WJCMS

Wasit Journal of Computer and Mathematics Science

Journal Homepage: https://wjcm.uowasit.edu.iq/index.php/WJCM
e-ISSN: 2788-5879 p-ISSN: 2788-5887

A Graph-Aware Recurrent Fusion Deep Ensemble Architecture for Fake News Detection

Karrar Kanaan Hasan[1,*], Jaafar Salman Mujaysir Alsafi[2] and Hussein Mashaan Meteab Alzuwaydee[3]

[1]Department of Computer Science, First Al-Mutafawiqeen Secondary School, Directorate of Education of Thi-Qar, Ministry of Education, Nasiriyah, Thi-Qar, Iraq.
[2]Department of Information Technology Division, Vocational Education Department, Thi-Qar Education Directorate, Thi-Qar, Iraq.
[3]Department of Informatics and Electrical Engineering, Széchenyi István University, Győr, Hungary.

*Corresponding Author: Karrar Kanaan Hasan

DOI: https:// doi.org/10.31185/wjcms.500
Received 01 March 2026; Accepted 27 March 2026; Available online 31 March 2026

ABSTRACT: With the rapid spread of fake news on social media platforms, it has become essential to develop effective detection mechanisms that overcome the limitations of content-only or propagation-only approaches. This study aims to design a unified framework that jointly models textual semantics and diffusion characteristics for improved misinformation detection. To achieve this, we propose GRAFT-FND, a Graph-Aware Recurrent Fusion Deep Ensemble architecture that integrates contextual word embeddings (Word2Vec, BERT, and BERTweet) with recurrent neural networks (RNN/GRU/LSTM/BiLSTM) and graph-based node embedding methods (Node2Vec and DeepWalk) within a fusion-aware learning module. Extensive experiments conducted on the Twitter15 and Twitter16 benchmark datasets using 10-fold cross-validation demonstrate that the proposed framework consistently outperforms baseline and recent state-of-the-art models, with the fusion mechanism and propagation-aware representations contributing significantly to performance improvement. The results indicate that jointly modeling semantic and structural information enhances the ability to capture complex misinformation patterns and improves generalization across datasets. In conclusion, the proposed framework provides a robust and scalable solution for fake news detection in social media environments. Future work is recommended to investigate adversarial robustness, real-time deployment, and the integration of multimodal data sources.

Keywords: Fake News Detection; Graph Neural Networks; Recurrent Neural Networks; Node Embedding; Fusion Learning; Social Media Analytics.

1. INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

The media ecosystem has undergone a digital transformation, which has profoundly transformed the production, distribution, and consumption of information. Source: Social networking sites facilitate immediate communication among end users and connect, enabling broadcasting of significant pieces of information to the world instantly [1],[2]. Although this transformation entails improved reach and democratization of knowledge exchange, it has also enabled turbocharged propagation of misinformation and disinformation with potential adverse socio-political and economic implications [3],[4]. Misinformation diffuses from social accounts to a higher extent based on complicated interaction sequences like reposts, mentions and replies–which generates dynamic diffusion structures over social networks [5] Conventional fake news detection methods focus on the textual content using linguistic features, statistical representations, or deep contextual embeddings [6]. While these approaches achieve high degrees of semantic understanding, they are generally incapable of discerning the propagation mechanisms that differentiate false information from truthful content. In contrast, graph-based approaches capture structural diffusion patterns but often discard rich textual semantics.

*Corresponding author: karar_kanaan@utq.edu.iq

As a result, there is an increasing demand for integrated frameworks jointly modelling both semantic content and propagation dynamics to obtain more reliable and generalizable fake news detection [7].

1.2 RESEARCH GAP

Though there have been great advancements in the detection of fake news, there are still many limitations that capabilities of current systems. Firstly, most existing approaches take textual and propagation features as separate components, leading to ineffective interaction of modalities and suboptimal performance. While some propose hybrid models, most of them concatenate features from heterogeneous modalities without learning deep cross-modal relationships [8]. These learn and become better models to build on as they only improve over time. Second, incorporation of state-of-the-art (SOTA) models like large language models (LLMs) with graph neural networks (GNNs) through contrastive learning-based framework or tasks showed increase in representation learning capability. However, they generally incur heavy computation overhead or are not integrated in a systematic way with recurrent architectures, thus less feasible for real-world social media environments [9],[10].

