An automatic framework to continuously monitor multi-platform information spread

Zhouhan Chen
New York University
zhouhan.chen@nyu.edu

Kevin Aslett
New York University
kma412@nyu.edu

Jen Rosiere Reynolds
New York University
jr5505@nyu.edu

Juliana Freire
New York University
juliana.freire@nyu.edu

Jonathan Nagler
New York University
jonathan.nagler@nyu.edu

Joshua A. Tucker
New York University
joshua.tucker@nyu.edu

Richard Bonneau
New York University & Flatiron Institute
bonneau@nyu.edu

ABSTRACT

Identifying and tracking the proliferation of misinformation, or fake news, poses unique challenges to academic researchers and online social networking platforms. Fake news increasingly traverses multiple platforms, posted on one platform and then re-shared on another, making it difficult to manually track the spread of individual messages. Also, the prevalence of fake news cannot be measured by a single indicator, but requires an ensemble of metrics that quantify information spread along multiple dimensions. To address these issues, we propose a framework called Information Tracer, that can (1) track the spread of news URLs over multiple platforms, (2) generate customizable metrics, and (3) enable investigators to compare, calibrate, and identify possible fake news stories. We implement a system that tracks URLs over Twitter, Facebook and Reddit and operationalize three impact indicators – *Total Interaction*, *Breakout Scale* and *Coefficient of Traffic Manipulation* – to quantify news spread patterns. Using a collection of human-verified false URLs, we show that URLs from different origins have different propensities to spread to multiple platforms, cover different topics, while exhibit similar retweet patterns. We also demonstrate how our system can discover URLs whose spread pattern deviate from the norm, and be used to coordinate human fact-checking of news domains. Our framework provides a readily usable solution for researchers to trace information across multiple platforms, to experiment with new indicators, and to discover low-quality news URLs in near real-time.

KEYWORDS

misinformation, cross platform, fake news

1 INTRODUCTION

The COVID-19 pandemic has increased the consumption of news via social media. For example, [1] a recent global survey found that, since the beginning of COVID-19, 43% of consumers increased time spent on YouTube, 40% on Facebook and 23% on Twitter. As people spend more time consuming news from online platforms, the volume of online misinformation has also increased, resulting in the World Health Organization declaring an Infodemic [20]. To mitigate misinformation and promote high-quality content, it is important for us to first understand where information originates and how it spreads. Two major technical challenges remain. First, information is often posted on one platform and shared on another, but recent work in cross-platform news spread focus on single events, which are ad-hoc and not scalable [4, 18]. Second, there is no unified approach to measure and quantify information spread. Different measurements result in different estimations of misinformation prevalence. For example, [19] points out that depending on the chosen datasets and metrics, the amount of misinformation on Twitter can range between 1% to 70%. Measuring the prevalence of fake news with a single indicator is inadequate.

In this paper, we propose a framework called Information Tracer that contributes three major improvements to previous work. First, we define a unified data collection pipeline to trace and visualize data from multiple platforms. Second, we support a multi-pronged approach that uses multiple indicators to measure information spread. Third, we provide a user interface to enable researchers to comparatively identify URLs with unusual metrics, and to facilitate fact-checkers by contextualizing URL spread across multiple platforms. We implement Information Tracer to track URLs over three platforms – Twitter, Facebook, and Reddit, the most popular mobile social networking platforms in the United States as of September 2019 [16]. To quantify information spread, we operationalize three **impact indicators** – *Total Interaction*, *Breakout Scale*, and *Coefficient of Traffic Manipulation*. Finally, we create a web interface to visualize both raw data and aggregated statistics[1].

We also present three real-world applications to demonstrate the capability of Information Tracer. In Application One, we investigate three main questions using a collection of fake news URLs from four origins (Twitter, Facebook, YouTube, News outlets):

(1) Do URLs from different origins have different likelihood to spread across multiple platforms?
(2) Do they have different Twitter retweet traffic patterns?
(3) Do they cover different topics?

[1] Our user interface is available at https://informationtracer.com/

We find that URLs from Facebook are less likely to spread over multiple platforms; URLs from different platforms cover different false stories; and there is no significant difference in retweet patterns.

Application Two (A2) and Three (A3) include human oversight and interaction, so called human-in-the-loop capability, to our framework. In A2, we demonstrate how Information tracer can assist humans to identify URLs whose impact indicators deviate from the sample average. In A3, we instruct human coders to fact-check qualities of news domains with the help of Information Tracer. We show that our system can potentially reduce the time it takes to discover previously unknown low-quality news sites.

The paper is organized as the following: Section 2 details each component of Information Tracer system. Section 3 applies our framework in three three real-world settings. Section 4 discusses the limitation of our research. We examine related work in Section 5, and conclude the paper in Section 6.

2 INFORMATION TRACER SYSTEM

On a high level, Information Tracer consists of three components – **data collector**, **data aggregator**, and **data visualizer**; these modules collect data, generate summary statistics, and enable visualizing data respectively. Figure 1 shows the system architecture. In this section, we detail how we implement each component.

Although we implement our framework with a particular set of configurations that help us answer our research questions, our proposed framework is customizable and users can design their own metrics to better answer other questions. A metric can be a simple count, or a numerical output from a machine learning model. Our framework is also extendable – users can integrate additional sources (social media platforms, weblogs, messaging software) into the system, without altering the overall data pipeline.

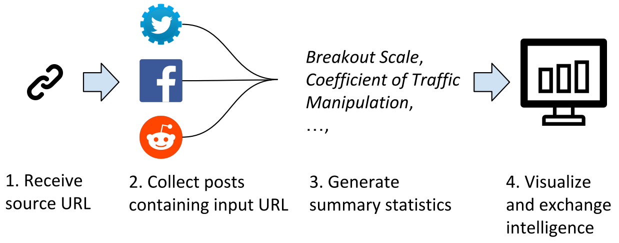

Breakout Scale,
Coefficient of Traffic
Manipulation,
...,
1. Receive source URL
2. Collect posts containing input URL
3. Generate summary statistics
4. Visualize and exchange intelligence

Figure 1: Information Tracer Architecture. The input to the system is a valid URL. Upon receiving the URL, our system automatically collects posts containing this URL from designated social networking platforms, generates aggregated statistics, and finally presents results on a web interface.

