Misinformation detection based on news dispersion

Vasileios Moschopoulos [*], Maria Tsourma [*] [†], Anastasios Drosou [*], and Dimitrios Tzovaras [*]

[*]*Information Technologies Institute, Centre for Research and Technology Hellas (CERTH),*
Thermi-Thessaloniki, Greece

[†]*National and Kapodistrian University of Athens, Greece*
Email: vmoscho@iti.gr, mtsourma@iti.gr, drosou@iti.gr, dimitrios.tzovaras@iti.gr

***Abstract*—Misinformation dispersion is of critical importance nowadays, especially in the media sector. With the increase of data sources and the rise of information dispersion within social media platforms, users face the dilemma of whether the information they come across is true or false. Technology has advanced significantly in recent years, aiming to detect misinformation by improving existing algorithms and introducing new features to achieve its early detection. In this paper, a Propagation, and opinion-related Feature-enhanced GNN-based framework (PFGNN) is presented. The aim of this framework is to capture all the correlations among the collected information and identify covered patterns in the collected data, in order to efficiently detect false and real information through the creation of an information propagation tree. This framework receives as input an enhanced propagation tree, including information collected from social media platforms, and news sources, along with propagation-related features extracted from the graph's analysis, news representation, as well as social media users' posts, in an attempt to capture user-specific behavior. This method has been evaluated on two data sets (i.e. PolitiFact and GossipCop), while has been compared with a baseline method using a form of user preferences information capture to tackle the misinformation detection challenge.**

I. Introduction

The spread of misinformation has become a significant problem in modern society, especially with the rise of social media and online news platforms. The increasing prevalence of social media platforms has exacerbated the problem, as misinformation can spread rapidly and reach a large audience. Misinformation is defined as "intentionally and verifiably false information that is spread widely and rapidly through communication channels with the potential to cause harm" [1]. This definition emphasizes both on the intentional nature of misinformation (i.e., it is created with the purpose of deceiving or misleading people), as well as its potential to cause harm or increase violence [2]. Thus, identifying and thwarting misinformation has become crucial in fields including journalism, public administration, and information management.

In journalism, more specifically, misinformation detection is an important task for journalists, as it allows them to maintain their credibility, uphold journalistic standards, and promote accurate reporting and media literacy. In media, misinformation is spread on a wide scale since when an event occurs, journalists have limited time and resources for the information collection and filtering process. This leads to the dissemination of false information, which can have serious consequences in politics and in other fields of society, while can threaten also social cohesion.

Detecting misinformation is a challenging task due to its complexity. Misinformation articles or posts can be intentionally crafted to look like legitimate pieces of information and may contain subtle differences that make them difficult to detect [3] [4]. Therefore, using not only the text itself but relying also upon additional information that concerns the user or the information dispersion can lead to uncovering distinct patterns of real and false information propagation. In literature, Extended Artificial Intelligence (AI) and Machine learning (ML) methods have been implemented to tackle the misinformation detection problem [5] [6]. In addition, graph-based neural networks have been implemented also in this scope, aiming to capture a wider extent of user's behavior and the information propagation path [7] [8].

This paper presents a graph-based module enhanced with propagation and opinion-related features, called Propagation and opinion-related Feature-enhanced GNN-based module (PFGNN). This module aims to detect misinformation using as input a propagation graph depicting the dispersion of a piece of information among social media platforms and news websites, including also extended propagation features on each node. The list of input features leveraged by the module includes the textual information available for each news piece either on social media or in news articles, and also additional propagation and opinion-related features concerning the news representation, extracted from the analysis of the propagation graph. The contribution of this paper is the:

- Proposal of a Graph-based neural network that can efficiently use the information provided as input in order to increase its accuracy in misinformation detection.
- Evaluation of the proposed model's efficiency with the use of 16 propagation-related features, resulting from the propagation tree analysis, along with user characteristics collected from social media networks.
- Experimentation with the combination of different Graph Neural Network (GNN) layers (i.e., TransformerConv and SAGE) to evaluate the accuracy of the implemented module.

