

True or False? Detecting False Information on Social Media Using Graph Neural Networks

Samyo Rode-Hasinger¹, Anna Kruspe^{1,2} and Xiao Xiang Zhu¹

¹Technical University of Munich

²Technische Hochschule Nürnberg

{samyo.rode, xiaoxiang.zhu}@tum.de, anna.kruspe@th-nuernberg.de

Abstract

In recent years, false information such as fake news, rumors and conspiracy theories on many relevant issues in society have proliferated. This phenomenon has been significantly amplified by the fast and inexorable spread of misinformation on social media and instant messaging platforms. With this work, we contribute to containing the negative impact on society caused by fake news. We propose a graph neural network approach for detecting false information on Twitter. We leverage the inherent structure of graph-based social media data aggregating information from short text messages (tweets), user profiles and social interactions. We use knowledge from pre-trained language models efficiently, and show that user-defined descriptions of profiles provide useful information for improved prediction performance. The empirical results indicate that our proposed framework significantly outperforms text- and user-based methods on misinformation datasets from two different domains, even in a difficult multilingual setting.

1 Introduction

The spread of misinformation on social media is a growing problem that can hardly be tackled without the help of AI-based detection methods due to the large amount of data and its complexity. This is evident in crisis situations such as the COVID-19 pandemic (Naeem and Boulos, 2021) or Russia’s attack on Ukraine where a sheer flood of fake news has exacerbated the situation causing great insecurity and harm among the people.

Previous work has focused primarily on the verification of news content, taking into account user profiles and propagation patterns in social networks. However, in real life scenario, news articles are not always freely available, and matching user-generated content from social media to published articles is often hard to accomplish (Shu et al., 2017). Therefore, we propose a method for au-

tomatic fake news detection that is based only on data available on social media. We introduce a unified framework with graph neural networks (GNNs) that leverages short text messages, user profile information and social network properties. As a case study, we train and evaluate our model on mono- and multilingual social media content from Twitter. To this end, we jointly model the heterogeneous graph structure of the data formed by users, retweeters and their tweets, and cast the verification task as a node classification problem. We exploit self-defined profile descriptions from Twitter users and retweeters as well as the tweets’ text to create initial user and tweet node features. Unlike previous approaches which use pre-trained word embeddings to encode text features (Monti et al., 2019) or learn word-level features during training (Lu and Li, 2020), we utilize state-of-the-art context-aware multilingual representations from Sentence-BERT (Reimers and Gurevych, 2019). Since we avoid expensive fine-tuning of the text encoders, we make our model efficient and easily applicable. Finally, we train our system in an inductive setting, boosting its capability to reliably predict new unseen instances without the need of re-training.

2 Related Work

Text- or content-based fake news detection models have been greatly enhanced by the advancement of pre-trained language models (Hossain et al., 2020; Kaliyar et al., 2021; Panda and Levitan, 2021; Tzifafas et al., 2021). Since GNNs leverage news propagation patterns and user network information, they are particularly suitable for social media data. However, GNNs have only recently been introduced for the detection of false information in social networks.

Monti et al. (2019) collect news stories and Twitter content, and are the first to employ a GNN architecture to model text and user features together with the social network properties for fake news

detection. Lu and Li (2020) extract user and tweet features, and model user propagation paths with GRU- and CNN-based models. A graph convolutional network (GCN) is used to learn interactions among users who share the same content. Chandra et al. (2020) independently train a text encoder for news content and a graph encoder for modelling the follower-following network of users spreading a new articles. Han et al. (2020) cast fake news detection as a graph classification task, extract features from Twitter’s user objects only and use GNNs to compute the dissemination of news content among multiple users. The authors tackle the problem of new, unseen data by using techniques from continual learning. Finally, Dou et al. (2021) propose a GNN-based method for user preference-aware fake news detection which exploits historical user posts for node generation and models propagation patterns of news articles among respective retweeters.