2.1 Component One: data collector

The goal of the data collector is to parse queries submitted by end users, then collect posts that match those queries from a list of platforms. For the scope of this paper, we restrict the query to a valid URL, and we consider three platforms – Twitter, Facebook and Reddit. We focus on the URL because it has a well-defined structure, is indexed by all three platforms, and serves as a unique identifier of news stories.

URL sanitization and normalization. Before we make API calls to each platform, we sanitize and normalize the input URL to maximize the number of matched posts on our three platforms. We sanitize a URL in following ways:

- Remove prefixes *http://*, *https://* and *www.*. For example, query *https://www.cnn.com/XYZ* will become *cnn.com/XYZ*. This ensures that we match all posts that refer to article *cnn.com/XYZ*.
- Remove query parameters. A query parameter is any substring that follows a "?". They are usually appended at the end of URL for tracking purposes. We strip query parameters to normalize the input URL with only a few exceptions. For example, a YouTube video has a canonical form *https://www.youtube.com/watch?v=VideoID*, in which "?" is important and cannot be removed. We maintain an allowlist of such domains.

Twitter Collection Our Twitter search is powered by Twitter Academic Track API[2]. This API provides us with access to Twitter's full-archive tweet corpus. As of February 17, 2021, the API imposes a cap of 10,000,000 tweets per month. Due to this rate limit, we have to be judicious about how we collect tweets. Our strategy is to collect influential tweets that receive a high level of interactions such as retweets and replies, and to avoid collecting tweets with low interaction (tweets along the "long tail"). This intuition comes from a previous study on Twitter user characterization, which finds that a small number of influential users control most of conversation diffusion [3]. Because the definition of "influential" is subjective, we introduce five tunable parameters that can be specified by users during query submission – minimum number of retweets (*min_retweets*), minimum number of replies (*min_replies*), maximum number of original tweets (*max_originals*), maximum number of retweets (*max_retweets*), and maximum number of replies (*max_replies*). The following is our data collection protocol:

1. Given a *URL=q*, *min_retweets=x*, *min_replies=y*, we construct a special URL – https://twitter.com/search?q=min_retweets:x%20min_replies:y%20url:q&f=live. This URL returns us matched original tweets, with at least x retweets and y replies. We use a Python Selenium headless browser to automatically visit this URL, scroll down the page, and extract *max_originals* number of tweets, or until there is no result. We have to use a headless browser to automate this process because the two search parameters (*min_retweets*, *min_replies*) are not available via the API.
2. Then for each original tweet with *id=TweetID*, we use full-archive search endpoint [3] to collect retweets and replies. We set *query=status/TweetID* to retrieve all quoted tweets and retweets of quoted tweets. We set *query=conversation_id:TweetID* to match all replies of the original tweet. We collect *max_retweets* and *max_replies* number of results.

Our Twitter collection module is thus customizable: by tuning each threshold one can collect more or less tweets, and adapt to different questions and API rate limits. For example, to collect all matched retweets and replies one can set *min_retweets=0*, *min_replies=0*, *max_originals=∞*, *max_replies=∞*, *max_retweets=∞*. In practice, we strongly recommend setting thresholds to avoid

[2] https://developer.twitter.com/en/solutions/academic-research
[3] https://api.twitter.com/2/tweets/search/all

burning API usage. These settings should be application-specific, and thus, we present use cases in Section 3.

Facebook Collection. We use Crowdtangle to collect Facebook public posts containing the input URL. Crowdtangle is a tool that collects and aggregates engagement data of Facebook, Instagram and Reddit posts. It provides API to journalists and academic researchers. We use the search API to collect Facebook posts containing the input URL. The API returns up to 1,000 posts. To collect influential posts, we use the *sort* parameter to retrieve posts with the highest score. The score is a metric designed by Crowdtangle to indicate if a post "overperforms." [4] Importantly, Crowdtangle does not index every single Facebook page. According to Crowdtangle's documentation[5], as of February 24, 2021, more than six million Facebook pages, groups, and verified profiles are indexed. This includes "all public Facebook pages and groups with more than 100K likes, all US-based public groups with 2k+ members, and all verified profiles," and therefore misses private groups and pages.

Reddit Collection. Similar to Facebook data collection, we use Crowdtangle to collect the top 1,000 Reddit posts containing the input URL sorted by the "overperform" score. Crowdtangle indexes more than 20,000 of the most active sub-reddits, and adds more sub-reddits on an ongoing basis.

To summarize, due to limitations from each API endpoint, we are not able to retrieve *every* post that matches a query. Specifically, private posts are unavailable, and posts from less popular groups may not be indexed yet. We argue that the omission of those low-interaction posts are acceptable because they do not play a significant role in spreading information. From a resource allocation perspective, storing only popular posts (cutting off the long tail) saves storage space, and improves data processing speed.

2.2 Component Two: data aggregator

The goal of data aggregator is to distill intelligence from heterogeneous cross-platform data sources. It achieves this goal by calculating summary statistics to quantify information spread. In this paper, we refer to those statistics as *impact indicators*, as they indicate the relative impact of a URL on one or more platforms. Over the years many indicators have been proposed and explored. In this paper, we operationalize three indicators – **Total Interaction**, **Breakout Scale** [9], and **Coefficient of Traffic Manipulation (CTM)** [8].

We choose those measurements because they are compatible with our dataset. Specifically, Breakout Scale requires multi-platform data to measure information spread, CTM requires retweet data to measure Twitter traffic pattern, and Total Interaction requires total number of interactions of every post. All three types of data are available in our collection. We want to point out that our framework is indicator-agnostic. The indicators we operationalized may be more helpful on one dataset but less on another. We now introduce each indicator in detail.

2.2.1 *Total Interaction.*

Interaction count is a simple yet effective measurement to quantify the popularity of a post. This metric has proven to be useful in recent studies to quantify fake news spread during COVID-19 [2, 7, 12]. For each URL, we define its total interaction as the summation of total interactions of every Twitter, Facebook, and Reddit post. We define the post-level total interaction as:

- Twitter post. The total number of retweets, replies and likes.
- Facebook post[6]. The total number of reactions, shares and comments.
- Reddit post. The total number of upvotes and comments.

2.2.2 *Breakout Scale.*

Breakout Scale is originally proposed as a comparative model for measuring and calibrating Information Operations (IOs) based on "data that are observable, replicable, verifiable, and available from the moment they were posted. [9]" It measures how many platforms an IO percolates to, and assigns an IO to one of six categories, as shown in Table 1.

