Proceedings of the Nineteenth International AAAI Conference on Web and Social Media (ICWSM 2025)

Labeled Datasets for Research on Information Operations

Ozgur Can Seckin[*†1], **Manita Pote**[†1], **Alexander C. Nwala**[†2], **Lake Yin**[1], **Luca Luceri**[3], **Alessandro Flammini**[1], **Filippo Menczer**[1]

[1]Observatory on Social Media, Indiana University, Bloomington
[2]William & Mary, Williamsburg, Virginia
[3]University of Southern California, Information Sciences Institute, Los Angeles

Abstract

Social media platforms have become a hub for political activities and discussions, democratizing participation in these endeavors. However, they have also become an incubator for manipulation campaigns, like information operations (IOs). Some social media platforms have released datasets related to such IOs originating from different countries. However, we lack comprehensive control data that can enable the development of IO detection methods. To bridge this gap, we present new labeled datasets about 26 campaigns, which contain both IO posts verified by a social media platform and over 13M posts by 303k accounts that discussed similar topics in the same time frames (control data). The datasets will facilitate the study of narratives, network interactions, and engagement strategies employed by coordinated accounts across various campaigns and countries. By comparing these coordinated accounts against organic ones, researchers can develop and benchmark IO detection algorithms.

Introduction

Originally perceived as tools to democratize information, social media platforms have also evolved into channels for the dissemination of conspiracy theories and questionable information (Lazer et al. 2018; Vosoughi, Roy, and Aral 2018). The proliferation of inauthentic accounts (Shao et al. 2018; Yang and Menczer 2024), political sock puppets (Woolley and Howard 2018a), and state-sponsored operators (Badawy et al. 2019) exacerbates the vulnerability of social media to misleading narratives and propaganda. Deceptive, orchestrated campaigns, known as *information operations* (IOs), have been defined as coordinated efforts to manipulate or corrupt public debate within a target audience for a strategic goal (Facebook 2021).

Accounts participating in IOs employ various strategies, ranging from artificial amplification for promoting content to targeted attacks on specific account communities. IOs can be complex, sophisticated efforts characterized by several elements, including:

1. **Domains**: The primary focus can vary widely across operations, but in most cases it is political.
2. **Goals**: IOs can encompass a wide range of objectives, such as advancing narratives about politics, amplifying pro- or anti-government content, and spreading propaganda and/or disinformation.
3. **Targets**: The audiences subject to manipulation. These can vary in scale, ranging from small groups to entire countries and geopolitical regions.
4. **Tactics**: The methods employed to achieve an IO's strategic goals. These tactics can range from simple actions (e.g., spamming) to complex ones involving coordination, obfuscation, or impersonation of political figures.
5. **Platforms**: The activities associated with IOs can extend across multiple online platforms.
6. **Users**: The accounts involved in IOs may range from small to large sets of accounts, human-operated or automated, and may involve fake or compromised profiles.

One of the first documented IOs was uncovered in a South Korean social media platform in 2012 (Keller et al. 2020). Since then, IOs have been reported globally and have emerged as a global threat (Bradshaw and Howard 2017; Woolley and Howard 2018b; Stanford Internet Observatory 2021). A well-known case is the interference in the 2016 U.S. Presidential Election by the Russian Internet Research Agency (IRA) (Senate Select Committee on Intelligence 2019). As a result of the potential adverse effects of IOs, major social media platforms like Twitter (now X), Facebook, and Reddit started releasing reports and data on IOs identified on their platforms.

Research on IOs has mainly utilized Twitter data to characterize IO accounts, uncover their tactics, and propose methods for their detection Nwala, Flammini, and Menczer (2023); Luceri et al. (2024); Cima et al. (2024); Saeed et al. (2024). This research requires control data that include the activity of a baseline or negative class, i.e., legitimate accounts engaged in conversations similar to those of IO campaigns. The availability of control datasets for multiple IO campaigns is critical for at least two reasons. First, existing datasets are obsolete, private, or tied to a specific campaign, hindering the development of detection models that can generalize across IOs with varying origins, contexts, and levels of sophistication (Badawy et al. 2019; Cima et al. 2024). Second, following the shutdown of the Twitter API (Murtfeldt et al. 2024), it has become prohibitively expensive for

*Corresponding author. Email: oseckin@iu.edu.
†Equal contributions.

