Identifying Coordinated Networks Promoting Media Narratives from Unreliable Wikipedia Sources

Raghav Jain*
Indian Institute of Technology, Patna

Jhagrut Lalwani*
Veermata Jijabai Technological Institute
jhagrutpradeep@gmail.com

Jay Gala*
Indian Institute of Technology Madras

Deep Gandhi*
University of Alberta

Jash Mehta*
Georgia Institute of Technology

Dhara Mungra*
Bombura Inc.

Deep Patel
D J. Sanghvi College of Engg.

Swapneel Mehta
New York University

ABSTRACT

There has been a growing occurrence of false and misleading information including low-quality news on events of civic volatility like elections, public health crises, and wars. The spread of such unreliable content is often achieved by employing groups of accounts that promote news narratives through coordinated campaigns. We present a content-agnostic method to analyze the influence of Wikipedia's perennial sources. This method relies on publicly available engagement and network data and is capable of identifying previously unknown accounts that are part of coordinated networks promoting potentially unreliable narratives. We track articles between 2020 and 2022 published by two Russian media outlets, Russia Today and Sputnik News, and investigate the networks of accounts involved in amplifying their content to a wider audience on the popular microblogging website, Twitter. In view of the need for open-source intelligence systems to track the external impact of news providers in order to quantify the harms from Wikipedia's unreliable sources, we present a platform containing the results of our analysis, where external actors including the public can access the detailed analysis of content-sharing networks that we develop without reliance on the substantive nature of the content.

KEYWORDS

Wikipedia, News, Social Network, Coordinated Inauthentic Behavior, Graph Theory, Misinformation

1 INTRODUCTION

In recent years, the rise of social media misinformation has become a growing concern due to its potential to spread false information and manipulate public opinion [4, 7, 10, 13, 16]. Social media has connected people from all over the world and enabled them to share ideas and news quickly and easily. However, this has also created an environment that can be used to spread false information and propagate conspiracy theories with relative ease[11, 15]. Disinformation has been used to manipulate political perspectives, with modern tools like ChatGPT [1] shown to be successful in delivering persuasive messages to the population. While neutral venues like Wikipedia provide labels to assess the reliability of these media providers, it is important to quantify their external impact and measure the risks they present for the public. We develop a novel audit to analyze the dissemination of information over the social media platform. This audit is analogous to those conducted in complex financial transactions to monitor the flow of funds between entities. Such an audit provides the public with insight into how organizations are using social networks for communication. These audits involve a systematic review of the flow of information and its origin to ensure transparency and integrity. Additionally, by gathering data from a sample of social networks and analyzing the content, we can gain insight into the overall landscape and more accurately evaluate how certain types of information are being spread. This research focuses on an audit of two Russian state-backed media outlets, Russia Today and Sputnik News, to investigate the content-sharing mechanisms and networks of actors involved in propagating misleading information. These outlets have been widely documented as having been involved in disseminating false or politically-biased information and classified as unreliable by Wikipedia and the US State Department. Our audit conducted a thorough analysis of the activities of these networks by tracking their activity on the platform since 2008, enabling us to accurately identify coordinated inauthentic networks of actors engaged in spreading misleading information.

Prior Art: Recent research has explored various techniques to effectively analyze and interpret data from social media platforms for various applications.[12] utilized a data mining approach to investigate the phenomenon of fake news on social media while others identified rumors on social media platforms. [14] took an approach that focused on investigating the accuracy of fake news and related factuality issues. [6] presents a survey of stance detection in social media posts.

2 METHODOLOGY

We obtained three datasets from the FOCUSdata Project [3]. This project analyzed English content posted by official media and foreign ministry websites of Russia, Iran, China, and North Korea. These datasets comprised of 410, 000 articles that were published by official sources like Sputnik, and Russia Today from 2004 to 2022. We selected a sample of 25,000 articles from these 3 datasets that

*Authors contributed equally to this research. Correspondence to swapneel.mehta@nyu.edu

[1]https://chat.openai.com

specifically mention Ukraine as the topic of interest. We scraped tweets containing the URL of a subset of 8000 articles with the help of academic and developer access to Twitter API [2]. This data comprised of 50M tweets posted by 14M users having 145M connections.

3 EXPERIMENTS

3.1 Visualization

When the URL to a news article is entered on our dashboard [3], a detailed analysis is presented, and the key information is visualized. Firstly, a summary of the article is provided along with a word cloud and link to related articles. Then we provide a manipulation score which is calculated using 4 factors: source, language, coordination, and bot-like activity. Each of these factors are accounted for independently and then their average is reported as the manipulation score. We report our misinformation score by rating the likelihood of misinformation based on the source and identifying propagandistic words and phrases with the help of AI techniques. Additionally, we consider inauthentic sharing behavior such as trolls and bots in order to detect propaganda activity. We provide tweet statistics, like top 10retweeted users, hashtags, languages, and time series bar charts; plus a network graph to show connections between accounts that shared a URL within a short period of time. Lastly, we give an overview of bot activity related to a specific article, with the overall score from Botometer [17].

