MEFaND: A Multimodel Framework for Early Fake News Detection

Asma Sormeily, Sajjad Dadkhah, *Senior Member, IEEE*, Xichen Zhang, and Ali A. Ghorbani, *Senior Member, IEEE*

***Abstract*—Alongside social media platforms' rise in popularity, fake news circulation has increased, highlighting the need for more practical methods to detect this phenomenon. The constantly evolving format of fake news makes it difficult for approaches that rely on a single modality of news to generalize the different types of false news. Furthermore, earlier approaches require extensive propagation data to determine the veracity of news, which can be challenging to collect in the early stages of news dissemination. Thus, we propose a multimodal early fake news detection approach that leverages latent insights into both news content and propagation knowledge. We design a multimodule architecture using graph neural networks (GNNs) to represent edge-enhanced and node-enhanced propagation graphs and bidirectional encoder representations from transformers (BERTs) to generate contextualized representations of news content. Our approach tackles the challenge of early detection in a more realistic scenario, accessing early propagation data in a single social media post and short-length news content. Moreover, we conduct comprehensive studies on user characteristics using statistical techniques to identify attributes with strong discriminative capability for identifying false news. We also analyze temporal and structural properties of fake news propagation graphs to demonstrate distinguishable patterns of false and real news behavior. Our model outperforms several state-of-the-art methods, achieving an impressive F1-score of 99% and 96% on two public datasets. The individual contribution of various components in our model to the final performance is also measured, which can be insightful for future research on multimodal false news detection.**

***Index Terms*—Bidirectional encoder representations from transformers (BERTs), enhanced propagation graphs, fake news behavior, graph neural networks (GNNs), multimodal early fake news detection (MEFaND), social media user characteristics.**

I. Introduction

In recent years, social media platforms have experienced an unparalleled surge in their number of users. It is estimated that the global user base for social media is expected to reach 3.43 billion by the close of 2023 [1]. The increasing trend of social media popularity is proving to be advantageous for both individuals and business owners. In addition to the entertainment value of social media, users can leverage these platforms to reach large audiences, share their organic content, and consume information, including local and global news.

While the growing popularity of social media obtains opportunities for people, it has become a suitable place to intentionally mislead and fool users with fake news for various purposes. Recently, misinformation or false information being disseminated and reported as factual has become increasingly prevalent in the mainstream media. Research suggests that 45% of British audiences are exposed to counterfeit news stories online daily [2].

The pervasiveness of disinformation has evolved into a significant issue that poses a hazard to both individuals and society [3]. Moreover, the proliferation of false information can have negative impacts across multiple domains, including, but not limited to, the economy, democratic systems, journalism, and the exercise of freedom of speech [4], [5]. As one example, the COVID-19 epidemic caused the deaths of nearly 800 people due to misinformation.[1] Additionally, the dissemination of counterfeit information via social media and digital news platforms can profoundly impact the public's perception, as demonstrated by various studies in the field [6]. A particularly striking illustration is the influence of fake news and disinformation on the outcome of the 2016 U.S. presidential election, which has been extensively documented in the literature [7].

Due to the overall adverse effects of exposure to false news, detecting and mitigating misleading information on online news sites and social media are crucial. It is imperative that online fake news can be detected early, before its proliferation into the public consciousness and the subsequent assimilation by online users. This is because once an individual's perception of an issue has been formed, it is often challenging to alter, even if the initial impression was inaccurate [8].

While various fact-checking websites aim to identify instances of fake news,[2,3,4] a significant level of human involvement is essential for this process. Furthermore, the unsubstantial volume of fake news, rumors, and misinformation disseminated daily on social media often render fact-checking websites incapable of detecting and interpreting such content promptly.

Numerous studies have focused on automatically identifying fake news on social media platforms and news outlets by

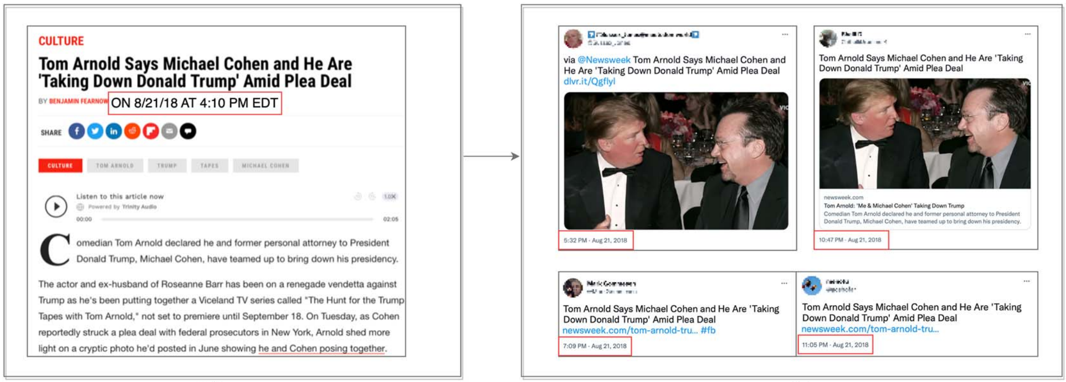
Manuscript received 14 March 2023; revised 1 November 2023 and 4 January 2024; accepted 15 January 2024. Date of publication 9 February 2024; date of current version 2 August 2024. *(Corresponding author: Sajjad Dadkhah.)*

Asma Sormeily, Sajjad Dadkhah and Ali A. Ghorbani are with the Canadian Institute for Cybersecurity, Department of Computer Science, University of New Brunswick, Fredericton, NB E3B 5A3, Canada (e-mail: sdadkhah@unb.ca).

Xichen Zhang is with the Sobey School of Business, Saint Mary's University, Halifax, NS B3H 3C3, Canada.

Digital Object Identifier 10.1109/TCSS.2024.3355300

[1]https://www.bbc.com/news/world-53755067
[2]https://www.snopes.com
[3]https://www.factcheck.org
[4]https://fullfact.org



(a) (b)

Fig. 1. Sample news from the GossipCop dataset. Spread of (a) fake news from a news agency to (b) users on Twitter; red rectangles highlight the timestamp of publishing news and associated tweets.

analyzing various aspects of false information, including news content (e.g., [9], [10], and [11]), author and disseminator traits (e.g., [12] and [13]), and the social setting in which news is spread (e.g., [14], [15], and [16]).

Furthermore, multiple studies (e.g., [17], [18], and [19]) have sought to enhance the fake news identification models' robustness to new inputs by integrating multiple modalities of news (i.e., textual content, social context, and visual content). In addition, discovering disinformation at an early stage is another topic that has received considerable attention in the academic literature recently (e.g., [19], [20], and [21]). Moreover, natural language processing (NLP) developments have allowed for exceptional gains in the efficacy of false news detection algorithms based on textual knowledge of news sources (e.g., [22] and [23]).

Nevertheless, identifying fake news is an intricate and challenging endeavor due to the complicated nature of human behavior involved in its creation, dissemination, and reception [24]. Due to fake news' intricacy, the present investigations into false news have their own set of complications and restrictions that must be addressed.

Specifically, previous research has primarily focused on utilizing a singular modality, such as using textual information in techniques based solely on the content (e.g., [9] and [10]) or using dissemination information in merely propagation-based methods (e.g., [15]). The fast-changing format, language, and subject of fake news make it difficult for models that rely on only textual data to generalize, and models that only rely on propagation patterns may be inadequate for recently distributed news without a broader social context [21].

Additionally, in previous research (e.g., [19], [20], and [25]), information from multiple posts related to a single news item was used to predict its veracity, incorporating aspects such as user responses, characteristics of engaged users, and hierarchical propagation networks derived from several posts written about a news story, as shown in Fig. 1. However, data collection is challenging in real-world situations, particularly during the early stages of news distribution. To address this challenge, our work aims to minimize the amount of data required to distinguish false news by exploiting the information in a single social media post.

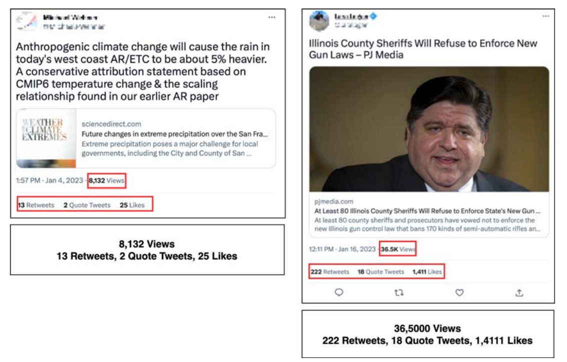


Fig. 2. Two instances of news posts on Twitter with a large number of views but comparatively few retweets and replies.

As another challenge for early identification models, they have yet to consider the accelerated rate at which fake news spreads on social communication websites compared to genuine news. Previous studies, including those by Lazer et al. [24] and Vosoughi et al. [14], highlight the requirement to consider temporal distinctions in fake news propagation patterns. Our experiments in Section IV-B demonstrate that fake news spreads significantly quicker, more profoundly, and more expansively than trustworthy news. Furthermore, the number of people affected by news is higher than those who reacted and contributed to its propagation path [14], [24]. Fig. 2 shows news stories[5,6] that attracted many readers but prompted relatively small responses or retweets. Detecting fake news in a timely manner is crucial to prevent its harmful and pervasive effects, which can spread rapidly within a few hours of its publication. Hence, early identification models must be capable of effectively detecting fake news within this narrow time window.

In this research, we delineate the field of "fake news detection" as a domain focused on the identification of false or misleading information after its dissemination on social media, often employing various techniques from the existing literature.

[5]https://www.factcheck.org/2023/01/democrats-misleadingly-suggest-widespread-gop-support-for-fairtax -bill-unlikely-to-become-law/

[6]https://www.factcheck.org/2023/01/its-too-soon-to-attribute-the-california-storms-to-climate-change-experts-say/

However, conventional approaches within the realm of fake news detection may not fully consider the implications of news articles gaining increased exposure and readership over time. These approaches may lack an appreciation for external temporal factors, such as political influence, evolving social media dynamics, and other contextual elements, which can impact the detection of fake news, particularly in relation to the temporal aspects inherent in news propagation.

In contrast, the concept of "early detection of false news" places a heightened emphasis on strategies aimed at achieving more accurate detection outcomes within a compressed timeframe. The urgency in promptly identifying false news stems from the realization that as more individuals are exposed to misleading information, there is an elevated risk of such narratives gaining traction and potentially influencing the course of events for various actors' benefit.