We find the Breakout Scale framework appealing as it allows us to quantify how many platforms a URL is popular over. To operationalize this framework, we use total interaction as a proxy for popularity. Formally, for each URL u, we denote the total number of interactions it receives on platform p as $interaction_p$. We then set a threshold t, if $interaction_p > t$, we consider u to be popular on platform p. The final Breakout Scale for u is the total number of popular platforms.

2.2.3 *Coefficient of Traffic Manipulation (CTM).*

We also compute fine-grained indicators that quantify platform-specific patterns. Because we only have page and group level statistics for Facebook and Reddit posts, we focus on summarizing Twitter traffic here, for which we have full access. CTM is a comparative model that allows one to compare different Twitter traffic flows "against measurable criteria and assess which of those movements appear to have been subject to manipulation." [8]

Originally, CTM was a weighted average of three measurements: the average number of tweets per user (m_1), the percentage of retweets as a proportion of total tweets (m_2), and the proportion of tweets generated by top fifty accounts (m_3). After analyzing real-world Twitter traffic containing manipulated hashtags, the authors concluded that m_1 and m_3 are more informative to identify manipulated traffic. In our implementation, we modify and define CTM as a tuple of two values: average number of tweets per user, and proportion of tweets generated by top 10% accounts. We focus on percentage instead of top fifty accounts as, in our experiments, we find tweet threads with fewer than fifty accounts.

We want to note here that a high CTM *does not* always imply traffic manipulation. For example, a tweet thread with high CTM could be caused by authentic users who are engaged in the conversation and replied many times. Similarly, a tweet corpus with low CTM might be manipulated by a sophisticated bot campaign, in which each bot only creates one tweet, thus evading this metric. In Section 3 we show how to use our system to discover the cause of high CTM.

2.3 Component Three: data visualizer

Data visualization is a key element of both validating this platform and enabling needed human interaction. Thus we aim to facilitate real-time exchange of cross-platform data and intelligence. We

[4] https://help.crowdtangle.com/en/articles/3213537-crowdtangle-codebook

[5] https://help.crowdtangle.com/en/articles/1140930-what-data-is-crowdtangle-tracking

[6] This definition is adopted from https://help.crowdtangle.com/en/articles/1184978-crowdtangle-glossary

Table 1: Definition of Breakout Scale. Our system can automatically derive Category 0 to 3 based on data collected from three platforms.

Category	Definition	Can we operationalize?
0	Popular on zero platform	Yes
1	Popular on one platform	Yes
2	Popular on two platforms	Yes
3	Popular on three or more platforms	Yes
4	Popular on multiple media (social media, mainstream, offline)	Not yet
5	Celebrity amplification	Not yet
6	Require policy change	Not yet

propose and implement two main data visualizations – a summary page and an item-wise detail page.

2.3.1 Summary page visualization. The summary page allows investigators to compare, calibrate and identify data points (in our case URLs) with unusual spread patterns. We currently use a scatter plot[7] to visualize all three impact indicators. Investigators can identify an interesting quadrant, zoom in, and click on individual point (which represents a URL) to navigate to the detail page.

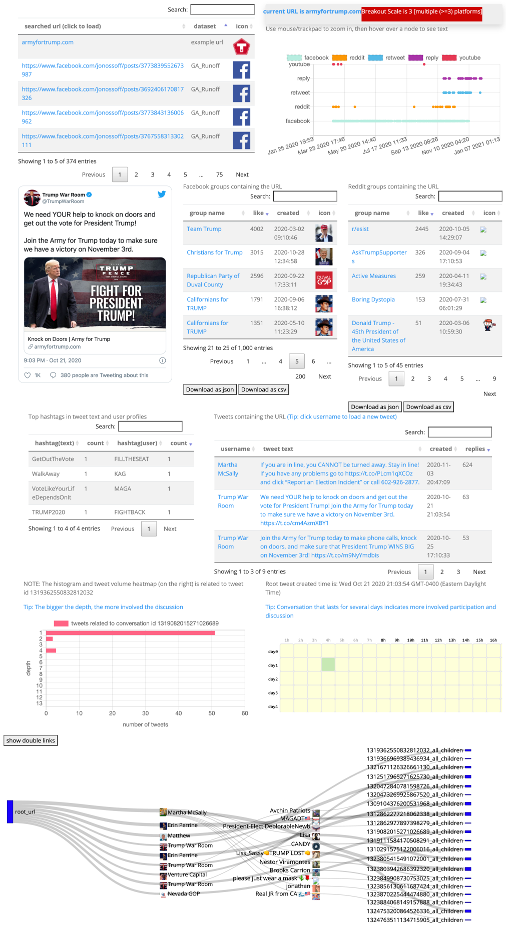
2.3.2 Item-wise Detail visualization page. The detail page allows investigators to visualize individual posts from different platforms, and explore how posts interact with each other along multiple dimensions, such as temporal, network, and contextual. Figure 2 is a rendering of one detail page that contextualizes the spread pattern of URL armyfortrump.com. Those visualizations provide answers to questions such as when the URL is shared on each platform, who posted it, and how users who share the URL interacted with each other via retweet and reply.

3 REAL WORLD APPLICATIONS

We now introduce three real-world applications (denoted as *A1*, *A2* and *A3*). *A1* uses Information Tracer to understand and compare how fake news URLs from different origins spread over three platforms. We focus on four origins for the fake-news we will trace: Twitter, Facebook, Youtube and News domains. *A2* and *A3* incorporate human-in-the-loop intelligence. For A2, we uses Information Tracer to discover URLs with unusual impact indicators. For A3, we instruct human coders to assess qualities of news domains using our system. In the rest of the section, we first introduce our data sources, then explain each application.

3.1 Overview of datasets

Google Fact Check Dataset (abbr. *Google FN*). The Google Fact Check Dataset is a repository of false claims, fact checked by journalists around the world. The dataset has been adopted by many fact checkers around the world, including those verified by International Fact Checking Network (IFCN). It also powers fact checking features behind Google Search, Google News and Bing Search [8].



