

Exploring Fake News Detection with Heterogeneous Social Media Context Graphs^{*}

Gregor Donabauer and Udo Kruschwitz

Information Science, University of Regensburg, Regensburg, Germany
`{gregor.donabauer,udo.kruschwitz}@ur.de`

Abstract. Fake news detection has become a research area that goes way beyond a purely academic interest as it has direct implications on our society as a whole. Recent advances have primarily focused on text-based approaches. However, it has become clear that to be effective one needs to incorporate additional, contextual information such as spreading behaviour of news articles and user interaction patterns on social media. We propose to construct heterogeneous social context graphs around news articles and reformulate the problem as a graph classification task. Exploring the incorporation of different types of information (to get an idea as to what level of social context is most effective) and using different graph neural network architectures indicates that this approach is highly effective with robust results on a common benchmark dataset.

Keywords: Fake News Detection · Social Media Networks · Graph Machine Learning

1 Introduction

Detecting online information disorders is an important step to guarantee free expression and discourse, e.g. [7]. A traditional approach to address this problem is manual fact-checking by experts, which is highly labour-intensive, but more recently automated methods that make use of external knowledge bases and natural language processing (NLP) have been proposed [29,10], sometimes as a tool to support the fact-checker [20]. Recent studies have demonstrated the growing importance of social media as a source of information for many people [21,25]. In this context it has been pointed out that a range of information types extracted from social media and the relations between such entities are useful for detecting fake news, e.g., [26,19]. Relations between news items and related information signals have been modelled to extract additional features by applying graph neural networks (GNNs) on graph structures (as discussed below), but so far this has typically either been limited to individual social media features or in combination with features that are not taken from social media (which makes it hard to assess the relevance of *social context only*).

^{*} version as accepted at ECIR 2023 with author information, Github link and Acknowledgements added

To link different types of information they can be modelled as heterogeneous graphs [21], and such networks have the capability of introducing multiple types of relations around a news article within a *single data structure*, which allows to see each graph as an independent data point. However it is not clear if such data structures are superior compared to homogeneous graphs that model only one relation type, as spreading and interaction behaviour between true and fake news generally heavily differs on social media [28].

There is a substantial body of work that demonstrates the use of GNNs, but it remains an open question as to how suitable they are for the use-case of fake news detection with *social media context features only*. It is also unclear which social media signals are most effective here and if there are significant differences between GNN architectures. To investigate these questions, our contributions can be summarized as follows: **(1)** We introduce **HetSMCG**, a methodology to construct extensive **H**eterogeneous **S**ocial **M**edia **C**ontext **G**raphs around news articles fusing multiple types of information in a single graph and reformulate fake news detection as a *graph classification problem*; **(2)** We evaluate a range of experimental settings exploring information sources taken from *social media only* and different GNN architectures to understand how the task can effectively be solved; **(3)** We compare how performance between heterogeneous multi-relation graphs and homogeneous single-relation graphs differs; **(4)** We make all our code and in-depth experimental results available to the community to foster reproducibility and sharing of resources.¹

2 Related Work

One way of classifying the different research directions at a high level is to distinguish *content-based* methods that focus on the stories themselves, *context-based* approaches tapping into, e.g., social media signals and *intervention-based* approaches [23]. We limit our focus on the context-based paradigm with graph-based representations in our brief contextualization within existing related work.

The majority of context-based approaches for fake news detection use graph structures to model relations, e.g. [4] created user-article graphs and utilized the neighborhood structure to embed news nodes that can then be classified. However, the number of information types is limited to users and news items. In similar problem formulations, user-post interaction graphs are considered [18,22]. Mehta et al. [16] treat fake news detection as a node classification task based on a large graph of users, news and publishers (though they do not model differences in edge types and do not limit information to social media context). In follow up work they also introduce link prediction in this graph structure to find previously unavailable connections [17]. Lu and Li [14] also consider user graphs to extract social interactions using GNNs. In combination with recurrent networks, that leverage retweet patterns, they classify related news articles. Similar approaches that use temporal retweet patterns have been proposed [8,27]. Formulating fake