This study seeks to address some of the limitations inherent in prior research by introducing a framework for early false news detection that takes into account both textual news content, such as headlines and tweet text which is posted by users, and the early propagation patterns observed within the initial hours of a news article's dissemination.

To accomplish this objective, we employ a transformer-based language model to extract contextually rich and sensitive representations of news content. Additionally, we construct propagation graphs that incorporate temporal information and friendship connections between authors and spreaders, i.e., users who reshare the source tweet. These constructed graphs serve as the foundation for training a multimodule graph neural network (GNN), which is proficient in effectively modeling the early propagation dynamics of both false and authentic news.

Additionally, this study attempts to address the following research questions using various experiments on our proposed model.

1) *RQ1*: What are the temporal and structural differences between accurate and inaccurate news propagation patterns on social media? Specifically, how faster false news travels in comparison with legitimate news?
2) *RQ2*: How effective are selected node-level user features in distinguishing genuine and false news posts?
3) *RQ3*: To what extent does combining content and transmission patterns enhance the speed and accuracy of early fake news identification?
4) *RQ4*: What is the individual contribution of each element of the proposed framework to the overall performance?

Furthermore, the main contribution of this research can be summarized as follows.

1) We proposed a multimodal framework for early false news detection that: a) represents news content using a powerful encoder-based language model; b) leverages limited propagation knowledge by employing various GNN architectures to generate informative representations of early propagation networks; and c) enhances the ability to identify false news in a more practical setting, yielding a noteworthy F1-score of 99% and 96% on two distinct real-world datasets.
2) Conducted a thorough analysis of user characteristics in social media platforms to identify informative features with strong discriminative capability for detecting false news. We validated the effectiveness of the identified feature sets through a comprehensive ablation study and various statistical techniques, such as Pearson's correlation, Spearman's correlation, Chi-square, and Student's t-test.
3) Analyzed temporal and structural properties of fake news propagation networks to reveal distinguishable patterns between false and authentic news propagation, discovering that fake news spreads significantly faster, deeper, and more extensively than trustworthy news.
4) Incorporated relational, sentimental, and temporal information of news spreaders and authors to enhance the construction of news propagation graphs, resulting in a more accurate representation of the propagation dynamics and enabling the identification of influential spreaders and potential sources of false news.

In the following sections, we present a systematic literature review of fake news detection works in Section II. In Section III, we provide a detailed description of our proposed model architecture and its components. Section IV presents the results of our experiments and evaluations of the model. In Sections V and VI, we offer some final thoughts and suggestions for potential avenues for future research, as well as a conclusion of our research work.

II. Literature Review

In this section, we divide the field of false news detection research into three broad categories: content-based methods, propagation-based methods, and hybrid methods. Finally, we present a brief overview of ongoing efforts to detect fake news in its early stages.

A. Content-Based Approaches

In content-based fake news detection methods, various informative knowledge from news content, including latent or hand-crafted text features, is extracted to distinguish between fraudulent and honest news. This section briefly reviews knowledge-, style-, and transformer-based content models.

1) Knowledge Based: Knowledge graphs (KGs) are graph-based data structures that express organized relationships between entities. In the context of false news identification, KGs have been employed to represent the associations between entities in news content. The representations are then compared with a pregenerated KG to determine the credibility of the news. Authors [26] use edge prediction algorithms to predict missing links in the KG and compare the predicted links with those in the news articles.

Despite their effectiveness in research, knowledge-based approaches face obstacles in practice, such as dependence on data quality and availability and the inability to detect nuanced types of false news.

2) Style Based: Style-based false news identification models attempt to capture distinctive characteristics of authors' writing styles to detect fake news effectively [27]. Rubin and Lukoianova [9] employ rhetorical structure theory (RST) and vector space model to identify distinctive characteristics between deceptive and legitimate news.

In a transdisciplinary research work focused on the style of news stories, Zhou et al. [10] represent news articles throughout various language levels based on recognized social and forensic theoretical studies. A supervised learning method is applied to categorize news articles based on the computational features of the content. Qian et al. [28] propose a neural network architecture (TCNN-URG) that uses word-level and sentence-level representations to extract semantic information from news articles. Additionally, TCNN-URG performance is enhanced by a generative model that learns the nature of user replies and creates responses to new articles enabling early identification of fake news.

In a recent study, Zhu et al. [29] make significant strides in the field of style-based fake news detection by tackling domain-related challenges, particularly in the realm of multidomain fake news detection within the complex landscape of interconnected news domains. Their work underscores the imperative to transcend conventional single-domain approaches.

Implementing style-based approaches requires human labor and expertise in feature engineering. Additionally, the effectiveness of these models is contingent upon the dataset used for training.

3) Transformer Based: Transformers are a new architecture in NLP designed to handle tasks sequentially while coping with long-distance interdependence. Recent studies have leveraged these models to address various NLP problems, such as content-based fake news classification, through pretrained transformer-based language models, such as bidirectional encoder representations from transformer (BERT) [30] and GPT-2 [31]. Liu et al. [32] propose a deep neural network model based on the BERT language model for spotting relatively short hoaxes. Another group of researchers, Jwa et al. [33] propose a framework for identifying fake news based on the BERT language model. This method identifies falsehoods by analyzing headlines and the articles they reference. Models that rely solely on textual data are insufficient in detecting fake as malicious actors constantly adapt their writing style and language to evade detection, rendering feature-based methods that rely on language-specific cues ineffective. Therefore, developing multimodal approaches that integrate multiple modalities is necessary to overcome these limitations and accurately detect fake news.

B. Propagation-Based Approaches

Empirical studies have established that the contextual dynamics surrounding news dissemination offer valuable insights for discerning between false and authentic news sources. This is attributed to the notable disparity in the propagation patterns observed between fake and genuine news. Furthermore, it is noteworthy that the contemporary arsenal of AI-powered techniques designed to manipulate visual or textual content of news articles faces significant challenges when attempting to influence these contextual dynamics.

In line with this perspective, Shu et al.'s research [34] delves into the intricate relationships between publisher bias, news stance, and user engagement within the context of social media. To address these complex correlations, they propose the innovative trirelationship fake news detection framework (TriFN).

In [13], a disinformation detection framework (FNED) is developed using multiple deep learning mechanism that incorporates user comments on posts and the profile features extracted from news spreaders and authors.

In another examination of false news detection relying on social context, Bian et al. [15] proposed a novel approach utilizing graph convolutional networks (GCNs) [35]. The proposed method, referred to as bidirectional graph convolutional networks (Bi-GCNs), explores the intricacies of both the propagation and dispersion of rumor cascades through top–down and bottom–up analysis. In another propagation-based study, Nguyen et al. [16] capture the high-fidelity embeddings of social context through a graph learning mechanism to detect fake news with promising results.

Additionally, Shu et al. [36] examine both implicit and explicit user attributes by looking into tweets from users' previous timelines and profile information. By integrating the characteristics of news spreaders, Shu et al. demonstrate the effectiveness of the obtained user features in determining whether a piece of news is authentic.

To develop a pattern-based false news detection approach, Zhou and Zafarani [12] characterize and quantify several patterns of fake news, including trends of news propagation, spreaders, and connections among the spreaders at several network levels. These patterns' existence is validated and interpreted by referring to empirical investigations and social psychological theories.

Even though the social context of news offers essential knowledge to separate truth from falsity, linguistic, and stylistic cues are not represented in the news dissemination statistics despite being reflected in the content. In addition, external factors, such as social media algorithms, may affect the social context of news and lead to inaccuracies in fake news detection.

C. Hybrid Approaches

In hybrid approaches, more than one news modality is analyzed to differentiate between truth and deception. By incorporating more modalities, the model can gain a more comprehensive understanding of the news item, thereby increasing its accuracy and generalization performance in spotting false news.

The false news identification model outlined in [17] incorporates two key components: an unsupervised approach to learning domain embeddings and a supervised approach to classifying news in a domain-agnostic manner. These components work together to detect false information by leveraging domain-specific knowledge as well as cross-domain knowledge within the content and context of news articles. Palani et al. [37] incorporate text-based and visual components of news to form a multimodal feature vector using a BERT model for textual features and a CapsNet model for visual features. When combined, these factors yield a complete data representation for evaluating the reliability of the news.

Another research work proposed by Wei et al. [38] leverages discriminant feature representations from both news textual and visual content, while taking reconstruction into account, which allows for the capture of modality- and event-invariant information.

Graph-aware coattention networks (GCANs) [18] is another research work that falls into hybrid approaches, which exploits both tweet content and the sequence of retweeters to classify misinformation. GCAN represents propagation paths using CNN, GRU, and GCN neural networks and a GRU-based encoder to incorporate source tweet information at the word level.

In [19], the authors suggest a hierarchical coattention network (dEFEND) collecting explainable top-k critical phrases and user replies for false news detection, capitalizing on news content and user comments. dEFEND incorporates four components: a news content encoder that can learn representations of textual news information at the word and sentence levels. Those components incorporate an encoder for user responses, coattention between sentences and comments, and a false news prediction element to classify news records as false or legitimate news.

Ensuring sufficient and valid data is a significant obstacle to developing efficient false news detection models, and this difficulty increases as more data are required. Although multimodal approaches assure a comprehensive representation of news, most existing works are limited by their requirement for large amounts of data from various posts rather than one post written and shared by limited number of users.

D. Early Detection of Fake News

Researchers in recent fake news detection works endeavor to solve the early identification challenge, restrain false news' reach, and prevent its further spread. The goal is to detect false news early by considering the limited propagation information within a specific time delay of the original news being published.

Propagation2vec, a deep neural network model proposed by Silva et al. [25], reconstructs full propagation knowledge from limited data collected during the early phases of the spread of disinformation on social media. For identifying false news, Propagation2vec employs a hierarchy-based attention method to place more weight on nodes and cascades that contain useful information.

In [20], Davoudi et al. present a hybrid approach to early false news identification, incorporating both propagation tree and instance network analysis. Their model encompasses three aspects: dynamic analysis utilizing a recurrent neural network to encode evolution patterns, static analysis utilizing a fully connected (FC) network to represent comprehensive properties, and structural analysis through the application of the node2vec [39] algorithm as a network embedding technique.

In [19], the authors study various hierarchical propagation network characteristics from structural, temporal, and linguistic aspects and demonstrate the proposed features' practicality in identifying fake news. The proposed hierarchical propagation network feature (HPFN) performs comparably well in the early detection deadlines (12 to 96 h) of news publication time.

