Selected Papers of #AoIR2023:
The 24th Annual Conference of the
Association of Internet Researchers
Philadelphia, PA, USA / 18-21 Oct 2023

REVEALING MULTIMODAL COORDINATION IN SOCIAL MEDIA: A CASE STUDY OF PRO-RUSSIAN INFLUENCE CAMPAIGNS REVEALING MULTIMODAL

Guangnan Zhu
Queensland University of Technology

Timothy Graham
Queensland University of Technology

Daniel Whelan-Shamy
Queensland University of Technology

Robert Fleet
Queensland University of Technology

Abstract

Coordinated online disinformation campaigns are by nature difficult to detect. In response, communication scholars have developed a range of methods and analytical frameworks to discover and analyse disinformation campaigns. The use of social network analysis (SNA) to find and map coordinated behavioural patterns has become increasingly popular and demonstrated effective results. However, these methods are designed for text and behavioural but miss an important aspect of disinformation campaigns: image-based coordination. This paper examines this gap by introducing a multimodal network analysis framework to map coordinated retweeting behaviour and coordinated image-sharing. We show that image-based coordination is both widespread and different in structure to other forms of coordination. This is important because it highlights a major gap in research, where text-based computational methods are not suited to detecting and analysing the scale and scope of visual disinformation on platforms. To address this, we suggest new methods to complement existing approaches, using deep learning models and advanced SNA to detect image similarity. The paper concludes with a reflection of limitations and suggestions for the next steps.

Suggested Citation (APA): Zhu, G., Graham, T., Whelan-Shamy, D., Fleet, R. (2023, October). *Revealing image-based coordination in social media: A case study of pro-Russian influence campaigns*. Paper presented at AoIR2023: The 24th Annual Conference of the Association of Internet Researchers. Philadelphia, PA, USA: AoIR. Retrieved from http://spir.aoir.org.

Introduction

The growing prevalence of online influence campaigns that spread false or misleading information on social media has become one of the main threats to platform integrity, trust, and safety. New and emerging scholarship has begun to focus on the group-wise interactions and organizational mechanisms in these influence campaigns, which platforms characterize as ‘coordinated inauthentic behaviour’ (CIB) (Gleicher, 2019) - a group of diverse actors act together in organized ways to exert and amplify influence on a large number of targeted users on online social media. Nevertheless, previous studies on CIB mainly focus on text-based coordinated behaviour and ignore the influence and harms driven by the semantic false content embedded in images.

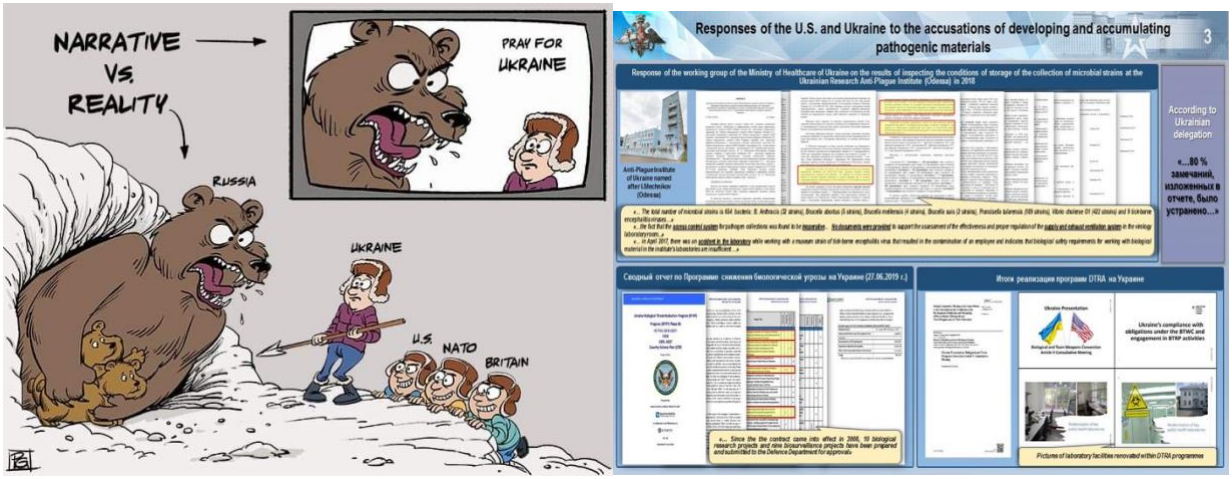

Figure 1. Pro-Russian images shared in the influence campaigns from the Russia-Ukraine War.