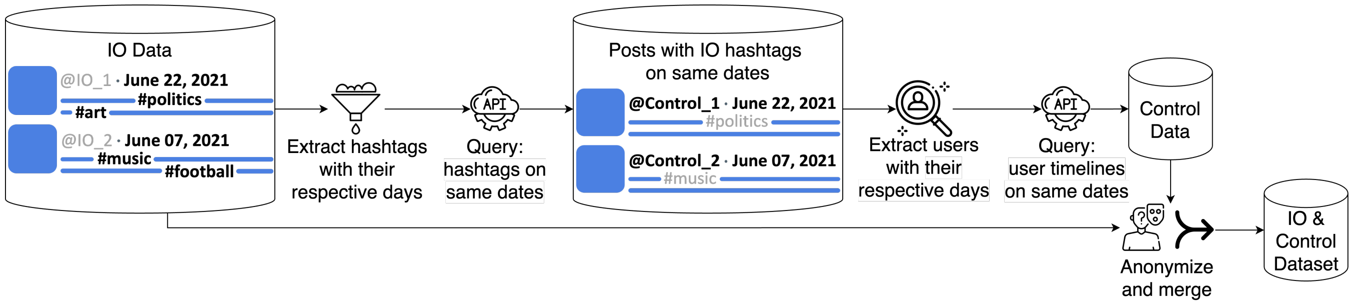

Figure 1: Data Collection and Curation Pipeline

researchers to collect new data related to IOs. Nevertheless, control datasets for multiple IO campaigns remain underdeveloped. To bridge this gap, here we introduce new datasets that include both IO data and related control data covering legitimate accounts involved in online discussions across 26 distinct, verified IO campaigns from different countries. Our datasets are anonymized to preserve user privacy. Such comprehensive, accessible, and contextually enriched datasets provide an invaluable resource for researchers to analyze and characterize IOs in various contexts, as well as develop new methodologies to detect and counter IOs.

The datasets are available at https://doi.org/10.5281/zenodo.14141549.

Related Work

Characterizing IOs and Their Tactics

Previous research has explored the activity of IO accounts, such as state-sponsored trolls targeting the #BlackLivesMatter movement (Stewart, Arif, and Starbird 2018) and the 2016 U.S. Election (Badawy et al. 2019), and the differences in their activities across campaigns (Zannettou et al. 2019b). Some studies have shown how IO accounts leverage inauthentic or automated accounts to increase their prominence and artificially amplify messages (Linvill and Warren 2020; Elmas 2023), while being resilient to large-scale shutdown (Merhi, Rajtmajer, and Lee 2023). Researchers have reported on different tactics used by IO accounts, such as trolling (Zannettou et al. 2019a), flooding through political cartoons (Fecher et al. 2022), hashtag hijacking (Ong and Cabañes 2018), deletion of content to avoid detection (Torres-Lugo et al. 2022), disinformation and propaganda (Woolley and Howard 2018b), political memes (Rowett 2018; Zannettou et al. 2020; Ng, Moffitt, and Carley 2022), and advertising or paid digital influencers (Ong and Cabañes 2018).

Detecting Inauthentic Coordinated Behaviors

A variety of unsupervised and supervised machine-learning models have been developed to identify messages and coordinated accounts linked to IOs. Unsupervised methods to detect coordination include multi-view modularity clustering Uyheng, Cruickshank, and Carley (2022), a Bayesian approach based on narrative and account characteristics (Smith, Ehrett, and Warren 2024), and network-based models analyzing similarities in sharing activities (Pacheco et al. 2021; Luceri et al. 2024; Cima et al. 2024). Some supervised learning approaches classify posts to determine if a message is part of an IO. These methods employ linguistic features to train off-the-shelf machine learning algorithms (Addawood et al. 2019; Im et al. 2020) or large language models (Luceri, Boniardi, and Ferrara 2024). Other classifiers attempt to differentiate between IO accounts and organic accounts based on their behaviors. Methods for this task have employed inverse reinforcement learning (Luceri, Giordano, and Ferrara 2020), sequences of account actions (Nwala, Flammini, and Menczer 2023; Ezzeddine et al. 2023), and Hawkes modeling (Kong et al. 2023). Coordinated account detection has also exploited content (Alizadeh et al. 2020) and generative models (Sharma et al. 2021). Recent machine-learning models have leveraged features extracted from cross-campaign data, such as third-party applications and reposting patterns, to detect accounts from previously unseen campaigns (Saeed et al. 2024).

Collecting IO Control Datasets

Combating information operations requires accessible data about inauthentic coordinated activity. Early examples of such data appeared after 2016, when Congress investigated Russian interference in the U.S. Election. The investigation uncovered malicious trolls and bots linked to the IRA that spread biased information, prompting social media platforms to address coordinated and fake activities. The House Intelligence Committee released redacted PDF files of 3,517 Facebook ads bought by the IRA.[1] Meta also shared blog posts and threat reports to inform users about its efforts against misinformation.[2] Reddit identified 944 suspicious accounts possibly linked to the IRA[3] along with a transparency report.[4]

Although these datasets offer a good descriptive understanding of coordinated behavior, they fall short of helping differentiate between coordinated and regular accounts. Researchers have collected control datasets for specific cam-