Another research work to detect false news stories that exploits content and social context associated with news is proposed by Raza and Ding [21]. The suggested framework is based on a transformer architecture with two modules, an encoder to embed the knowledge extractable from news text and social context and a decoder to predict future behavior based on past observations for early detection purposes. As a method of addressing the label shortage, the suggested model employs a weak supervision labeling method.

Existing research works focusing on the early identification of false news are also subject to the restrictions mentioned in Section II-C, depending on the modality and input data requirement. Moreover, more than the conventional time thresholds used in the literature to detect fake news may be needed to counteract the rapid spread of false information.

III. METHODOLOGY

A. Problem Formulation

We characterize the problem of early false news identification as follows: let S be a set of labeled news stories $S = \{s_1, s_2, s_3, ..., s_n\}$. Each news record $s_i \in S$ consists of a tuple $\langle C_i, U_i, y_i\rangle$ where C_i is the short-length news content (i.e., source tweet and short news title) consisting of k words: $C_i = \{w_1^i, w_2^i, w_3^i, ..., w_k^i\}$. U_i represents s_i's propagation path, which is the set of users who post/repost the news story on social media (i.e., engaged users). Each U_i consists of l users: $U_i = \{(u_1^i, X_1^i, t_1^i), (u_2^i, X_2^i, t_2^i), ..., (u_l^i, X_l^i, t_l^i)\}$ in which each $u_l^i \in U_i$ is associated with a set of d dimensional feature vectors: $X_l^i \in \mathbb{R}^{d\times l}$ and t_l^i is the timestamp of the user u_l^i creating/sharing the news story. And finally, $y_i \in \{0, 1\}$ is the news label ($y = 1$ true news and $y = 0$ false news). A directed propagation graph $G_i = (N_i, E_i, X_i, L_i)$ is constructed for each news story using the set of users U_i and corresponding features for each user X_l^i. N_i is the set of vertices/nodes corresponding to user U_i, E_i is the set of edges/links representing how a news story is transmitted from one user to another, X_i is the node-level features extracted from users' profile information, and L_i is the edge-level features derived from temporal, relational, and sentimental user information. Our goal is to "classify each unlabeled news records $s_i \in S$ as false or authentic within a time window Δ which refers to the detection deadline defined by researchers (i.e., 5 h) where the short textual content C_i and a partial sequence of engaged users U_i are available for."

B. Multimodal Early Fake News Detection (MEFaND)

The overview of our proposed MEFaND is presented in Fig. 3. For each news, its early propagation graph and corresponding textual content are considered as the input of the MEFaND framework. Subsequently, MEFaND employs two distinct modules, a propagation information encoder and a transformer-based textual encoder, to produce rich representations of propagation and textual knowledge. These representations are then concatenated (F: Concatenate), merging the

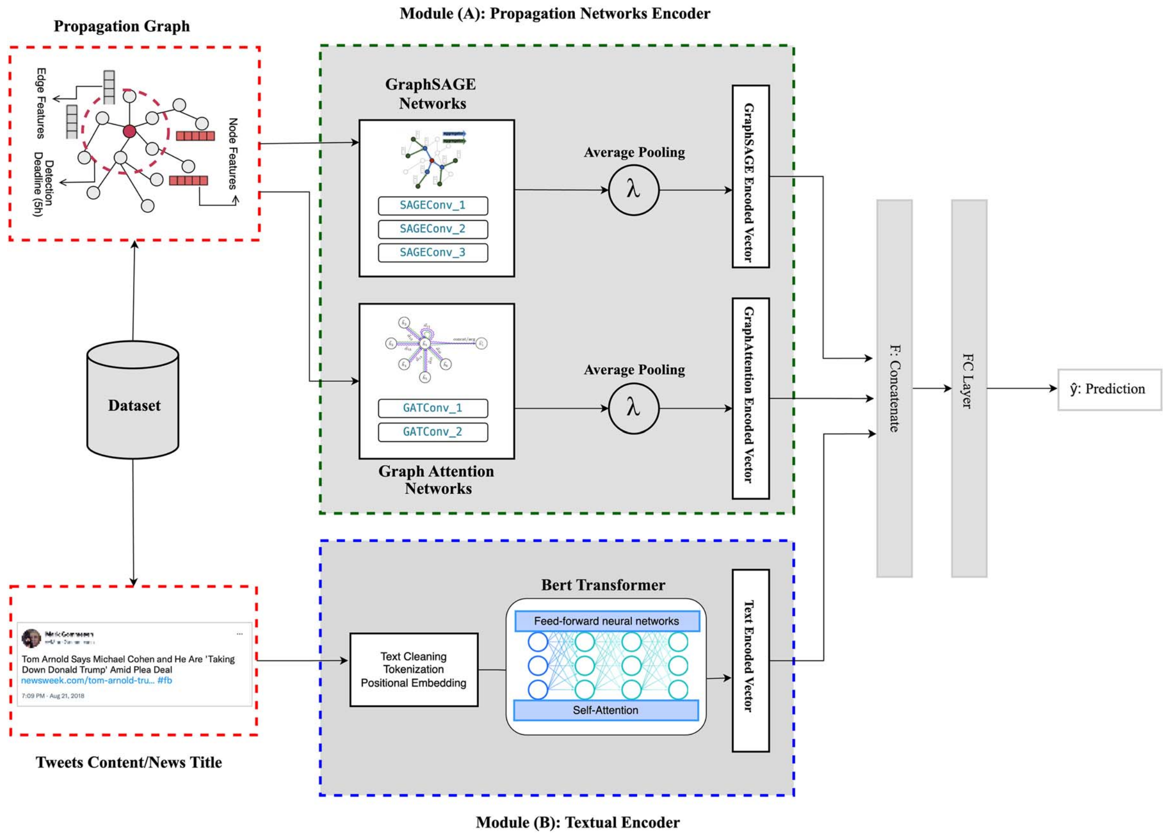


Fig. 3. Overview of our proposed MEFaND model.

distinct modalities into a unified feature vector, which is crucial for capturing the full spectrum of characteristics associated with fake news. The concatenated vector is further processed through a FC layer, which outputs the final prediction, determining the veracity of the news story. Details on the propagation information encoder and textual encoder modules, as well as the concatenation and FC layers, are presented in Sections III-D and III-E, respectively.

C. Data Preprocessing

1) Feature Selection: This section describes our suggested approach for feature selection, which aims to appropriately emphasize the substantial effect of author and spreader traits on false news identification. This knowledge is extracted and analyzed to enrich our model performance to distinguish fraudulent news.

The full distribution of the dataset is prohibited due to constraints imposed by Twitter's privacy regulations and news publishers' copyright protections. User engagement data and user-specific information are also unavailable owing to Twitter's policy restrictions.

Nevertheless, we effectively harnessed the Twitter API to access and classify social context-related features into three primary categories. Since our study places a strong emphasis on integrating social context as a dependable modality for fake news detection, it is conceivable to implement our approach in practical scenarios, provided that no limitations related to Twitter's privacy policies and news publishers' copyright restrictions exist.

TABLE I
LIST OF SELECTED TEMPORAL NODE-LEVEL FEATURES

Attribute	Type	Description
Account Age	Float	The interval of time between the user account creation and the tweet/retweet action.
Retweet Distance	Float	The time difference between the creation time of the source tweet and the time of retweet action by news spreaders.
Retweet Rank	Float	Considering a tweet spread in a sequence of timestamps, we can assign a rank to each spreader based on their index in the sequence.

The next section discusses the specifics of various types of features.

a) Node-level features: We extract insightful information from user profiles to construct a set of node-level features, enhancing the informativeness of news propagation graphs. The selected features fall under three categories: text-based, implicit profile, and textual, which are briefly described in the following sections.

Temporal features: To capture the knowledge of temporal differences in real and false news authors and spreaders, we extract informative temporal features as described in Table I.

TABLE II
LIST OF SELECTED PROFILE IMPLICIT NODE-LEVEL FEATURES

Attribute	Type	Description
Verified	Binary	Whether the user has been authenticated and verified by Twitter.
Followers Count	Integer	The number of people following the user.
Friends Count	Integer	The number of people who the user follows.
GEO	Binary	Whether the user inserted geographic data.
Favorite Count	Integer	The number of statuses that the user liked since created their account.
Statuses Count	Integer	The number of statuses (including retweets) published by the user since created their account.
Listed Count	Integer	The number of times the user was added to a public list by other users.
Profile Banner	Binary	Whether the user exchanged the default profile banner with an uploaded photo.
Profile Image	Binary	Whether the user exchanged the default profile image with an uploaded photo.

Profile implicit features: Profile implicit characteristics are derived from the user's behavior and other indirect information on the platform. The selected profile implicit features and a brief explanation of each are listed in Table II.

Textual features: Text-based user features refer to any text content generated by users. These features are insightful to distinguish users between real and fake news engagement based on their differences in linguistic and writing styles. These features are extracted from user timeline tweets and text-based components of their profiles. The data are collected using the Twitter API, covering a period of 6 months to a year, depending on the availability of tweets before the user's engagement in either genuine or fake news. Additionally, other text-based information of user profiles, such as the bio, location, and website, is extracted as text-based features. Table III shows a list of selected text-based user features and their corresponding descriptions.

b) Edge-level features: In order to provide a more precise illustration of the propagation graphs, we develop edge-level features using relational, temporal, and sentimental information in the adjacent nodes. Table IV shows the specific edge-level features we incorporate in our analysis and the corresponding descriptions.

c) Feature scaling method: After computing node-level and edge-level attributes, we need to standardize the range of all independent variables in the dataset. We use standard scaler to scale the feature sets by normalizing each feature and scaling its variance to reduce the impact of noise in propagation data on our analysis, i.e.,

$$\hat{X}_i = \frac{X_i - \overline{X}}{\sigma}. \quad (1)$$

For each $\hat{X}_i$ from a sample, its normalized version can be determined from (1), where the sample standard deviation is σ and sample mean is $\overline{X}$.

TABLE III
LIST OF SELECTED TEXTUAL NODE-LEVEL FEATURES

Attribute	Type	Description
Description Length	Integer	The length of the user-generated string describing their account (description).
Total Name Length	Integer	The length of the name added to the length of the screen name with which users identify themselves.
Profile Emojis Count	Integer	The number of emojis used in their description, website, and location text input divided by the total number of words.
Profile Entities Count	Integer	The total number of hashtags, user mentions, and URL links in the user description.
Profile Sentiment Score	Float	The overall sentiment score of description, website, and location text inputs computed using VADER.[7]
Statuses Entities Count	Float	The sum of user mentions, URL links, and hashtags divided by total tokens in timeline tweets.
Statuses Emojis Count	Float	The number of emojis in timeline tweets divided by the number of words in timeline tweets.
Statuses Influence Score	Float	The total number of times the timeline tweets have been liked, retweeted, replied to, and quoted by other users is divided by the number of timeline tweets for the user.
Statuses Sentiment Score	Float	The compound sentiment score estimated by VADER of timeline tweets.
Statuses Positive Words Count	Float	The proportion of positive words to all words in timeline tweets.
Statuses Negative Words Count	Float	The proportion of negative words to all words in timeline tweets.