This is problematic given that images nowadays increasingly play a dominant role in political mobilisation as image-based content more easily conveys and expresses people’s sentiments and opinions (Johnson, 2007) in the context of the current online sphere (Heiskanen, 2017). In fact, Images are easy o manipulate and those that contain problematic and/or false content often escape traditional detection (Thomson et al., 2022). Similarly, images that do not represent false or misleading information can be de- or re-contextualized to support and spread malinformation (Wardle, 2022), evading the detection of both researchers and platforms. The widespread of visual influence campaigns such as meme warfare (Merrin, 2019) and fake 'fact-checking' infographics (Romero, 2022) are frequent practices of coordinated disinformation campaigns and conspiracy theory groups (Krafft & Donovan, 2019). For instance, several pro-Russia images in the form of memes and infograph are posted and shared in the pro-Putin (#IStandWithPutin) communities (Figure 1).

This paper presents a novel study in studying image-based coordination network. We examine the recent breakthroughs in social network analysis for detecting coordinated activity on social media platforms (Keller et al., 2019; Schoch et al., 2022) and open-source toolkits (Giglietto et al., 2020; Graham, 2020; Righetti, 2021). Given these approaches only make use of text and URL-based data as well as behavioural activities

(e.g., retweeting and replying) and are poorly suited to image-based content and online image-sharing behaviours, this study poses and answers three research questions:

1. How can coordination network analysis be modified to incorporate image-based content?
2. How does the volume and network structure differ between coordinated image sharing and text-based tweeting network?
3. How can we develop methods for detecting and analysing sets of similar images shared in influence campaigns?

Methods

To address our research question and construct image-based coordinated networks for Social Network Analysis, we have enhanced existing 'coordination network analysis' methodologies. Our extension allows for the identification of accounts that post not only similar textual content but also share visual content, all within a short time frame (e.g., within 60 seconds). In our approach, we process both textual and visual data, converting them into embedding feature vectors. For the textual data, we utilize a pretrained BERT-based (Bidirectional Encoder Representations from Transformers) text-embedding model. For the visual data, we employ a pretrained Vgg16 model (pretrained on the ImageNet Dataset). To integrate these two types of data, we deploy a multimodal model, fusing the two embeddings into a single joint multimodal vector that represents the original posts. This method addresses the limitation in previous RGB Histogram feature vectors (Pacheco et al., 2020) by obtaining more hierarchical and no hand-crafted features. Following this, we calculate the cosine similarity between vectors in each pair of posts. Actors with higher similarity values that also meet our extended criteria for coordination are then connected and mapped. The final step of our process is achieved through the use of a toolkit developed by Graham (2020), ensuring a comprehensive and robust analysis of coordinated network activities.

We evaluate our methods through a case study of Russian visual disinformation sharing in a coordinated fashion on Twitter by over 78487 pro-Russia and pro-Putin accounts, as well as fringe bot- and troll-like accounts, on Twitter that posted text-only or image content (n=13140) with the hashtag #IStandWithPutin in 2022 during the start of the Russia-Ukraine War.