The rest of the paper is organized as follows. Section 2 presents the related work available in the literature on misinformation detection. The Propagation-related feature-enhanced GNN model proposed in this paper is presented in detail in section 3, including also the presentation of the information propagation tree (aka knowledge graph) designed. Section 4 includes the experimental settings used for the proposed

979-8-3503-3959-8/23/$31.00 ©2023 IEEE

model's evaluation along with the evaluation results. Section 5 presents a discussion of the evaluation outcomes resulted and the conclusions of the paper along with future work.

II. Related work

The detection of misinformation has become a topic of increasing interest in recent years, and a number of approaches have been proposed in the literature. Traditional approaches to misinformation detection have often relied on textual analysis and machine learning algorithms [9]. One promising solution to this problem is using Extended Artificial Intelligence (AI) and Machine learning (ML) methods, along with Natural Language Processing (NLP) techniques for the automated misinformation detection [10] [11].

In earlier works on fake news classification, hand-crafted features extracted primarily from news textual data, as well as features related to users that spread or publish news items were used. One negative side of such pure linguistic features is that tend to lose their discriminative ability, as fake news language similarity to real news seems to increase over time [12]. These extracted features are treated as input to traditional machine learning classifiers [13].

However, these methods have limitations in capturing the complex relationships between entities in a news article or social media post. Graph-based methods, on the other hand, have emerged as a promising approach for misinformation detection, due to their ability to model the complex relationships between entities in a news article or social media post.

In this category, the User Preference-aware Fake News Detection (UPFD) model [7] was used to evaluate the effect of user endogenous preferences, by taking also as input a series of graph convolutional encoding layers. In more detail, for each user node that has participated in the diffusion of a news piece, an initial vector representation is created from the user's past historical post data (i.e., embedding encoding of older tweets). The exogenous preferences are extracted, as usual, from news propagation graph connections. In [14], the fact that not all tweet nodes should have the same weight during the pooling operation is stated, so a context attention pooling is being used, instead of a normal mean pooling layer, resulting in improved classification. In study [15], the authors utilize a hypergraph model connecting news items with three types of hyperedges, in an attempt to capture news-level group connections for the identification of false information.

Authors in [16] use a combination of 3 graph feature extraction modules. The first extracts a set of static (global) graph features. The second module extracts a similar feature vector by feeding similar graph features extracted at specific timestamps to LSTM layers, while the final module performs structural analysis on the graph using node2vec as an encoder.

Graph-based methods can also be used to model the propagation of misinformation on social media platforms. The spread of misinformation can be modeled as a diffusion process on a network, with nodes representing users or news articles and edges representing the propagation of information between nodes [17] [18].

In addition, in literature the use of features extracted from the analysis of the propagation tree has been conducted [19] [20], showing that these parameters affect the efficiency of the corresponding AI model used for the misinformation detection task. However, the results of this analysis are based also on the information included within each dataset, and therefore further analysis is required in order to examine if these features can be used in conjunction with additional features in order to enhance the accuracy of the results. Based on the aforementioned literature results, and aims to evaluate the use of user characteristics along with propagation-based features on the misinformation detection task, this paper proposes a graph-based neural network using as input not only textual information from social media posts but also user profile characteristics and propagation-based information extracted from the propagation-graph analysis. At this point, it should be mentioned that this paper proposes the initial results of this research topic.

III. Propagation-related feature-enhanced GNN model

This section presents the PFGNN method proposed, the implementation of the knowledge graph used, along with the features provided as input in the model. At this point, it should be defined that the labels resulting from the proposed network are two, corresponding to true and false information.

A. *Propagation tree construction*

Initially, all the raw input data collected from both social media and news articles are represented in the form of a graph. The root node represents the source news item while starting from this node, the complete news dissemination on social media (i.e. Twitter) sequence is unfolded. This unweighted graph structure contains a series of tweets, retweets, and replies, that were part of this news social media spread in one form or another. Two tree nodes are connected if the one is a retweet or a reply of the other, while the root node is connected to all the tweet nodes that actually tweeted it. For each node, tweet text and author user profile information are accessible through the official Twitter API [21]. Twitter's social media platform was selected due to the great volume of information posted by the majority of users.