TABLE IV
LIST OF SELECTED EDGE-LEVEL FEATURES

Attribute	Type	Description
Follower Relation	Binary	Whether there is a following or follower relationship between the target node and the source node (author).
Retweet Distance	Float	The interval of each adjacent node retweet/tweet time.
Overall Sentiment Score	Float	The average sentiment score of the profiles of the two adjacent nodes.

d) Feature smoothing method: Due to the presence of highly noisy and extreme values in the propagation data, we utilize the Savitzky–Golay filter [40] to smooth the normalized feature sets. The Savitzky–Golay filter is a widely used digital filter that can decrease noise in data while maintaining the overall structure of features. This filter operates by fitting polynomial functions to the data with the smallest sum of squares, according to the following equation:

$$Y_i = \sum_{j=1}^{M} W_j X_{i-j} \quad (2)$$

[7]https://github.com/cjhutto/vaderSentiment

Algorithm 1 Propagation Graph Construction Algorithm

```
Input: Propagation path U
Output: Propagation graph G
1: function GRAPHCONSTRUCTION
2:     G.addNode(node = author, feat = X_author, t= t_author)
3:     for user in U do
4:         if user follows author then
5:             G.addNode(node = user, feat = X_user, t= t_user)
6:             G.addEdge(src=author, trg=user, feat = L_author-user)
7:             firstDepthNodes.add(user)
8:     remainedNodes = U - firstDepthNodes
9:     for user in firstDepthNodes do
10:        target = loadNearestNode(remainedNodes, user)
11:        G.addNode(node = target, feat = X_target, t= t_target)
12:        G.addEdge(src=user, trg=target, feat=L_user-target)
13:        secondDepthNodes.add(user)
14:        remainedNodes.remove(target)
15:     for user in remainedNodes do
16:        G.addNode(node=user, feat=X_user, t=t_user)
17:        G.addEdge(src=author, trg=user, feat=L_author-user)
18:        secondDepthNodes.add(user)
```

where Y_i denotes the smoothed feature, X_{i-j} denotes the original noisy feature at index $i-j$, W_j denotes the filter coefficients, and M is the number of coefficients used in the filter. The coefficients and window size can be adjusted to balance the tradeoff between smoothing and retaining the inherent structure in the data. To optimize this balance for our analysis, we selected a window size of 5 and a polynomial order of 2 based on previous experimentation and empirical observations of the data.

2) Enhanced Propagation Graphs Construction: In microblogging platforms, information dissemination occurs through the network of relationships between users, where information flows from followers to followees. To effectively capture this flow of information, we develop a method for constructing propagation graphs that takes into account the friendship connections between authors and spreaders, as well as the temporal aspect of information diffusion, by incorporating the timestamps of retweets (described in Algorithm 1). The proposed algorithm creates a propagation graph for a news sample using its propagation path. The graph is initialized with the author of the news as the source node (line 2 in Algorithm 1). The set of associated node-level features (e.g., X_{author}) and the timestamp of node creation (e.g., t_{author}) are assigned to each new node added to the graph. To identify the author's immediate neighbors, the algorithm iterates over the propagation path and creates edges from the author to users who have a friendship relationship with the author (lines 3–7 in Algorithm 1). The set of selected edge-level features is also assigned to each constructed edge between two adjacent nodes (e.g., the edge-level features connecting $author$ to $user$ is denoted as $L_{\text{author-user}}$). The algorithm then moves on to the first-degree users, creating edges between first-degree users and the other users depending on their respective timestamps (lines 9–14 in Algorithm 1). Finally, the remaining users are connected to the author to complete the propagation graph.

D. Propagation Information Encoder

Propagation information of news has informative properties to facilitate false news identification, as confirmed by our experiments on propagation networks in Section IV-B and prior studies [14], [41], [42]. Exploiting the distinguishable knowledge latent in the social context of authentic and fraudulent news, we design a propagation encoder as shown in Fig. 4.

The propagation representation module leverages the limited amount of knowledge available during the early phases of fake news dissemination to create an encoded vectorial representation of the propagation graph structures.

Accordingly, the input data for the propagation encoder module are the edge- and node-enhanced propagation graphs discussed in Section III-C.

To achieve high precision in representing propagation graphs during the early stages of news dissemination, we feed the model *early* propagation graphs, which contain data accessible within a five-hour time frame (e.g., Fig. 5). This time frame aligns with the conventional detection deadline reported in [25].

1) GraphSAGE Representation: The propagation encoder module processes early propagation graphs by first sending them through a three-layer GraphSAGE (SAmple and aggreGatE) layer. The choice of three layers for the GraphSAGE representation is based on its utility in handling dynamic graphs and its suitability for the early identification of false news in evolving news propagation scenarios. GraphSAGE, introduced in [43], is a GNN architecture that generates node embeddings by aggregating the embeddings of their neighboring nodes and is highly suitable for dynamic graphs. GraphSAGE eliminates the need for retraining the model when new nodes are added to the graph, rendering it a suitable approach for the early identification of false news in the ever-evolving landscape of news dissemination on social media platforms. Fig. 4 demonstrates how the GraphSAGE architecture is realized by three distinct GraphSAGE layers, each of which performs an operation on the output of the preceding layer to produce a new node embedding. The representation is then pooled over the entire graph to form a global representation average pooling. Dropout [44] layers are applied between the GraphSAGE layers and the final representation to reduce overfitting. The operation of each GraphSAGE layer can be described as follows:

$$h_u^{(l+1)} = \text{aggregate}\left(\text{ReLU}\left(\text{MLP}\left(h_v^{(l)}\right)\right) v \in \mathcal{N}u\right) \quad (3)$$

where $h_u^{(l)}$ is the node embedding of node u at layer l, $\mathcal{N}_u$ is the list of adjacent nodes for node u, MLP$(\cdot)$ is a multilayer perceptron, and aggregate$(\cdot)$ is a nodewise aggregation function.

To improve the model's ability to learn complex relationships in data and reduce the risk of the vanishing gradient problem, we add an elementwise application of the rectified linear unit (ReLu) activation function on the output of each SAGEConv layer as expressed by the formula $f(x) = \max(0, x)$.

Finally, a graph pooling operation takes the average of all node representations in the graph and outputs a graph-level representation $\bar{z}$. The equation for this operation is defined as: $\bar{z} = 1/N \sum_{i=1}^{N} z_i$ where z is the node representation.

2) Graph Attention Representation: Second representations of early propagation graphs are learned through a two-layer graph attention network (GAT) [45].

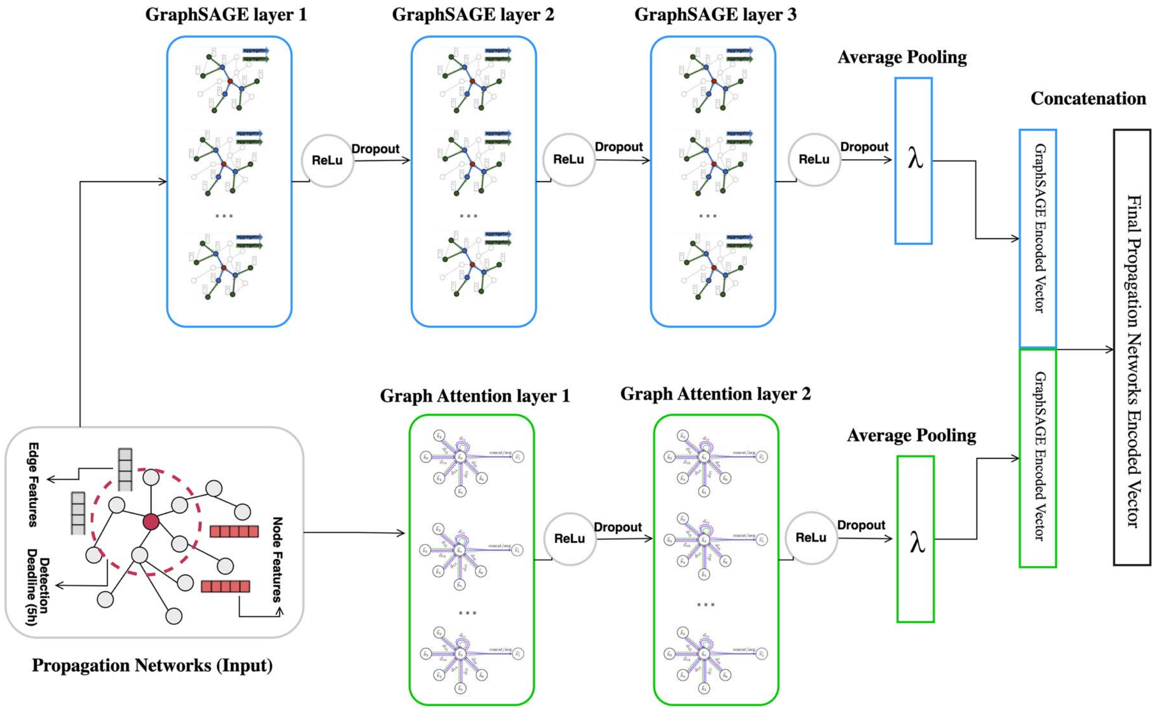


Fig. 4. Overview of propagation information encoder.

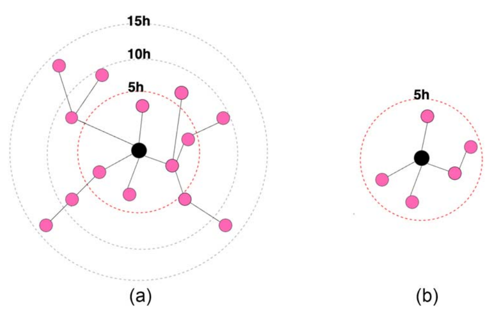


Fig. 5. (a) Complete and (b) early propagation graphs. Black nodes: the author of the tweet; pink nodes: the spreaders of the tweet; solid edges: information flow from one node to another; and dashed concentric circles: time stamps.