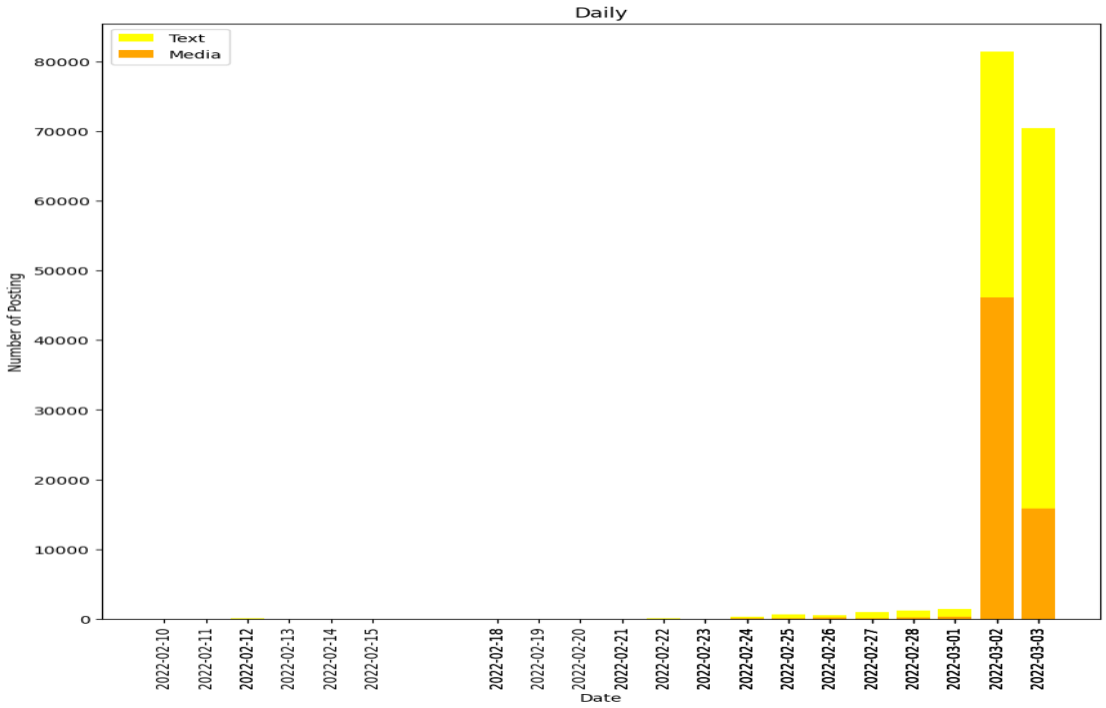

Figure 2. Comparison on number of posted tweets among text-only (yellow) and multimodal (orange) content

Results

The descriptive comparison Figure 2 demonstrates the obvious difference in number between text-only tweets and ones containing images during the period at the start of the Russian invasion. We generate two coordination networks, one text-based and one multimodal network. We define the similarity threshold as 0.9 for both networks to indicate the highly similar pairs of posts. We also define the time window as 60 seconds and edge weight larger than 2 to detect pairs of nodes (accounts) in the network who must have posted similar images at least twice within 60 seconds of each other. All colored clusters within the multimodal coordination network share a pro-Russian sentiment while the sentiment within the text-based network is much the same. The text-based coordination network consists of 322 nodes and 320 edges while the multimodal coordination network consists of 188 and 286 extra nodes and edges, respectively. The accounts in the multimodal network are more likely to share media content (e.g. images, videos, etc.) than in the text-based network. The results illustrate that the multimodal method effectively supplements the text-based detection method and expands the original method in detecting image-based coordination and more importantly, the multimodal level.

