Graph Neural Network for Fake News Detection and Classification of Unlabelled Nodes at MediaEval 2022

Asep Maulana[1,*,†], Konstantin Pogorelov[1,†], Daniel Thilo Schroeder[1,†] and Johannes Langguth[1,2,†]

[1]*Simula Research Laboratory, Norway*
[2]*University of Bergen, Norway*

Abstract
In this paper we describe our approach to fake news detection for the MediaEval 2022 challenge that has run for the third time. As in the previous editions, the goal of the challenge is the detection of misinformation tweets, but in this edition, both text and graph data are provided.

We focus on the classification of unlabelled nodes/users in the graph by utilizing graph neural networks to classify them as either fake news spreader or just an ordinary node i.e. non fake news spreader. Apart from those labels, the classification apply for unlabelled nodes in conspiracy theories related to COVID-19 in nine different categories. Furthermore, graph based node classification detection for whole categories will be done since this will lead to more comprehensive classification analysis rather then just to label them either as a spreader or non spreader of fake news.

1. Introduction

During the course of the COVID-19 pandemic, a large amount of misinformation of various kinds that were observed in online and offline media. A particular example of this misinformation are conspiracy theories related to the origin, nature, and treatment of COVID-19. Irrational and or harmful conspiracy theories spread widely in many online media, have resulted in negative effects in the real world. Accordingly, our aim here is to study new ways of detecting such content, or people, in the Twitter network who are suspected to be misinformation or misinformation spreaders along with any other users that have a connection to those misinformation spreaders. For our analysis and prediction we specialized in the structure of the network with their relationship and neglecting the content of the text or sentence embedding to those users. The reason for using the networks is that conspiracy content can be difficult to detect by pure text analysis, since many such ideas are communicated via hidden or implied meaning, codes, or intentional misspellings such as *plANdemic* instead of *pandemic*.

Analysis tools based on text-only classification challenges have been developed by other researchers in the field [1, 2, 3]. There are already many methods for automatic news analysis and fake content detection in the social media and news analysis field that cover a wide range of approaches, including knowledge graphs, diffusion models, and natural language processing [4, 5, 6, 7]. These methods typically rely on labeled data. Consequently, several such datasets have been published in recent years [8, 9, 10, 11, 12, 13, 14, 15].

MediaEval'22: Multimedia Evaluation Workshop, January 13–15, 2023, Bergen, Norway and Online
*Corresponding author.
†These authors contributed equally.
asep@simula.no (A. Maulana); konstantin@simula.no (K. Pogorelov); daniels@simula.no (D. T. Schroeder); langguth@simula.no (J. Langguth)

Our task specialized in detecting and classifying misinformation spreader in the context of users relationship in the network.

2. Approach

Our focus work will be on the second subtask with the provided data of a graph with $1,679,011$ vertices and $268,694,698$ edges, along with $1,913$ and 830 vertex labels for the development and test set respectively.

In this subtask, we have an undirected graph generated from social network data where the nodes are users and the link between them represent as connections in the context of contact or communication occurred among each other. The nodes in this graph have a set of attributes, including number of followers, location, as well as some texts posted by that user. Some users are labeled as misinformation spreaders based on manually annotated tweets, and some are labeled as non-misinformation spreaders. On the other hand, we have the data regarding the users without any labels and attribute in which we need to predict and classify them either belong to misinformation spreader or non-misinformation spreader. Our task here is to predict and classify the unlabelled users(nodes) in the graph based on their connection to the labeled users as well as their attributes that related to the topic of learning on graphs [16] [17].

Before we continue with the task of prediction for unlabelled users, it is good to understand more about users or individuals that are spreading fake news. Let us assumes that individuals spreading false messages tend to be organized in so-called homophile networks. This kind of homophile networks are more strongly connected internally than externally, which relates to echo chambers and related concepts which can lead to resistance to ideas coming from outside the homophile network [?]. However, because Twitter does not allow access to true retweet cascades (i.e. for each retweet list the specific account whose existing retweet caused the new retweet) but instead returns just a list of retweeters for a particular tweet, this challenge offers the subgraphs of Twitters' follower network that were induced by these retweeters.

For our task prediction, we try to optimize the work by exploring the users with only have the labels and attributes as reference for prediction and classification of unlabelled users [18]. In this context we cut down the the graph to two subgraphs consisting of a subgraph of labelled users and subgraph of unlabelled users. The cutting down of the subgraph conducted from main graph that consist of of $1,679,011$ vertices and $268,694,698$ edges along with their connection to all other nodes. Here is the step of the task for prediction and classification of unlabelled nodes:

1. Construct the subgraph of labelled nodes(with their internal connection) by cutting down it from main graph. This subgraph will be used as a reference for prediction of unlabelled nodes.
2. Construct the subgraph of unlabelled nodes(with their internal connection) by cutting down it from main graph. This unlabelled subgraph is a subject of task prediction.
3. Apply GNN to train the labelled of subgraph and save the model of that training
4. Use the model above to predict unlabelled nodes and classify them either as misinformation spreader or non-misinformation spreader
5. After classifying them and find new label for unlabelled nodes, merge the nodes from labelled and unlabelled nodes and construct new subgraph. In this subgraph put the label as a result of prediction in previous step to unlabelled nodes and train this subgraph to investigate the accuracy of the classification.
6. with different run, we generate the prediction according to the smallest difference of accuracy in the training of labelled subgraph and the training of labelled merge subgraph.