Third, previous studies conduct less exploration into the fusion mechanisms and rarely incorporate a comprehensive ablation analysis to quantify the contributions of semantic, structural and fusion components. Consequently, the exact mechanisms of action of joint modeling strategies are still poorly understood [11].

Such a systematic representation mechanism can overcome those limitations and hence it is necessary to build a compact, time-efficient, as well as interpretable framework capable of fusing semantic texts and propagation topology into one complete learning architecture.

1.3 OBJECTIVES OF THE STUDY

The challenges identified so far inform this study's development of a strong, unified fake news detection framework making use of semantic and structural information. The main aims of this study are:

- A deep learning architecture is designed to be hybrid that integrates the contextual textual embeddings with propagation-aware representations based on graphs.
- To learn meaningful relations between textual and structural features, we devise a fusion-aware mechanism.
- By applying the suggested framework on benchmark datasets, we can test the efficiency of our approach herein against baseline and state-of-the-art models.
- Ablation studies and performance comparisons to understand the contributions of each component.
- To offer a scalable and generalizable solution applicable to real-world social media settings.

1.4 CONTRIBUTIONS

To overcome the above challenges, we propose Graph-Aware Recurrent Fusion Deep Ensemble Architecture for fake news detection. Our main contributions can be summarized as follows:

- Retaining positional information with LSTM and CNN embeddings in a hybrid semantic representation for deep learning transformers
- Structural diffusion modeling through a propagation-aware graph embedding module which encodes various structural diffusion patterns into node embedding.
- Additional model, a Model that integrates fusion-aware learning semantics and structure surface representation relations in one stage.
- The effectiveness of the proposed architecture is confirmed through extensive experimental evaluation as well as ablation studies and cross validation.
- Validation of the reported performance improvements between recent state-of-the-art models.

1.5 RESEARCH QUESTIONS

This study aims to answer the following research questions:
(Q1) What is the impact of advanced textual semantic embeddings on fake news detection performance?
(Q2) How do propagation diffusion embeddings influence classification effectiveness?
(Q3) Does fusion-aware joint modeling provide measurable improvements over independent modeling strategies?
(Q4) How sensitive is the proposed architecture to dataset characteristics and distributional differences?
(Q5) How does the proposed model compare with current state-of-the-art approaches in terms of accuracy, robustness, and generalization?

1.6 PAPER ORGANIZATION

The remainder of this paper is structured as follows. In section 2 we provide related work on methods for fake news detection, which may be text based, graph based or fusion based. Section 3 explains our method comprised of the feature representation, the graph modeling depending on the architecture GRAFT-FND and the algorithmic framework. In Section 4, we share the datasets used along with our experiment setups including ablation studies and hyperparameter tuning and state-of-the-art comparisons. Section 5 concludes the paper, and suggests future research directions.

2. RELATED WORK

2.1 TEXT-BASED FAKE NEWS DETECTION

Traditionally, the detection of fake news has focused on textual content and is heavily reliant on classical natural language processing methods, including n-grams, TF–IDF [12] representations or handcrafted linguistic features (such as writing style and semantic patterns) [13]. With the advent of deep learning, methods like Word2Vec, and GloVe allowed the mapping of sparse textual features to dense semantic representations [14],[15], resulting in enhanced contextual information. In contrast, more recent transformer-based architectures like BERT, RoBERTa and their domain-specific counterparts such as BERTweet have been shown to excel in successfully capturing contextual dependencies and long range semantic relationships within social media text [16].

But even with such advances, text-only models have inherent limitations. Fake news articles are usually handcraft and made to look similar as a real news text, which weakens the discriminative ability of semantic-based approaches. To address this problem, recent work employs large language models (LLMs) by using special transformers like GPT and instruction-tuned to increase contextual reasoning and semantic robustness [17]. Work based on this approach enhances the detection performance for low reservation environment using robust contextual knowledge and zero/few-shot capabilities. Solution: Textual Solutions Fall Short As powerful as text-generative AIs such as LLMs are in advancing any need for a wealth of text-centric applications, such approaches fail to embody the underlying social process and propagation patterns which lead to misinformation [18].