¹ <https://github.com/doGregor/Graph-FakeNewsNet>

news detection as graph classification has been applied by Dou et al. [8]. They calculated user representations by leveraging timeline tweet embeddings and use the outcome as node features to model news propagation paths associated with articles. Another propagation-based approach was introduced by Song et al. [27]. In their work graph structured temporally evolving retweet patterns have been used to classify whether an article is fake or not. Two other approaches left are most similar to our work [21,13]. Both are constructing large heterogeneous graphs of articles, users, social media features like tweets, and publishers to classify news nodes as either true or fake (again not limited to social media context). Incoming news nodes need to be integrated into this grown data structure to produce accurate results. With our approach we aim at proposing a method where each article can be seen as an independent data point.

We conclude that graph structures are a powerful way to include different dimensions of information and thus to improve fake news detection. However, linking many different entity types (and not just one) in independent data structures has barely been investigated yet (and even less if we only consider information types taken from social media in a heterogeneous way) and we aim at contributing to filling this gap with our research.

3 Methodology and Experiments

3.1 Dataset and Graph Construction Method

The only common benchmark dataset that provides rich social context information in network-like structures is FakeNewsNet [24] which is why we adopt it. FakeNewsNet consists of two different datasets, *Politifact* and *GossipCop*. Both datasets are from different domains (PolitiFact for political news and GossipCop for entertainment stories) which allows to distinguish between domain-specific performances. Twitter’s terms and condition state that tweets and other information cannot be publicly released directly.² We therefore recrawl all data (thereby acquiring a dataset that will slightly differ from what others have used so far).

We construct a heterogeneous social media context graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ around each news article. Thus, each snapshot consists of a set of disjoint vertex sets $\mathcal{V} = \mathcal{V}_N \cup \mathcal{V}_T \cup \mathcal{V}_U$ where $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset, \forall i \neq j$. \mathcal{V}_N represents news articles and $|\mathcal{V}_N| = 1$ for each graph. Node features for this type of vertex are obtained using BERT-base document embeddings, i.e. $\forall v_n \in \mathcal{V}_N : v_n \in \mathbb{R}^{768}$. All embeddings are generated using flairNLP³. Set \mathcal{V}_T includes all tweets that refer to the original news article, all tweets retweeting those posts and the latest timeline tweets of each user. Since in our experiments we are incrementally increasing the amount of information types in the heterogeneous graphs, the number of nodes in \mathcal{V}_T varies depending on the experimental setup. Node features are a concatenation of textual BERT-base document embeddings, as well as retweet count and favorite count, i.e. $\forall v_t \in \mathcal{V}_T : v_t \in \mathbb{R}^{770}$. Finally, \mathcal{V}_U is the set of user nodes where features

² <https://developer.twitter.com/en/developer-terms/agreement-and-policy>

³ <https://github.com/flairNLP/flair>

are a concatenation of BERT-base profile description embeddings, follower count, friends count, favorites count and statuses count, i.e. $\forall v_u \in \mathcal{V}_U : v_u \in \mathbb{R}^{772}$. We also evaluate graphs with text embeddings only and leave out other features (e.g. number of favorite count) in that case. Edges connecting the nodes are satisfying constraints according to the node types they link together. More specifically, we use at most three types of edges: tweets citing news articles $((v_t, \tau_{TN}, v_n) \in \mathcal{E} \rightarrow v_t \in \mathcal{V}_T, v_n \in \mathcal{V}_N)$, users posting tweets (which also applies to users posting retweets and users posting timeline tweets) $((v_u, \tau_{UT}, v_t) \in \mathcal{E} \rightarrow v_u \in \mathcal{V}_U, v_t \in \mathcal{V}_T)$ and tweets retweeting tweets $((v_t, \tau_{TT}, u_t) \in \mathcal{E} \rightarrow v_t \in \mathcal{V}_T, u_t \in \mathcal{V}_T)$. The graphs fuse all types of information (the news piece as well as a range of different social media context features) without any prior aggregation step.

