

A Detection System for AI-Generated Profiles and Influence Campaigns

Nikhil Singla, Patrick Gerard, Saif Islam, Sujit Ponguluri, Kian Boon Tan

University of Southern California

Abstract

The rise of fake and AI-generated social media accounts involved in coordinated influence operations (IOs) poses a growing threat to the integrity of online discourse. We propose a dual-detection framework that identifies both AI-generated social media profiles and coordinated inauthentic behavior across platforms. Our architecture combines Support Vector Machine (SVM) classifiers trained on user-profile and behavioral patterns with graph-based models that incorporate textual similarities across user profiles. After training our models on historical datasets from multiple campaigns, including campaigns associated with China, Venezuela, and Russia, we evaluate the system on real-world social media data from the 2024 U.S. presidential election. These experiments show that the proposed framework outperforms the reported baseline methods. We conclude by presenting a scalable and adaptable framework for defending against evolving disinformation threats on social media platforms.

1 Introduction

This project proposes a multi-platform detection system designed to identify AI-generated social media profiles using machine learning techniques. With the growing threat of disinformation campaigns and coordinated inauthentic behavior online, detecting such accounts is important for preserving digital trust. The team uses a hybrid detection architecture that combines transformer-based language models, anomaly-detection algorithms, and graph neural networks to analyze textual content, behavioral patterns, and network structures. The project aims to develop a robust, scalable, and adaptable detection system using datasets from historical information operations and real-world social media posts collected from platforms such as Twitter and Truth Social. This effort addresses the immediate challenge of detecting AI-generated content while

contributing to the broader goal of securing social media ecosystems against evolving threats.

2 Literature Review

2.1 A Systematic Review of Machine Learning Approaches for Detecting Deceptive Activities on Social Media: Methods, Challenges, and Biases (Liu et al., 2024)

In this paper, the authors rigorously evaluate machine-learning approaches for detecting deceptive activity on social media. Using the Prediction model Risk Of Bias ASsessment Tool (PROBAST) (Wolff et al., 2019), they identify key sources of bias, as well as limitations related to evaluation metrics, data preprocessing, and model generalization across social media environments. These limitations provide a foundation for our work by identifying gaps that our approach seeks to address.

PROBAST evaluates machine-learning approaches through a series of questions targeting common limitations in four domains: sample selection and representativeness, data preprocessing, model development, and model evaluation (Wolff et al., 2019). The authors identify several recurring issues, including a focus on specific events or user groups during data sampling, inadequate handling of linguistic nuances, inconsistent hyperparameter tuning, and overreliance on accuracy as a performance metric in imbalanced datasets.

The paper also raises broader challenges for detecting deceptive social media content, including the unstructured nature of online discourse, the increasing sophistication of social media bots, class imbalance between legitimate and deceptive content, and limited generalization across platforms. These challenges motivate our dual-detection approach.

2.2 Linguistic Cues to Deception: Identifying Political Trolls on Social Media (Addawood et al., 2019)

This paper examines how linguistic patterns can be used to identify Russian troll accounts on Twitter during the 2016 election. The authors focus on deceptive communication strategies used to manipulate public opinion. Using a dataset of 13 million election-related posts, the authors build a model grounded in linguistic theory to detect subtle signs of deception. Their analysis shows that troll accounts often rely on persuasive language and simplify their messages through fewer nouns and shorter phrases. The authors use these patterns to train a classifier that achieves an 82% F1 score and 88% recall. Language-based features related to persuasion and message complexity are particularly useful for distinguishing trolls from legitimate users. The paper also describes challenges such as mislabeled data and potential dataset bias. Nevertheless, it demonstrates the value of linguistic analysis for detecting coordinated misinformation campaigns. Addawood et al.'s methodology advances the application of linguistic analysis to coordinated influence operations. Their approach draws on established deception-detection theories to create a framework tailored to political discourse on social media. By examining linguistic features across five dimensions—certainty, immediacy, cognitive complexity, non-immediacy, and persuasion—they identify patterns in troll accounts that are directly relevant to our work. The researchers find that Russian troll accounts exhibit consistent linguistic behaviors that differ from those of legitimate users. Beyond the persuasive language and simplified messaging described above, these accounts show reduced linguistic diversity, greater use of present-tense verbs, and greater reliance on first-person pronouns to create a sense of immediacy. These patterns suggest strategic communication choices designed to maximize influence while reducing the cognitive effort required of readers. Their findings on feature importance are particularly relevant to our research. The authors demonstrate that language-based features related to message complexity and persuasive content are more discriminative than metadata features alone. This insight informs our dual-detection approach, which incorporates both linguistic and behavioral signals to improve detection across evolving influence campaigns.