2.2 GRAPH-BASED FAKE NEWS DETECTION

To address the shortcomings of content-limited models, most people have adopted graph-based approaches, specificity for generalization, which enable modeling propagation structure of information through social networks. GNNs, such as GCN and GAT have already been widely applied to leverage local/global structural dependencies in propagation graphs [19],[20]. These models exploit diffusion characteristics showing differences between real and fake information by considering user interactions, reposting behavior, and examining the underlying networks structure (topology) [21],[22].

More recent work has proposed more advanced graph learning frameworks, including heterogeneous graph attention networks and the modeling of dynamic graphs to embed temporal evolution and multi-relational interactions between social network users [23],[24]. Recurrent Bayesian evolution models allow us to approximate the complex propagation behaviors in our training data, improving detection accuracy by modeling both structural constraints and temporal signals. Moreover, node embedding models like Node2Vec and DeepWalk are still effective as they focus on maintaining the topology of a network by leveraging random walk methods [25].

Recently, the graph methods based on contrastive learning has attracted a lot of attention. These methods learn generalizable representations by enforcing agreement among positive node pairs whilst discriminating against the negative samples. For example, hybrid contrastive graph neural networks have been proposed to dualistically learn discorded propagation behaviors under noisy and sparse data conditions, which has successfully boosted the generalization and robustness of fake news detection [26]. However, most graph-based models remain orthogonal to semantic encoders which hampers their ability to fully characterize various features of misinformation.

2.3 FUSION-BASED ARCHITECTURES

To take advantage of semantic and structural information at the same time, a few fusion-based architectures have been developed to combine textual with graph-based representations. Early fusion approaches largely resorted to simple concatenation or late-fusion-based mechanisms, aggregating outputs from independent models without capturing deep cross-modal interdependencies [25]. Although such methods outperform unimodal systems, they usually fail to appropriately encode complex dependencies between textual information and propagation dynamics.

Over the last few years, there has been a move to more sophisticated hybrid architectures that combine LLMs based textual encoders with graph neural networks in a single architecture. As a result, the hybrid LLMs–GNN models facilitate joint learning of semantic and structural representation for significantly improved detection performance and robustness [3],[6],[7]. Moreover, various attention-based fusion mechanisms and multi-view learning frameworks have proposed for the modeling to learn how much different modalities contribute dynamically [30], which could further align with the varying degrees of textual ambiguity and propagation signals.

Another line of work in this viewpoint is about contrastive multimodal learning where we instead learn joint representations via aligning the textual and graph representations to a shared embedding space. This allows for a more consistent feature representation and helps the model to focus on subtle differences between fake or real news. [26]. Although these methods utilize sophisticated architectures, most of the current fusion approaches are either loosely coupled or fall short of a comprehensive ablation analysis which sacrifices interpretability and shed light on cross-modal interactions.

While remarkable advancements have been achieved in text-oriented, graph-based, and hybrid methods, several issues are still unsolved. Firstly, existing methods struggle to combine advanced LLMs-based semantic representations and propagation-driven graph modeling within a single optimized framework. Second, most existing fusion methods do not learn interaction-aware representations that capture deep dependencies across modalities; they simply concatenate information from different modalities. Third, few contrastive learning are utilized and insufficient interpretability analysis has limited the robustness of current models and their generalization.

To mitigate these issues, this work proposes a Graph-Aware Recurrent Fusion Deep Ensemble architecture that cohesively combines several semantic, structural and fusion-aware learning components to improve robustness against overfitting, thereby addressing the overkill problem in fake news detection.

3. PROPOSED METHODOLOGY

3.1 FEATURE REPRESENTATION

3.1.1 Text Preprocessing and Word Embeddings

The proposed framework relies on a dual representation strategy that models both textual semantics and structural propagation information. In this section, we describe the textual feature representation process, including preprocessing, tokenization, document modeling, and embedding construction.