[1] github.com/simonw/russian-ira-facebook-ads-datasette
[2] about.fb.com/news/tag/coordinated-inauthentic-behavior/
[3] www.reddit.com/wiki/suspiciousaccounts/
[4] www.reddit.com/r/announcements/comments/8bb85p/reddits_2017_transparency_report_and_suspect/

paigns released by Twitter, allowing for a more nuanced comparison between these groups. For example, Badawy et al. (2019) collected tweets based on a list of hashtags and keywords related to the 2016 U.S. Presidential Election. Alizadeh et al. (2020) curated control data for multiple IO campaigns by combining random account IDs and accounts who followed at least five American politicians. Vargas, Emami, and Traynor (2020) collected four control groups for IO campaigns in 2018–2019: a political community (U.S. Congress and UK Parliament members), a non-political community (academic and security researchers), a group based on a trending hashtag, and popular accounts selected by a random walk through the follower network. Smith, Ehrett, and Warren (2024) curated control datasets for four campaigns by querying specific keywords associated with each campaign within the same period. Cima et al. (2024) compiled a control dataset for two IO campaigns. They selected the top hashtags used by the coordinated accounts in each campaign and collected all tweets that included at least one of these hashtags during the last 4 months of the campaigns. Similarly, Guo and Vosoughi (2022) introduced control datasets for 28 IO campaigns from 14 countries and spanning 2015–2018. The authors included the tweets containing the top hashtags used by the coordinated accounts in each month and gathered the tweets from a 1% sample of real-time tweets provided by the Internet Archive.[5]

While existing control datasets are valuable, they have significant limitations we aimed to address when curating the control datasets presented in this paper. First, they typically include control accounts who have posted on similar topics to those of the IO accounts, but they fail to include posts from these control accounts on unrelated topics. In contrast, our datasets include posts from control accounts discussing similar topics as well as other posts from their timelines, providing a more comparable set of posts to those of IO accounts. Second, while most previous work provides control datasets for only a small number of IO campaigns and countries, our datasets include data for 26 campaigns from multiple countries, involving operations backed by several states. Third, some existing datasets were assembled using a small sample of public posts. Our datasets have 100% coverage of control data. Finally, most previous datasets provide only IDs and need to be "re-hydrated" using platform APIs that are currently inaccessible to researchers. We provide anonymized data that complies with platform privacy policies and does not need require re-hydration.

IO Datasets

We curated control datasets for a number of IO datasets. We started from a public archive containing data on state-sponsored IOs, made available by a social media platform on their transparency website. The campaigns spanned several years and countries.

After taking down the IO campaigns, the platform released corresponding datasets. We focus on 26 of the campaigns, attributed by the platform to 16 state actors. Each campaign is identified by a state actor (e.g., Russia or Catalonia) and a number to distinguish campaigns from the same state actor. A campaign contains records about the entire timelines of the IO accounts associated with the campaign — whether or not each post in these timelines is part of the campaign. For example, a hijacked account might have been repurposed to become part of a campaign. For each campaign, we gathered these records from multiple files after consulting documentation by the platform, in the form of a README file and blog post, and cross-referencing the numbers of IO accounts.

We found four groups of campaigns that could be merged according to platform documentation:

1. **Venezuela_1** and **Venezuela_2**
2. **China_1** and **China_2**
3. **Russia_1** and **Russia_4**
4. **Iran_2**, **Iran_3**, and **Iran_4**

However, we deliberately refrained from merging these datasets, leaving it for researchers to do so if necessary or beneficial for their investigations.

Control Datasets

What constitutes a good control dataset for a given IO campaign is debatable. One of the main tactics of IO actors is to coordinate efforts to get engagement, interaction, and trust from organic accounts, who can be both targets and unwitting collaborators in pushing their agenda (Starbird 2019). However, not every organic account is susceptible to IO influence attempts. Some can discuss similar topics without endorsing the IO messages. Therefore, a suitable control would consist of accounts who engage in the same topics at the same time, but who are not part of the IO. One method to capture such topics is by extracting hashtags used in the IOs.

Data Curation

Control Data Collection

Fig. 1 illustrates our data collection and curation pipeline. We first gathered all the hashtags used by IO accounts in a given IO campaign to identify control accounts discussing similar topics. Note that this approach is followed separately for each of the campaigns, e.g., the control datasets for **China_1** and **China_2** were collected independently. Subsequently, we utilized these hashtags as queries to identify accounts that had posted on the same dates and used the same hashtags as the IO accounts, using the platform's Application Program Interface (API). In the final step, we reconstructed the daily timelines of the control accounts by extracting up to 100 messages posted on the same dates as the IO accounts. For instance, if an IO account posted with the hashtag `#election` on September 10, 2019, we created the corresponding control account list by identifying accounts who also used that hashtag on the same date. We then pulled their timelines from the API for that specific date. We collected control data for each of the 26 IO campaigns. Our selection of these campaigns was in part dictated by API and computational limitations.