2.3 Coordinated Activity Modulates the Behavior and Emotions of Organic Users: A Case Study on Tweets about the Gaza Conflict (Dey et al., 2024)

This paper examines how coordinated users influence organic users on Twitter during the 2023 Gaza conflict. Using a dataset of 3.5 million tweets from more than 1.3 million users, the authors find that coordinated accounts spread content across the network, with more than one-third of their messages reshared by regular users. The study also identifies a gradual increase in engagement with coordinated content alongside increases in pessimism, sadness, and fear among organic users. These findings suggest that repeated exposure to coordinated activity can contribute to more extreme viewpoints. Dey et al.'s research provides critical insights into how coordinated campaigns impact organic user behavior beyond just spreading information. Their methodological innovation lies in combining network analysis with sentiment tracking over time, allowing them to measure the tangible effects of influence operations on the broader ecosystem. The researchers document several dynamics that inform our detection framework. They observe a measurable “emotional contagion” effect in which engagement with coordinated content correlates with increasing negative sentiment among users who repeatedly interact with coordinated accounts. Specifically, they report increases in pessimism, sadness, and fear. Particularly valuable is their development of metrics for quantifying coordination impact, including network transmission rates and emotional-trajectory mapping. They identify a multiphase response in which users initially exposed to coordinated content show moderate changes, followed by a plateau and then a secondary surge after continued interaction. This pattern suggests that influence operations may achieve their greatest impact through cumulative exposure rather than a single interaction, informing our analysis of temporal coordination patterns.

2.4 Unmasking the Web of Deceit: Uncovering Coordinated Activity to Expose Information Operations on Twitter (Luceri et al., 2024)

Recent work by Luceri et al. provides methodological advances for detecting coordinated influence operations on social media platforms. Their research, funded by DARPA and presented at the

2024 Web Conference, addresses the challenge of detecting increasingly sophisticated influence campaigns in the age of generative AI and large language models. Drawing on a dataset of 49 million tweets from verified influence campaigns originating in six countries—China, Cuba, Egypt, Iran, Russia, and Venezuela—the researchers develop an approach that outperforms traditional detection methods. Their methodology constructs a unified similarity network, termed a “Fused Network,” that integrates multiple behavioral signals rather than relying on individual coordination patterns. This approach parallels our Fused Behavioral Graph methodology and provides relevant empirical support. Their unsupervised machine-learning model, which uses network-topology features, achieves higher precision than traditional approaches for detecting influence campaigns. Their finding that different nation-state actors employ distinct coordination strategies is particularly relevant to our research; for example, co-retweeting is more prevalent in Cuban and Venezuelan campaigns than in Russian operations. This supports our use of graph-based coordination detection alongside profile-level authenticity assessment across actors and platforms.

3 Methodology

Our approach uses a machine-learning system with two connected modules. The first module detects users who may be involved in coordinated influence operations by analyzing their behavior within a network structure. We begin by constructing user-similarity graphs based on shared behavioral signals such as retweets, shared URLs, and hashtag patterns. In addition to standard baselines such as co-retweet and co-URL graphs, we introduce two enhanced variants: a fused behavioral graph that combines multiple signals, and a fused + bio graph that also incorporates semantic similarity between user bios using BERT-based embeddings. Once the graphs are constructed, we use Node2Vec to generate 128-dimensional node embeddings for each user. These embeddings encode structural information about users’ positions and connections in the network, allowing the model to capture patterns consistent with coordinated activity. We then train a multilayer perceptron (MLP) classifier on these embeddings to predict whether a user is part of an influence operation. The model is trained using labeled data from previously identified IO

campaigns, including Chinese, Russian, and Iranian state-backed accounts. During inference, each user receives an IO-likelihood score between 0 and 1 that reflects the model’s confidence in the user’s involvement in coordinated behavior.