Cleaning Pipeline

For example, raw social media texts often have noise such as URLs, emojis, special characters, repeated spaces and name mention. Hence a structured cleaning pipeline is utilized pre-modeling. These preprocessing steps involve the removal of hyperlinks, punctuation and unwanted symbols, converting everything to lowercase and normalizing the spacing. In the case of static embeddings like Word2Vec, it is possible to optionally remove stop-words in order to keep the vocabulary reduced. This mechanism enhances the homogeneity of representation and reduces sparsity, especially for short and less formal posts.

Tokenization

The data is first cleaned, and then tokenized to get a set of word tokens for every document. A vocabulary of all unique tokens in the data set is built from which, a unique numerical index is assigned to each token. Due to the varying lengths of documents, we must apply padding and take in a fixed maximum sequence length for all samples. Allows for batch processing in Neural Network training.

Document-to-Token Matrix

The whole corpus becomes a document-to-token matrix, in which each row is a padded sequence of token indices for one document.

Let n be the number of documents and k be the maximum sequence length, then the document matrix is defined as: $\mathbf{D} \in \mathbb{N}^{n\times k}$ Each element in this matrix corresponds to the integer index of a token in the vocabulary.

Word Embedding Generation

In order to represent semantic relations of tokens, we use three embedding techniques:

- Word2Vec (Skip-Gram): learns word embeddings by surrounding context words. It contains local semantic and syntactic relations.
- BERT: A deep bidirectional, unsupervised language representation.
- BERTweet: A tweet-specific variant of BERT that is pre-trained on Twitter data and thus better suited to short social media texts.

Embedding Matrix

For a vocabulary of size $m$and embedding dimension s, the embedding matrix is defined as: $\mathbf{W} \in \mathbb{R}^{m\times s}$ In this matrix, each row represents the vector embedding of an individual token. The document-to-token matrix is converted into a dense embedding tensor, in which each token index is replaced by its corresponding embedding vector during training.

Essentially, the textual representation component letter-informs raw text into dense semantic vectors, involving a number of indispensable steps from preprocessing to tokenization and matrix construction, along with embedding generation. The embeddings are then fed into the deep recurrent modelling part of our architecture [26].

3.1.2 Graph Representation and Node Embeddings

To model information diffusion, the social network is represented as a directed graph: $G = (V, E)$

where Vdenotes nodes (users/posts) and Erepresents interactions such as reposts or replies. An edge $(u, v) \in E$indicates that node $v$propagates information from node u.

Propagation Tree Modeling

For each source post, a propagation tree $T = (V_T, E_T) \subseteq G$ is constructed, where the root corresponds to the original post and child nodes represent subsequent interactions. This structure captures diffusion depth and branching patterns associated with misinformation spread.

Node2Vec

Node2Vec learns structural embeddings through biased second-order random walks. The transition bias is defined in Equation (1):

$$\alpha_{p,q}(t,v) = \begin{cases} \frac{1}{p}, & d_{t,v} = 0 \\ 1, & d_{t,v} = 1 \\ \frac{1}{q}, & d_{t,v} = 2 \end{cases} \tag{1}$$

where p and q control exploration behavior. The objective function is formulated as shown in Equation (2):

$$\mathcal{L} = \sum_{u \in V} \log P(N(u) \mid f(u)) \tag{2}$$

DeepWalk

DeepWalk applies uniform random walks and Skip-Gram learning. Its objective is given in Equation (3):

$$\mathcal{L} = \sum_{u \in V} \sum_{v \in N(u)} \log P(v \mid f(u)) \tag{3}$$

with probability modeled via softmax.

Node Embedding Matrix

The learned representations form a node embedding matrix: $\mathbf{N} \in \mathbb{R}^{|V| \times s}$ where s is the embedding dimension. These embeddings encode structural and diffusion characteristics and are used as input to the propagation modeling component of the architecture [27].

3.2 GRAFT-FND ARCHITECTURE

GRAFT-FND is a unified deep ensemble framework with a semantic textual modeling and propagation-aware graph representation, as shown in Figure 1. Core Model: The core model mainly consists of two assembling branches, a Text branch that represents text via embedding (e.g. word embeddings) followed by RNN/GRU/LSTM/BiLSTM steps to learn a Textual Semantic Embedding and, with the Propagation Branch learns semantic representations from propagation patterns in networks. Node2Vec or DeepWalk node embeddings which directly captures these and then it will pass the dense transform to get the Propagation Diffusion Embedding. In a Fusion Module, both representations are concatenated and then passed through a fully connected layer that learns the joint Fusion-Aware Graph–Text representation to be used as input for classification via softmax. In this double-stream architecture, contextual semantics and structured propagation behavior can be optimized from the growing block in a framework.