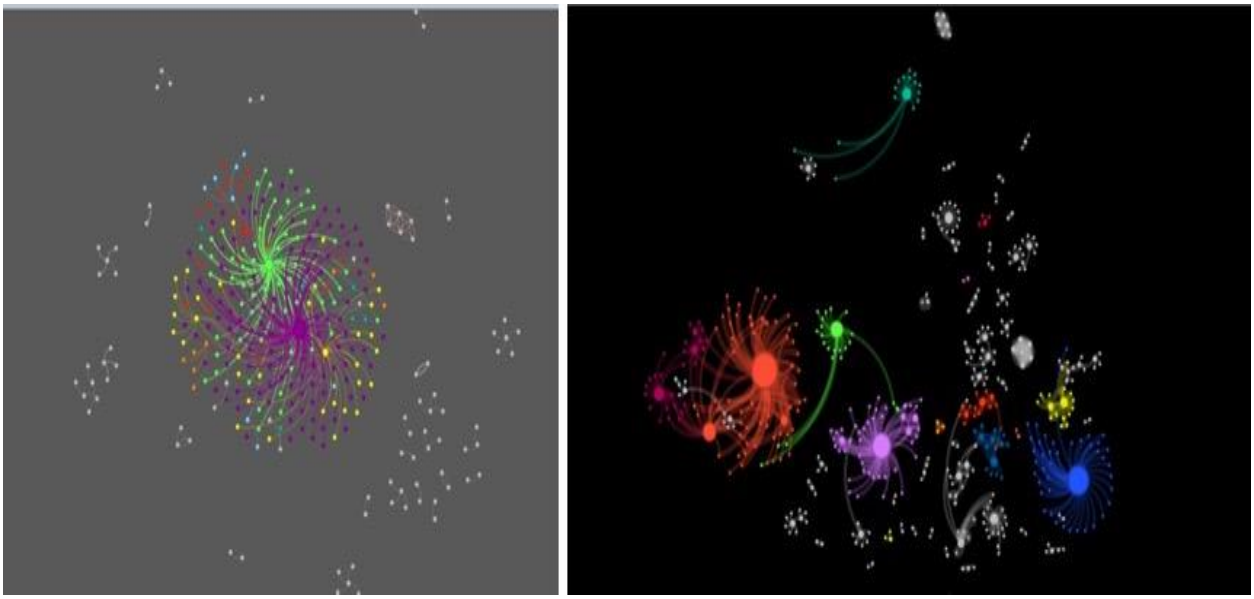

Figure 3. A text-based and a multimodal coordination network generated from the #IStandWithPutin dataset.

Conclusion

Image-based online communication by posting and sharing on social media has become an important mechanism in influence campaigns. The new multimodal coordination detection method utilises non-textual modes of data and information and introduces network-based analysis methods for detecting and studying multimodal content. Although it is an ongoing project, the results illustrate the irreplaceable key role of multimodal methods in the practice of more complete and multimodal social media network studies and also serve to further improve coordination detection in disinformation campaigns. We also suggest deploying the image clustering method in future work to partition the images into semantically meaningful clusters to interpret their implicit relationship and use clusters as the indicator for sameness instead of explicit similarity measurement. This method can help us conduct the analysis of memetic templates such as image macros. Apart from image and text, there are more unexplored forms and modes of information, where we look forward to future work leveraging more advanced methods to study more diverse coordinated activities such as online activism and participatory culture.

References

Giglietto, F., Righetti, N., & Marino, G. (2019). Understanding coordinated and inauthentic link sharing behavior on facebook in the run-up to 2018 general election and 2019 european election in italy.

Gleicher. (2018). Coordinated Inauthentic Behavior Explained. https://newsroom.fb .com/news/2019/08/cib-uae-egypt-saudi-arabia/. (Accessed: 2021-10-15)

Graham, T. (2020). Coordination network toolkit. Queensland University of Technology. (Software). Retrieved from ttps://doi.org/10.25912/RDF 1632782596538

Heiskanen, B. (2017). Meme-ing electoral participation. European journal of American studies, 12(12-2).

Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political astroturfing on twitter: How to coordinate a disinformation campaign. Political Communication, 37(2), 256–280.

Krafft, P. M., & Donovan, J. (2020). Disinformation by design: The use of evidence collages and platform filtering in a media manipulation campaign. Political Communication, 37(2), 194-214.

Merrin, W. (2019). President troll: Trump, 4Chan and memetic warfare. In Trump's media war (pp. 201-226). Palgrave Macmillan, Cham.

Pacheco, D., Flammini, A., & Menczer, F. (2020). Unveiling coordinated groups behind white helmets disinformation. In Companion proceedings of the web conference 2020 (pp. 611–616).

Righetti, N. (2021). CooRTweet: an R package to detect coordinated behavior on Twitter. Retrieved from: https://www.nicolarighetti.net/digital-media/coortweet-an-r-package-to-detect-coordinated-behavior-on-twitter/

Romero, L. (2022). How ‘War on Fakes’ uses fact-checking to spread pro-Russia propaganda. Politifact. Retrieved from: https://www.politifact.com/article/2022/aug/08/how-war-fakes-uses-fact-checking-spread-pro-russia/

Schoch, D., Keller, F. B., Stier, S., & Yang, J. (2022). Coordination patterns reveal online political astroturfing across the world. Scientific reports, 12(1), 1–10.

Starbird, K., Arif, A., & Wilson, T. (2019). Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1–26.

T.J. Thomson, Daniel Angus, Paula Dootson, Edward Hurcombe & Adam Smith (2022) Visual Mis/disinformation in Journalism and Public Communications: Current Verification Practices, Challenges, and Future Opportunities, Journalism Practice, 16:5, 938-962, DOI: 10.1080/17512786.2020.1832139

Wardle, C. (2020, Sep 22). Understanding Information disorder. First Draft. Retrieved from: https://firstdraftnews.org/long-form-article/understanding-information-disorder/