The core idea underlying GAT is to learn a weighted aggregation of the node neighbors in a graph, where the weighting mechanism is learned end-to-end with the network. By allowing the network to zero in on the most pertinent information and dampening the impact of noisy information in news propagation, the attention mechanism boosts the efficacy of early false news detection. The choice of two layers has been made to strike a balance between model complexity and performance since additional layers could escalate the computational demands related to edge-level feature extraction and attention weight calculations, without a guaranteed improvement in results.

The GAT representation of node u in the $(l+1)$th layer is expressed as follows:

$$h_u^{(l+1)} = \sum_{v \in \mathcal{N}(u)} \alpha_{uv} \cdot \mathrm{MLP}\left(h_v^{(l)}\right) \tag{4}$$

where the attention weight α_{uv} can be computed by

$$\alpha_{uv} = \frac{\exp(\mathrm{LeakyReLU}(\mathrm{MLP}(h_u^{(l)}, h_v^{(l)}, e_{uv})))}{\sum_{k \in \mathcal{N}(u)} \exp(\mathrm{LeakyReLU}(\mathrm{MLP}(h_u^{(l)}, h_k^{(l)}, e_{uk})))}. \tag{5}$$

Here, $h_u^{(l)}$ and $h_v^{(l)}$ are the hidden representations of nodes u and v at the lth layer, $\mathcal{N}(u)$ is the collection of neighbors of node u, and e_{uv} is the edge attribute between nodes u and v.

The ReLu and Dropout layers follow each GAT layer, and an average pooling mechanism is applied to the final representation that creates graph-level outputs, described in Section III-D1.

E. Textual Information Encoder

We utilize the BERT language model as the encoding mechanism for news textual information. BERT is a pretrained transformer-based architecture that has delivered state-of-the-art outcomes in various downstream NLP tasks, including document classification [30]. After being exposed to a vast amount of material during training, the model is able to acquire nuanced linguistic representations. Our goal in using BERT is to enhance the effectiveness of news content classification by accurately encoding the semantic meaning of the content.

As illustrated in Fig. 6, the news content first undergoes some preprocessing steps. To get a cleaner version of the text, we remove special characters, digits, and other unnecessary information from the news text. The BERT model was furthermore used for tokenization and positional embedding of the news content. Tokenization involves splitting the input text into subwords or tokens, allowing the model to handle out-of-vocabulary words and handle morphological variations.

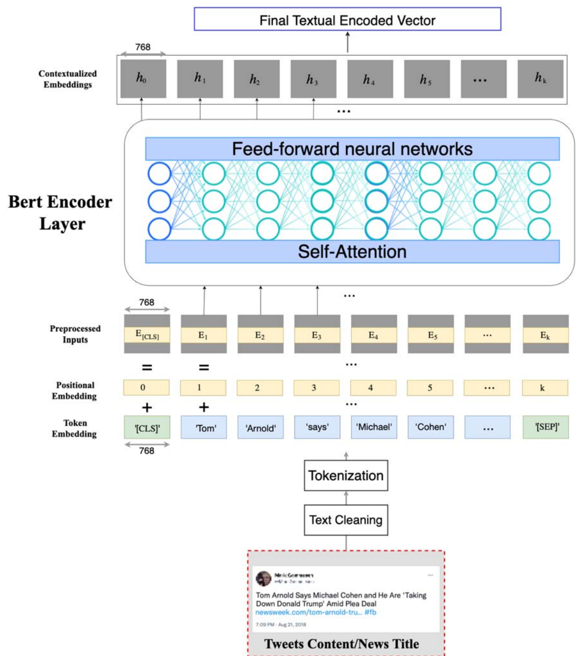


Fig. 6. Overview of textual encoder module.

The positional embeddings provide information about the position of each token within the input sequence, which in turn helps the model to differentiate between tokens' relative importance. The combination of tokenization and positional embedding allowed for a more accurate encoding of the semantic meaning of the news content, leading to improved performance.

The preprocessed input sequences are analyzed using a transformer-based architecture, specifically a multihead attention mechanism. In the multihead attention mechanism, the attention scores for each head h are computed through the utilization of a query matrix Q, a key matrix K, and a value matrix V, which can be described as

$$\text{Attention}(Q, K, V) = \left[\text{head}_1 || \cdots || \text{head}_h\right] W^O \tag{6}$$

where W^O matrix serves to capture the underlying relationship between the concatenated heads, and subsequently, the final attention scores are obtained for each head as

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{7}$$

and W_i^Q, W_i^K, and W_i^V are the weight matrices for the query, key, and value representations, respectively. The attention scores are used to weigh the values, providing the model with a compact representation of the most relevant information for each query. The model's multihead attention mechanism allows it to focus on several pieces of the input sequence simultaneously, boosting its performance and enabling it to better grasp the interdependencies between the various input components.

The final textual encoded vector will be fed to the next module in the model, the News Classifier.

F. News Classifier

The News Classifier component of the model integrates the output representations from the propagation encoder and textual encoder. The integrated representation is then processed through a linear layer, represented as

$$y = \text{Linear}(x_{\text{gnn}} \oplus x_{\text{txt}}) \tag{8}$$

where x_{gnn} represents the output of the propagation encoder and x_{txt} represents the output of the textual encoder, and $\oplus$ symbolizes the concatenation of these two representations. The final output of the linear layer y is then utilized to predict whether the news record is authentic or fraudulent.

IV. Experiments and Evaluation

This section discusses the evaluation results of our experiments in terms of experimental settings, propagation graphs, user characteristics, and model performance, aiming to address the research questions mentioned in Section I.

A. Datasets

1) Datasets Characteristics: The dataset used in our study is extracted from FakeNewNet[8] dataset which is a repository with news content, social context, and spatiotemporal information. We conduct in-depth experiments on these two publicly available and real-world datasets, PolitFact[9] and GossipCop[10] outlined in [3], [34], and [46].

The selection of these datasets for our study was driven by the following considerations.

1) *Complete components for each news record*: The datasets both consist of the same components, including the news title, the URL address of the news, the tweet IDs that were written revolving around the news, and the news classes (label) verified with fact-checking websites PolitiFact and GossipCop. However, it is noteworthy that the datasets also contain image data. Despite the presence of image information within the datasets, we emphasize that our research primarily focuses on the analysis of textual content and propagation aspects of news articles. As can be observed from a sample of the GossipCop dataset in Fig. 1, once a news record is published on a news outlet, within a period, users on Twitter start to propagate the news along with additional statements on their Twitter pages. We use the dataset information and adopt Twitter API[11] and retrieve retweet lists, retweeter information, and time creation of retweets for all tweets.
2) *Alignment with fake news definition*: These datasets also comply with the definition of fake news as was expressed in [4] as "deliberately and verifiably false news publicized by a news venue" as all news records are accurately labeled by fact-checking websites as false or legitimate statements.

[8]https://github.com/KaiDMML/FakeNewsNet
[9]https://www.politifact.com
[10]https://www.gossipcop.com
[11]https://developer.twitter.com

TABLE V
DESCRIPTIVE STATISTICS OF THE DATASET; AVAILABLE TWEETS ASSOCIATED WITH EACH NEWS RECORD AND AVAILABLE USERS WHO TWEETED/RETWEETED TWEETS EXTRACTED USING TWITTER API

Dataset	*PolitiFact*	*GossipCop*
#Fake News	269	1269
#Real News	230	2466
#Fake Tweets	3011	2946
#Real Tweets	3055	3029
#Engaged Users in Fake Tweets	148 433	115 142
#Engaged Users in Real Tweets	101 438	108 295

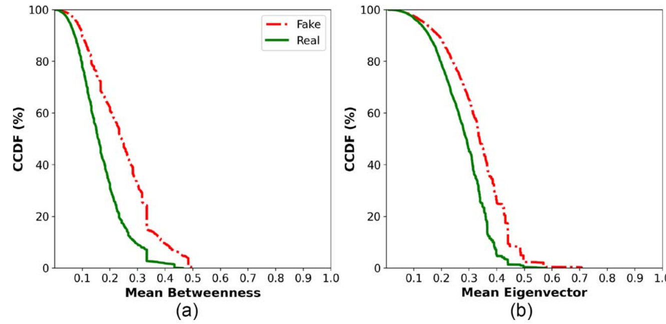


Fig. 7. Complementary cumulative distribution functions (CCDFs) of legitimate and fraudulent news propagation graphs. (a) Mean BC. (b) Mean EC.

2) Dataset Imbalance Correction: As can be observed in descriptive statistics of the datasets indicated in Table V, each dataset contains multiple news records in two classes, fake and legitimate. Employing Twitter API, we extract each record's associated tweets and retrieve engaged users in the related tweets. Many tweets were inaccessible due to user suspension, tweet deletion, and other causes, which imbalanced the final dataset for each class.

Given the scope of our research, which primarily emphasizes the analysis of propagation data, we opted for a relatively straightforward approach, utilizing a random under sampling technique to equalize the sample sizes between the largest and smallest classes. While these techniques, as established in the literature, are generally considered robust even in the presence of varying levels of noise and class imbalance [47], it is important to note that more advanced or tailored methods for handling imbalanced datasets containing social context could potentially be explored in future research to further enhance the robustness of our approach.

B. Propagation Graphs Analysis

To extract the architectural and temporal distinctions across fake and genuine news dissemination trends (*RQ1*), we first examine the centrality metrics of spreaders in the two classes and then investigate the disparities in network size and network growth over time.

Experiments described in the following sections were run using a combined version of the GossipCop and PolitiFact datasets to broaden the applicability of the findings and improve the reliability of the drawn conclusions.

1) Centrality Measures Analysis: The term *centrality* in network analysis pertains to how influential a node is within a network by quantifying its interconnectedness. More precisely, the centrality of node i in the social network $\mathbf{g}$ is $c_i(\mathbf{g})$ and can be computed by the function $\mathbf{c} : G(n) \to \mathbb{R}^n$ [48]. We examine real and false news spreaders by employing two critical centrality measurements, betweenness centrality (BC) and eigenvector centrality (EC).

The significance of a node in linking other nodes in the network is assessed by its BC score [49]. This metric operates under the premise that a node with a higher BC score plays a crucial role in transmitting information from one segment of the network to another as it serves as a linking mechanism among other nodes. Moreover, EC algorithm was proposed by Bonacich and Lloyd [50] to quantify the transitive influence of nodes in a network by assigning relative scores to all nodes. The premise behind the EC algorithm is that for node i, connections to high-scoring neighbors increase EC score $\mathbf{c_i^{eig}}$, whereas i's links to less interconnected nodes decrease c_i^{eig}. We compute the average of EC scores for all nodes in each false and legitimate propagation graph to apprehend the distinctions between these two classes.