[5] archive.org/details/twitterstream

Campaign	Years		Accounts		Posts		Reposts		Replies		Hashtags		URLs		Mentions	
	IO	Ctrl	IO	Ctrl	IO	Ctrl	IO	Ctrl	IO	Ctrl	IO	Ctrl	IO	Ctrl	IO	Ctrl
Armenia	14-20	14-20	31	1,767	72,960	69,118	522	35,167	263	6,253	866	26,859	89,989	58,022	607	31,127
Bangladesh	09-18	10-18	11	929	26,212	35,974	1,020	16,640	310	2,504	950	19,113	23,532	28,831	618	16,541
Catalonia	11-19	11-19	76	2,607	9,489	115,520	6,096	95,436	2,201	9,540	699	25,049	2,226	75,634	2,260	39,974
China_1	08-19	08-19	699	42,120	1,898,108	1,772,646	356,054	1,026,404	178,496	205,897	25,024	226,561	553,156	915,030	139,206	524,836
China_2	07-19	08-19	191	35,806	1,708,078	1,705,976	671,070	338,756	223,878	293,122	74,879	176,105	629,535	526,831	170,215	588,095
Cuba	10-20	10-20	503	30,099	4,802,243	1,353,088	3,341,163	626,977	151,121	128,785	88,022	178,692	1,586,187	777,179	141,107	391,511
Ecuador	10-19	10-19	787	21,138	700,240	770,289	577,749	456,703	46,159	67,271	32,283	145,667	135,890	483,906	49,739	237,738
Egypt_UAE	12-19	12-19	240	370	214,898	14,712	71,883	8,164	11,584	1,211	33,864	5,174	155,877	7,690	10,851	4,972
Ghana_Nigeria	14-20	14-20	60	1,166	39,964	43,445	16,310	30,304	8,697	4,995	8,340	15,247	15,130	29,345	11,168	28,054
Iran_1	10-18	12-18	660	5,015	1,122,936	199,853	232,337	84,206	339,350	17,306	105,598	55,882	612,899	137,711	296,150	72,244
Iran_2	09-20	11-20	389	6,615	1,302,012	329,781	502,373	148,585	30,405	30,822	90,857	50,394	928,318	140,738	44,628	72,115
Iran_3	11-19	14-16	210	2,240	1,963,141	91,297	1,609,566	48,025	222,737	7,596	156,904	18,144	388,313	41,499	102,584	22,787
Iran_4	08-19	13-19	2,519	7,612	254,781	306,804	90,324	181,129	62,800	27,974	33,806	76,313	76,287	201,689	37,031	105,735
Iran_5	20-20	20-20	104	1,247	2,450	42,955	551	30,996	663	5,706	148	14,280	1,166	27,725	1,024	28,314
Iran_6	09-20	10-20	209	16,885	560,571	729,241	100,446	467,772	89,139	61,339	37,964	147,664	462,034	554,717	56,718	240,109
Qatar	10-20	13-20	29	19,481	220,254	805,934	163,470	548,761	5,647	70,332	16,347	125,488	27,757	385,300	23,036	162,002
Russia_1	09-18	09-18	3,293	31,303	1,826,344	1,823,859	697,697	406,516	49,039	138,448	60,876	176,935	1,115,538	910,924	188,637	342,223
Russia_2	10-18	11-18	361	2,175	920,761	90,291	712,800	36,534	72,380	9,968	82,946	25,553	460,929	51,324	137,222	36,283
Russia_3	19-20	20-20	5	780	1,368	31,741	494	23,135	72	4,572	205	9,238	803	19,129	242	17,770
Russia_4	09-20	10-20	24	36,533	68,914	1,776,226	21,775	976,712	5,587	168,996	9,439	322,534	42,268	1,251,577	8,118	485,809
Russia_5	13-20	14-20	51	13,077	26,684	683,219	3,746	443,987	2,870	67,512	6,640	163,071	21,308	546,679	3,534	225,216
Spain	19-19	19-19	216	1,681	56,712	62,778	27,042	50,549	8,985	5,589	1,413	15,792	10,670	41,796	3,558	33,268
Thailand	15-20	18-20	455	2,549	21,385	121,738	11,199	109,226	6,844	5,157	1,615	17,797	2,871	66,251	1,432	36,824
UAE	11-19	11-19	3,658	10,545	1,325,530	382,262	644,486	233,446	99,152	38,579	92,920	94,113	402,945	229,095	105,744	125,456
Venezuela_1	15-18	16-18	578	6,327	984,980	261,310	1,416	164,960	5	31,784	12,472	41,680	1,645,110	182,546	305	87,065
Venezuela_2	12-19	12-19	33	3,865	569,453	156,989	28,959	99,952	125	18,190	930	33,510	892,852	113,081	563	62,706

Table 1: Descriptive Statistics. For IO and control data in each dataset, we report the time frame, numbers of accounts, posts, replies, reshares, unique hashtags, URLs, and mentioned accounts. All years are in the 2000s. Counts for hashtags, URLs, and mentioned accounts reflect unique values.

We removed from the IO data the columns without corresponding fields in the control data from the API. Similarly, we removed from the control data any fields without corresponding columns in the IO data. Finally, we aligned the field names to ensure consistency between IO and control data.

The merged datasets have the following fields: *postid, post_text, application_name, post_language, in_reply_to_postid, in_reply_to_accountid, post_time, accountid, account_profile_description, follower_count, following_count, account_creation_date, is_repost, reposted_accountid, reposted_postid, hashtags, urls, account_mentions, is_control*. Records from IO campaigns are marked as `False` in the *is_control* column, while control records are marked as `True`.