3. Results

As a result of our work, we obtained 2 runs for Graph-based Detection subtasks. The result was -0.0085 and -0.0084 which means that it is still under the random classifier performance. We conclude that such simple approaches are not suitable for the structure-oriented classification tasks.

4. Discussion and Outlook

While the first subtask of the challenge was identical to the 2021 edition [19], the graph based detection constitutes a harder problem. Here we have tried to apply a graph neural network (GNN) for node property prediction and classification [18]. It assumes that GNNs with semi supervised learning can be utilized to predict some unlabelled nodes/vertices in the graph based on others nodes which have label and attributes in the graph. While it is likely that the approach is viable, our chosen method failed to provide any useful classification. In future work, we will study a modification of the model that we have applied, as well as utilize different Graph Neural Network architectures.

5. Acknowledgements

This work was funded by the Norwegian Research Council under contract #303404 and has benefited from the Experimental Infrastructure for Exploration of Exascale Computing (eX3), which is financially supported by the Research Council of Norway under contract #270053.

References

[1] Q. Do, Jigsaw unintended bias in toxicity classification (2019).
[2] Toxic comment classification challenge - identify and classify toxic online comments, 2018. URL: https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/.
[3] A. Mungekar, N. Parab, P. Nima, S. Pereira, Quora insincere question classification, National College of Ireland (2019).
[4] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, Automatic detection of fake news, arXiv preprint arXiv:1708.07104 (2017).
[5] T. Le, S. Wang, D. Lee, Malcom: Generating malicious comments to attack neural fake news detection models, arXiv preprint arXiv:2009.01048 (2020).
[6] L. Cui, H. Seo, M. Tabar, F. Ma, S. Wang, D. Lee, Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 492–502. doi:10.1145/3394486.3403092.
[7] D. de Beer, M. Matthee, Approaches to identify fake news: A systematic literature review, in: T. Antipova (Ed.), Integrated Science in Digital Age 2020, Springer International Publishing, Cham, 2020, pp. 13–22.
[8] M. Nabil, M. Aly, A. Atiya, Astd: Arabic sentiment tweets dataset, in: Proceedings of the 2015 conference on empirical methods in natural language processing, 2015, pp. 2515–2519.
[9] W. Y. Wang, " liar, liar pants on fire": A new benchmark dataset for fake news detection, arXiv preprint arXiv:1705.00648 (2017).
[10] F. K. A. Salem, R. Al Feel, S. Elbassuoni, M. Jaber, M. Farah, Fa-kes: A fake news dataset around the syrian war, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 13, 2019, pp. 573–582.

[11] S. Dhoju, M. Main Uddin Rony, M. Ashad Kabir, N. Hassan, Differences in health news from reliable and unreliable media, in: Companion Proceedings of The 2019 World Wide Web Conference, 2019, pp. 981–987.
[12] E. Dai, Y. Sun, S. Wang, Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 14, 2020, pp. 853–862.
[13] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media, arXiv preprint arXiv:1809.01286 8 (2018).
[14] A. Ghenai, Y. Mejova, Fake cures: user-centric modeling of health misinformation in social media, Proceedings of the ACM on human-computer interaction 2 (2018) 1–20.
[15] K. Nakamura, S. Levy, W. Y. Wang, r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection, arXiv preprint arXiv:1911.03854 (2019).
[16] W. Hu, M. Fey, M. Zitnik, Y. Dong, H. Ren, B. Liu, M. Catasta, J. Leskovec, Open graph benchmark: Datasets for machine learning on graphs, Advances in neural information processing systems 33 (2020) 22118–22133.
[17] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, arXiv preprint arXiv:1609.02907 (2016).
[18] J. Tang, R. Liao, Graph neural networks for node classification, in: L. Wu, P. Cui, J. Pei, L. Zhao (Eds.), Graph Neural Networks: Foundations, Frontiers, and Applications, Springer Singapore, Singapore, 2022, pp. 41–61.
[19] K. Pogorelov, D. T. Schroeder, S. Brenner, J. Langguth, Fakenews: Corona virus and conspiracies multimedia analysis task at mediaeval 2021, in: Multimedia Benchmark Workshop, 2021, p. 67.