B. *PFGNN Module's architecture*

The module presented in this paper, named Propagation-related Feature-enhanced GNN-based framework (PFGNN), is built with the capability to encode users' social media interactions during the dissemination of a piece of information, in order to enhance the final classification performance of the misinformation classifier (Figure 1). The model incorporates three types of encoded vector representations from the input propagation graph structure, namely i) news article text encoding, ii) propagation and opinion-related feature vector, as well as iii) an embedding representing the user's behavior by encoding the textual data they have shared, related with a piece of information, along with personal information available in the user's profile. For the encoding of this information, a GNN layer has been used. The concatenation of these three distinct vectors is provided as input in the misinformation detection classification module. These vectors essentially let the model take into account the source news textual data (i), explicit graph structure information (ii), and its involved social media users' interactions, content, and personal information, respectively(iii).

In greater detail, for news article text encoding, the text included in the source news is transformed into embeddings using the word2vec pre-trained model. These embeddings are then fed to a fully-connected Neural Network layer (FC-NN) for dimensionality reduction, followed by a dropout layer that produces a final news article embedding.

For the second type of input information concerning the graph and opinion-related features, a features vector is constructed using information collected from the analysis of the propagation graph using two methods. The first method concerns the calculation of the social media propagation graph-related features, as introduced in [16], including features that capture propagation-related statistics (e.g. number of no-engagement tweets). The second method concerns the extraction of opinion-related features, reflected through the sentiment analysis of the collected information. The opinion-related features were introduced in the network in order to determine whether changes in public opinion on a topic affect the change in news type (i.e. from real tag to false). For sentiment analysis, the VADER sentiment analysis tool [22] has been used.

A detailed list of the propagation and opinion-related features along with a short description is presented in Table I. Features 1 to 12 from Table I are calculated based on the original tweet propagation graph. For the extraction of features 13-16, a new opinion-related knowledge graph is implemented. On this knowledge graph, the node connections rely on the sentiment that users express in their posts included in the initial propagation graph. This type of information can be valuable in the case where a news item's veracity has changed during the social media dispersion. More specifically, this opinion-related knowledge graph includes all reply nodes of the original propagation graph, while its edges are formed by all 2-tuples of nodes that have absolute text sentiment score differences between the respective nodes less than a threshold. This threshold is set to 0.5 for the case of the VADER sentiment scores [22] that is studied in this paper. This feature set is, also, being normalized per feature on a scale of 0 to 1, before it is fed to the model so that these features are on a similar scale to the other profile and textual features. After the extraction of all features, they are used to synthesize the Propagation and opinion-related feature vector.

Finally, regarding the extraction of the user behavior's embedding, the information collected about the users' interaction with each piece of information collected from the social media platform is fed to a graph neural network (GNN) layer, followed by a global max pooling operation, to obtain a final graph embedding representation vector. Each user profile vector of length 10 is initialized with respective user-related attributes extracted from its social media account profile, namely account verification status, geo-spatial positioning status, followers, followees and statuses count, time of account creation and finally word count in the screen name and description. These profile vectors are then normalized accordingly per feature in the scale of [0,1]. The GNN layers employed for the experiments of the current paper are the GraphSAGE [23] and TransformerConv [24], but it can be in general any graph convolutional layer. As a global graph pooling operator, the max pooling operation is selected, as it performed better compared to the corresponding global mean operator in our evaluation.

TABLE I. PROPAGATION AND OPINION-RELATED FEATURE SET (GRAPH STRUCTURE FEATURES VECTOR)

Number	Short description
1	Maximum tweet out-degree
2	# of tweets without engagements
3	# of engagements in a tweet with the maximum out-degree
4	# of normal and leader users in the tweet with the maximum out-degree
5	Time diff. between the first tweet and the last tweet/retweet/reply in the propagation tree
6	Average time difference between two consecutive tweets
7	Average time difference between two consecutive retweets
8	Average time difference between the tweet post time and the first retweet time
9	Time diff. between the first tweet and the tweet with the maximum out-degree
10	Average sentiment scores of the first layer reply/replies
11	Average sentiment scores of replies in the deepest cascade
12	Sentiment ratio
13	The difference of sentiment scores between the two communities of the opinion-related knowledge graph with the most controversy
14	Local node connectivity
15	Total sentiment score of nodes forming the shortest path between the nodes with the maximum and minimum scores
16	The total sentiment of nodes forming the maximum independent set of the opinion-related knowledge graph

After this 3-way extraction process, the three respective vectors produced as described previously, are concatenated to form a final vector which is fed to a series of three fully connected linear layers, before a final softmax layer that produces probability scores for the 2 target classes of the misinformation detection task. The ReLU activation function [25] is used, for the fully connected NN layers.