3.2 Experimental Setup

We construct graphs by incrementally increasing the amount of included data as follows: **(1)** We start with only considering tweets related to the news article; **(2)** We add user profiles of the people who posted the tweets to the graphs; **(3)** We include the five latest timeline posts of those users; **(4)** We add retweets related to the initial Twitter posts; **(5)** We combine all types of information mentioned so far. For simplicity we set the number of timeline tweets to five to keep the relation between number of all nodes and number of timeline tweet nodes in the graph manageable. The in-depth exploration of different settings is left as future work. To keep the structure of graphs consistent we exclude samples where news data are no longer available. Graphs with fewer than five vertices per node type are skipped to make sure that at least some social media features are available (other studies sometimes even set higher limits of e.g. a minimum of 15 tweets [15]). Furthermore, we make all edges undirected to improve message passing during graph convolution. For all experiments we use 5-fold cross-validation and report the average scores as results. We keep the overall number of graphs and the cross-validation split consistent for all setups and datasets to obtain comparable results. The number of graphs per dataset amounts to 483 for PolitiFact (real: 235; fake: 248), 12,214 for GossipCop (real: 10,067; fake: 2,147) and thus 12,697 for the full FakeNewsNet dataset (real: 10,302; fake: 2,395) which consists of a mixture of the two subsets.

We are also flattening all graphs into non-heterogeneous structures to understand how such a setup performs compared to the one described so far. Since homogeneous graphs do not explicitly model disjoint sets of vertices and use a single adjacency matrix, we have to unify all feature dimensions and we do that by either pruning all nodes to a dimension of 768 (only text) or zero-padding all nodes to a dimension of 772. We classify each graph using two-layer vanilla graph neural networks and evaluate three different types of graph convolution: (1) GraphSAGE; (2) Graph Attention Convolution (GAT); (3) Heterogeneous Graph Transformer (HGT). For details on the models we refer to our GitHub.

All models are trained with the same hyperparameter setting to obtain comparable results. We use 20 train epochs, a batch size of 16 and a learning-rate of $8e^{-5}$. As the number of real and fake graphs is unbalanced in terms of GossipCop

and the full FakeNewsNet dataset, we use class weights during experiments with those data. Our implementations are based on PyTorch Geometric⁴ [9]. Training and testing are executed using a single NVIDIA GTX 1080 GPU with an overall graphical memory size of 11GB. Even though we use high-dimensional node features and a large number of graphs during our experiments, training and inference in the described setup only take about 15 minutes on a single GPU (with respect to the highest number of graphs).

4 Results

In line with common practice in NLP, we report precision, recall and macro F1 scores for all setups [12]. Detailed results are reported on our GitHub. As previously mentioned, the values are the average results obtained by 5-fold cross validation. In general, excluding social features like retweet count and number of followers yields better results than concatenating them with text embedding to initially represent the graph nodes. We observe the highest scores for Politifact with setup (5) (0.979 macro F1 and accuracy), for GossipCop with setup (3) (0.972 macro F1 and 0.983 accuracy) and the full FakeNewsNet dataset with setup (5) (0.966 macro F1 and 0.979 accuracy), only considering textual features. Interestingly, we get good results for Politifact from the point of adding retweets. Timeline tweets instead are not important. For GossipCop we observe very similar performance for all setups after once including tweets and users. Here, the other features (retweets and timeline tweets) do not significantly change the model performance. As the full FakeNewsNet dataset mainly consists of GossipCop, we can observe very similar results here. The best performing GNN convolution type is throughout all setups HGT.