Search:

searched url (click to load)	dataset	icon
armyfortrump.com	example url	
https://www.facebook.com/jonossoff/posts/3773839552673987	GA_Runoff	f
https://www.facebook.com/jonossoff/posts/3692406170817326	GA_Runoff	f
https://www.facebook.com/jonossoff/posts/3773843136006962	GA_Runoff	f
https://www.facebook.com/jonossoff/posts/3767558313302111	GA_Runoff	f

Showing 1 to 5 of 374 entries

Previous 1 2 3 4 5 … 75 Next

current URL is armyfortrump.com Breakout Scale is 3 [multiple (>=3) platforms]

Use mouse/trackpad to zoom in, then hover over a node to see text

Trump War Room @TrumpWarRoom

We need YOUR help to knock on doors and get out the vote for President Trump!

Join the Army for Trump today to make sure we have a victory on November 3rd.

Knock on Doors | Army for Trump
armyfortrump.com

9:03 PM · Oct 21, 2020

♡ 1K 380 people are Tweeting about this

Facebook groups containing the URL

Search:

group name	like	created	icon
Team Trump	4002	2020-03-02 09:10:46	
Christians for Trump	3015	2020-10-28 12:34:58	
Republican Party of Duval County	2596	2020-09-22 17:33:11	
Californians for TRUMP	1791	2020-09-06 16:38:12	
Californians for TRUMP	1351	2020-05-10 11:23:29	

Showing 21 to 25 of 1,000 entries

Previous 1 … 4 5 6 … 200 Next

Download as json Download as csv

Reddit groups containing the URL

Search:

group name	like	created	icon
r/esist	2445	2020-10-05 14:29:07	
AskTrumpSupporters	326	2020-09-04 17:10:53	
Active Measures	259	2020-04-11 19:34:43	
Boring Dystopia	153	2020-07-31 06:01:29	
Donald Trump - 45th President of the United States of America	51	2020-03-06 10:59:30	

Showing 1 to 5 of 45 entries

Previous 1 2 3 4 5 … 9 Next

Download as json Download as csv

Top hashtags in tweet text and user profiles

Search:

hashtag(text)	count	hashtag(user)	count
GetOutTheVote	1	FILLTHESEAT	1
WalkAway	1	KAG	1
VoteLikeYourLifeDependsOnIt	1	MAGA	1
TRUMP2020	1	FIGHTBACK	1

Showing 1 to 4 of 4 entries Previous 1 Next

Tweets containing the URL (Tip: click username to load a new tweet)

Search:

username	tweet text	created	replies
Martha McSally	If you are in line, you CANNOT be turned away. Stay in line! If you have any problems go to https://t.co/PLcm1qXCOz and click "Report an Election Incident" or call 602-926-2877.	2020-11-03 20:47:09	624
Trump War Room	We need YOUR help to knock on doors and get out the vote for President Trump! Join the Army for Trump today to make sure we have a victory on November 3rd. https://t.co/cm4AzmXBY1	2020-10-21 21:03:54	63
Trump War Room	Join the Army for Trump today to make phone calls, knock on doors, and make sure that President Trump WINS BIG on November 3rd! https://t.co/m9NyYmdbis	2020-10-25 17:10:33	53

Showing 1 to 3 of 9 entries Previous 1 2 3 Next

NOTE: The histogram and tweet volume heatmap (on the right) is related to tweet id 1319362550832812032

Tip: The bigger the depth, the more involved the discussion

Root tweet created time is: Wed Oct 21 2020 21:03:54 GMT-0400 (Eastern Daylight Time)

Tip: Conversation that lasts for several days indicates more involved participation and discussion

show double links

facebook
reddit
retweet
reply
youtube
Jan 25 2020 19:53
Mar 23 2020 17:46
May 20 2020 14:40
Jul 17 2020 11:33
Sep 13 2020 08:26
Nov 10 2020 04:20
Jan 07 2021 01:13
FIGHT FOR PRESIDENT TRUMP!
TRUMP PENCE
tweets related to conversation id 1319082015271026689
depth
1 2 3 4 5 6 7 8 9 10 11 12 13
0 10 20 30 40 50 60
number of tweets
1h 2h 3h 4h 5h 6h 7h 8h 9h 10h 11h 12h 13h 14h 15h 16h
day0
day1
day2
day3
day4
root_url
Martha McSally
Erin Perrine
Matthew
Trump War Room
Erin Perrine
Trump War Room
Venture Capital
Trump War Room
Nevada GOP
Avchin Patriots
MAGADT🇺🇸
President-Elect DeplorableNewb
Lisa
CANDY
Liss_Sassy👊TRUMP LOST👊
Nestor Viramontes
Brooks Carrion
please just wear a mask 🌵🌹
jonathan
Real JR from CA 🇺🇸
1319362550832812032_all_children
1319366969389436934_all_children
1321671126326661130_all_children
1312517965271625730_all_children
1320472840781598726_all_children
1320473269925867520_all_children
1309104376200531968_all_children
1312862277218062338_all_children
1312862977897398279_all_children
1319082015271026689_all_children
1319111584170508291_all_children
1310291575122006016_all_children
1323805415491072001_all_children
1323803942686392320_all_children
1323849908730753025_all_children
1323856130611687424_all_children
1323870225444474880_all_children
1323884068149157888_all_children
1324753200864526336_all_children
1324763511134715905_all_children

Figure 2: Information Tracer User Interface. This detail page visualizes the spread pattern of URL armyfortrump.com. To understand when people share the website, one can examine the scatter plot (first row). To understand who shares, one can inspect top Facebook groups, top Reddit groups, and top original tweets (second and third row). Finally, to understand how conversation unfolds, one can check the tweet retweet and reply networks (bottom row).

We collect all US-based claims from Google Fact Check Dataset during 2020. To do so, we first download all claims from the web portal[9]. We sort all claims by fact checking organizations, and manually check the origin of top 30 organizations (which account for more than 90% of all claims). We identify six organizations that operate in the United States – politifact.com, factcheck.org, washingtonpost.com, usatoday.com, nytimes.com, poynter.org. Then for each claim from each organization, we examine the API structure, and extract URLs from field *entry["itemReviewed"]*, which are URLs that point to the source of fake news. If the URL is archived, we run another script to extract the original URL from archived page. In the end, we extract **1427** unique URLs.

[7] Our summary page is available here: https://informationtracer.com/intelligence.

