Assembling coordinated patterns in the study of dis/misinformation campaigns

Keywords: Disinformation, Coordinated Activity, Social Network, Social Media, COVID-19

Extended Abstract

The rapid development of social media technology and platforms has greatly promoted the construction of online communities across countries and regions. Global information exchange on the platforms enables people's interaction to be infinitely extended. However, it has become a breeding ground for the spread of false and harmful information that not only violates the rules of social platforms and government regulations but also ethical and civil norms. It also challenges the traditional fact-checking scholars and authorities who utilise large-scale text-based classification methods, as the line between fact and falsehood is not always straightforward, and binary classification is usually controversial (Molina et al., 2021) as 'false dichotomies' would strengthen and polarize debates among the actual and online communities in different contexts (Escandon et al., 2021).

Owing to this controversy and the complexity of distinguishing false content from factual at scale, new and emerging scholarship has begun to focus on the group-wise interactions and organisational mechanisms in these information campaigns among different communities. Among the growing suite of mixed methods approaches, social network analysis (SNA) has emerged as a de-facto technique for detecting and analysing disinformation campaigns on platforms such as Twitter. By collecting large-scale digital datasets, practitioners examine the structure of interactions (network ties) between accounts (network nodes) involved in influence campaigns to map and understand their scale and scope. Recent breakthroughs are grounded in social network analysis (SNA) for detecting coordinated activity (Keller et al., 2019; Schoch et al., 2022) and open-source toolkits (Giglietto et al., 2018; Graham et al., 2020). Studies of 'coordinated inauthentic behaviour' have examined how users in some online communities often have suspiciously similar behaviours or post highly similar content to exert influence over users outside the community (Schoch et al., 2022; Gleicher, 2019). Detecting coordination provides overwhelming evidence to help scholars identify suspicious activities (such as troll farms, botnets, etc.) in regard to disinformation or misinformation campaigns and understand how they amplify influence and cultivate online communities (Starbird et al., 2019).

Our study organically combines widely used text classification models with coordinated network detection, and we extend current coordination network analysis methods to enable detection of accounts that all spread semantically similar content repeatedly within a brief time window of each other (e.g., within 60 seconds). We achieve this goal by combining topic modeling technique Latent Dirichlet Allocation (LDA) (Blei et al., 2003) and its variants seeded-LDA (Lu et al., 2010) with the state-of-the-art coordinated network analysis toolkit (Graham, 2020) to construct coordinated topical network (**CTN**) and provide fine-grained qualitative analysis and comparison on the themes discussed in different coordinated communities. This innovative work effectively addresses the limitations of previous text-based models in understanding complex dis- and misinformation campaigns and online discourse by integrating network patterns and qualitative analysis of communities into previous text-only methods. It provides more convincing and multimodal evidence, enabling scholars to examine and contextualize the content and structure of spreading information for further dis- and misinformation classification.

To examine the methods, this study focuses on online discussions and activities that surround the April 2020 CNN interview of Melinda Gates and her statements on the impact of Covid-19 on the developing world on Twitter. In the interview, she said that coronavirus will be horrible in the developing world and showed her concerns on how Africa would handle the global pandemic. The interview triggered multiple responses, claims and interpretations across many social media platforms, with users from the African continent, the diaspora, and other actors engaging in a global conversation.

We deploy the CTN methods to detect and analyse disinformation campaigns on a dataset of 64,399 tweets surrounding the interview and Covid-19 in Africa on Twitter collected using keywords-based approach. Our preliminary results find that a large set of accounts share the same links in a short period, indicating potential "co-linking" activity (Graham, 2020; Giglietto et al., 2019) to share similar contents or topics to amplify influence and shape online discourse. Figure 1 illustrates the main clusters where accounts share the same links: the red cluster shares an article that describes a hidden agenda behind the vaccination program by "powerful Western philanthropists targeting the Global South". The article shared by the teal clusters questions COVID vaccination due to the past public health crises. Further investigation identifies that the accounts in the cluster are mainly from Africa and India.

In the next stage of the project, we will compare our results with some widely text-only disinformation and misinformation classification models (LDA, Random Forest, and BERT). Additionally, to enrich the dataset with cross-regional texts representing the contexts of several West and South-African countries, as well as the United States, we performed community detection with Louvain algorithm on a network of retweets. We identified 5 modularity classes corresponding to the communities of interest, and collected additional 429,537 tweets from 22,114 Twitter conversations users from these communities participated in.