IV. EVALUATION RESULTS

This section presents the experimental setup used for the evaluation of the proposed method, along with the results of this process.

A. Dataset & evaluation metrics

FakeNewsNet dataset is used for all the experiments, a public dataset of news articles, along with their corresponding propagation networks on Twitter, which have been labeled as either "fake" or "real" by fact-checking organizations. This dataset includes two subsets: Politifact and GossipCop, which cover political domain news articles, and celebrity gossip news, respectively. From both datasets, a balanced subset split has been conducted (Table II), before the implementation of the propagation graph. The performance of each model is evaluated using the accuracy and F1 metrics on the test set. The negative log-likelihood is used as loss during the experiments and the learning rate value is set to 10^{-3} for all versions. For the PolitiFact subset of the dataset, a 110-epoch training loop is conducted, while for GossipCop the train lasts 90 epochs, respectively. For each version that is evaluated, a 5-fold cross-validation is performed [1] and we average the result for every

[1]Folds creation: random permutation of index array $\{0,1,...,total_{news}\}$ with SEED=0 and split into equal sized chuncks-folds

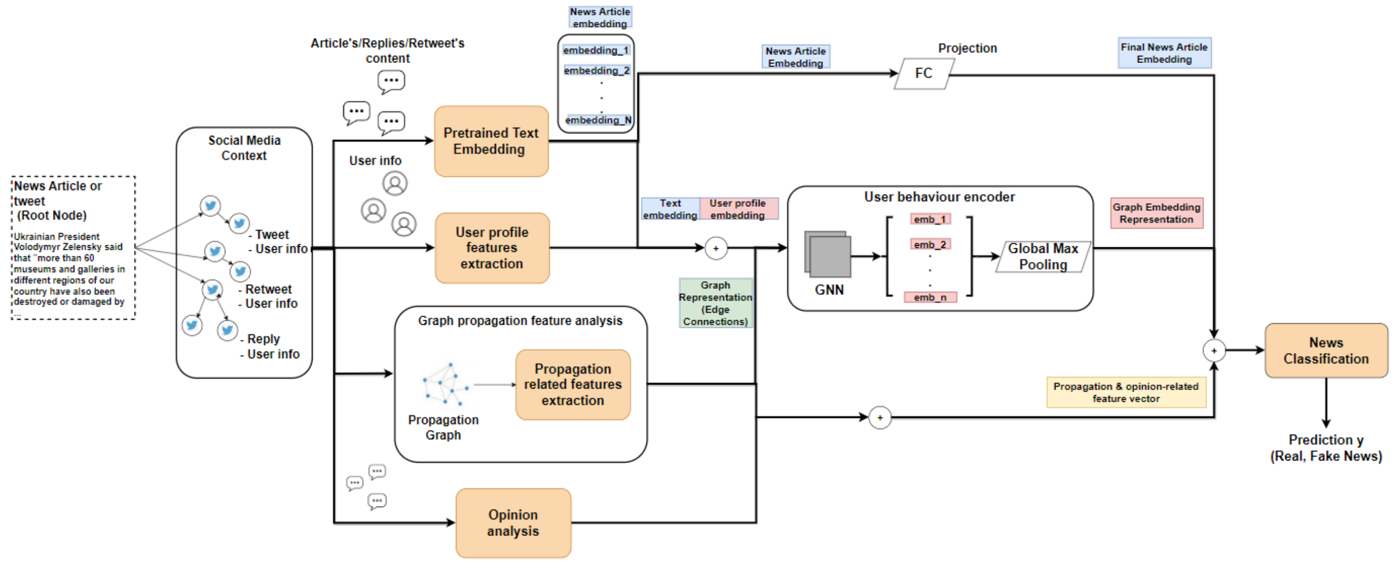

Fig. 1. Propagation-based feature -enhanced GNN (PFGNN) module's architecture

metric.

