, Cultural, and Behavioral Modeling - 11th International Conference, SBP-BRiMS 2018, Proceedings, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, Germany, 2018, pp. 389–394.
[4] M. Bossetta, The digital architectures of social media: Comparing political campaigning on facebook, Twitter, instagram, and snapchat in the 2016 U.S. election, J. Mass Commun. Q. 95 (2) (2018) 471–496.
[5] H. Lin, L. Qiu, Two sites, two voices: Linguistic differences between facebook status updates and tweets, in: P.L.P. Rau (Ed.), Cross-Cultural Design. Cultural Differences in Everyday Life, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 432–440.
[6] O. Papakyriakopoulos, J. Medina Serrano, S. Hegelich, The spread of COVID-19 conspiracy theories on social media and the effect of content moderation, Harvard Kennedy Sch. (HKS) Misinf. Rev. (2020).
[7] N. Velásquez, R. Leahy, N.J. Restrepo, Y. Lupu, R. Sear, N. Gabriel, O.K. Jha, B. Goldberg, N.F. Johnson, Online hate network spreads malicious COVID-19 content outside the control of individual social media platforms, Sci. Rep. 11 (11549) (2021).
[8] J. Lukito, Coordinating a multi-platform disinformation campaign: Internet research agency activity on three U.S. social media platforms, 2015 to 2017, Polit. Commun. 37 (2) (2020) 238–255.
[9] B. Nimmo, C. François, C. Eib, L. Ronzaud, M. Smith, T. Lederer, J. Carter, E. McAweeney, IRA in Ghana: Double Deceit, Tech. Rep., Graphika, 2020.
[10] S. Fischer, Reddit bans subreddit group "r/DonaldTrump", Axios (2021).
[11] Z. Chen, K. Aslett, J. Rosiere, J. Reynolds, J. Nagler, J. Tucker, R. Bonneau, An automatic framework to continuously monitor multi-platform information spread, in: CEUR Workshop Proceedings, Vol. 2890, CEUR-WS, 2021.
[12] K. Hunt, B. Wang, J. Zhuang, Misinformation debunking and cross-platform information sharing through Twitter during Hurricanes Harvey and Irma: a case study on shelters and ID checks, Nat. Hazards 103 (1) (2020) 861–883.
[13] O. Yağan, D. Qian, J. Zhang, D. Cochran, Conjoining speeds up information diffusion in overlaying social-physical networks, IEEE J. Sel. Areas Commun. 31 (6) (2013) 1038–1048.
[14] K. Starbird, T. Wilson, Cross-platform disinformation campaigns: Lessons learned and next steps, Harvard Kennedy Sch. (HKS) Misinf. Rev. (2020).
[15] Y. Golovchenko, C. Buntain, G. Eady, M. Brown, J. Tucker, Cross-platform state propaganda: Russian trolls on Twitter and YouTube during the 2016 U.S. presidential election, Int. J. Press/Polit. 25 (3) (2020) 357–389.
[16] L. Xing, K. Deng, H. Wu, P. Xie, H.V. Zhao, F. Gao, A survey of across social networks user identification, IEEE Access 7 (2019) 137472–137488.
[17] R. Zafarani, H. Liu, Connecting corresponding identities across communities, in: Proceedings of the Third International ICWSM Conference, 2009, pp. 354–357.
[18] J. Liu, F. Zhang, X. Song, Y.-I. Song, C.-Y. Lin, H.-W. Hon, What's in a name? An unsupervised approach to link users across communities, in: Proc. 6th ACM Int. Conf. Web Search Data Mining, 2013, pp. 495–504.
[19] W. Ahmad, R. Ali, Understanding users display-name consistency across social networks, Int. J. Eng. Adv. Technol. (IJEAT) 8 (5S3) (2019) 471–476.
[20] Y. Li, Y. Peng, W. Ji, Z. Zhang, Q. Xu, User identification based on display names across online social networks, IEEE Access 5 (2017) 17342–17353.
[21] V. Sharma, C. Dyreson, LINKSOCIAL: Linking user profiles across multiple social media platforms, in: 2018 IEEE International Conference on Big Knowledge (ICBK), 2018, pp. 260–267.
[22] T. Iofciu, P. Fankhauser, F. Abel, K. Bischoff, Identifying users across social tagging systems, in: Proceedings of the International AAAI Conference on Web and Social Media, Vol. 5, 2021, pp. 522–525.
[23] J. Ma, Y. Qiao, G. Hu, Y. Huang, M. Wang, A.K. Sangaiah, C. Zhang, Y. Wang, Balancing user profile and social network structure for anchor link inferring across multiple online social networks, IEEE Access 5 (2017) 12031–12040.
[24] D. Pacheco, P.-M. Hui, C. Torres-Lugo, B. Truong, A. Flammini, F. Menczer, Uncovering coordinated networks on social media: Methods and case studies, in: Proc. International AAAI Conference on Web and Social Media (ICWSM), Vol. 15, 2021, pp. 455–466, [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/18075.
[25] D. Pacheco, A. Flammini, F. Menczer, Unveiling coordinated groups behind white helmets disinformation, in: Companion Proceedings of the Web Conference 2020, Association for Computing Machinery, New York, NY, USA, ISBN: 9781450370240, 2020, pp. 611–616, [Online]. Available: https://doi.org/10.1145/3366424.3385775.
[26] C. Cao, J. Caverlee, K. Lee, H. Ge, J. Chung, Organic or organized? Exploring URL sharing behavior, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15, Association for Computing Machinery, New York, NY, USA, 2015, pp. 513–522.
[27] T. Magelinski, L. Ng, K. Carley, A synchronized action framework for detection of coordination on social media, J. Online Trust Saf. 1 (2) (2022).
[28] E. Ferrara, H. Chang, E. Chen, G. Muric, J. Patel, Characterizing social media manipulation in the 2020 U.S. presidential election, First Monday 25 (11) (2020) [Online]. Available: https://firstmonday.org/ojs/index.php/fm/article/view/11431.
[29] K. Sharma, E. Ferrara, Y. Liu, Characterizing online engagement with disinformation and conspiracies in the2020 U.S. presidential election, in: Proceedings of the Sixteenth International AAAI Conference on Web and Social Media, Association for the Advancement of Artificial Intelligence, 2022, pp. 908–919.
[30] A. Burton, D. Koehorst, Research note: The spread of political misinformation on online subcultural platforms, Harvard Kennedy Sch. (HKS) Misinf. Rev. (2020).
[31] Center for an Informed Public, Digital Forensic Research Lab, Graphika, Stanford Internet Observatory, The long fuse: Misinformation and the 2020 election, Stanford Digit. Repos.: Elect. Integr. Partnersh. (2021) [Online]. Available: https://purl.stanford.edu/tr171zs0069.
[32] Z. Sanderson, M.A. Brown, R. Bonneau, J. Nagler, T.J. A., Twitter flagged Donald Trump's tweets with election misinformation: They continued to spread both on and off the platform, Harvard Kennedy Sch. (HKS) Misinf. Rev. (2021).