The results for the setups with homogeneous graphs are all significantly lower at $p < 0.05$ compared to their equivalent with heterogeneous graphs. This observation holds true for truncated and padded homogeneous graphs for all three datasets. We demonstrate how modelling disjoint sets of nodes and multiple types of edges improve the representation of social context and lead to overall significantly better fake news detection performance.

For comparison, we use two recently published approaches as strong baselines. As we excluded some of the articles from our experiments (due to missing social context information), we rerun both baseline systems with the same data used in our experiments. There are many other studies reporting results on FakeNewsNet (e.g. [1,8,6,27]) but we picked the two most competitive, namely: (1) **CMTR** [11]: BERT classified texts that were preprocessed using summarization techniques as well as additional social media features like comments. (2) **HetTransformer** [13]: node classification in a heterogeneous graph featuring user, post and news nodes using an encoder-decoder transformer model. We use our results obtained by the same setup (HGT and all types of information) for comparison against the baseline scores even if this setup does not always reach

⁴ https://github.com/pyg-team/pytorch_geometric

the highest performance in each setup (to avoid cherry-picking). Table 1 reports competitive and robust results on all datasets and a new benchmark performance on the Politifact subset (though this not statistically significant).

| Approach | Politifact | | Gossipcop | | FakeNewsNet | |
|------------------------|--------------|--------------|----------------|----------------|----------------|----------------|
| | F1 | ACC | F1 | ACC | F1 | ACC |
| CMTR [11] | 0.965 | 0.965 | 0.854** | 0.922** | 0.859** | 0.920** |
| HetTransformer [13] | 0.900** | 0.900** | 0.994** | 0.997** | 0.985** | 0.991** |
| HetSMCG (our approach) | 0.979 | 0.979 | 0.969 | 0.982 | 0.966 | 0.979 |

Table 1. F1 and accuracy scores for *Politifact*, *GossipCop* and the full *FakeNewsNet* dataset. Significant differences to our approach (at $p < 0.05$ and Bonferroni correction) are marked with with **.

5 Discussion

An interesting observation is that the results for the Politifact dataset improve when we include retweet data in our network structures. It has already been pointed out that fake news tend to have more retweets than real news [24], and the results provide some support for the utility of this feature. Moreover, it has also previously been reported that misinformation in general spreads more effectively due to its emotionalizing content [2]. Due to the small size of the dataset (real: 239; fake: 261), this could have a large influence on classification performance. The much larger GossipCop and full FakeNewsNet datasets (real: 12,610; fake: 3,185) seems to be less influenced by such information type-specific characteristics. Since we assign retweets to the same node type as article-related tweets and user timeline tweets we address the problem of information type-specific performance gaps. An observation that applies to every dataset is that including all types of information leads to the best performing setup. For our experiments we also find that there are only marginal differences between the different GNN architectures while HGT gives the highest performance.

We should also point out some limitations of our work. Even though we compare our results against strong, recently published baselines one needs to be careful when reporting and interpreting improvements over existing systems, e.g. [5]. Generalisability of insights is a related issue [3]. Providing access to all code and data in our Github repository helps addressing this concern.

6 Conclusion

We have demonstrated how including a variety of social media context information can improve fake news detection. By modelling the problem as a graph

classification task using heterogeneous graph data structures we achieve competitive results on a real-world dataset. Incorporating all the contextual information at our disposal leads to the overall best performances.

Acknowledgments

This work was supported by the project COURAGE: A Social Media Companion Safeguarding and Educating Students funded by the Volkswagen Foundation, grant number 95564.