TABLE II. FakeNewsNet dataset statistics

Data Distribution	PolitiFact	GossipCop
# Total news samples	314	5464
# True labelled news	157	2732
# Fake labelled news	157	2732

B. Results

To evaluate the effectiveness of the additional graph-extracted features in a GNN-based model, the feature-enhanced model is compared with another GNN-based method that employs graph convolutional modules, namely the UPFD [7]. For this model, the architecture and input feature set is the same as the one presented in [7]. More specifically, each node of the input graph structure that is fed to the GNN layer, includes an embedding formed from the concatenation of the user-historical tweet text data with the user profile vector extracted with the same procedure as in section III. The aggregated tweet historical encoded text data, are generated with a pre-trained word2vec model and are provided by the paper authors publicly [2].

From the results presented in Table III and IV, it is observed that versions that utilize the propagation-related features from the propagation tree perform better on the task of misinformation detection, both on the PolitiFact and the GossipCop datasets. In more detail, the PFGNN model performs consistently better than the UPFD model in both datasets. The propagation-based features were used to enhance the proposed network, as they provide extract additional information to the network regarding the information spreading pattern. As stated also in [19], the analysis and use of the temporal aspects of a propagation tree for the misinformation detection task, depicts useful information about how users tend to spread false and real information.

[2]https://github.com/safe-graph/GNN-FakeNews

Regarding the use of two different GNN layers, the initial scope for evaluating the different operators was to identify any differences that might occur in the accuracy results once adding extended features on the news piece propagation graph. According to the results, there is not any noticeable difference between the SAGE and TransformerConv modules in this setup, as each use case achieves similar results on both layers. The same applies to the use of weights on graph edges for the TransformerConv versions. However, based on the results, it can be concluded that the performance of the operators depends highly on the data included in each use case and on the length of the graphs created.

V. Conclusion

In this paper, a graph-based neural network that receives as input a news pieces propagation graph along with extended features was presented. The proposed network receives as input information extracted from the news piece posted on social media platforms along with the user's social engagement attributes and propagation-related features. The aim of this work was to detect the fake information dispersed on social media platforms by considering the propagation path of the information dispersion, along with the interactions that users have with the initial news piece posted.

For the evaluation of this network, two datasets that are available in the literature were used, PolitiFact and GossipCop. The textual information required in order to complete the propagation graphs of both datasets was collected using the Twitter API. For the textual representation in the proposed method, the word2vec text embedding method has been used.

As discussed earlier, the additional features presented, apart from the user social engagement ones, seem to have a better impact on the model's performance and more specifically to increase its accuracy. This paper presents the initial results of this topic's research. As a next step, the authors of this paper have conducted further experiments on using BERT

TABLE III. RESULTS ON **POLITIFACT** FOR MISINFORMATION DETECTION

GNN layers	Model	ACC (%)	F1 score
TransformerConv	UPFD	89.3	88.96
	PFGNN	**91.91**	**91.71**
GraphSAGE	UPFD	90.38	89.95
	PFGNN	**91.97**	**91.79**

TABLE IV. RESULTS ON **GOSSIPCOP** FOR MISINFORMATION DETECTION

GNN layers	Model	ACC (%)	F1 score
TransformerConv	UPFD	97.98	97.99
	PFGNN	**98.11**	**98.12**
GraphSAGE	UPFD	97.98	97.99
	PFGNN	**98.07**	**98.08**

embeddings for both the current and historical posts included in the propagation tree. An additional step that has been planned to be evaluated as a next step is to identify the features that can uncover misinformation dispersion patterns on the early detection task, through the analysis of these information propagation networks, while using the identified features from this analysis as extended features on the proposed model's architecture.

ACKNOWLEDGMENT

This work is supported by the European Unions Horizon2020/Horizon Europe Research and Innovation Program through the ETAPAS project (Grant Agreement No. 101004594). The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the European Commission.

REFERENCES

[1] M. Zimdars and K. McLeod, *Fake news: understanding media and misinformation in the digital age*. MIT Press, 2020.

[2] A. Mishra and S. Samu, "Impact of fake news on social image perceptions and consumers' behavioral intentions," *Journal of Consumer Marketing, (ahead-of-print)*, 2021.