The second module examines profile information and statistics to determine whether a profile is AI-generated or contains synthetic traits. It primarily uses text extracted from the user’s bio and several activity-related metrics. For example, it calculates a user-specific activity score from follower count and the number of posts. These features are grouped by activity level using ranges based on a bell curve applied to the full dataset. The resulting features are passed to a linear-kernel Support Vector Machine (SVM) for training. This model performs binary classification using default settings. Although it is designed to be simple and lightweight, it uses TF-IDF features and numerical metadata to support classification.

This module complements the graph-based IO detection module. Together, the two modules produce the final output.

3.1 Data Preprocessing

Our training data for the authenticity module consists of fake and real profiles extracted from the social media platform formerly known as Twitter and labeled according to class.

During data preprocessing, we remove columns unrelated to activity metrics or post content because the API used to collect the dataset is comprehensive. We also clean missing values and convert them to the expected format to streamline training. The text is processed by removing stop words, links, URLs, hashtags, and usernames. We convert all text to lowercase and remove extra spaces. Finally, because the data may contain Markdown tags, our filters remove any such tags from the text.

3.2 Feature Construction

Baseline Networks. We implement commonly used network construction techniques (Martha et al., 2013; Bollen et al., 2011; Cinelli et al., 2021):

- **Co-Retweet:** An undirected edge is placed between users if they retweet the same content.
- **Co-URL:** Connects users who share the same URLs, capturing shared external references.

- **Fast Retweet:** Captures temporal coordination by connecting users who retweet the same post within a short time window.
- **Hashtag Sequence:** Connects users who use similar or identical hashtag patterns in similar temporal order.

Our Methods. We introduce two additional similarity networks:

- **Fused Behavioral Graph:** This network creates an edge between any two users if an edge exists between them in any baseline behavioral graph—specifically, the Co-Retweet, Co-URL, or Hashtag Sequence network. We use this to increase coverage by capturing alignment across multiple behavioral signals (e.g., retweets and hashtags).
- **Fused + Bio Network:** This method extends the Fused Behavioral Graph by incorporating additional semantic similarity between users’ bios. We embed user bios using a BERT-based model. This is then used as additional information for the model.

3.3 Model Evaluation and Results

Our evaluation focuses on two main components: the graph-based IO detection module and the fake user detection module. Each is tested independently using a held-out 10% test set, with results averaged over five runs for stability. Models are evaluated using both F1 and AUC to account for class imbalance and overall discrimination power.

Graph-based IO Detection. We benchmark several graph construction methods and GNN architectures on labeled datasets from prior IO campaigns, including China, Venezuela, and Russia. Table 2 reports test performance across these settings. The fused behavioral graph and fused + bio network consistently outperform traditional baselines such as co-retweet or co-URL graphs. In terms of modeling, we find that Graph Attention Networks (GAT) yield the highest F1 scores in most cases. However, Node2Vec embeddings followed by a simple Multi-Layer Perceptron (Node2Vec+MLP) offer nearly comparable performance while significantly reducing training and inference time. Given the potential for real-time deployment, we adopt Node2Vec+MLP as our default configuration.

Fake User Detection. Benchmarking for the Fake User Detection module is done by using our model to make predictions on manually labeled data and then evaluating its precision, recall, and F1 scores. We test the model on over 500 data points and continue refining from the training set, until it performs well on this testing set.

Table 1: Performance of the fake user detection module (SVM with TF-IDF + metadata).