Except for Lu and Li (2020) and Han et al. (2020), the above-mentioned approaches incorporate extensive text content from news articles. We follow the approach of Lu and Li (2020) by addressing the challenge of classifying short and noisy text documents, but instead of applying GNNs to user networks only, we propose a unified way of modelling user interactions and user-created content with a graph-based approach. Moreover, none of the previous works address multilingual aspects of social media messages and user origins. We show that our approach significantly outperforms text- and user-based baselines even in a multilingual setting.

3 Methodology

3.1 Graph Representation

We model Twitter users, their social media posts (tweets), and their interrelations by building a network graph with nodes and edges. We denote the set of user nodes by \mathcal{U} and the set of tweet nodes by \mathcal{T} . We establish connections, i.e., graph edges, between users and their tweets e^T and between tweets and users who re-posted (retweeted) a tweet e^R . For a given dataset, we construct a heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a set of nodes $\mathcal{V} = \{\mathcal{U} \cup \mathcal{T}\}$ and a set of edges $\mathcal{E} = \{e^T \cup e^R\}$. We note that an explicit connection between users and retweeters is not necessary, because these interactions are learned by means of the depth of our network. Likewise, the model can learn relations between tweets from users and retweeters if a

tweet-retweeter connection exists.

3.2 Tweet Nodes

Let $t_i \in \mathcal{T} = \{t_1, t_2, \dots, t_N\}$ be a tweet in a given dataset of size N . We generate the initial tweet nodes for our network graph by encoding the tweet’s text with a pre-trained language model. We preprocess the text by replacing URLs with the ‘HTTPURL’ token, e-mail addresses with the ‘EMAIL’ token and user mentions with the ‘@USER’ token. We also convert emojis into their corresponding string shortcodes.¹ We use Sentence-BERT (SBERT) to generate a vector representation \mathbf{v}_{t_i} of each processed tweet $t_i \in \mathcal{T}$. Specifically, we test two multilingual embedding models of different sizes from the SentenceTransformers library² trained in a teacher-student setting (Reimers and Gurevych, 2020): 1. `distiluse-base-multilingual-cased-v1` which is based on Multilingual Universal Sentence Encoder (mUSE) (Chidambaram et al., 2019; Yang et al., 2019) and a distilled version of mBERT (Sanh et al., 2019). This model supports 15 languages and has an embedding dimension $d_D = 512$. 2. `paraphrase-multilingual-mpnet-base-v2`, which was trained using `paraphrase-mpnet-base-v2` (Song et al., 2020) as teacher and the base version of XLM-RoBERTa (Conneau et al., 2020) as student model. It supports 50+ languages and has an embedding dimension $d_M = 768$.

3.3 User Nodes

Each tweet t_i is authored or retweeted by a user u_j on Twitter. The set of users in each dataset is defined as $\mathcal{U} = \{u_1, u_2, \dots, u_M\}$, where M is the total number of unique users. M includes the number of authors and the number of retweeters. It should be noted that a user u_j can be the author and retweeter of one or more tweets at the same time. To initialize the user nodes in our network graph, we generate a vector representation \mathbf{v}_{u_j} of the user’s `description` attribute contained in the user object.³ Again, we use preprocessing and the two pre-trained multilingual models from SentenceTransformers introduced in Sec. 3.2. To distinguish our systems with different initial tweet and

¹<https://pypi.org/project/emoji/>

²<https://www.sbert.net/>

³<https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/user>

user node representations in our experiments, we use the prefixes ‘Distiluse-’ and ‘Mpnet-’.

3.4 Model

Our proposed fake news detection framework has a 2-layer GNN at its core and takes as input the heterogeneous graph described in Sec. 3.1. We initialize the user and the tweet nodes with their corresponding embeddings $\mathbf{v}_u \in \mathbb{R}^d$ and $\mathbf{v}_t \in \mathbb{R}^d$ which we do not fine-tune during training.