References

1. Azri, A., Favre, C., Harbi, N., Darmont, J., Noûs, C.: Monitor: A multimodal fusion framework to assess message veracity in social networks. In: Bellatreche, L., Dumas, M., Karras, P., Matulevičius, R. (eds.) *Advances in Databases and Information Systems*. pp. 73–87. Springer International Publishing, Cham (2021)
2. Bakir, V., McStay, A.: Fake news and the economy of emotions. *Digital Journalism* **6**(2), 154–175 (2018). <https://doi.org/10.1080/21670811.2017.1345645>, <https://doi.org/10.1080/21670811.2017.1345645>
3. Bowman, S.: The dangers of underclaiming: Reasons for caution when reporting how NLP systems fail. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. pp. 7484–7499. Association for Computational Linguistics, Dublin, Ireland (May 2022), <https://aclanthology.org/2022.acl-long.516>
4. Chandra, S., Mishra, P., Yannakoudakis, H., Nimishakavi, M., Saeidi, M., Shutova, E.: Graph-based modeling of online communities for fake news detection. *CoRR abs/2008.06274* (2020), <https://arxiv.org/abs/2008.06274>
5. Church, K.W., Kordoni, V.: Emerging trends: Sota-chasing. *Natural Language Engineering* **28**(2), 249–269 (2022). <https://doi.org/10.1017/S1351324922000043>
6. Das, S.D., Basak, A., Dutta, S.: A heuristic-driven uncertainty based ensemble framework for fake news detection in tweets and news articles. *CoRR abs/2104.01791* (2021), <https://arxiv.org/abs/2104.01791>
7. Dori-Hacohen, S., Sung, K., Chou, J., Lustig-Gonzalez, J.: Restoring healthy online discourse by detecting and reducing controversy, misinformation, and toxicity online. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. p. 2627–2628. Association for Computing Machinery, New York, NY, USA (2021), <https://doi.org/10.1145/3404835.3464926>
8. Dou, Y., Shu, K., Xia, C., Yu, P.S., Sun, L.: User preference-aware fake news detection. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. p. 2051–2055. Association for Computing Machinery, New York, NY, USA (2021), <https://doi.org/10.1145/3404835.3462990>
9. Fey, M., Lenssen, J.E.: Fast graph representation learning with PyTorch Geometric. In: *ICLR Workshop on Representation Learning on Graphs and Manifolds* (2019)
10. Guo, Z., Schlichtkrull, M., Vlachos, A.: A Survey on Automated Fact-Checking. *Transactions of the Association for Computational Linguistics* **10**, 178–206 (02 2022). https://doi.org/10.1162/tac1_a_00454

11. Hartl, P., Kruschwitz, U.: Applying automatic text summarization for fake news detection. In: Proceedings of the Language Resources and Evaluation Conference. pp. 2702–2713. European Language Resources Association, Marseille, France (June 2022)
12. Jurafsky, D., Martin, J.: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 3rd (draft) edn. (2021), <https://web.stanford.edu/~jurafsky/slp3/>, <https://web.stanford.edu/~jurafsky/slp3/>
13. Li, T., Sun, Y., Hsu, S.I., Li, Y., Wong, R.C.W.: Fake news detection with heterogeneous transformer (2022). <https://doi.org/10.48550/ARXIV.2205.03100>, <https://arxiv.org/abs/2205.03100>
14. Lu, Y.J., Li, C.T.: GCAN: Graph-aware co-attention networks for explainable fake news detection on social media. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 505–514. Association for Computational Linguistics, Online (Jul 2020). <https://doi.org/10.18653/v1/2020.acl-main.48>, <https://aclanthology.org/2020.acl-main.48>
15. Lukasik, M., Cohn, T., Bontcheva, K.: Point process modelling of rumour dynamics in social media. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). pp. 518–523. Association for Computational Linguistics, Beijing, China (Jul 2015). <https://doi.org/10.3115/v1/P15-2085>, <https://aclanthology.org/P15-2085>
16. Mehta, N., Goldwasser, D.: Tackling fake news detection by interactively learning representations using graph neural networks. In: Proceedings of the First Workshop on Interactive Learning for Natural Language Processing. pp. 46–53. Association for Computational Linguistics, Online (Aug 2021). <https://doi.org/10.18653/v1/2021.internlp-1.7>, <https://aclanthology.org/2021.internlp-1.7>
17. Mehta, N., Pacheco, M., Goldwasser, D.: Tackling fake news detection by continually improving social context representations using graph neural networks. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1363–1380. Association for Computational Linguistics, Dublin, Ireland (May 2022), <https://aclanthology.org/2022.acl-long.97>
18. Min, E., Rong, Y., Bian, Y., Xu, T., Zhao, P., Huang, J., Ananiadou, S.: Divide-and-conquer: Post-user interaction network for fake news detection on social media. In: Proceedings of the ACM Web Conference 2022. p. 1148–1158. WWW ’22, Association for Computing Machinery, New York, NY, USA (2022). <https://doi.org/10.1145/3485447.3512163>, <https://doi.org/10.1145/3485447.3512163>
19. Mosca, E., Wich, M., Groh, G.: Understanding and interpreting the impact of user context in hate speech detection. In: Proceedings of the Ninth International Workshop on Natural Language Processing for Social Media. pp. 91–102. Association for Computational Linguistics, Online (Jun 2021). <https://doi.org/10.18653/v1/2021.socialnlp-1.8>, <https://aclanthology.org/2021.socialnlp-1.8>
20. Nakov, P., Corney, D., Hasanain, M., Alam, F., Elsayed, T., Barrón-Cedeño, A., Papotti, P., Shaar, S., Da San Martino, G.: Automated Fact-Checking for Assisting Human Fact-Checkers. In: Zhou, Z.H. (ed.) Proceedings of the Thirtieth