Metric	Mean Score	Support
F1 Score	0.91	267 + 293
AUC	0.92	560

Overall, these results support the use of our dual-detection framework: the graph-based module captures coordination and structural anomalies, while the profile-based module flags synthetic users based on text and behavior. Together, they provide a more comprehensive view of potential threats than either approach could achieve alone.

4 Applied Deployment: Detecting Threat Actors in the 2024 U.S. Election

To evaluate our framework in a high-stakes, real-world context, we deployed our dual-detection system on the 2024 U.S. Presidential Election Twitter Dataset (Balasubramanian et al., 2024). This dataset captures millions of tweets from politically engaged users and offers a timely lens into the dynamics of electoral discourse and potential manipulation efforts during a critical democratic event. Spanning from May 1, 2024 through November 20, 2024, a summary of the dataset’s core attributes is shown in Table 3.

Who Are The High-Risk Actors? We first applied our system to flag users exhibiting either high coordination likelihood or synthetic profile traits. Specifically, users were labeled as high-risk if their IO-likelihood score exceeded 0.95, as generated by our graph-based model trained on known influence operation datasets, or if they were classified as likely AI-generated or fake by our behavioral-text classifier.

To reduce false positives and focus our analysis on the most anomalous accounts, we filtered the flagged users to include only those whose IO-likelihood scores fell above the 99th percentile. As shown in Figure 1, the distribution of normalized IO scores is heavily right-skewed, with the

Table 2: Performance of similarity networks for the China, Venezuela, and Russia campaigns. The proposed methods outperform baseline approaches. Temporal decay performs well due to the clustered nature of posts in the original campaigns.

Method	Similarity Network	China		Venezuela		Russia	
		F1	AUC	F1	AUC	F1	AUC
Baseline	Co-Retweet	0.71 ± 0.02	0.60 ± 0.03	0.91 ± 0.02	0.74 ± 0.01	0.82 ± 0.02	0.84 ± 0.01
	Co-URL	0.76 ± 0.02	0.43 ± 0.06	0.84 ± 0.02	0.60 ± 0.04	0.86 ± 0.02	0.85 ± 0.01
	Fast Retweet	0.50 ± 0.01	0.38 ± 0.02	0.67 ± 0.02	0.54 ± 0.02	0.81 ± 0.01	0.83 ± 0.01
	Hashtag Sequence	0.51 ± 0.04	0.46 ± 0.03	0.58 ± 0.05	0.42 ± 0.03	0.83 ± 0.02	0.85 ± 0.01
Ours	Fused Behavioral Graph	0.82 ± 0.03	0.74 ± 0.01	0.92 ± 0.01	0.82 ± 0.03	0.88 ± 0.04	0.88 ± 0.01
	Fused + Bio Network	0.82 ± 0.02	0.76 ± 0.01	0.89 ± 0.02	0.82 ± 0.01	0.89 ± 0.02	0.88 ± 0.03

Table 3: Overview of the 2024 U.S. Presidential Election Twitter Dataset.

Attribute	Value
User Count	854,682
Post Count	26,016,477

99th percentile corresponding to a score of approximately 0.9737. We apply this stricter threshold to ensure that our downstream narrative and engagement analyses are concentrated on the users most indicative of coordinated or inauthentic activity, as determined by our model.

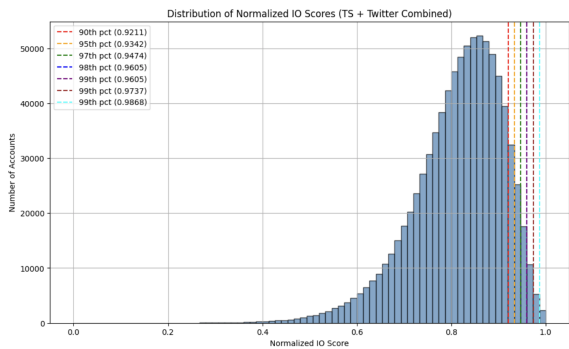


Figure 1: Distribution of normalized IO scores across Twitter and Truth Social users. Vertical lines represent key percentiles.