Next, we project the embeddings into a lower dimensional space $\mathbf{h}_{v_i} \in \mathbb{R}^{128}$ using a separate fully-connected layer followed by a ReLU activation function for each node type $v_i \in \mathcal{V}$.

For computing the node representations, we implement the GraphSAGE operator (Hamilton et al., 2017) according to the PyTorch Geometric⁴ library, add a ReLU non-linearity, and apply it to all edge types specified in \mathcal{E} (Sec. 3.1). One operation step of the GraphSAGE convolution with the mean-based aggregator at layer k is defined as:

$$\mathbf{h}'_v = \text{ReLU}(\mathbf{W}_1^k \mathbf{h}_v + \mathbf{W}_2^k \cdot \text{mean}_{n \in \mathcal{N}(v)} \mathbf{h}_n)$$

where $n \in \mathcal{N}(v)$ is a node in the neighborhood of v , \mathbf{h}_n its hidden representation, and \mathbf{W}_1^k and \mathbf{W}_2^k are the weight matrices at the k -th layer. Additionally, we ℓ_2 -normalize the output features of each node and group the features generated by different relations by summation.

We test one variant of our proposed network by replacing the SAGE operator with a graph attention network (GAT) (Veličković et al., 2017). By employing a self-attention mechanism (Vaswani et al., 2017), GAT learns different parameters for different nodes in a neighborhood and has been utilized in previous works (Monti et al., 2019; Chandra et al., 2020).

For the final binary node classification of ‘real’ and ‘fake’ tweets, we feed the tweet representations $\mathbf{h}_t \in \mathbb{R}^{128}$ learned by the GNN into a 2-layer feed-forward network with a ReLU non-linearity after the hidden layer and a logistic sigmoid function after the final layer.

4 Experimental Setup

4.1 Datasets

We collect two published fake news datasets which provide social media context: *FakeNewsNet* (Shu

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

et al., 2020), and a multilingual dataset related to COVID-19 (Alam et al., 2021b) which we refer to as *Covid-19-Disinfo*.

FakeNewsNet is a popular dataset for automated fake news detection which contains English news articles from two fact-checking websites together with related content from Twitter. For our study, we use the ‘fake’ and ‘real’ tweets compiled from PoliFact⁵ available at the *FakeNewsNet* data repository website.⁶ We hydrate the tweet objects via Twitter’s API using tweepy.⁷ As many tweets have been deleted since the date of the publication of the dataset (Balestrucci and De Nicola, 2020), we end up with a total size of 289,602 ‘real’ and 111,101 ‘fake’ (unique) tweets, which is 72.54% and 67.38% of the original dataset size, respectively. To prevent data leakage and bias during training and evaluation, we remove similar tweet objects from the dataset by normalizing the tweets’ text (incl. lowercasing, see Sec. 3.2) and applying exact-duplicate filtering according to Alam et al. (2021a). This results in a total number of 282,643 instances, with 233,071 tweets being annotated as ‘real’ and 49,572 as ‘fake’. In order to counteract the impact of an unbalanced dataset, we randomly sample 49,000 tweets from each class label. Finally, we randomly split all instances into 70% train, 10% validation and 20% test sets.

Covid-19-Disinfo is a multilingual Twitter dataset related to the spread of false information during the COVID-19 pandemic. The dataset was compiled for fine-grained disinformation analysis and contains various independent classification tasks formulated in the form of questions. We choose the binary classification task ‘Q2’ which is designed for detecting false information. When downloading the tweet objects via the Twitter API, we face similar issues as mentioned above. From the total number of 9,583 tweet IDs (Q2 task) we were able to hydrate only 8,810 unique tweet objects from Twitter, resulting in a predefined train, validation and test split of 6,462, 602 and 1,746 tweet objects, respectively.