Before making the dataset publicly accessible, it is imperative to honor the privacy of users. Therefore, we anonymized all personally identifiable information (PII) through a one-way hashing algorithm. This includes account IDs, post IDs, URLs and usernames. The last two items are hashed, even when occurring within post text and profile description. Location information is removed. We ensured consistent anonymization so that accounts mentioned in IO and control posts can be linked through their hashed IDs while still safeguarding privacy.

Each dataset was segmented into files of 50,000 posts each. This resulted in a total of 703 files across all campaigns, with the largest campaign (**Cuba**) consisting of 124 files.

Descriptive and Coverage Statistics

Table 1 provides a quantitative overview of the datasets. The dataset with the longest duration spans 12 years (**China_2**), while the shortest lasts less than a year (**Spain** and **Iran_5**). Note that the first post in a dataset does not necessarily represent the beginning of a campaign, since IO accounts could exist and post before they were involved in a campaign.

Account and post counts vary significantly from one campaign to another, as do the numbers of reposts and replies. This indicates that the levels of engagement differ across campaigns. We also observe that the number of control accounts always exceeds that of IO accounts, while the activity per IO accounts tends to be higher. Note, however, that this is likely a consequence of the different way in which the control data was collected — individual days rather than full timelines.

Fig. 2 reports on the coverage of IO campaigns by control data in terms of accounts, hashtags, and time. Account coverage refers to the portion of accounts mentioned, replied to, or reposted by IO accounts who were also mentioned, replied to, or reposted by control accounts. Account coverage varied between 3% (**Iran_3**) and 58% (**Catalonia**), with a median of 17%. The median coverage of IO hashtags was 31%, with a range from 3% (**Iran_3**) to 73% (**Venezuela_2**). Finally, the median date coverage of IO data was 44%, ranging from 15% (**Egypt_UAE**) to 84% (**Ecuador**).

Discussion

Researchers can leverage the datasets introduced in this paper to characterize different dimensions of inauthentic coordinated activity. This includes analyzing differences between the narratives, behavioral patterns, network structures, and temporal activities of IO accounts, across campaigns and countries. Researchers can also investigate the tactics employed by different campaigns, for example how IO accounts engage with their targets. In addition, the datasets