- International Joint Conference on Artificial Intelligence, IJCAI-21. pp. 4551–4558. International Joint Conferences on Artificial Intelligence Organization (8 2021). <https://doi.org/10.24963/ijcai.2021/619>, survey Track
21. Nguyen, V.H., Sugiyama, K., Nakov, P., Kan, M.Y.: Fang: Leveraging social context for fake news detection using graph representation. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. p. 1165–1174. CIKM '20, Association for Computing Machinery, New York, NY, USA (2020). <https://doi.org/10.1145/3340531.3412046>, <https://doi.org/10.1145/3340531.3412046>
 22. Rode-Hasinger, S., Kruspe, A., Zhu, X.X.: True or false? detecting false information on social media using graph neural networks. In: Proceedings of the Eighth Workshop on Noisy User-generated Text (W-NUT 2022). pp. 222–229. Association for Computational Linguistics, Gyeongju, Republic of Korea (Oct 2022), <https://aclanthology.org/2022.wnut-1.24>
 23. Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., Liu, Y.: Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)* **10**(3), 1–42 (2019), publisher: ACM New York, NY, USA
 24. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: Fakenews-net: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data* **8**(3), 171–188 (2020). <https://doi.org/10.1089/big.2020.0062>, <https://doi.org/10.1089/big.2020.0062>, PMID: 32491943
 25. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: A data mining perspective. *SIGKDD Explor. Newsl.* **19**(1), 22–36 (sep 2017). <https://doi.org/10.1145/3137597.3137600>, <https://doi.org/10.1145/3137597.3137600>
 26. Shu, K., Wang, S., Liu, H.: 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. p. 312–320. WSDM '19, Association for Computing Machinery, New York, NY, USA (2019). <https://doi.org/10.1145/3289600.3290994>, <https://doi.org/10.1145/3289600.3290994>
 27. Song, C., Shu, K., Wu, B.: Temporally evolving graph neural network for fake news detection. *Information Processing & Management* **58**(6), 102712 (2021). <https://doi.org/https://doi.org/10.1016/j.ipm.2021.102712>, <https://www.sciencedirect.com/science/article/pii/S0306457321001965>
 28. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. *Science* **359**(6380), 1146–1151 (2018). <https://doi.org/10.1126/science.aap9559>, <https://www.science.org/doi/abs/10.1126/science.aap9559>
 29. Zhou, X., Zafarani, R.: A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Comput. Surv.* **53**(5) (sep 2020). <https://doi.org/10.1145/3395046>, <https://doi.org/10.1145/3395046>