Table 4: Mean per-post engagement for high-IO vs. random users.

Metric	High-IO Mean	Random Mean
Like Count	21.02	20.03
Reply Count	1.95	1.90
Repost Count	23.38	17.07

While average engagement metrics between high-risk and random users are superficially simi-

lar, deeper analysis reveals more pronounced differences in tail behavior. Figure 2 shows the complementary cumulative distribution function (CCDF) of repost counts for both groups. The high-IO user distribution exhibits a markedly heavier tail, indicating that a subset of these accounts consistently generates content with far higher repost volume than typical users.

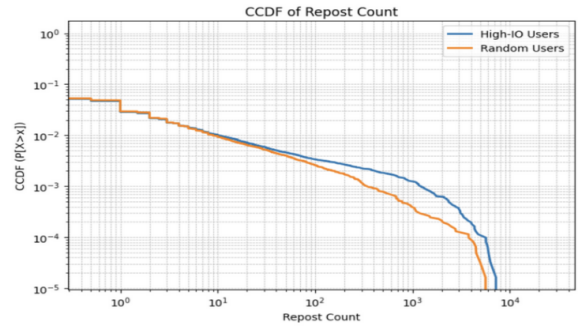


Figure 2: CCDF of repost counts for high-IO users vs. random users. High-IO users are overrepresented among extremely reposted content.

This heavy-tailed behavior highlights the amplification potential of high-risk accounts: while most users receive limited engagement, a small fraction of IO-flagged actors disproportionately dominate the attention economy. These actors may be positioned to reinforce divisive narratives, drive virality, and shape platform-wide discourse.

extbfNarratives Driven by High-Risk Actors. To analyze the narratives promoted by flagged users, we embedded their tweets using a multilingual variant of MPNet fine-tuned for semantic similarity. This embedding strategy captures both explicit and subtle similarities across linguistically diverse posts, enabling narrative grouping across users and platforms. Following prior work in multilingual narrative tracking (Hanley et al., 2024), we define a

narrative as a cluster of posts centered on a shared thematic concern—e.g., distrust in election procedures or delegitimization of political figures—that evolves across time and communities.

We apply DP-Means clustering to the embedded tweets, using cosine distance as the similarity metric with a threshold of 0.35 (i.e., minimum similarity 0.65). This value was selected to preserve semantic coherence without merging distinct themes, and was validated via human evaluation in line with methods from (Hanley and Durumeric, 2024).

The resulting narrative clusters reveal distinct patterns of messaging and amplification among high-risk users. Compared to a random sample of non-flagged users, high-IO accounts are disproportionately focused on a narrower set of emotionally charged and politically polarizing narratives. Table 5 summarizes the lean and dominant themes among high-IO versus random users.

Table 5: Narrative lean and topical breakdown by user group.

User Type	Right	Left	Neutral/Mixed
High-IO Users	32	7	10
Random Users	27	10	13

While both groups engage in political discourse, high-IO users are noticeably more focused in the topics they push. We see this in two key metrics: narrative entropy and the narrative Gini index.

Narrative entropy captures how many different narratives a user participates in. Lower entropy means a user tends to stay within a narrower range of topics. High-IO users have slightly lower entropy than random users (2.2 vs. 2.4), suggesting they stick to a few core narratives instead of bouncing between issues.

The Gini index measures how unevenly a user distributes posts across clusters. A higher Gini score indicates that a user posts heavily in a small number of narratives. Here, high-IO users are reported to have a Gini value of 0.83 versus 0.85 for random users, a comparison that should be checked before publication.

Together, these metrics suggest that high-IO users are not merely more active, but may also be more deliberate. Instead of reacting to a wide range of issues, they repeatedly push specific storylines, often the most polarizing ones. This behavior is less about generating new ideas and more about

amplifying familiar grievances that already exist in the ecosystem.