[3] G. Sansonetti, F. Gasparetti, G. D'aniello, and A. Micarelli, "Unreliable users detection in social media: Deep learning techniques for automatic detection," *IEEE Access*, pp. 213 154–213 167, 2020.

[4] C. L. M. Jeronimo, L. B. Marinho, C. E. Campelo, A. Veloso, and A. S. da Costa Melo, "Fake news classification based on subjective language," in *In Proceedings of the 21st International Conference on Information Integration and Web-based Applications Services*, 2019, pp. 15–24.

[5] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques," in *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, I. Traore, I. Woungang, and A. Awad, Eds. Cham: Springer International Publishing, 2017, pp. 127–138.

[6] L. Borges, B. Martins, and P. Calado, "Combining similarity features and deep representation learning for stance detection in the context of checking fake news," *Journal of Data and Information Quality (JDIQ)*, vol. 11, pp. 1–26, 2019.

[7] Y. Dou, K. Shu, C. Xia, P. S. Yu, and L. Sun, "User preference-aware fake news detection," 2021. [Online]. Available: https://arxiv.org/abs/2104.12259

[8] A. Benamira, B. Devillers, E. Lesot, A. K. Ray, M. Saadi, and F. D. Malliaros, "Semi-supervised learning and graph neural networks for fake news detection," in *In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2019, pp. 568–569.

[9] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD explorations newsletter*, pp. 22–36, 2017.

[10] A. R. Merryton and G. Augasta, "A survey on recent advances in machine learning techniques for fake news detection," *Test Eng. Manag*, pp. 11 572–11 582, 2020.

[11] A. M. Braşoveanu and R. Andonie, "Integrating machine learning techniques in semantic fake news detection," *Neural Processing Letters*, pp. 3055–3072, 2021.

[12] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Information Sciences*, vol. 497, pp. 38–55, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025519304372

[13] M. Aldwairi and A. Alwahedi, "Detecting fake news in social media networks," *Procedia Computer Science*, vol. 141, pp. 215–222, 2018, the 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2018) / The 8th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2018) / Affiliated Workshops. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050918318210

[14] Z. Chen, P. Wu, and L. Pan, "Leveraging diversity-aware context attention networks for fake news detection on social platforms," in *2022 International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 01–08.

[15] U. Jeong, K. Ding, L. Cheng, R. Guo, K. Shu, and H. Liu, "Nothing stands alone: Relational fake news detection with hypergraph neural networks," 2022. [Online]. Available: https://arxiv.org/abs/2212.12621

[16] M. Davoudi, M. R. Moosavi, and M. H. Sadreddini, "Dss: A hybrid deep model for fake news detection using propagation tree and stance network," *Expert Systems with Applications*, vol. 198, p. 116635, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417422001245

[17] T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, and J. Huang, "Rumor detection on social media with bi-directional graph convolutional networks," 2020. [Online]. Available: https://arxiv.org/abs/2001.06362

[18] C. Song, Y. Teng, Y. Zhu, S. Wei, and B. Wu, "Dynamic graph neural network for fake news detection," *Neurocomputing*, vol. 505, pp. 362–374, 2022.

[19] K. Shu, D. Mahudeswaran, S. Wang, and H. Liu, "Hierarchical propagation networks for fake news detection: Investigation and exploitation," in *In Proceedings of the international AAAI conference on web and social media*, vol. 14, 2020, pp. 626–637.

[20] W. S. . L. H. Shu, K., "Understanding user profiles on social media for fake news detection," in *In 2018 IEEE conference on multimedia information processing and retrieval (MIPR)*, 2018, pp. 430–435.

[21] K. Makice, *Twitter API: Up and running: Learn how to build applications with the Twitter API*. O'Reilly Media, Inc., 2009.

[22] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 216–225, May 2014. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14550

[23] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," 2018.

[24] Y. Shi, Z. Huang, S. Feng, H. Zhong, W. Wang, and Y. Sun, "Masked label prediction: Unified message passing model for semi-supervised classification," 2021.

[25] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "Fndnet–a deep convolutional neural network for fake news detection," *Cognitive Systems Research*, vol. 61, pp. 32–44, 2020.