[8] https://developers.google.com/fact-check/tools/api

[9] https://datacommons.org/factcheck/download

IFCN COVID-19 Fake News Dataset (abbr. *IFCN FN*). Our second dataset contains 8,627 false claims compiled by fact checkers among IFCN. The earliest entry is from 1/5/2020, and the latest entry is from 8/26/2020. Each entry contains a URL that points to the source of false claim. However, there are several special cases:

1. Shortened URLs. For those URLs, we write a script to resolve the final landing URL.
2. Non-URL texts. Some URLs are plain texts such as *web page removed, WhatsApp Chain (no link), There is no link. It is an e-mail.* We remove those entries.
3. Duplicated URLs. We only keep the first entry.

After the cleanup, the dataset has 4178 unique URLs. We then use the country column to select URLs whose column value is "United States." In the end our IFCN dataset has 501 URLs.

Tracing Information Cross-platform After we compile two datasets, we use Information Tracer to collect posts containing those URLs from three platforms, as described in Section 2. For Twitter collection, we set *min_retweets=10, min_replies=2, max_originals=50*, and *max_replies=max_retweets=20,000.* The two minimum thresholds filter out low-information tweets, and the three maximum thresholds prevent us from burning API quota. Finally, to calculate the Breakout Scale, we define the breakout threshold to be 100, which means a URL is considered popular on a platform if its total number of interactions from that platform is above 100. We experimented different thresholds and found the resulting trends to be consistent.

3.2 *Application 1*: understanding how fake news URLs spread across platforms

Application 1 demonstrates the core utility of Information Tracer, which is its ability to quantify information spread over multiple platforms. To facilitate further discussion, we categorize each fake news URL into one of four origins: Twitter, Facebook, YouTube and News domains, and consider how URL from each origin is shared on three platforms – Twitter, Facebook, Reddit. Here, Twitter and Facebook can be ***both origin and destination platforms***. When we say the origin of URL *A* is Twitter, we simply mean *A* is created on Twitter (i.e., *A* is a tweet). When we say URL *B* breaks out on Twitter, we mean there is a high number of tweets that contain URL *B*, while B can originate from any platform. Table 2 shows number of URLs from each origin in IFCN and Google datasets. Specifically, the definition of each origin is:

1. **Twitter**. URL has a pattern *twitter.com/username/tweetid*
2. **Facebook**. URL has a pattern *facebook.com/username/type/id*, or *facebook.com/photo?fbid=id. type* can be *posts* or *videos.*
3. **Youtube**. URL is a YouTube video. For example: *youtube.com/watch?v=videoid.*
4. **News domain**. URL is a news article. For example: *breitbart.com/link-to-article*

Given this taxonomy and our multi-dimensional indicators, we investigate three questions regarding fake news URLs from different origins. We start by analyzing **multi-platform** patterns (using Breakout Scale and Total Interaction), then comparing **single-platform** traffic pattern (using CTM), and finally understanding **contents of fake news** from each origin using unsupervised topic modeling.

Table 2: Overview of IFCN and Google datasets, separated by origins of the URL.

Dataset	# URLs Twitter	# URLs Facebook	# URLs Youtube	# URLs News	Total
IFCN FN	65	197	47	192	501
Google FN	241	747	127	312	1427

Table 3: Comparison of median impact indicators of URLs from different origins. URLs from Facebook are the least likely to spread over multiple platforms.

Dataset	avg. tweet per user	% tweets from top 10% users	breakout scale	total interactions
(I)Twitter	1.05	16	1	994
(I)FB	1.03	14	0	0
(I)Youtube	1.08	18	0	1637
(I)News	1.06	16	1	2080
(G)Twitter	1.06	15	1	544,444
(G)FB	N/A	N/A	0	0
(G)Youtube	1.04	18	0	272
(G)News	1.07	17	1	251

3.2.1 *Q1: do URLs from different origins have different likelihoods of breaking out over multiple platforms?* Using the Breakout Scale, we plot the percentage of URLs within each origin that spread on 0, 1, 2 and 3 platforms, shown in Figure 3. We find fake news URL originating from Facebook (a Facebook page or image) are the least likely to spread over two or more platforms. Specifically, more than 90% of URLs from Facebook do not break out on other platforms. In contrast, 40% of URLs from Twitter, YouTube and News domains break out on more than one platform, and 20% of URLs from Twitter and YouTube break out on all three major platforms. This suggests that when fake news is generated on Facebook, it is more likely to stay within the platform. When a fake news URL travels across platforms, it is more likely to be a tweet, a YouTube video, or a news article.

3.2.2 *Q2: do URLs from different origins receive different number of interactions, and have different Twitter traffics?* We calculate the median value of Total Interaction and Coefficient of Traffic Manipulation (CTM), listed in Table 3. We use median value instead of mean because the distribution of each indicator is highly skewed by extremely large values.

The value of total interaction is heavily influenced by the scale and other aspects of the underlying dataset and API availability. For example, in IFCN and Google datasets, Facebook URLs have a median total interaction of zero, which indicates that more than half Facebook URLs come from Facebook groups with low interactions and are not indexed by Crowdtangle API. In addition, in the Google FN dataset, fake news URLs from Twitter have a median total interaction of 544,444, a value way larger than interaction from other origins. Upon further investigation, we find that most

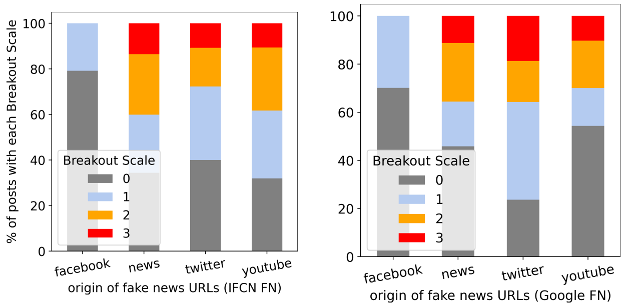


Figure 3: Percentage of fake news URLs that break out on 0, 1, 2, or 3 platforms, separated by origin. If the origin is a News domain, YouTube or Twitter, a URL is more likely to spread over two or more platforms.

of tweets in the dataset spread political fake news, and are created by high-profile accounts that receive unusually high number of interactions, such as *@realDonaldTrump* (suspended account of Former President Donald Trump, 88 million followers at the time of suspension) and *@seanhannity* (TV Host for Fox News, 5.3 million followers as of February 2021).