The dominant narratives promoted by high-IO users revolved around delegitimizing President Biden, often through conspiratorial or inflammatory claims about his mental fitness or alleged corruption. Many clusters also pushed claims of election fraud, especially around mail-in ballots and vote counting processes. Vice President Kamala Harris was frequently targeted as well, with posts questioning her competence or qualifications. A number of high-engagement posts expressed pro-Trump sentiment, particularly in the aftermath of the assassination attempt, framing him as a victim of political persecution. These themes were often layered with broader economic grievances, blaming inflation, immigration, or Democratic leadership for worsening financial conditions.

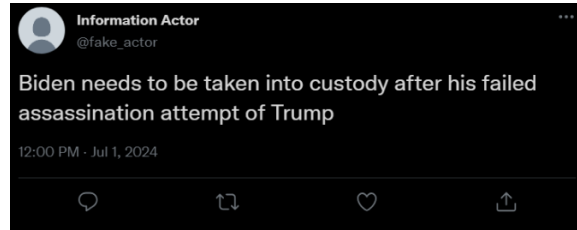


Figure 3: Example of a high-risk post falsely implicating President Biden in a failed assassination attempt (name and other metadata have been omitted). This post exemplifies the inflammatory and conspiratorial tone seen in several narrative clusters.

As illustrated in Figure 3, many posts within these clusters adopt aggressive framing and exploit emotionally resonant tropes to attract engagement. Importantly, high-risk users appear to target both far-right and anti-GOP leftist audiences—showing signs of narrative adaptation across ideological boundaries. For instance, one cluster contained posts portraying the GOP as ineffective or complicit, aiming to sow intra-party discord from a leftist perspective.

Taken together, this evidence suggests that high-risk actors do not invent new issues; they selectively amplify pre-existing, divisive discourse to erode trust in institutions, deepen polarization, and create a perception of crisis. Their strategic targeting and narrow focus may increase their potential to serve as influential hubs in the information ecosystem.

5 Conclusion

This project developed a two-part detection system that flags both fake or AI-generated accounts and those likely involved in coordinated influence operations. By combining a behavioral-text classifier with a graph-based model trained on known IO networks, we aimed to capture not just individual anomalies, but accounts that play a broader role in shaping and amplifying political narratives.

When applied to real data from the 2024 U.S. presidential election, the system surfaced a focused group of users who stood out both in their behavioral patterns and the types of content they promoted. These users were not merely statistical outliers: they consistently posted at high volume, engaged heavily with a narrow set of themes, and helped circulate emotionally charged, partisan narratives. Many of their posts centered on election fraud, economic insecurity, and attacks on political figures, often framed in ways that reinforced public distrust and deepened division.

Rather than introducing new ideas, many of these users focused on amplifying existing grievances already circulating in political discourse. This kind of activity can be easy to miss if models only look for bots or misinformation in isolation. Our findings suggest that tracking coordinated behavior and narrative focus together gives a clearer picture of how influence operations actually work in practice.

5.1 Limitations

The limitations which we have identified for this project fall under the following three categories:

1. False Positives Detection. Although our model flags conservatively, it should be noted that there is still the possibility of false positives, which result in our model flagging legitimate profiles to be AI / fake profiles. Given that our decision threshold is already rather high, one possibility is to integrate some level of post-processing into the system by including manual verification via human actors. If we were to apply our models for the purposes of real-time fake account detection and banning, we could also integrate a mechanism for the owners of real profiles to contest such decisions.

2. Lack of Non-Textual Content. Although we have included a variety of social media content over the course of training our models, the data which we have trained our models on is limited to the tex-

tual content supplied by social media accounts (e.g. post content etc.). Given that the content generated on social media is highly unstructured (Liu et al., 2024), including both user-generated multimedia content in the form of videos or photos, as well as site-generated content in the form of metadata such as profile or post creation timestamps, we find that being able to incorporate such data would have provided a significant boost to the accuracy of our models in detecting AI / fake profiles.