As can be observed from Fig. 7(a), falsehood spreaders have more significant BC scores in comparison with truth spreaders. Spreaders with higher BC scores substantially influence other nodes in obtaining the news due to their control over the flow of information. For instance, 36% of fake news networks have an average BC value of 0.3, whereas only 9% of legitimate news networks have the same BC score.

Likewise, Fig. 7(b) describes the dissimilarities between bogus and valid news spreaders' average EC scores. 62% of authentic news networks hold an average of 3.0 EC value, while 81% of unauthentic news networks hold the exact EC value of 3.0. News spreaders with higher eigenvector scores are linked to more other nodes that often have high scores, which conveys that the information they pass through the network will reach more people.

The findings emphasize the significance of timely detection of false news as it was observed that fake news spreaders possess more significant potential in disseminating misinformation to a larger audience compared to truth spreaders. These observations concordantly substantiate the structural disparity between fake and genuine news propagation graphs.

2) Graph-Level Analysis: Our research here focused on quantifying the extent and development of news networks to better understand their unique qualities. The network size, defined as the number of unique users in a network, was analyzed using CCDFs. The results, shown in Fig. 8(a), reveal a notable difference between the network sizes of fake and real news. Specifically, while only 6% of real news graphs reached 1000 people, 20% of false news networks engaged 1000 people. These results highlight the structural differences between false and genuine information propagation networks.

Furthermore, we investigate the temporal dynamics of propagation graphs by measuring the average time of real and false reposts, as shown in Fig. 8(b). The results indicate that false

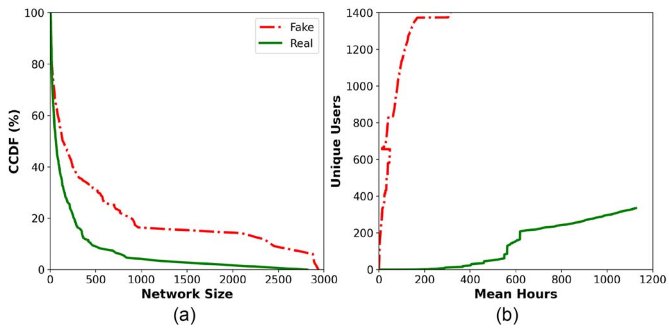


Fig. 8. Network-level analysis in fake and real news networks. (a) CCDFs of network size. (b) Descriptive statistics on users who retweeted fake and real news in a two weeks time window.

news networks propagate at a more rapid pace compared to authentic news, leading to the attraction of a greater number of users on average. By the time accurate news graphs had attracted less than 50 users, fake news networks had already attracted over 1200 users on average. These findings provide strong evidence of the pernicious and rapid spread of false information compared to factual content.

3) Node-Level Features Analysis: In this section, we perform a correlation analysis on the feature set discussed in Section III-C1 as well as conducting two hypothesis tests with the aim of evaluating the discriminatory power of the features in classifying fake news, thus addressing research question 2 (*RQ2*). We furthermore perform two *hypothesis tests* on user features carrying continuous and categorical values to compare the differences in each sample.

a) Pearson's correlation: Under the premise that both quantitative variables are normally distributed and measured on interval scales, Pearson's correlation coefficient r_{pearson} is a measure of the strength and direction of the linear association between two variables. The correlation between node-level features and the target label is computed using Pearson's value. The Pearson's correlation coefficient only represents the linear relationship between the corresponding variables; thus, even a very weak Pearson's coefficient does not reject the existence of a nonlinear relationship between features and the target class. Moreover, as previously noted in [51], the correlation between variables in problems related to human behavior typically exhibits low magnitudes. The results, as presented in Table VI, show a weak to moderate correlation between selected node-level characteristic and the news class, with r_{pearson} values ranging from 0.084 (profile text sentiment score) to 0.346 (followers counts). These results suggest a weak linear correlation between feature values and the news class, which can boost classification accuracy.

b) Spearmans' rank correlation: As previously stated, Pearson's correlation is limited in its assessment of the linear relationship between two variables. Thus, we employ another technique called Spearman's correlation coefficient, denoted as r_{spearman}, to assess the monotonic relation between spreaders' feature values in authentic and fraudulent news samples. Spearman's correlation coefficient is a nonparametric rank statistical technique that computes the strength and direction of the monotone association between two rated values. As Table VI illustrates, all user characteristics exhibit $|r_{\text{spearman}}|$ values ranging from 0.067 (profile text sentiment score) to 0.387 (listed count). The findings provide evidence for a monotonic relationship between the feature values and the news category, suggesting the feature set's informativeness in detecting false news.

TABLE VI

SUMMARY OF NODE-LEVEL FEATURES ANALYSIS. $r_{\text{pearson}}(y_0, y_1)$ AND $r_{\text{spearman}}(y_0, y_1)$ DENOTE PEARSON'S AND SPEARMAN'S CORRELATION COEFFICIENT VALUES OF FEATURES AND THE TARGET LABEL OF NEWS SPREADERS (y_0 = FAKE AND y_1 = TRUE), RESPECTIVELY. $t - \text{test}(y_0, y_1)$ AND $X^2(y_0, y_1)$ REPRESENT THE STUDENT'S T-TEST AND CHI-SQUARE RESULTS COMPARING AUTHENTIC AND FRAUDULENT NEWS SPREADERS' FEATURES, RESPECTIVELY

Attribute	$r_{\text{pearson}}(y_0, y_1)$	$r_{\text{spearman}}(y_0, y_1)$	$t - \text{test}(y_0, y_1)$		$X^2(y_0, y_1)$	
			p	t	p	X^2
Retweet Rank	**−0.178**	**−0.167**	0.005	2.56	-	-
Retweet Distance	**−0.155**	**−0.161**	**0.0**	11.7	-	-
Account Age	**0.192**	**0.177**	**0.0**	−25.72	-	-
Verified	0.142	**0.281**	-	-	**0.0**	172.89
Followers Count	**0.346**	**0.381**	**0.0**	4.77	-	-
Friends Count	0.124	**0.192**	**0.0**	14.1	-	-
GEO	**0.213**	0.117	-	-	**0.0**	115.08
Favorites Count	**0.12**	0.093	**0.0**	13.19	-	-
Statuses Count	0.1	**0.143**	**0.0**	48.66	-	-
Listed Count	**0.217**	**0.387**	**0.0005**	3.46	-	-
Profile Banner	0.101	−0.117	-	-	**0.0002**	13.70
Profile Image	0.128	**0.159**	-	-	0.0084	5.98
Description Length	**−0.193**	**−0.114**	**0.0009**	3.29	-	-
Total Name Length	**−0.117**	**−0.123**	0.0096	3.43	-	-
Profile Emojis Count	0.093	0.085	**0.0**	7.12	-	-
Description Entities Count	**−0.182**	**0.144**	0.006	−2.69	-	-
Profile Text Sentiment Score	0.084	0.067	**0.0**	6.08	-	-
Status Entities Count	**0.271**	**0.255**	**0.0**	−8.11	-	-
Status Emojis Count	**−0.205**	**0.222**	0.003	2.52	-	-
Status Public Metrics Count	**0.322**	**0.26**	0.0	−10.51	-	-
Status Sentiment Score	**0.239**	**0.136**	**0.0**	−12.64	-	-
Status Positive Words Count	**0.169**	0.098	**0.0008**	−3.34	-	-
Status Negative Words Count	0.099	0.092	**0.0**	−5.05	-	-

Note: The bold values indicate statistical significance, suggesting that these attributes are notably different between authentic and fraudulent news spreaders and may be critical in distinguishing them.

c) Student's t-test: One of the most prevalent statistical hypothesis tests is the Student's t-test which compares the means of two observed samples. Under the assumption of the *null hypothesis*, which states that there is no statistically significant difference between the means of features associated with the engaged users in fake and authentic news, a two-sample t-test is carried out to determine the statistical significance and discriminatory power of the features in classifying fake news. The critical value of t for a two-tailed t-test at a significance level of 0.05 and with an infinite degree of freedom is 1.96, according to the Student's t table. The results of the $t - \text{test}(y_0, y_1)$ in Table VI demonstrate that all user attributes exhibit a $|t|$ value larger than 1.96 and a p-value smaller than $\alpha = 0.05$, which refutes the null hypothesis and indicates a difference in the mean of false and genuine news user characteristics.

d) Chi-squared test: The Chi-square goodness of fit test (X^2) is performed on categorical (binary) user features. For this test, we assume the null hypothesis as if there is no substantial distinction between the distributions of categorical variables among false and authentic news spreaders. The $X^2(y_0, y_1)$ column in Table VI reveals that all categorical features exhibit chi-square values greater than 3.84 and p-values smaller than $\alpha = 0.05$, indicating a rejection of the null hypothesis and demonstrating a significant difference in the distributions of categorical variables between fake and genuine news spreaders.

C. Baseline Approaches

To compare the performance of the proposed approach, we implement eight state-of-the-art methods in online fake news detection as baseline methods. The details of each baseline approach are discussed as follows.

1) *RST* [9]: This content-based study uses RST as a framework to compare the coherence and structure of misleading and honest narratives.
2) *TCNN-URG* [28]: A neural network architecture that utilizes both word-level and sentence-level representations to extract semantic information from news articles. The performance is improved by a generative model that learns the nature of user replies and creates responses to new articles.
3) *Theory driven model (TDM)* [10]: A news categorization method that represents news content based on sociological and forensic theories.
4) *Propagation2vec* [25]: A deep neural network model that reconstructs full propagation knowledge from limited data and uses a hierarchy-based attention method to prioritize useful information for false news identification.
5) *dEFEND* [19]: A hierarchical coattention network that collects critical phrases from both news content and user replies for false news detection.
6) *CB-fake* [37]: A multimodal feature vector approach that uses BERT for textual features and CapsNet for visual features.
7) *Network-based fake news detection (NFND)* [12]: A pattern-based false news detection technique that characterizes and quantifies fake news dissemination, spreaders, and network linkages.
8) *DSS* [20]: A hybrid model that concurrently and dynamically uses the propagation tree and the instance networks. The model is broken down into three distinct analyses: dynamic, static, and structural.