3.2 Topic Modeling

We applied topic modeling to a dataset of nearly 25, 000 articles that were clustered into 198 topics using the Top2Vec model [1]. The content of these articles were provided as input to a pretrained propaganda detection model [2] with labels including *loaded language, flag waving, causal oversimplification,* and *appeal to authority* [8]. We developed a metric using a count vector for the topics promoting propagandist speech, identified by a manual process. In ongoing work, we are automating this process by using Scattertext [5] to identify the users whose conversations constitute outliers when compared to the rest of the corpus followed by a rank-based method to detect potential propaganda being promoted by a collection of similarly identified outliers.

3.3 Co-ordination Scores

We use a heuristic metric based on the time interval of sharing or reacting to a URL on Twitter to determine whether an article has been involved in coordination activity. We employ a technique analogous to the Coordination Detection Framework [9] which tries to detect coordinated activities based on interactions between several users. We expect to extend this to work across several media providers and social networks in order to drive impact on the media landscape externally using the collective wisdom of the Wikimedia community.

[2]`https://developer.twitter.com/en/docs/twitter-api`
[3]`https://parrot.report`

4 CONCLUSION

In conclusion, the rise in social media and disinformation has become a major concern for our society. It is essential to put in place effective rules and regulations to monitor the spread of information, and an audit of social media platforms can be a powerful tool to do so. Our research has demonstrated the potential of using an audit to identify coordinated networks that are disseminating media from outlets that Wikipedia lists as unreliable and allows the Wikimedia community to view the external impact and influence of these providers via a quantitative lens.

ACKNOWLEDGMENTS

We are supported generously by the Parrot Journalism AI Team comprising Simone DiStefano, Alessandro Alviani, Ademola Bello, and Venetia Menezes at Ippen Digital and The Sunday Times.

REFERENCES

[1] Dimo Angelov. 2020. Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470* (2020).
[2] Anton Chernyavskiy, Dmitry Ilvovsky, and Preslav Nakov. 2020. Aschern at SemEval-2020 Task 11: It Takes Three to Tango: RoBERTa, CRF, and Transfer Learning. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*. International Committee for Computational Linguistics, Barcelona (online), 1462–1468. https://doi.org/10.18653/v1/2020.semeval-1.191
[3] Scott Fisher and Graig Klein. 2020. *Russia's Sputnik News, China's Global Times, North Korea's Korean Central News Agency (KCNA)*. FOCUSdata Project.
[4] Matthew Hindman and Vlad Barash. 2018. Disinformation, and influence campaigns on twitter. *Knight Foundation: George Washington University* (2018).
[5] Jason Kessler. 2017. Scattertext: a Browser-Based Tool for Visualizing how Corpora Differ. In *Proceedings of ACL 2017, System Demonstrations*. Association for Computational Linguistics, Vancouver, Canada, 85–90. https://aclanthology.org/P17-4015
[6] Dilek Küçük and Fazli Can. 2020. Stance Detection: A Survey. *ACM Comput. Surv.* 53, 1, Article 12 (feb 2020), 37 pages. https://doi.org/10.1145/3369026
[7] Josephine Lukito. 2020. Coordinating a multi-platform disinformation campaign: Internet Research Agency activity on three US social media platforms, 2015 to 2017. *Political Communication* 37, 2 (2020), 238–255.
[8] Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2020. A survey on computational propaganda detection. *arXiv preprint arXiv:2007.08024* (2020).
[9] Diogo Pacheco, Pik-Mai Hui, Christopher Torres-Lugo, Bao Tran Truong, Alessandro Flammini, and Filippo Menczer. 2021. Uncovering Coordinated Networks on Social Media: Methods and Case Studies. *ICWSM* 21 (2021), 455–466.
[10] Niels Nagelhus Schia and Lars Gjesvik. 2020. Hacking democracy: managing influence campaigns and disinformation in the digital age. *Journal of Cyber Policy* 5, 3 (2020), 413–428.
[11] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer. 2017. The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592* 96 (2017), 104.
[12] Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond News Contents: The Role of Social Context for Fake News Detection. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining* (Melbourne VIC, Australia) *(WSDM '19)*. Association for Computing Machinery, New York, NY, USA, 312–320. https://doi.org/10.1145/3289600.3290994
[13] Fabio Tagliabue, Luca Galassi, and Pierpaolo Mariani. 2020. The "pandemic" of disinformation in COVID-19. *SN comprehensive clinical medicine* 2, 9 (2020), 1287–1289.
[14] James Thorne, Andreas Vlachos, Oana Cocarascu, Christos Christodoulopoulos, and Arpit Mittal. 2018. The Fact Extraction and VERification (FEVER) Shared Task. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*. Association for Computational Linguistics, Brussels, Belgium, 1–9. https://doi.org/10.18653/v1/W18-5501
[15] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *science* 359, 6380 (2018), 1146–1151.
[16] Tom Wilson and Kate Starbird. 2020. Cross-platform disinformation campaigns: lessons learned and next steps. *Harvard Kennedy School Misinformation Review* 1, 1 (2020).
[17] Kai-Cheng Yang, Emilio Ferrara, and Filippo Menczer. 2022. Botometer 101: Social bot practicum for computational social scientists. *arXiv preprint arXiv:2201.01608* (2022).