In summary, the main contribution is integrating advanced disinformation campaign analytic methods to complement existing text-only detection frameworks with network patterns and extend previous SNA+Text methods with more contextually grounded evidence of coordinated activities. This approach is especially useful as it can offer insights for data collection and dataset construction processes that depend on using keywords as queries (for example, when accessing data through Twitter API), since these datasets often have strong thematic and semantic dependencies that confuse and bias further classification.

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. the Journal of machine Learning research, 3, 993–1022.

Escandón, K., Rasmussen, A.L., Bogoch, I.I. et al. COVID-19 false dichotomies and a comprehensive review of the evidence regarding public health, COVID-19 symptomatology, SARS-CoV-2 transmission, mask wearing, and reinfection. BMC Infect Dis 21, 710 (2021). https://doi.org/10.1186/s12879-021-06357-4

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

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.

Lu, B., Ott, M., Cardie, C., & Tsou, B. K. (2011, December). Multi-aspect sentiment analysis with topic models. In 2011 IEEE 11th international conference on data mining workshops (pp. 81-88). IEEE.

Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). "Fake news" is not simply false information: A concept explication and taxonomy of online content. *American behavioral scientist*, *65*(2), 180-212.

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.

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.

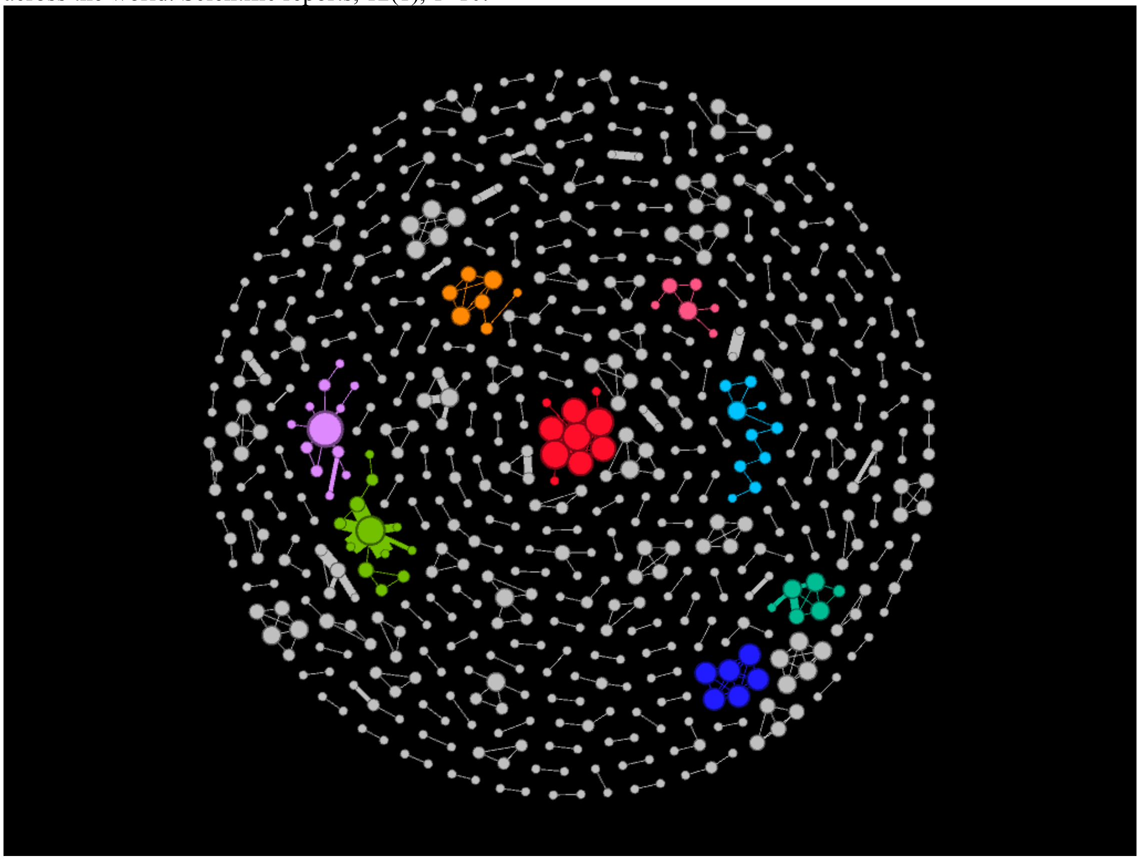

Figure 1. Co-link networks of tweets, with nodes representing accounts and edges representing sharing of the same link within a minute. Node size represents degree, and the colour clusters correspond to most shared links.