D. Experimental Setup

The implementation of the proposed model is executed utilizing PyTorch[12] [52], a highly regarded open-source machine-learning framework founded on the Python[13] programming language and the Torch library [53]. We employ a 7 : 3 cross-validation split, with 70% of the data utilized for training and 30% for evaluating the model's performance on unseen data. We apply spatial Dropout to each layer in the propagation information encoder to avoid overfitting.

To achieve greater flexibility and precision, our model undergoes pretraining on a range of news propagation graphs captured at different time intervals (i.e., 2, 4, 8, and 16 h) and diverse sources of news content (i.e., tweet text and news title). By exposing the model to varied sources of textual content and propagation data, we compel it to learn the patterns in data more comprehensively.

TABLE VII
OPTIMAL VALUES OF HYPERPARAMETERS FOR OUR PROPOSED MODEL AND THE EXPERIMENTAL RANGES OF EACH

Hyperparameters	**Optimal Value**	**Experimental Range**
Batch Size	2	$2, 4, 8$
Propagation Encoder Hidden Size	128	$32 - 256$
Textual Encoder Hidden Size	768	–
Propagation Encoder Attention Heads	8	$2 - 16$
Textual Encoder Attention Heads	12	–
Weight Decay (L2 Regularization)	10^{-5}	$10^{-6} - 10^{-4}$
Number of Epochs	120	$50 - 150$
Overall Dropout Rate	0.2	$0.1 - 0.6$
Initial Learning Rate	10^{-5}	$10^{-6} - 10^{-4}$
Optimizer	AdamW	Adadelta, SGD
Loss Function	Cross Entropy Loss	Squared Hinge Loss
Learning Rate Scheduler	ReduceLR On Plateau	Step LR, Exponential LR

To monitor the convergence of the model, we use a learning rate scheduler *ReduceLROnPlateau* which adjusts the learning rate based on a factor of 0.1 which reduces the learning rate by 10% when the monitored metric does not improve for 4 four consecutive epochs. All models were evaluated using conventional effectiveness criteria for classification problems, including accuracy and F1-score. The hyperparameter details, including optimal values and ranges that were experimentally explored for finding the optimal values, are presented in Table VII.

E. Results

1) Overall Performance: We thoroughly examine our proposed model's effectiveness on two real-world datasets, namely PolitiFact and GossipCop. Our assessment involves a comparative analysis with a range of state-of-the-art models, as detailed in Section IV-C. Our suggested model, MEFaND, achieves an F1-score of 99% and 96%, respectively, on the PolitiFact and GossipCop datasets, proving its superiority over several existing state-of-the-art baselines as demonstrated in Table VIII.

Specifically, MEFaND achieves superlative performance by utilizing only the limited propagation information collected within 5 h of the news publication. In comparison, Propogation2vec, another baseline with a similar detection deadline of 5 h, falls behind MEFaND by a considerable margin, as our model outperforms Propogation2vec by 10% in both accuracy and F1-score.

The DSS model exhibits a 2% advantage over MEFaND in terms of accuracy on the GossipCop dataset. However, the F1-score, a comprehensive measure of a model's performance incorporating precision and recall, is more relevant. MEFaND outperforms other models, including DSS, on both datasets,

[12]https://pytorch.org
[13]https://www.python.org

TABLE VIII
EVALUATION OF OUR PROPOSED METHOD AND EXISTING FAKE NEWS DETECTION MODELS USING GOSSIPCOP AND POLITIFACT DATASETS WITH FOCUS ON EARLY DETECTION CAPABILITY (E), CONTENT-BASED (C) AND PROPAGATION-BASED (P) APPROACHES, AND THEIR DETECTION DEADLINES (Δ) IF APPLICABLE (∞ DENOTES THE PROPAGATION-BASED METHODS LEVERAGING COMPLETE PROPAGATION DATA)

Baselines		**Type**			**PolitiFact**		**GossipCop**	
Name	Δ	C	P	E	Acc	F1	Acc	F1
RST	-	X		X	.607	.569	.531	.512
TCNN-URG	∞	**X**	**X**	**X**	.712	.810	.736	.603
TDM	-	X		X	.892	.892	-	-
Propagation2Vec	**5 h**		X	X	.897	.893	.892	.874
dEFEND	∞	X	X		.904	.928	.808	.755
UPF	∞		X		.909	.904	.966	.966
CB-Fake	-	X		X	.930	.920	.920	.840
NFND	∞		X	X	.929	.932	-	-
DSS	12 h		X	X	.987	.988	**.981**	.913
MEFaND	**5 h**	**X**	**X**	**X**	**.995**	**.996**	.960	**.968**

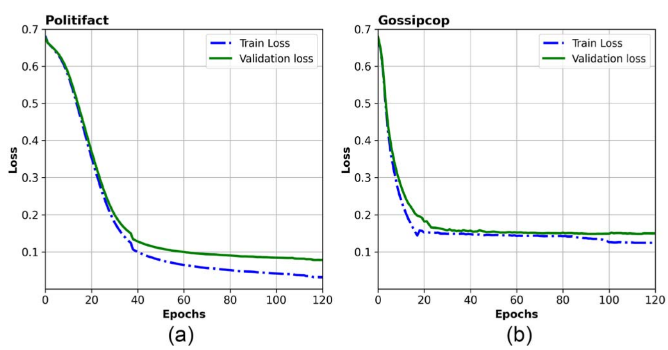


Fig. 9. Our proposed model's learning curves on (a) PolitiFact and (b) GossipCop datasets.

as indicated by the F1-score, implying that it is better at discriminating between genuine and fabricated news. Furthermore, the DSS model reports a minimum detection time of 12 h, which is more than twice as long as MEFaND's detection deadline. As a result, although DSS may have a slightly higher accuracy, MEFaND's superior F1-score and faster detection time make it the more desirable choice for identifying fake news.

MEFaND not only surpasses several other baselines in terms of performance but also the quantity of data that our model employs to identify false news is substantially smaller than other baselines, as it relies solely on the analysis of short news content and the chain of news disseminators.

Moreover, our model's training and validation results are graphically represented in Fig. 9. Throughout several iterations, the training and validation losses have steadily decreased to reach the optimal minimum, proving that the model is not overfitting or underfitting. Therefore, the model can generalize well to new data, and its performance is not confined to the training dataset.

These findings support the hypothesis that the integration of both textual and propagation graph characteristics of news enhances the speed and accuracy of fake news detection models, which addresses research question 3 (*RQ3*).

TABLE IX
LIST OF ABLATION STUDY VARIATIONS AND THE BRIEF DESCRIPTION OF EACH

Variation	**Description**
MEFaND	The complete framework.
MEFaND-*PNFP*	The propagation information encoder was altered by excluding node-level implicit profile features.
MEFaND-*PNFT*	The propagation information encoder was altered by excluding node-level textual features.
MEFaND-*PNFM*	The propagation information encoder was altered by excluding node-level temporal features.
MEFaND-*PEF*	The propagation information encoder was altered by excluding edge-level features.
MEFaND-*PGAT*	The propagation information encoder was altered by removing the graph attention module.
MEFaND-*PSAGE*	The propagation information encoder was altered by removing the GraphSAGE module.
MEFaND-*PPG*	The propagation information encoder was altered by replacing the constructed propagation graphs with conventional propagation graphs.
MEFaND-*TTI*	The news title was removed from the data fed into the textual information encoder.
MEFaND-*TTE*	The tweet text was excluded from the input data of the textual information encoder.

Note: The bolded metrics denote superior performance, signifying the highest accuracy, F1 scores, and optimal detection timelines, thus identifying the most proficient models for fake news detection within the PolitiFact and GossipCop datasets.

2) Ablation Study: Concerning research question 4 (*RQ4*), an ablation study is performed to assess the relative prominence of the various components of the proposed model. Multiple independent variants were generated by selectively eliminating components, and their relative F1-scores were calculated for comparison with the baseline (i.e., MEFaND) as represented in Table IX.

The ablation study, presented in Fig. 10, highlights the critical role of each component in the MEFaND framework's performance. The results indicate that removing components leads to a decrease in the F1-score, with a range of 1%–7% on PolitiFact and 1%–8% on GossipCop. The properties inherent to individual nodes and edges in the propagation encoder were found to contribute by over 7% and 2%, respectively, to the F1-score in both datasets. The GraphSAGE and graph attention layers were also found to have a 2%–4% contribution. The combination of tweet text and news title improved performance by more than 7%. It is worth noting that results were measured by a five-hour detection deadline to examine the overall contribution of each component in the early stages of news propagation.

V. DISCUSSION OF LIMITATIONS

A. Further Modalities Integration

One potential area for improvement in our proposed approach is the integration of additional modalities, such as visual content. Visual elements, such as images and videos, can carry valuable information for distinguishing between false and true

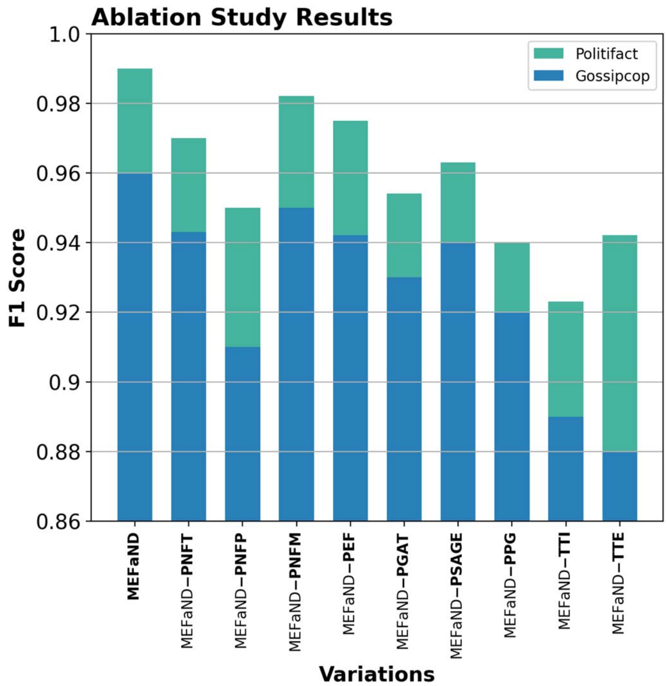


Fig. 10. Comparison of our proposed model with various variations.

news. Given the increasing use of AI techniques for manipulating visual content, incorporating these modalities alongside textual content and propagation data could enhance our model's ability to detect fake news. Future research can explore methods for effectively extracting and utilizing visual features to improve detection accuracy.