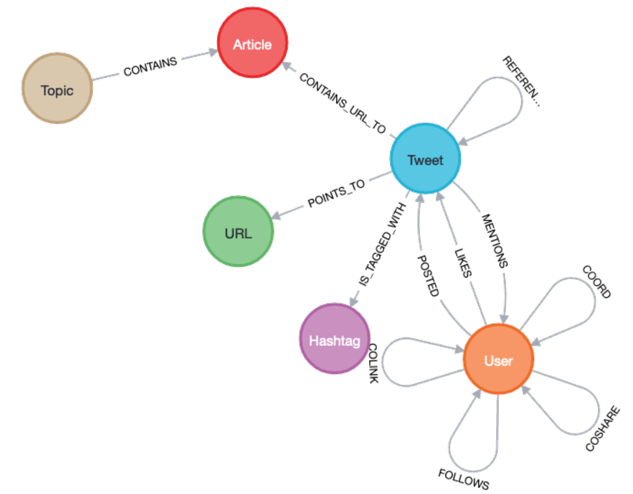

Figure 3: The Schema of the Graph Database that the Collected Data is Ingested into for Low-latency Querying Over Entity-Relationships.

Figure 4: Comparing per-article Novel Accounts in relation to Botometer Scores for their 'bot-like' activity for 27 randomly sampled articles

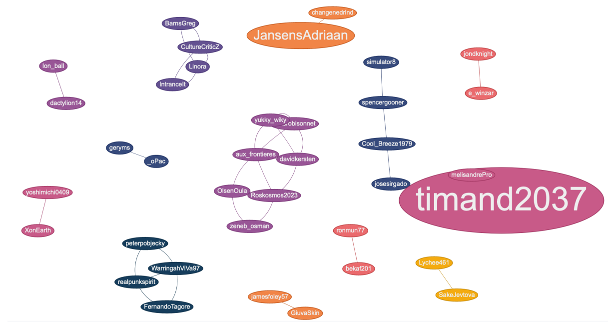

Figure 1: An example of a graph showing accounts that shared a specific URL and have been observed to co-share content within very short time intervals, which is a possible indicator of bot activity. The size of the node corresponds to the total engagement with their tweet.

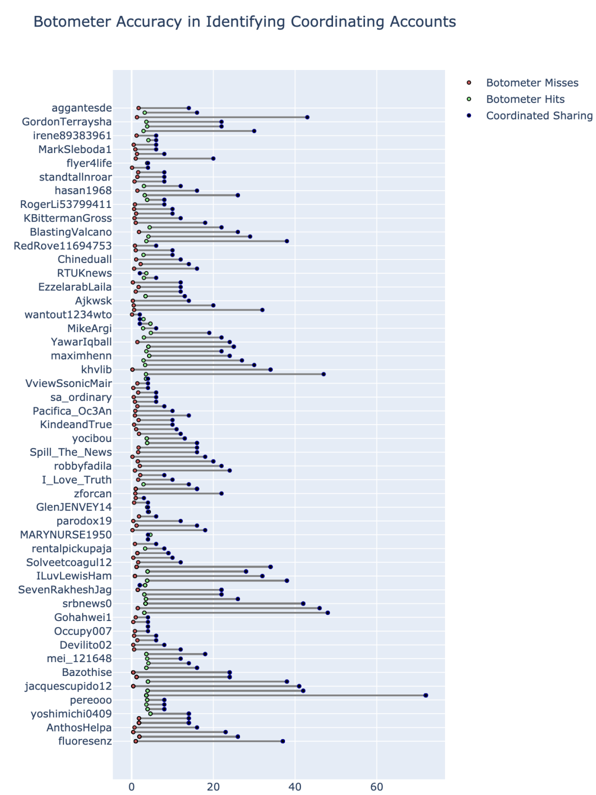

Figure 2: Sample of 160 Accounts Spreading RT/Sputnik news: Botometer Scores for 'bot-like' activity do not identify them behavior despite clear coordinated link-sharing behavior (0-100)