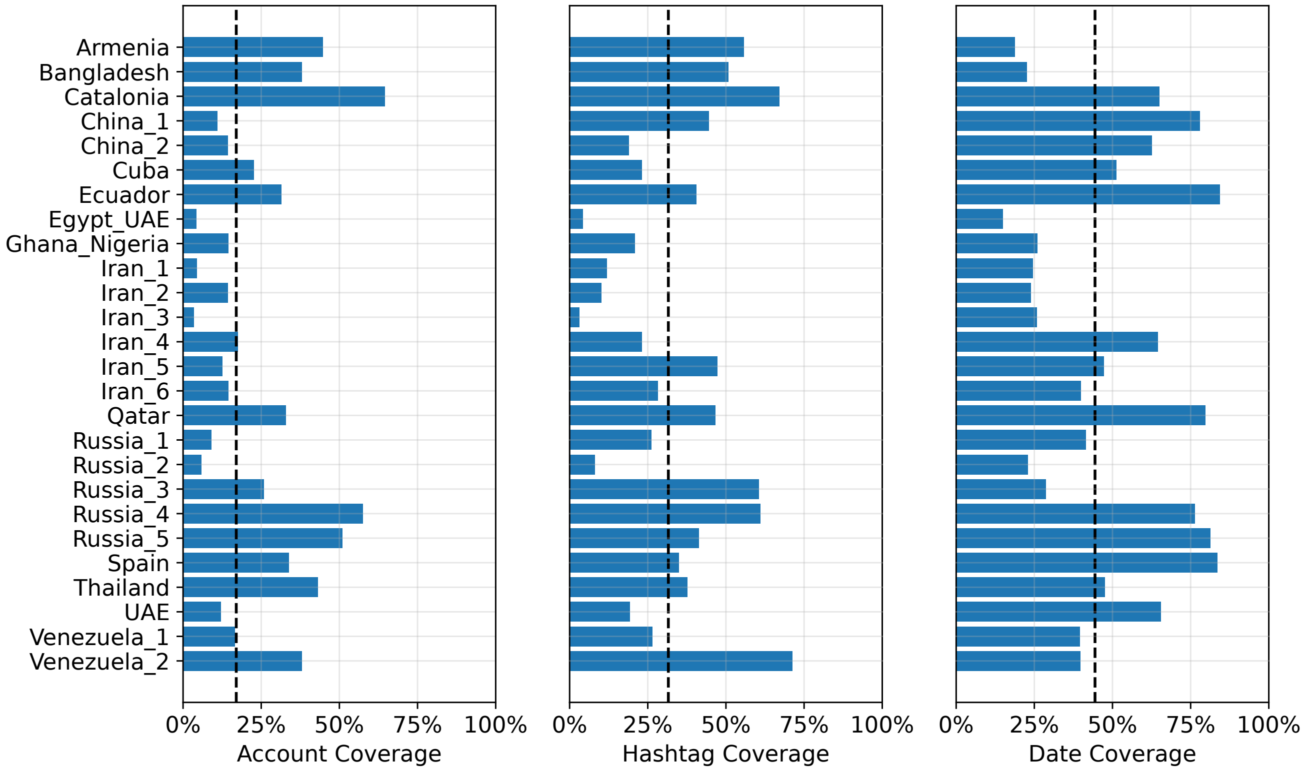

Figure 2: Data Coverage. Left: Percentage of accounts mentioned/replied to/reposted by IO accounts that can be found in the control dataset. Middle: Percentage of IO hashtags that were also used by control accounts. Right: Percentage of days that IO accounts posted and that were covered by control accounts. Dashed lines indicate median values.

contribute systematic control data for IO campaigns, empowering researchers to design and evaluate methods for their detection.

There are some limitations in the way the control data was collected. The choice to condition the collection of control accounts on the co-sharing of IO hashtags means that the quality of the control sample depends on IO hashtags being a quality proxy for IO content, which is not always the case. For example, some IO accounts may have employed tactics that did not involve the use of any hashtags. In such cases, our datasets may lack accurate control counterparts for those accounts.

Even when a campaign does actively use hashtags, some of the hashtags may be very popular and generic. Using such hashtags to select control accounts can introduce a sampling bias towards more active accounts. Because of the inclusion of popular hashtags, control data may have non-IO coordinated activity, which would be inaccurately labeled in the dataset. As an example, suppose an IO account used the hashtag `#crypto`. The control data may include accounts managed by a spammer to push cryptocurrency manipulation. While these accounts would be engaged in coordinated inauthentic activity, they would not be labeled as IO in the dataset, therefore an evaluation would wrongly mark them as false positive errors if they were correctly identified by a detection algorithm.

While the entire user timeline is present for IO accounts, control account timelines are cropped at 100 posts and exclude posts that occurred after the date on which they met the inclusion criteria. This results in a significant misalignment in the active days for individual control accounts compared to individual IO accounts. Although the posts by control accounts have reasonable *collective* coverage of the IO activity dates (Fig. 2), a *single* control account timeline tends to cover relatively short periods while the IO account timelines can span years. Research who wish to mitigate the effect of this bias might selectively choose IO and control accounts with matching temporal activity.

In addition to temporal bias, our coverage of IO and control hashtag use is asymmetric. Control accounts may cover topics not present among IO accounts since we do not exclude control content that is not matched to IO hashtag use.

All of these discrepancies might impact detection algorithms or descriptive studies aiming to differentiate IO and control accounts.

Finally, while IOs can spread across multiple social media platforms (Wilson and Starbird 2021), our datasets are focused on a single platform. Considering how campaigns unfold on multiple online social networks is a valuable direction for future studies.

Contributions

OCS, MP, and ACN contributed to the conceptualization of the project, investigation, data curation, validation and analysis, and writing of the manuscript. OCS prepared the figures. LY contributed to data anonymization. LL conceived a consistent anonymization strategy to link IO and control posts and edited the manuscript. FM and AF contributed to the conceptualization of the project, acquired funding, supervised the project, and reviewed and edited the manuscript.

Ethical Statement

The collection and release of the dataset was deemed exempt from review by the Indiana University IRB (protocol 1102004860) as it contains de-identified public data with minimal or no risks to the subjects. Although PII has been anonymized to protect privacy, sensitive information (such as political opinions, racial or ethnic origins, or religious beliefs) may still be inferred from the content of posts. If combined with other data sources, this could lead to unintended re-identification. Therefore, researchers should handle the data responsibly, ensure compliance with ethical guidelines, and refrain from attempting to de-anonymize or link the data with other datasets in a manner that could compromise individual privacy. Finally, we allow researchers to download the datasets in small segments to comply with platform terms.

Acknowledgments

We are grateful to David Axelrod and Filipi Nascimento Silva for helpful discussions on the limitations of the datasets. This work was supported in part by the Knight Foundation, Craig Newmark Philanthropies, and Volkswagen Foundation. The funders had no role in study design, data collection, and analysis, the decision to publish, or the preparation of the manuscript.

References

Addawood, A.; Badawy, A.; Lerman, K.; and Ferrara, E. 2019. Linguistic Cues to Deception: Identifying Political Trolls on Social Media. In *Proc. Intl. AAAI Conf. on Web and Social Media*, volume 13, 15–25.

Alizadeh, M.; Shapiro, J. N.; Buntain, C.; and Tucker, J. A. 2020. Content-based Features Predict Social Media Influence Operations. *Science Advances*, 6(30).

Badawy, A.; Addawood, A.; Lerman, K.; and Ferrara, E. 2019. Characterizing the 2016 Russian IRA Influence Campaign. *Social Network Analysis and Mining*, 9: 1–11.