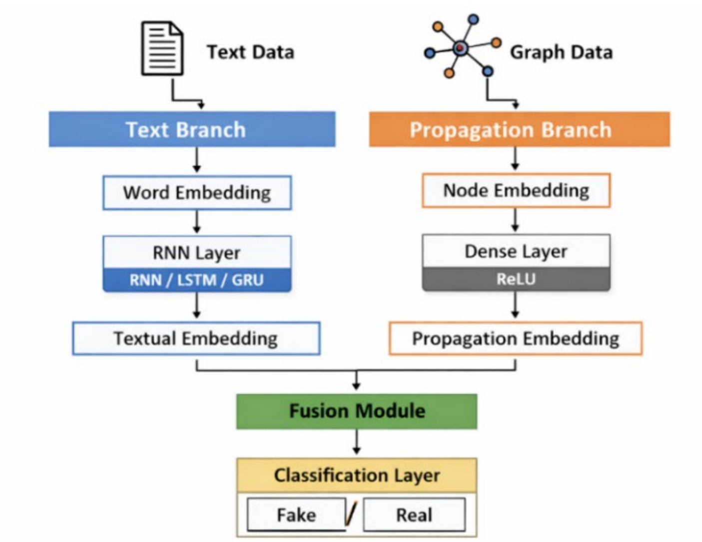

FIGURE 1. Overall Architecture of the Proposed GRAFT-FND Framework. (Designed by the authors)

3.2.1 Text Branch

The Text Branch aims to capture semantic representations from textual data. The first layer is an input layer that takes padded token sequences as its inputs from preprocessing. These tokens are then mapped to dense vector representations through a word embedding layer that employs Word2Vec, BERT, or BERTweet models to capture the contextual meanings of words. After embedding sequences, they are serialized to a recurrent neural network layer for modeling sequential dependencies and contextual flow within the text. Depending on the configured stateful RNN, this layer can consist of a combination of RNN, GRU, or LSTM and their bidirectional versions (BiRNN, BiGRU, BiLSTM) to capture forward as well as backward contextual information. The property of the recurrent layer is that at any timestep (d; d = 1, 2,...D), it defines a hidden representation based on the current input and previous hidden state, but due to the temporally ordered structure of textual data, its final hidden representation r_z at D^{th} timestep contains high-level features that represent Document's semantic information and which will be considered as Textual Semantic Embedding and fed into fusion module where mutual learning takes place among propagation-based features [28].

3.2.2 Propagation Branch

The Propagation Branch serves to characterize structural diffusion patterns on social networks. First, a graph input layer is created to represent the constructed social graph where users interact with each other (through reposts, replies, mentions). From the graph, they implement node embedding techniques (such as Node2Vec or DeepWalk) to generate a low-dimensional vector representation that capture structural relationships and propagation behavior. Once we have embedded the nodes, we end up with a collection of features which can be optimized and converted to a more discriminatively expressive space through additional layers as for example in this case where we feed these node embeddings into a fully connected dense layer using ReLU activation. The result of this process is known as the Representation Spread Embedding (RSE), which summarizes how information flows through your graph, which will subsequently be concatenated with the textual semantic representation and used to classify at the final classification layer [29].

3.2.3 Fusion Ensemble Module

Fusion Ensemble Module All outputs from each branch, i.e., Text Branch and Propagation Branch are concatenated over all their dimensions to produce a fused joint representation. ed by jetsing text/contextual and structural information in a concatenation layer, aggregate to get a joint feature vector Further extracting n-gram features from the textual or data point vectors removing the rich plan context helps capture domain-level relation between stages and within plan. This output is then passed to a fully connected dense layer with ReLU Activation function, which helps the model understand complex interactions of the semantic content and diffusion trends. We call this representation Fusion-Aware Graph–Text Embedding, as it encodes diverse information of each modality. At the end of this learning process, we used a classification layer by applying softmax activation functions and outputs the class probabilities for each reality indicator (fake or real) [30].