We extend *FakeNewsNet* and *Covid-19-Disinfo* with 73,722 and 57,966 unique retweeter objects, respectively. Thus, we obtain a total number of 147,690 unique user objects for *FakeNewsNet* and 62,598 unique user objects for *Covid-19-Disinfo*.

⁵<https://www.politifact.com/>

⁶<https://github.com/KaidMML/FakeNewsNet>

⁷<https://www.tweepy.org/>

Model	FakeNewsNet	Covid-19-Disinfo
Mpnet-Tweet	.8817 (.0025)	.4101 (.0274)
Distiluse-Tweet	.8618 (.0013)	.3791 (.0243)
Mpnet-Tweet-User	.8696 (.0010)	.4135 (.0192)
Distiluse-Tweet-User	.8650 (.0003)	.3310 (.0228)
Mpnet-GAT	.9241 (.0016)	.4252 (.0193)
Distiluse-GAT	.9351 (.0006)	.3889 (.0325)
Mpnet-SAGE	.9370 (.0008)	.4868 (.0172)
Distiluse-SAGE	.9467 (.0015)	.4421 (.0055)

Table 1: Mean F_1 scores (‘fake’ class) and standard deviation (\pm) of 5 runs on the test sets of *FakeNewsNet* (Politifact) and *Covid-19-Disinfo*. **Bold**: Best overall performance for each dataset.

4.2 Baselines

We use two baseline models to compare the performance of our proposed GNN model each with two input feature variations.

Tweet Neural Network. We encode the tweets’ text adopting the same embedding models described in Sec. 3.2. We then compute a prediction for each instance with a 3-layer feed-forward network similar to the prediction network in our GNN model. We also use a hidden size of 128, but add the tanh activation function (instead of ReLU) after each layer.⁸ We refer to these baselines as ‘Distiluse-Tweet’ and ‘Mpnet-Tweet’, depending on the embedding model.

Tweet-User Neural Network. We encode the tweets’ text and the description attribute of the user objects (see Sec. 3.3). We use two separate 2-layer feed-forward networks to obtain the hidden representations $\mathbf{h}_u, \mathbf{h}_t \in \mathbb{R}^{128}$ of user u_j who posted tweet t_i in the dataset. Again, we use the tanh activation function after each layer. Intuitively, the network should learn the interrelation between users and their messages. We compute $\mathbf{h}' = \mathbf{h}_u \oplus \mathbf{h}_t$, where \oplus is the concatenation operator, and use another fully-connected layer for the final prediction. We denote this baseline by ‘Tweet-User’ prepended by the embedding specifier.

5 Results and Analysis

For each model architecture, we report the mean F_1 of the positive class (‘fake’) of five runs with different random seeds. The results are listed in Table 1. Overall it can be observed that our proposed GNN model outperforms all baselines on both datasets, except for ‘GAT’ which is inferior with initial ‘Distiluse’ features and only marginally

⁸In our preliminary experiments, the tanh activation function performed slightly better than ReLU.

better with ‘Mpnet’ on *Covid-19-Disinfo*. Specifically designed for inductive graph representation learning, the SAGE module is more robust than GAT and can generalize better on unseen test data (Brody et al., 2021).

As for *FakeNewsNet*, Distiluse-SAGE outperforms all baseline architectures, Mpnet-SAGE and our proposed GNNs with the GAT operator. Among the baseline models, additional user information (Tweet-User) only helps with ‘Distiluse’ embeddings. However, the best results are achieved with ‘Mpnet’ representations. Since ‘Mpnet’ embeddings have a larger dimension, i.e., $d_M = 768$ vs. $d_D = 512$, they have the ability to capture more information within shallower networks. Among the GNN architectures initial ‘Distiluse’ nodes outperform their ‘Mpnet’ counterparts.