Bradshaw, S.; and Howard, P. 2017. Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation. Technical report, University of Oxford.

Cima, L.; Mannocci, L.; Avvenuti, M.; Tesconi, M.; and Cresci, S. 2024. Coordinated behavior in information operations on Twitter. *IEEE Access*.

Elmas, T. 2023. Analyzing Activity and Suspension Patterns of Twitter Bots Attacking Turkish Twitter Trends by a Longitudinal Dataset. In *Companion Proc. ACM Web Conf.*, 1404–1412.

Ezzeddine, F.; Ayoub, O.; Giordano, S.; Nogara, G.; Sbeity, I.; Ferrara, E.; and Luceri, L. 2023. Exposing Influence Campaigns in the Age of LLMs: A Behavioral-based AI Approach to Detecting State-Sponsored Trolls. *EPJ Data Science*, 12(1): 46.

Facebook. 2021. Threat Report: The State of Influence Operations 2017-2020. Technical report, Facebook.

Fecher, L.; Reich, T.; Taylor, J.; and Warren, P. 2022. Oh, The Places You'll Guo! The Tactics and Impact of a Chinese Multilingual Narrative Flooding Campaign through Political Cartoons. https://open.clemson.edu/cgi/viewcontent.cgi?article=1000&context=mfh_ci_reports.

Guo, X.; and Vosoughi, S. 2022. A large-scale longitudinal multimodal dataset of state-backed information operations on Twitter. In *Proc. of the Intl. AAAI Conf. on Web and Social Media*, volume 16, 1245–1250.

Im, J.; Chandrasekharan, E.; Sargent, J.; Lighthammer, P.; Denby, T.; Bhargava, A.; Hemphill, L.; Jurgens, D.; and Gilbert, E. 2020. Still Out There: Modeling and Identifying Russian Troll Accounts on Twitter. In *12th ACM Conf. on Web Science*.

Keller, F. B.; Schoch, D.; Stier, S.; and Yang, J. 2020. Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign. *Political communication*.

Kong, Q.; Calderon, P.; Ram, R.; Boichak, O.; and Rizoiu, M.-A. 2023. Interval-censored Transformer Hawkes: Detecting Information Operations using the Reaction of Social Systems. In *Proc. ACM Web Conf.*, 1813—-1821.

Lazer, D.; Baum, M.; Benkler, Y.; Berinsky, A.; Greenhill, K.; Menczer, F.; Metzger, M.; Nyhan, B.; Pennycook, G.; Rothschild, D.; Schudson, M.; Sloman, S.; Sunstein, C.; Thorson, E.; Watts, D.; and Zittrain, J. 2018. The science of fake news. *Science*, 359(6380): 1094–1096.

Linvill, D. L.; and Warren, P. L. 2020. Engaging with Others: How the IRA Coordinated Information Operation Made Friends. *Misinformation Review*.

Luceri, L.; Boniardi, E.; and Ferrara, E. 2024. Leveraging Large Language Models to Detect Influence Campaigns in Social Media. In *Companion Proc. of the 2024 ACM Web Conf.*

Luceri, L.; Giordano, S.; and Ferrara, E. 2020. Detecting Troll Behavior via Inverse Reinforcement Learning: A Case Study of Russian Trolls in the 2016 US Election. In *Intl. AAAI Conf. on Web and Social Media*, volume 14, 417–427.

Luceri, L.; Pantè, V.; Burghardt, K.; and Ferrara, E. 2024. Unmasking the web of deceit: Uncovering coordinated activity to expose information operations on Twitter. In *Proc. of the 2024 ACM Web Conf.*

Merhi, M.; Rajtmajer, S.; and Lee, D. 2023. Information Operations in Turkey: Manufacturing Resilience with Free Twitter Accounts. In *Proc. Intl. AAAI Conf. on Web and Social Media*.

Murtfeldt, R.; Alterman, N.; Kahveci, I.; and West, J. D. 2024. RIP Twitter API: A eulogy to its vast research contributions. *arXiv preprint arXiv:2404.07340*.

Ng, L. H. X.; Moffitt, J.; and Carley, K. M. 2022. Coordinated through a Web of Images: Analysis of Image-based Influence Operations from China, Iran, Russia, and Venezuela. *Preprint arXiv:2206.03576*.

Nwala, A. C.; Flammini, A.; and Menczer, F. 2023. A Language Framework for Modeling Social Media Account Behavior. *EPJ Data Science*, 12(1): 33.

Ong, J. C.; and Cabañes, J. V. A. 2018. Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines. Tech. rep., UMass Amherst. https://doi.org/10.7275/2cq4-5396.

Pacheco, D.; Hui, P.-M.; Torres-Lugo, C.; Truong, B. T.; Flammini, A.; and Menczer, F. 2021. Uncovering Coordinated Networks on Social Media: Methods and Case Studies. In *Proc. Intl. AAAI Conf. on Web and Social Media*, volume 21, 455–466.