For CTM, we do not find any difference among URLs from different origins. Specifically, median values of average-tweet-per-user range from 1.03 to 1.08, and median values of percent-tweets-from-top-10%-users range from 14 to 18. The fact that there is no difference on the aggregated level does not mean no difference on the individual level. In Section 3.3 we show how to identify individual URL whose indicators deviate from the norm.

3.2.3 Q3: do fake news URLs from different origins cover different topics? To better understand the substance of fake news, we investigate whether URLs from different origins cover different topics. To quantify topics, we use non-negative matrix factorization (NMF), an unsupervised clustering algorithm that factorizes a document-word matrix into a document-topic matrix and a word-topic matrix. Using both matrices, we can identify top words within each topic, and top topics a document belongs to. Previous work has used NMF to discover meaningful political topics from tweets censored by Turkish government [17].

In both IFCN and Google datasets, there is a “claim” column that summarizes the content of each false URL. The input to the NMF algorithm is thus a claim-word matrix, where each row is a claim, and each column is a unique word. The cell value is the tf-idf[10] weight of the word. We lower-case all words, choose a dictionary size of 5,000 (that is, our matrix has 5,000 columns in maximum), and remove all English stopwords. We experiment with different number of topics, and find that clustering claims of URLs into 6 topics give us meaningful and interpretable results.

Figure 4 shows proportion of fake news URLs belonging to each topic, and most frequent words per topic. We find that in the IFCN dataset, topic “5G causes coronavirus” is a popular YouTube topic (accounting for 14% of all YouTube URLs), “China bioweapon” is popular within news websites (35% of news URL), and Twitter users talk more about “President Trump” and his “administration.” In the Google FN dataset, we find more discussion about ”election fraud“ on Twitter (30% of Twitter URL), and more references to ”Kamala Harris“ on YouTube and Facebook. Those differences suggest that fake news topics is platform-specific. Instead of relying on a fixed list of keywords, social media platforms can adopt similar methods to discovered unknown topics from suspicious URLs, and use discovered keywords as technical signals to track and stop fake news spread at an early stage.

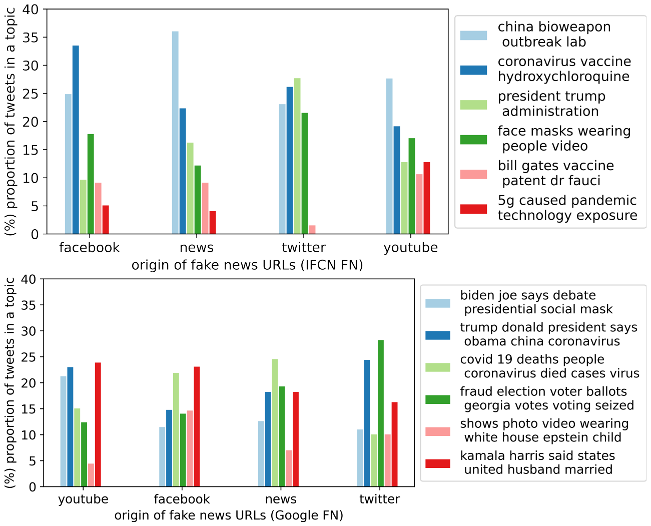


Figure 4: Percentage of fake news URLs that belongs to each topic, separated by origins of the URL. Each color represents a topic. The legend shows keywords that are most likely to appear within each topic. Topics are discovered using non-negative matrix factorization. We find that fake news originating from different platforms cover different topics. For example, in IFCN dataset, topic “5G causes coronavirus” is more discussed on YouTube than on other platforms, percentage-wise.

3.3 *Application 2*: investigating news stories with unusual spread patterns

The previous section shows how our framework can compare news spread patterns of various groups of URLs. Even though aggregated analysis is helpful to reveal trends or patterns, investigators may also want to examine individual data points. In this section we show a case study that uses Information Tracer to understand a URL whose impact indicator deviates from the sample mean. The URL (denoted as u_1) we consider is a YouTube video[11] from our IFCN dataset. It has an average-tweet-per-user (part of CTM) value of 2, the highest number among all URLs in IFCN dataset. Figure 5 shows that the URL is on the top-right quadrant, a clear outlier.

[10] We use Python sklearn to calculate tf-idf and run NMF: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.NMF.html

[11] The link to the video is https://youtube.com/watch?v=zFN5LUaqxOA. The video falsely claims that coronavirus is caused by 5G, and has already been removed by YouTube, but tweets containing the link are still available.

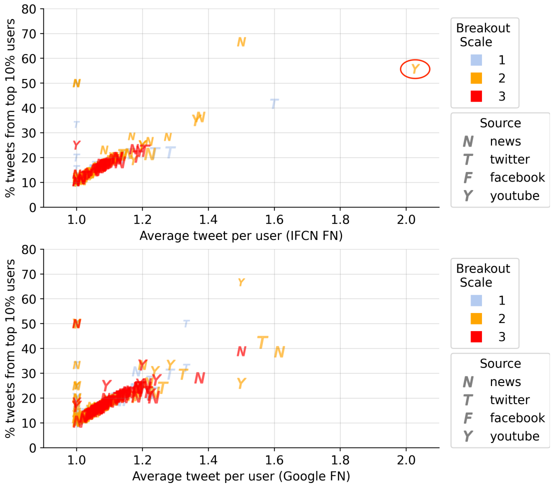
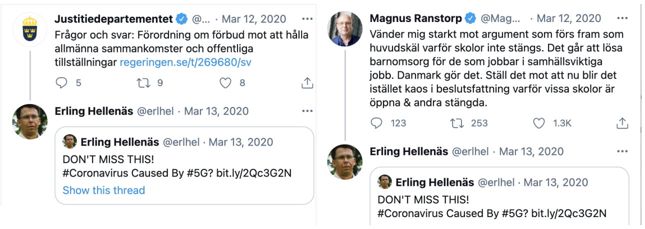


Figure 5: Multi-dimensional visualization of impact indicators. For each scatter plot, a marker represents one URL. The color of the marker reflects its Breakout Scale. The text of the marker reflects its origin. The size of the marker is in proportion to its total interaction in logarithmic scale.