Regarding *Covid-19-Disinfo*, all ‘Mpnet’ models outperform the ‘Distiluse’ models. The larger embedding size seems to be useful in scenarios where little training data is available. For our proposed approach, ‘Mpnet’ embeddings outperform ‘Distiluse’ representations by roughly 4 percentage points in both setups, i.e., ‘GAT’ and ‘SAGE’. The strongest model, Mpnet-SAGE, is more than 7 percentage points better than the strongest baseline, Mpnet-Tweet-User.

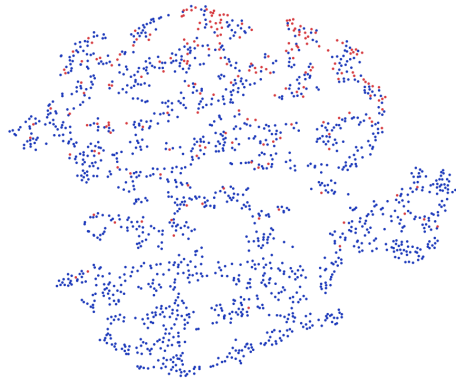


Figure 1: t-SNE (van der Maaten and Hinton, 2008) plot of *Covid-19-Disinfo* tweet embeddings (test set) generated by the baseline Mpnet-Tweet model. Fake tweets are in red.

In general, the results suggest that the *Covid-19-Disinfo* classification task is much harder than the *FakeNewsNet* task. Most likely this is due to the lack of sufficient training and validation data, since we use regularization methods to mitigate overfitting. Other reasons could be the multilingual character of the content and the domain-specific vocabulary which is difficult to capture for the pre-

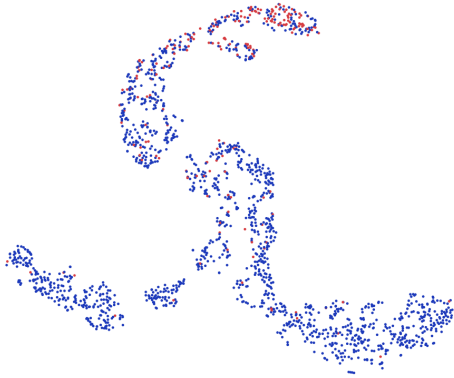


Figure 2: t-SNE plot of *Covid-19-Disinfo* tweet features (test set) generated by our MpNet-SAGE encoder. Fake tweets are in red.

trained language models. However, the final representations of ‘fake’ and ‘real’ tweets generated by our proposed GNN detection framework are more distinct than the baseline features, and, therefore, help to improve detection performance on both datasets (see Figs. 1 and 2).

Although we use publicly available data in a purely observational manner, we point out that our model may learn a ‘semantic bias’ (Shah et al., 2020) towards user-defined descriptions. Under certain conditions, this bias could lead to questionable results that are not intended.

Ablation Study In order to investigate the effect of the models’ input components to the results, we conduct a comparative study with the best performing model on the corresponding dataset. To this end, we either randomize tweet (‘SAGE (rnd tweets)’ or user (‘SAGE (rnd users)’ node features while keeping other model settings constant. The results of 5 runs (Table 2) indicate that in the balanced dataset scenario with sufficient examples (*FakeNewsNet*), pre-trained tweet nodes primarily contribute to the performance of Distiluse-SAGE. Yet, both node representations modelled with our proposed GNN lead to the significant performance gain. In the case of the more challenging *Covid-19-Disinfo* dataset, we observe for both model variations a sharp drop in performance to almost equal F_1 scores. This indicates that both input components equally contribute to the performance increase of our proposed model.