The complete training process of the proposed GRAFT-FND framework is presented in Algorithm 1. The model takes input the dataset, the built social graph, the selected word embedding model for text, the node embedding method used and recurrent configuration for Text Branch It returns a trained GRAFT-FND model that can be used to predict misinformation classes.

Algorithm 1. GRAFT-FND Training Procedure

Input:
Dataset $\mathcal{D} = \{(t_i, g_i, y_i)\}_{i=1}^{N}$,
Graph $G = (V, E)$,
Word embedding model WE,
Node embedding model NE,
Recurrent configuration RNN_{type}
Output:
Trained GRAFT-FND model $\mathcal{M}$
Initialize parameters of embedding layers, recurrent layer, dense layers, and softmax classifier
Preprocess all textual samples t_i (cleaning, tokenization, padding)
$E_{text} \leftarrow WE(t_i)$ // Generate word embeddings
$E_{node} \leftarrow NE(G)$ // Generate node embeddings from graph
for each training sample $i = 1$ to N do
  $h_{text}^{(i)} \leftarrow RNN_{type}(E_{text}^{(i)})$ // Text Branch
  $h_{prop}^{(i)} \leftarrow Dense_{ReLU}(E_{node}^{(i)})$ // Propagation Branch
  $h_{fusion}^{(i)} \leftarrow Concatenate(h_{text}^{(i)}, h_{prop}^{(i)})$
  $\hat{y}^{(i)} \leftarrow Softmax(Dense(h_{fusion}^{(i)}))$

```
  Update model parameters using loss ℒ(y^(i), ŷ^(i))
end for
Return trained model ℳ
```

4. EXPERIMENTAL STUDY

4.1 DATASET DESCRIPTION

In this section, we conduct experiments on the two popular benchmark datasets Twitter15 and Twitter16 to evaluate the performance of our proposed GRAFT-FND framework. These datasets have been extensively used for rumor and fake news detection as they provide the textual content along with their corresponding propagation tree structures from Twitter interactions. Twitter15 has 1,490 source events and Twitter16 has 818. An event contains the original source tweet as well as its reply and retweet propagation structure. Both datasets are labeled into four classes, including True, False, Unverified and Non-rumor. These datasets are especially suitable for evaluating graph-aware detection architectures because structural diffusion data is directly available. Table 1 presents a statistical summary of both datasets, including the number of events in all interactions, interaction volume, class categories used and structural features such as average tree depth and average number of nodes per propagation tree.

Table 1. Summary of Twitter15 and Twitter16 Datasets

Dataset	Events	Source Tweets	Replies/Retweets	Classes	Class Labels	Avg. Tree Depth	Avg. Nodes per Tree
Twitter15	1,490	1,490	331,612	4	True, False, Unverified, Non-rumor	4.2	223
Twitter16	818	818	204,820	4	True, False, Unverified, Non-rumor	3.8	250

4.1.1 Textual Data Analysis

The class distributions of both datasets are shown in Figure 2, which reveals a slight imbalance between the categories. The inherent imbalance in class distributions, as shown above, necessitates the evaluation of classifiers using macro-averaged performance metrics to provide a fair assessment across all classes. The distribution of document lengths is visualized in Figure 3 as most tweets being short, as expected for social media. However, there are some samples which could take a longer discussion making the tokens in different samples not similar. The histogram of word count is depicted by Figure 4 to examine the sparsity characteristics and frequency distribution of textual data. (Note that we see most samples have moderate token counts, and therewith we can choose a good maximum sequence length for the Text Branch padding and recurrent modeling in general.