3. Capturing Dominant Coordination over Subtle Coordination.

Given the nature of the dataset which we have trained our models on (e.g. accounts with the same retweeted content, accounts that retweet within a short time window etc.), our models are more focused on detecting dominant coordination patterns among IO actors. However, capable IO actors, especially those that employ significant resources (e.g. nation-states), are aware of the detectability of such methods and strive to make their operations more discreet. They employ methods such as offsetting the timing of their posts, or pushing shared narratives instead of the exact same textual content, to continue to perform their operations while still keeping themselves under the radar.

5.2 Next Steps

1. Incorporate Non-Textual data. Given the plethora of available information on social media networks outside of their textual content, we want to integrate some of these information over the course of training future models. For instance, taking note of semantic similarities across photo or video content can also serve to detect groups of actors which are pushing a shared agenda.

2. Detect Coordinated Timing Patterns. Following on the previous idea, we can also work on incorporating timestamp metadata to serve as a useful tool for detecting more subtle methods of coordination used by IO actors. Apart from identifying IO actors by the similarity of their content, taking note of the time when they make their posts, or the length of time between when they share each others' posts across the network, we can also identify groups of users that appear to be working together as part of an IO campaign.

3. Real-Time Detection and Narrative Monitoring. As we improve on the speed and accuracy of our models, we also want to start on expanding

our models to be capable of real-time detection and monitoring. This would allow us to capture IO actors in real-time and prevent them from achieving their agenda before too much damage has taken place. We believe that such a tool would be useful for both social media owners, who would benefit from the increased trust surrounding their network, and the public at large, as they would be less impacted by such IO campaigns aiming to push their own agendas or spread disinformation.

References

- Aseel Addawood, Adam Badawy, Kristina Lerman, and Emilio Ferrara. 2019. Linguistic cues to deception: Identifying political trolls on social media. In *Proceedings of the international AAAI conference on web and social media*, volume 13, pages 15–25.
- Ashwin Balasubramanian, Vito Zou, Hitesh Narayana, Christina You, Luca Luceri, and Emilio Ferrara. 2024. [A public dataset tracking social media discourse about the 2024 u.s. presidential election on twitter/x](#). Preprint, arXiv:2411.00376.
- Johan Bollen, Bruno Gonçalves, Guangchen Ruan, and Huina Mao. 2011. Happiness is assortative in online social networks. *Artificial life*, 17(3):237–251.
- Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9):e2023301118.
- Priyanka Dey, Luca Luceri, and Emilio Ferrara. 2024. Coordinated activity modulates the behavior and emotions of organic users: A case study on tweets about the gaza conflict. In *Companion Proceedings of the ACM Web Conference 2024*, pages 682–685.
- Hans WA Hanley and Zakir Durumeric. 2024. Partial mobilization: Tracking multilingual information flows amongst russian media outlets and telegram. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, pages 528–541.
- Hans WA Hanley, Deepak Kumar, and Zakir Durumeric. 2024. Specious sites: Tracking the spread and sway of spurious news stories at scale. In *2024 IEEE Symposium on Security and Privacy (SP)*, pages 1609–1627. IEEE.
- Yunchong Liu, Xiaorui Shen, Yeyubei Zhang, Zhongyan Wang, Yexin Tian, Jianglai Dai, and Yuchen Cao. 2024. [A systematic review of machine learning approaches for detecting deceptive activities on social media: Methods, challenges, and biases](#).
- Luca Luceri, Valeria Panté, Keith Burghardt, and Emilio Ferrara. 2024. [Unmasking the web of deceit: Uncovering coordinated activity to expose information operations on twitter](#). In *Proceedings of the ACM Web Conference 2024*, pages 2530–2541. ACM.
- VenkataSwamy Martha, Weizhong Zhao, and Xiaowei Xu. 2013. A study on twitter user-follower network: A network based analysis. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1405–1409.
- Robert F Wolff, Karel G M Moons, Richard D Riley, Penny F Whiting, Marie Westwood, Gary S Collins, Johannes B Reitsma, Jos Kleijnen, and Sue Mallett. 2019. [Probast: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration](#).