6 Conclusion

In this work, we present a simple, yet efficient GNN approach for the detection of fake news on social

Model	FakeNewsNet	Covid-19-Disinfo
SAGE (rnd tweets)	.6594 (.0138)	.2964 (.0318)
SAGE (rnd users)	.9102 (.0016)	.3019 (.0387)
MpNet-SAGE	–	.4868 (.0172)
Distiluse-SAGE	.9467 (.0015)	–

Table 2: Ablation results (mean F_1 scores (‘fake’ class) and standard deviation (\pm)) of best performing models randomizing either tweet or user node features. **Bold:** Best performance without random features.

media. Our model employs pre-trained language models to encode text features of social media messages and user profile descriptions. By jointly modelling the relations between users and their tweets and between users who shared similar content, our GNN architecture outperforms text-based models as well as models which combine text and user features from pre-trained language models. In addition, our model is able to apply its knowledge to unseen data without the need of re-training. We show that our approach has limitations in settings with insufficient training data. But with the right choice of initial node representations, the model still outperforms all baselines. In future work, we will investigate domain-adapted language models for initializing graph nodes. Further, we plan to evaluate our model on similar social media content, such as Reddit (Sakketou et al., 2022).

Acknowledgements

The work is jointly supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. [ERC-2016-StG-714087], Acronym: *So2Sat*), by the Helmholtz Association through the Helmholtz Excellent Professorship “Data Science in Earth Observation - Big Data Fusion for Urban Research” (grant number: W2-W3-100), and by the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab “AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond” (grant number: 01DD20001). We would like to thank Matthias Häberle and all the anonymous reviewers for their valuable input.

References

Firoj Alam, Hassan Sajjad, Muhammad Imran, and Ferda Ofli. 2021a. *Crisisbench: Benchmarking crisis-related social media datasets for humanitarian in-*

- formation processing. *Proceedings of the International AAAI Conference on Web and Social Media*, 15(1):923–932.
- Firoj Alam, Shaden Shaar, Fahim Dalvi, Hassan Sajjad, Alex Nikolov, Hamdy Mubarak, Giovanni Da San Martino, Ahmed Abdelali, Nadir Durrani, Kareem Darwish, Abdulaziz Al-Homaid, Wajdi Zaghoulani, Tommaso Caselli, Gijs Danoe, Friso Stolk, Britt Bruntink, and Preslav Nakov. 2021b. **Fighting the COVID-19 infodemic: Modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society.** In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 611–649, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Alessandro Balestrucci and Rocco De Nicola. 2020. **Credulous users and fake news: a real case study on the propagation in twitter.** In *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, pages 1–8.
- Shaked Brody, Uri Alon, and Eran Yahav. 2021. **How attentive are graph attention networks?** *ArXiv*, abs/2105.14491.
- Shantanu Chandra, Pushkar Mishra, Helen Yanakoudakis, and Ekaterina Shutova. 2020. **Graph-based modeling of online communities for fake news detection.** *ArXiv*, abs/2008.06274.
- Muthu Chidambaram, Yinfei Yang, Daniel Cer, Steve Yuan, Yunhsuan Sung, Brian Strope, and Ray Kurzweil. 2019. **Learning cross-lingual sentence representations via a multi-task dual-encoder model.** In *Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019)*, pages 250–259, Florence, Italy. Association for Computational Linguistics.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. **Unsupervised cross-lingual representation learning at scale.** In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.
- Yingtong Dou, Kai Shu, Congying Xia, Philip S. Yu, and Lichao Sun. 2021. **User preference-aware fake news detection.** In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '21*, page 2051–2055, New York, NY, USA. Association for Computing Machinery.
- William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. **Inductive representation learning on large graphs.** In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, page 1025–1035, Red Hook, NY, USA. Curran Associates Inc.
- Yi Han, Shanika Karunasekera, and Christopher Leckie. 2020. **Graph neural networks with continual learning for fake news detection from social media.** *ArXiv*, abs/2007.03316.
- Tamanna Hossain, Robert L. Logan IV, Arjuna Ugarte, Yoshitomo Matsubara, Sean Young, and Sameer Singh. 2020. **COVIDLies: Detecting COVID-19 misinformation on social media.** In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*, Online. Association for Computational Linguistics.
- Rohit Kumar Kaliyar, Anurag Goswami, and Pratik Narang. 2021. **Fakebert: Fake news detection in social media with a bert-based deep learning approach.** *Multimedia Tools Appl.*, 80(8):11765–11788.
- Diederik P. Kingma and Jimmy Ba. 2014. **Adam: A method for stochastic optimization.**
- Yi-Ju Lu and Cheng-Te Li. 2020. **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*, pages 505–514, Online. Association for Computational Linguistics.
- Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M. Bronstein. 2019. **Fake news detection on social media using geometric deep learning.** *ArXiv*, abs/1902.06673.
- Salman Bin Naeem and Maged N Kamel Boulos. 2021. **Covid-19 misinformation online and health literacy: A brief overview.** *International Journal of Environmental Research and Public Health*, 18.
- Subhadarshi Panda and Sarah Ita Levitan. 2021. **Detecting multilingual COVID-19 misinformation on social media via contextualized embeddings.** In *Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 125–129, Online. Association for Computational Linguistics.
- Nils Reimers and Iryna Gurevych. 2019. **Sentence-BERT: Sentence embeddings using Siamese BERT-networks.** In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.
- Nils Reimers and Iryna Gurevych. 2020. **Making monolingual sentence embeddings multilingual using knowledge distillation.** In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4512–4525, Online. Association for Computational Linguistics.
- Flora Sakketou, Joan Plepi, Riccardo Cervero, Henri Jacques Geiss, Paolo Rosso, and Lucie Flek.