Justitiedepartementet @... · Mar 12, 2020
Frågor och svar: Förordning om förbud mot att hålla allmänna sammankomster och offentliga tillställningar regeringen.se/t/269680/sv
5 9 8

Erling Hellenäs @erlhel · Mar 13, 2020

Erling Hellenäs @erlhel · Mar 13, 2020
DON'T MISS THIS!
#Coronavirus Caused By #5G? bit.ly/2Qc3G2N
Show this thread

Magnus Ranstorp @Mag... · Mar 12, 2020
Vänder mig starkt mot argument som förs fram som huvudskäl varför skolor inte stängs. Det går att lösa barnomsorg för de som jobbar i samhällsviktiga jobb. Danmark gör det. Ställ det mot att nu blir det istället kaos i beslutsfattning varför vissa skolor är öppna & andra stängda.
123 253 1.3K

Erling Hellenäs @erlhel · Mar 13, 2020

Erling Hellenäs @erlhel · Mar 13, 2020
DON'T MISS THIS!
#Coronavirus Caused By #5G? bit.ly/2Qc3G2N

Figure 6: Screenshots of two reply chains. Using Information Tracer, we find account *@erlhel* replied under multiple verified accounts, encouraging users to watch a Youtube video with false information. This repetitive tweeting pattern results in a high coefficient of traffic manipulation (CTM).

To understand why u_1 has a high CTM, we navigate to its detail page[12], study its retweet network, and find several accounts that repeatedly sent u_1 to targeted users. For example, Figure 6 shows Twitter user *@erlhel* sharing u_1 with verified accounts, while encouraging users to watch u_1. This spammy behavior boosts the average-tweet-per-user count. Even though we can not assess whether account *@erlhel* is human or bot, its behavior requires more intervention such as account warning or account suspension.

3.4 *Application 3*: assessing quality of unknown news domains

Another promising use case for Information Tracer is to facilitate human fact-checking. To assess the utility of our framework, we recruit 30 native English speakers from surgehq.ai, a platform that provides high-skill workforce. For each coder, we ask them to assess the veracity of a domain discovered by a content-agnostic classifier (CAC) [5]. The CAC works in two steps. In step one, it collects live tweets from Twitter Streaming API based on pre-defined keywords, extracts domains embedded in tweets, and clusters domains together if they are shared by similar users. In step two, CAC assigns a fakeness score to each domain, using features from HTML pages. We deployed a CAC from 10/29/2020 to 11/11/2020, using keyword "election." We set a clustering threshold of 0.6, and selected top 30 unlabeled domains sorted by the fakeness score[13]. For each domain, we used Information Tracer to collect social media posts containing URLs from that domain, and visualized results on our web interface. For instance, Figure 2 is the screenshot of social media presence of one discovered domain *armyfortrump.com*.

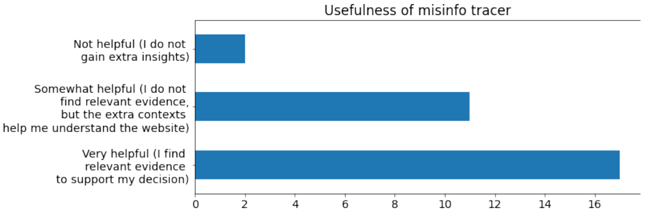


Figure 7: Perceived utility of Information Tracer from 30 human coders who use the system to assess factualness of news domains. 93% (28/30) coders find the tool helpful.

We then randomly assigned one domain to one coder, and asked everyone to assess the factualness of the domain with the help of Information Tracer. Specifically, we asked coders to look for following signals: is the domain shared across multiple platforms (e.g., Facebook, Twitter, Reddit)? If so what groups are sharing the domain on each platform? What hashtags do they use, are they verified? To teach coders how to navigate through our web interface, we also shared with them a detailed video instruction[14].

A comprehensive analysis of the CAC model accuracy based on ground-truth labels from human coders is beyond the scope of this paper. The result we want to highlight is people's perceived utility of Information Tracer. According to Figure 7, when asked "how helpful is the Information Tracer," 93% coders find it at least somewhat helpful, and 57% find it very helpful. In addition, we asked coders if they had any feedback about Information Tracer. One said *"I like it a lot except the node part is really hard to understand"*; the other pointed out *"It was easier to look at the page/Twitter page itself, but the recent tweets on Information Tracer gave a good idea of what the site would be."* We will improve our system based on those suggestions.

4 LIMITATION AND NEXT STEPS

Data access remains a bottleneck. Despite recent collaboration between academia and social media platforms, getting access to more accurate metrics remains a challenge. For example, [11] points out that social media platforms aggregate two types of metrics –

[12] The detail page is available on our web interface:https://informationtracer.com/?url=youtube.com/watch?v=zFN5LUaqxOA. We encourage investigators to explore the retweet and reply networks.

[13] To learn more about how we choose clustering threshold, and a detailed list of discovered domains from our deployed CAC, visit: https://zhouhanc.github.io/misinformation-discoverer/

[14] The 5-minute video instruction is available on Google Drive: https://drive.google.com/file/d/1Hqaql5MHlyUKWAwmF7_uKCNyg_ed_nfB/view?usp=sharing

impressions and expressions. Impressions are publicly available statistics such as number of retweets, replies and likes. Expressions are more fine-grained measurements such as "who scrolls what tweet thread for how many seconds." Impressions can be a better proxy to estimate the popularity of a post and to derive Breakout Scale. Unfortunately, current API does not expose impressions data. We hope to engage with platforms and deepen current collaboration.

Observational data versus experimental data. Even if we can collect all social media posts, a gap remains where people's online actions do not necessarily translate to real-world behavioral changes. For example, a story that receives more interactions may or may not change more people's behaviors. To measure behavioral change, controlled experiments are often required. We plan to introduce our framework to the broader political behavior research community. We also plan to collect alternative data sources such as direct web traffic log or responses from human subjects to validate our observation.

5 RELATED WORK

Information tracking tools Many open-source tracking systems have been built over the years. For example, Hoaxy is a system to visualize the spread of fact-checking claims [13]. FakeNewsTracker [15] is a similar framework to collect, analyze, and visualize tweets related to fake news claims. More recently, [14] build dashboard to analyze COVID-19 misinformation. [6] provides a more detailed list of open source tools that track misinformation. Limitations of current frameworks are (1) only focusing on a single platform (usually Twitter) and (2) not providing sufficient metrics to assess the impact of different news stories. Our framework aims to overcome those limitations.