[33] L.H.X. Ng, I.J. Cruickshank, K.M. Carley, Cross-platform information spread during the January 6th capitol riots, Soc. Netw. Anal. Min. 12 (1) (2022) 133, [Online]. Available: https://link.springer.com/10.1007/s13278-022-00937-1.

[34] D.R. Farine, A guide to null models for animal social network analysis, Methods Ecol. Evol. 8 (10) (2017) 1309–1320, [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12772.

[35] E.A. Hobson, M.J. Silk, N.H. Fefferman, D.B. Larremore, P. Rombach, S. Shai, N. Pinter-Wollman, A guide to choosing and implementing reference models for social network analysis, Biol. Rev. 96 (6) (2021) 2716–2734, [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/brv.12775.

[36] L. Yin, M. Brown, N. Baram, G. Eady, SMAPPNYU/urlexpander, 2018, [Online]. Available: https://github.com/SMAPPNYU/urlExpander.

[37] Bias and Factual Reporting Ratings, Media Bias / Fact Check, 2022, [Online]. Available: https://mediabiasfactcheck.com, Last accessed Jan. 2022.

[38] G. Weld, M. Glenski, T. Althoff, Political bias and factualness in news sharing across more than 100,000 online communities, in: Proceedings of the Fifteenth International AAAI Conference on Web and Social Media, Association for the Advancement of Artificial Intelligence, 2021, pp. 796–807.

[39] V. Patricia Aires, F. G. Nakamura, E. F. Nakamura, A link-based approach to detect media bias in news websites, in: Companion Proceedings of the 2019 World Wide Web Conference, WWW ’19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 742–745, [Online]. Available: https://doi.org/10.1145/3308560.3316460.

[40] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C.M. Valensise, E. Brugnoli, A.L. Schmidt, P. Zola, F. Zollo, A. Scala, The COVID-19 social media infodemic, Sci. Rep. 10 (16598) (2020).

[41] G.K. Shahi, A. Dirkson, T.A. Majchrzak, An exploratory study of COVID-19 misinformation on Twitter, Online Soc. Netw. Media 22 (2021) 100104, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2468696420300458.

[42] J.A. Veech, Significance testing in ecological null models, Theor. Ecol. 5 (4) (2012) 611–616, [Online]. Available: https://doi.org/10.1007/s12080-012-0159-z.

[43] How to Conduct Person of Interest Investigations Using OSINT and Maltego, Maltego, 2021, [Online]. Available: https://https://www.maltego.com/blog/how-to-conduct-person-of-interest-investigations-using-osint-and-maltego/.