Rowett, G. 2018. The Strategic Need to Understand Online Memes and Modern Information Warfare Theory. In *IEEE Big Data*.

Saeed, M. H.; Ali, S.; Paudel, P.; Blackburn, J.; and Stringhini, G. 2024. Unraveling the Web of Disinformation: Exploring the Larger Context of State-Sponsored Influence Campaigns on Twitter. *arXiv preprint arXiv:2407.18098*.

Senate Select Committee on Intelligence. 2019. Report of the Select Committee on Intelligence United States Senate on Russian Active Measures Campaigns and Interference in the 2016 U.S. Election Volume 2: Russia's Use of Social Media with Additional Views. https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf. Accessed: 2024-01-16.

Shao, C.; Ciampaglia, G. L.; Varol, O.; Yang, K.; Flammini, A.; and Menczer, F. 2018. The Spread of Low-credibility Content by Social bots. *Nature Communications*, 9: 4787.

Sharma, K.; Zhang, Y.; Ferrara, E.; and Liu, Y. 2021. Identifying Coordinated Accounts on Social Media through Hidden Influence and Group Behaviours. In *27th ACM SIGKDD Conf. on Knowledge Discovery & Data Mining*, 1441–1451.

Smith, D. H.; Ehrett, C.; and Warren, P. L. 2024. Unsupervised Detection of Coordinated Information Operations in the Wild. *arXiv preprint arXiv:2401.06205*.

Stanford Internet Observatory. 2021. Published reports of the Stanford Internet Observatory. https://github.com/stanfordio/publications.

Starbird, K. 2019. Disinformation's spread: bots, trolls and all of us. *Nature*, 571(7766): 449–450.

Stewart, L. G.; Arif, A.; and Starbird, K. 2018. Examining Trolls and Polarization with a Retweet Network. In *Proc. ACM WSDM Workshop on Misinformation and Misbehavior Mining on the Web (MIS2)*. https://api.semanticscholar.org/CorpusID:44033303.

Torres-Lugo, C.; Pote, M.; Nwala, A.; and Menczer, F. 2022. Manipulating Twitter through Deletions. In *Proc. Intl. AAAI Conf. on Web and Social Media*, 1029–1039.

Uyheng, J.; Cruickshank, I. J.; and Carley, K. M. 2022. Mapping State-sponsored Information Operations with Multi-view Modularity Clustering. *EPJ Data Science*, 11(1): 25.

Vargas, L.; Emami, P.; and Traynor, P. 2020. On the Detection of Disinformation Campaign Activity with Network Analysis. In *Proc. of the 2020 ACM SIGSAC Conf. on Cloud Computing Security Workshop*, 133–146.

Vosoughi, S.; Roy, D.; and Aral, S. 2018. The Spread of True and False News Online. *Science*.

Wilson, T.; and Starbird, K. 2021. Cross-platform Information Operations: Mobilizing Narratives & Building Resilience through both'Big'&'Alt'Tech. *Proc. of the ACM on Human-Computer Interaction*.

Woolley, S. C.; and Howard, P. N. 2018a. *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford University Press.

Woolley, S. C.; and Howard, P. N. 2018b. Conclusion: Political Parties, Politicians, and Computational Propaganda. In *Computational propaganda: Political parties, politicians, and political manipulation on social media*, 241–248. Oxford University Press.

Yang, K.-C.; and Menczer, F. 2024. Anatomy of an AI-powered Malicious Social Botnet. *Journal of Quantitative Description: Digital Media*, 4.

Zannettou, S.; Caulfield, T.; Bradlyn, B.; De Cristofaro, E.; Stringhini, G.; and Blackburn, J. 2020. Characterizing the Use of Images in State-Sponsored Information Warfare Operations by Russian Trolls on Twitter. *Proc. of the Intl. AAAI Conf. on Web and Social Media*.

Zannettou, S.; Caulfield, T.; De Cristofaro, E.; Sirivianos, M.; Stringhini, G.; and Blackburn, J. 2019a. Disinformation Warfare: Understanding State-sponsored Trolls on Twitter and their Influence on the Web. In *Companion Proc. WWW Conf.*, 218–226.

Zannettou, S.; Caulfield, T.; Setzer, W.; Sirivianos, M.; Stringhini, G.; and Blackburn, J. 2019b. Who Let the Trolls Out? Towards Understanding State-sponsored Trolls. In *Proc. 10th ACM Conf. on Web Science*, 353–362.

Ethical Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes
   (e) Did you describe the limitations of your work? Yes
   (f) Did you discuss any potential negative societal impacts of your work? Yes
   (g) Did you discuss any potential misuse of your work? Yes

(h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes

(i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes

2. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

(a) If your work uses existing assets, did you cite the creators? The IO data is no longer available to the public, therefore we do not name the source platform.

(b) Did you mention the license of the assets? The license of the new datasets is indicated in the data repository.

(c) Did you include any new assets in the supplemental material or as a URL? Yes, we provide the Zenodo URL and DOI for the datasets.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes, in the Ethical Statement section.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? Yes

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? Yes