Cross-platform misinformation spread Research shows that misinformation are increasingly spread over multiple platforms. Understanding where misinformation originates, and where it gets amplified, can help researchers design effective mitigation strategies [11]. Recently, [18] analyzes the disinformation campaign targeting the White Helmets group using Twitter and Youtube data. [4] studies how different types of news spread on 4chan and Reddit. [10] collected URLs from four platforms, Facebook, Twitter, Reddit, and 4chan, quantified information diffusion, and measured the impact of content moderation. As suggested by [19], previous research in tracing cross-platform news spread lacks a unified data collection pipeline and well-defined metrics. Our framework aims to fill this gap.

6 CONCLUSION

In this paper, we propose and implement Information Tracer, a framework to track and quantify information spread across multiple platforms. We operationalize three metrics – Total Interaction, Breakout Scale and Coefficient of Traffic Manipulation, and apply our framework on real world datasets. We find that fake news URLs with different origins have different likelihoods to spread over multiple platforms, with URLs from Facebook being the least likely to spread over multiple platforms. Finally, our real-world use cases demonstrate that Information Tracer can help investigators to identify abnormal spread patterns, facilitate fact-checking, and design better intervention strategies.

REFERENCES

[1] 2020. Global Online Content Consumption Doubles in 2020. https://doubleverify.com/newsroom/global-online-content-consumption-doubles-in-2020-research-shows/. [Online; accessed 22-February-2021].

[2] Manon Berriche and Sacha Altay. 2020. Internet users engage more with phatic posts than with health misinformation on Facebook. *Palgrave Communications* 6, 1 (2020), 71. https://doi.org/10.1057/s41599-020-0452-1

[3] David R. Bild, Yue Liu, Robert P. Dick, Z. Morley Mao, and Dan S. Wallach. 2015. Aggregate Characterization of User Behavior in Twitter and Analysis of the Retweet Graph. *ACM Trans. Internet Technol.* 15, 1, Article 4 (March 2015), 24 pages. https://doi.org/10.1145/2700060

[4] ANTHONY G. BURTON and DIMITRI KOEHORS. 2021 (accessed January 13, 2021. *Research note: The spread of political misinformation on online subcultural platforms.*

[5] Zhouhan Chen and Juliana Freire. 2020. Proactive Discovery of Fake News Domains from Real-Time Social Media Feeds. In *Companion Proceedings of the Web Conference 2020* (Taipei, Taiwan) *(WWW '20)*. Association for Computing Machinery, New York, NY, USA, 584–592. https://doi.org/10.1145/3366424.3385772

[6] The RAND Corporation. 2019. Tools That Fight Disinformation Online . https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search.html. [Online; accessed 13-January-2021].

[7] Andrea Moscadelli, Giuseppe Albora, Massimiliano Alberto Biamonte, Duccio Giorgetti, Michele Innocenzio, Sonia Paoli, Chiara Lorini, Paolo Bonanni, and Guglielmo Bonaccorsi. 2020. Fake News and Covid-19 in Italy: Results of a Quantitative Observational Study. *International journal of environmental research and public health* 17, 16 (08 2020), 5850. https://doi.org/10.3390/ijerph17165850

[8] Ben Nimmo. 2021 (accessed January 5, 2021. *Measuring traffic manipulation on Twitter.* https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/01/Manipulating-Twitter-Traffic.pdf

[9] Ben Nimmo. 2021 (accessed January 5, 2021. *THE BREAKOUT SCALE: MEASURING THE IMPACT OF INFLUENCE OPERATIONS.* https://www.brookings.edu/wp-content/uploads/2020/09/Nimmo_influence_operations_PDF.pdf

[10] ORESTIS PAPAKYRIAKOPOULOS, JUAN CARLOS MEDINA SERRANO, and SIMON HEGELICH. 2021 (accessed January 13, 2021. *The spread of COVID-19 conspiracy theories on social media and the effect of content moderation.*

[11] Irene V. Pasquetto, Briony Swire-Thompson, Michelle A. Amazeen, et al. 2021 (accessed January 8, 2021. *Tackling misinformation: What researchers could do with social media data.* https://doi.org/10.37016/mr-2020-49

[12] Cristina M Pulido, Laura Ruiz-Eugenio, Gisela Redondo-Sama, and Beatriz Villarejo-Carballido. 2020. A New Application of Social Impact in Social Media for Overcoming Fake News in Health. *International journal of environmental research and public health* 17, 7 (04 2020), 2430. https://doi.org/10.3390/ijerph17072430

[13] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2016. Hoaxy. *Proceedings of the 25th International Conference Companion on World Wide Web - WWW '16 Companion* (2016).

[14] Karishma Sharma, Sungyong Seo, Chuizheng Meng, Sirisha Rambhatla, and Yan Liu. 2020. COVID-19 on Social Media: Analyzing Misinformation in Twitter Conversations. arXiv:2003.12309 [cs.SI]

[15] Kai Shu, Deepak Mahudeswaran, and Huan Liu. 2019. FakeNewsTracker: a tool for fake news collection, detection, and visualization. *Computational and Mathematical Organization Theory* 25, 1 (2019), 60–71.

[16] Statista. 2020 (accessed January 8, 2021). *Most popular mobile social networking apps in the United States as of September 2019.* https://www.statista.com/statistics/248074/most-popular-us-social-networking-apps-ranked-by-audience/

[17] Rima S. Tanash, Zhouhan Chen, Tanmay Thakur, Dan S. Wallach, and Devika Subramanian. 2015. Known Unknowns: An Analysis of Twitter Censorship in Turkey. In *Proceedings of the 14th ACM Workshop on Privacy in the Electronic Society* (Denver, Colorado, USA) *(WPES '15)*. Association for Computing Machinery, New York, NY, USA, 11–20. https://doi.org/10.1145/2808138.2808147

[18] Tom Wilson and Kate Starbird. 2021 (accessed January 13, 2021. *Cross-platform disinformation campaigns: lessons learned and next steps.*

[19] Kai-Cheng Yang, Francesco Pierri, Pik-Mai Hui, David Axelrod, Christopher Torres-Lugo, John Bryden, and Filippo Menczer. 2020. The COVID-19 Infodemic: Twitter versus Facebook. arXiv:2012.09353 [cs.SI]

[20] John Zarocostas. 2020. How to fight an infodemic. *The Lancet* 395, 10225 (2021/02/19 2020), 676. https://doi.org/10.1016/S0140-6736(20)30461-X