B. Domain-Specific Adaptation

While our approach shows promise in the context of social media, its generalizability to other domains like politics, sports, and entertainment may be limited. Fine-tuning the textual information encoder for specific domains, especially those influenced by ethical considerations or political censorship, could enhance adaptability and performance. Addressing domain shift and labeling challenges, as suggested by recent works [29], may further enhance the model's versatility. Future investigations should focus on domain-specific adaptations and strategies for robust early fake news detection in diverse contexts.

C. Enhanced Scalability and Efficiency

Scaling our approach to handle larger datasets extracted from real-time social media stream is an important consideration for practical deployment. Optimizing the model for scalability and efficiency can make it more suitable for real-world applications. In addition, extending our approach to operate effectively across different social media platforms (i.e., Weibo) and languages (Chinese, Farsi, etc.) is a promising direction. Given the global nature of the fake news problem, a model capable of handling multilingual and multiplatform environments could have a significant impact.

D. Ethical Considerations

As generative AI-driven fake news generation and detection systems become more prevalent, ethical considerations gain more significance. Further research is necessary to examine the ethical implications of such techniques, especially in order to reduce the bias and improving the fairness of narratives before it becomes widespread.

E. User Analysis Importance in Early Fake News Detection

As discussed in Section III-C, understanding user behavior and susceptibility to the dissemination of fake news is crucial to combating misinformation and establishing a trustworthy media landscape.

In our node-level analysis, we emphasized the significance of implicit profile features. These features offer a notable advantage by being less vulnerable to manipulation by malicious actors, particularly in situations where misinformation is deliberately spread by a considerable volume of fake users or social bots.

Drawing from insights derived from social and psychological studies, such as the echo chamber effect, it is evident that individuals tend to gravitate toward like-minded users, leading to the formation of segregated communities on social media platforms. Our research findings discussed in Section IV-B further underscore the potential benefit of integrating supplementary network-based attributes within a single social media post into false news detection models.

Incorporating various aspects of author and spreaders analysis specifically in small communities on social media, enables the early detection of fake news while capitalizing on the intrinsic value of a single post, even when comprehensive propagation data may be limited.

In summary, these techniques allow identifying and grading users feasible to publish and propagate incorrect or misleading news before engaging in such practices. Adapting our model to dynamic environments, where news propagation through social media continuously evolve, is essential.

VI. Conclusion

In this work, we introduced MEFaND, a multimodal framework for early fake news identification that consists of three parts. The propagation information encoder simulates the initial stages of news circulation using two effective GNN architectures—GraphSAGE and GAT. To encode the semantic meaning of news articles, the textual information encoder employs the BERT language model. Lastly, the news classifier receives the encoders' output vectors and processes them through a linear layer to make a label prediction.

We conducted comprehensive experiments on two publicly available datasets, PolitiFact and GossipCop, demonstrating that our model outperforms several state-of-the-art methods with an impressive F1-score of 99% and 96%, respectively. The proposed MEFaND framework holds several advantages over existing approaches, including its ability to detect fake news with limited propagation data and generalize the different types of false news. Our analysis of user characteristics and propagation graphs demonstrates the effectiveness of our approach in early fake news detection. We also measured the individual contribution of various components in our model to the final performance, which can provide insights for future research on multimodal false news detection.

References

[1] "MS Windows NT fake news statistics." Accessed: Sep. 30, 2010. [Online]. Available: https://guardian.ng/technology/over-3-4bn-people-to-use-social-media-by-2023-says-report/

[2] "MS Windows NT fake news statistics." Accessed: Sep. 30, 2010. [Online]. Available: https://journolink.com/resources/post/319-fake-news-statistics-2019-uk-worldwide-data

[3] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explor. Newslett.*, vol. 19, no. 1, pp. 22–36, 2017.

[4] X. Zhou and R. Zafarani, "Fake news: A survey of research, detection methods, and opportunities," 2018, *arXiv:1812.00315*.

[5] K. Rapoza, "Can 'fake news' impact the stock market?" *Forbes*. Accessed: Feb. 26, 2017. [Online]. Available: https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market. 2017.

[6] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," *ACM Comput. Surveys (CSUR)*, vol. 51, no. 2, pp. 1–36, 2018.

[7] A. Bovet and H. Makse, "Influence of fake news in Twitter during the 2016 US presidential election," *Nature Commun.*, vol. 10, no. 1, 2019.

[8] A. Roets et al., "'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions," *Intelligence*, vol. 65, pp. 107–110, 2017.

[9] V. L. Rubin and T. Lukoianova, "Truth and deception at the rhetorical structure level," *J. Assoc. Inf. Sci. Technol.*, vol. 66, no. 5, pp. 905–917, 2015.

[10] X. Zhou, A. Jain, V. V. Phoha, and R. Zafarani, "Fake news early detection: A theory-driven model," *Digit. Threats, Res. Pract.*, vol. 1, no. 2, pp. 1–25, 2020.

[11] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, "A review of relational machine learning for knowledge graphs," *Proc. IEEE*, vol. 104, no. 1, pp. 11–33, Jan. 2016.

[12] X. Zhou and R. Zafarani, "Network-based fake news detection: A pattern-driven approach," *ACM SIGKDD Explor. Newslett.*, vol. 21, no. 2, pp. 48–60, 2019.

[13] Y. Liu and Y.-F. B. Wu, "FNED: A deep network for fake news early detection on social media," *ACM Trans. Inf. Syst. (TOIS)*, vol. 38, no. 3, pp. 1–33, 2020.

[14] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[15] T. Bian et al., "Rumor detection on social media with bi-directional graph convolutional networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 01, 2020, pp. 549–556.

[16] V.-H. Nguyen, K. Sugiyama, P. Nakov, and M.-Y. Kan, "FANG: Leveraging social context for fake news detection using graph representation," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 1165–1174.

[17] A. Silva, L. Luo, S. Karunasekera, and C. Leckie, "Embracing domain differences in fake news: Cross-domain fake news detection using multimodal data," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 1, 2021, pp. 557–565.

[18] Y.-J. Lu and C.-T. Li, "GCAN: Graph-aware co-attention networks for explainable fake news detection on social media," 2020, *arXiv:2004.11648*.

[19] K. Shu, D. Mahudeswaran, S. Wang, and H. Liu, "Hierarchical propagation networks for fake news detection: Investigation and exploitation," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 14, 2020, pp. 626–637.

[20] 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 Syst. Appl.*, vol. 198, 2022, Art. no. 116635.

[21] S. Raza and C. Ding, "Fake news detection based on news content and social contexts: A transformer-based approach," *Int. J. Data Sci. Analytics*, vol. 13, no. 4, pp. 335–362, 2022.

[22] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimedia Tools Appl.*, vol. 80, no. 8, pp. 11765–11788, 2021.

[23] J. F. Low, B. C. Fung, F. Iqbal, and S.-C. Huang, "Distinguishing between fake news and satire with transformers," *Expert Syst. Appl.*, vol. 187, 2022, Art. no. 115824.

[24] D. M. Lazer et al., "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.

[25] A. Silva, Y. Han, L. Luo, S. Karunasekera, and C. Leckie, "Propagation2Vec: Embedding partial propagation networks for explainable fake news early detection," *Inf. Process. Manage.*, vol. 58, no. 5, 2021, Art. no. 102618.

[26] B. Shi and T. Weninger, "Discriminative predicate path mining for fact checking in knowledge graphs," *Knowl.-Based Syst.*, vol. 104, pp. 123–133, 2016.

[27] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Inf. Process. Manage.*, vol. 57, no. 2, p. 102025, 2020.

[28] F. Qian, C. Gong, K. Sharma, and Y. Liu, "Neural user response generator: Fake news detection with collective user intelligence," in *Proc. IJCAI*, vol. 18, 2018, pp. 3834–3840.

[29] Y. Zhu et al., "Memory-guided multi-view multi-domain fake news detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 7, pp. 7178–7191, Jul. 2023.

[30] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[31] A. Radford et al., "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, no. 8, p. 9, 2019.

[32] C. Liu et al., "A two-stage model based on BERT for short fake news detection," in *Proc. 12th Int. Conf. Knowl. Sci., Eng. Manage.*, Athens, Greece, August 28–30, 2019, Springer, pp. 172–183.

[33] H. Jwa, D. Oh, K. Park, J. M. Kang, and H. Lim, "exBAKE: Automatic fake news detection model based on bidirectional encoder representations from transformers (BERT)," *Appl. Sci.*, vol. 9, no. 19, 2019, Art. no. 4062.

[34] K. Shu, S. Wang, and H. Liu, "Exploiting tri-relationship for fake news detection," 2017, *arXiv:1712.07709*.

[35] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[36] K. Shu, X. Zhou, S. Wang, R. Zafarani, and H. Liu, "The role of user profiles for fake news detection," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, 2019, pp. 436–439.

[37] B. Palani, S. Elango, and V. Viswanathan K, "CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT," *Multimedia Tools Appl.*, vol. 81, no. 4, pp. 5587–5620, 2022.

[38] P. Wei, F. Wu, Y. Sun, H. Zhou, and X.-Y. Jing, "Modality and event adversarial networks for multi-modal fake news detection," *IEEE Signal Process. Lett.*, vol. 29, pp. 1382–1386, 2022.

[39] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 855–864.

[40] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964.

[41] J. Ma, W. Gao, and K.-F. Wong, "Detect rumors in microblog posts using propagation structure via kernel learning," *Proc. 55th Annu. Meet. Assoc. Comput. Linguistics (ACL 2017)*, Vancouver, Canada, July 30–August 4, pp. 708–717. Available: https://ink.library.smu.edu.sg/sis_research/4563

[42] Y. Liu and Y.-F. Wu, "Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018.

[43] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, *arXiv:1706.02216*.

[44] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[45] P. Velicković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.

[46] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media," 2018, *arXiv:1809.01286*.

[47] A. Fernández, S. Garcia, F. Herrera, and N. V. Chawla, "Smote for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *J. Artif. Intell. Res.*, vol. 61, pp. 863–905, 2018.

[48] F. Bloch, M. O. Jackson, and P. Tebaldi, "Centrality measures in networks," 2016, *arXiv:1608.05845*.

[49] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977.

[50] P. Bonacich and P. Lloyd, "Eigenvector-like measures of centrality for asymmetric relations," *Social Netw.*, vol. 23, no. 3, pp. 191–201, 2001.

[51] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, vol. 2. New York, NY, USA: Springer, 2009.

[52] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

[53] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: A matlab-like environment for machine learning," in *Proc. NIPS Workshop BigLearn*, 2011.