2022. [Factoid: A new dataset for identifying misinformation spreaders and political bias](#). In *Proceedings of the Language Resources and Evaluation Conference*, pages 3231–3241, Marseille, France. European Language Resources Association.
- Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. [Distilbert, a distilled version of BERT: smaller, faster, cheaper and lighter](#). *CoRR*, abs/1910.01108.
- Deven Santosh Shah, H. Andrew Schwartz, and Dirk Hovy. 2020. [Predictive biases in natural language processing models: A conceptual framework and overview](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5248–5264, Online. Association for Computational Linguistics.
- Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. [Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media](#). *Big Data*, 8(3):171–188. PMID: 32491943.
- 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):22–36.
- Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. [Mpnnet: Masked and permuted pre-training for language understanding](#). In *NeurIPS 2020*. ACM.
- Georgios Tziafas, Konstantinos Kogkalidis, and Tommaso Caselli. 2021. [Fighting the COVID-19 infodemic with a holistic BERT ensemble](#). In *Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 119–124, Online. Association for Computational Linguistics.
- L.J.P. van der Maaten and G.E. Hinton. 2008. [Visualizing high-dimensional data using t-sne](#). *Journal of Machine Learning Research*, 9(nov):2579–2605. Pagination: 27.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, page 6000–6010, Red Hook, NY, USA. Curran Associates Inc.
- Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2017. [Graph attention networks](#).
- Yinfei Yang, Daniel Cer, Amin Ahmad, Mandy Guo, Jax Law, Noah Constant, Gustavo Hernández Ábrego, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. 2019. [Multilingual universal sentence encoder for semantic retrieval](#). *CoRR*, abs/1907.04307.

A Additional Training Configurations

We use PyTorch⁹ and the PyTorch Geometric library to build our models. We train our GNN framework for 300 epochs with early stopping. We optimize with Adam (Kingma and Ba, 2014) setting the learning rate to 0.005 and weight decay to 0.001. For regularization of the whole network, we use dropout with $p = 0.3$ before the first fully-connected layer, after each graph neural network layer, and after the hidden layer in the prediction network.

We train all baseline models for 100 epochs with early stopping and a batchsize of 64. We set the size of all hidden layers to 128. Again, we optimize with Adam setting the learning rate to 0.005 and weight decay to 0.001. We use dropout with $p = 0.5$ after the first hidden layer for regularization. All experiments are run on NVIDIA GeForce RTX 3090 24 GB GPUs.

⁹<https://pytorch.org/>