Title: Mitigation strategies against Fake News diffusion in online social platforms

Direction: (1) Anastasios Giovanidis (Sorbonne University & CNRS - LIP6),
(2) Vincent Gauthier (Télécom SudParis - SAMOVAR)

Context: Fake News (FN) are claims that are factually wrong and that have the intention to mislead an audience (see [1] and references therein). This malicious behaviour is extremely prevalent on social media platforms like Twitter and Facebook. In fact, the majority of people with online access will partly or fully be informed through these platforms, posing a serious risk on how social opinions are formed since FN are often merged with truthful information which makes them hard to assess by a victim. Indeed, very recent events confirm their power. For example, it has been reported a deluge of misinformation related to Covid-19 [2] and vaccines [3] occurred in social platforms. Equally important is the case of electoral campaigns: the authors of [4] showed that among 171 million tweets related to the US 2016 elections, 25% of them were FN; and similar observations have been observed in the French and Italian elections [10]. The word “infodemic” has been coined to describe this pandemic of FN, which indeed have been shown to propagate faster, and deeper than real news [5]. In sum, studying the spreading of FN and their quick detection is a subject of utmost importance.

Main Objectives: In this thesis we aim to study the principles that rule the diffusion of Fake News inside social platforms. In particular, we have the following two main objectives: (I) to characterise diffusion patterns and classify/predict FN from diffusion paths, and (II) to propose novel mitigation strategies by means of learning algorithms that can quickly spot FN from massive interactions traces.

Method and Novelty: In contradistinction to most literature devoted to the study of FN that heavily rely on natural language processing or sentiment analysis methods, in this work we propose to thoroughly study the diffusion dynamics of FN inside the social graph. Our hypothesis is that FN leave a signature that can be characterised by the structural and dynamical properties of interactions of the underlying social graph: namely, that FN will transit through specific diffusion paths or cause super-spreaders to be active at specific time intervals. Such structural signature has recently been incorporated into machine learning pipelines for FN detection [11, 12] with great success, yet a detailed study is still missing. The novelty of our approach will thus come from the use of original models of social platforms [6] as well as graph-related tools [9], both recently developed in LIP6, that can naturally find application in the analysis of diffusion and the proposal of mitigation control policies. Furthermore, the question how to fight against Fake News has not been sufficiently studied either, and there are no realistic policies suggested until now.

Specific Aims:

(I) Characterise the diffusion of FN: This part will answer the question why and how misinformation can trigger a large sharing cascade among the users on social media. To do so, we will study patterns in the way diffusion episodes are spread among users, and use these for classification and prediction. In this part of the work, questions related to user and/or activity clustering as well as assortativity are very pertinent, because there is a relation of FN spread with filter bubbles and polarisation of opinions [7]. To achieve this, we will adapt our models from [6] but we will also use ideas from the LIP6-ComplexNetworks team, which introduced Link Streams to model and study interactions over time and diffusion paths formed by them [9]. For classification and prediction purposes we will build upon current pipelines that leverage Graph Neural Networks.

(II) Study ways to mitigate FN: In this part we will mostly study centralised policies (from the social-platform perspective). Existing works [8] propose to dynamically incentivise users to spread true news. In our opinion, however, the platform has a more important role in the control of FN, because it can intervene on user feeds to either filter FNs or to promote real news. Such platform-based policies can be cast in the framework of Reinforcement Learning; the platform can have partial observation of the current situation (since not all FN can be identified) and intervene to correct with its own advertising or filtering policies. Alternatively, user coalitions (from mainstream or online media) could cooperate and adapt their actions to promote more intensively verified content.

DATA: There are many available datasets for FN related to their spread in social media. These come primarily from Twitter, see CREDBANK and FakeNewsNet in [1], Covid19 Infodemics Observatory in [2]. Two datasets: Recent Fake News and 2011 Tohoku earthquake and tsunami in [13].

REFERENCES

[1] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, (2017). “Fake News Detection on Social Media: A Data Mining Perspective,” SIGKDD Explor. Newsl. 19, 1 (June 2017), 22–36. DOI:https://doi.org/10.1145/3137597.3137600

[2] V. d'Andrea, O. Artime, N. Castaldo, P. Sacco, R. Gallotti, and M. De Domenico, (2022) “Epidemic proximity and imitation dynamics drive infodemic waves during the COVID-19 pandemic”, Phys. Rev. Research, vol.4, iss.1, pp.013158-(1-10), American Physical Society, doi: 10.1103/PhysRevResearch.4.013158

[3] A. Ghaddar, S. Khandaqji, Z. Awad, R. Kansoun (2022), “Conspiracy beliefs and vaccination intent for COVID-19 in an infodemic”, PLoS ONE 17(1): e0261559. https://doi.org/10.1371/

journal.pone.0261559

[4] A. Bovet, H.A. Makse (2019), “Influence of fake news in Twitter during the 2016 US presidential election”, Nature Communications, vol.10, no.7, doi: https://doi.org/10.1038/s41467-018-07761-2

[5] Vosoughi Soroush, Roy Deb and Aral Sinan, (2018) “The spread of true and false news online”, Science, vol. 359, no. 6380, pp.1146–1151, DOI: 10.1126/science.aap9559

[6] A. Giovanidis, B. Baynat, C. Magnien and A. Vendeville, (2021) "Ranking Online Social Users by Their Influence," in IEEE/ACM Transactions on Networking, vol. 29, no. 5, pp. 2198–2214, Oct. 2021, doi: 10.1109/TNET.2021.3085201

[7] Daron Acemoglu, Asuman Ozdaglar and James Siderius, (2022) “A Model of Online Misinformation”, National Bureau of Economic Research, Working Paper 28884, DOI 10.3386/w28884

[8] Mahak Goindani, Jennifer Neville, (2020) “Social Reinforcement Learning to Combat Fake News Spread”, Proceedings of The 35th Uncertainty in Artificial Intelligence Conference, in Proceedings of Machine Learning Research 115:1006–1016, https://proceedings.mlr.press/v115/goindani20a.html

[9] Matthieu Latapy, Tiphaine Viard, Clémence Magnien, “Stream graphs and link streams for the modeling of interactions over time”. Soc. Netw. Anal. Min. 8(1): 61:1–61:29 (2018)

[10] Francesco Pierri. (2020) “The Diffusion of Mainstream and Disinformation News on Twitter: The Case of Italy and France”. Companion Proceedings of the Web Conference 2020. Association for Computing Machinery, New York, NY, USA, 617–622. DOI:https://doi.org/10.1145/3366424.3385776

[11] Kaliyar, R. K., Goswami, A., & Narang, P. (2021). DeepFakE: improving fake news detection using tensor decomposition-based deep neural network. The Journal of Supercomputing, 77(2), 1015–1037.

[12] Dou, Y., Shu, K., Xia, C., Yu, P. S., & Sun, L. (2021, July). User preference-aware fake news detection. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 2051–2055).

[13] Murayama, T., Wakamiya, S., Aramaki, E., & Kobayashi, R. (2021). Modeling the spread of fake news on Twitter. Plos one, 16(4), e0250419.13