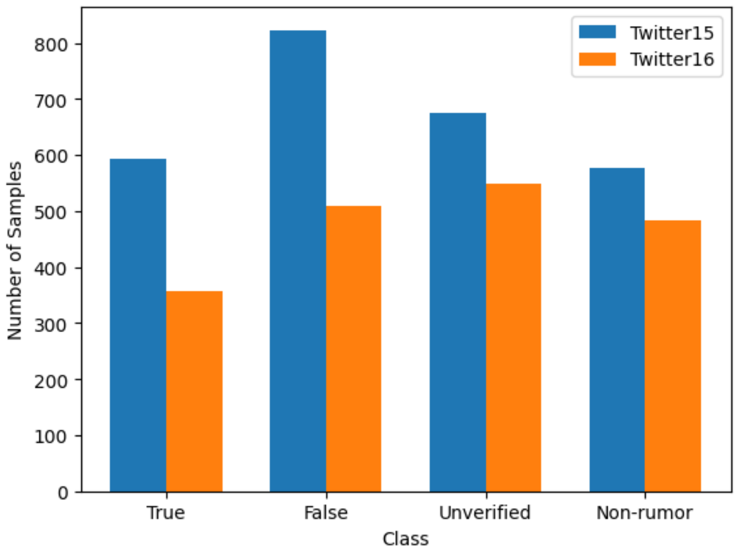

FIGURE 2. Class distribution of the Twitter15 and Twitter16 datasets

As evidenced by the class distribution depicted in Figure 2, there is a slight imbalance between the four different classes between both datasets. Such imbalance emphasizes the need to use macro-averaged evaluation metrics to guarantee fair performance assessment over all classes. Some classes, like “Unverified”, are also noted to have a smaller number of samples, which can make classifiying more difficult and lead to higher chances of mis-classifying elements in minority categories.

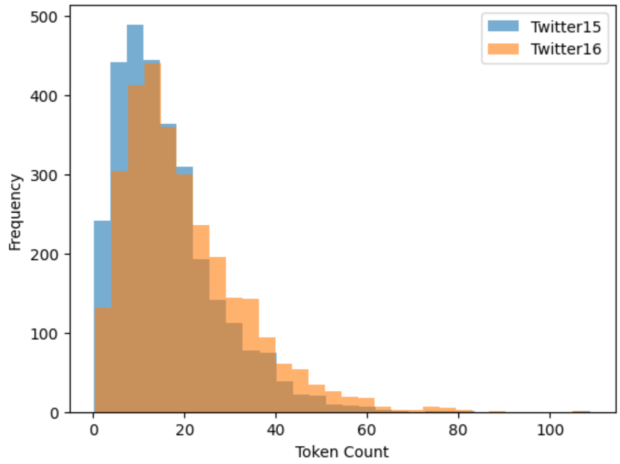

FIGURE 3. Distribution of token lengths in the Twitter15 and Twitter16 datasets

As shown in Figure 3, most of the textual samples are very short, which is common since some social media services have a length limit to its post. But longer sequences indicate that there is variation in user engagement as well as the richness of content. This variance allows us to comfortably choose an appropriate max sequence length in preprocessing, as it is about correlation between computational expense and information retention.

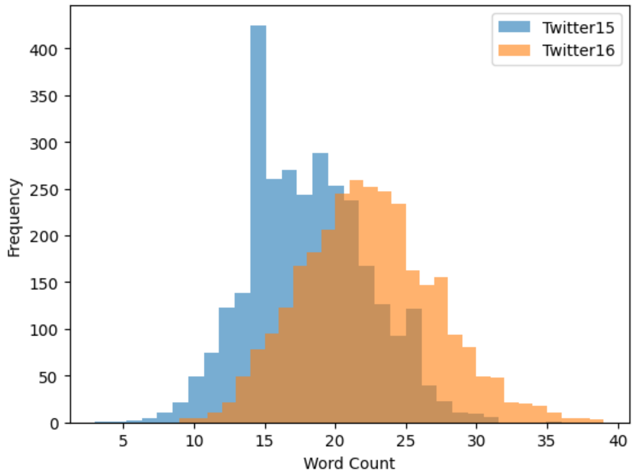

FIGURE 4. Word count distribution of the Twitter15 and Twitter16 datasets

Figure 4 illustrates that word count distribution follows a moderately skewed pattern, with the majority of samples being few tokenized. Therefore, BERT and BERTweet, which can better understand semantics in short texts, continue to win over their less-aware cousins due to the sparsity of what they work with.

4.1.2 Network Data Analysis

Besides textual aspects, we analyse the structural properties of the propagation networks. We visualize example propagation trees in Figure 5, which shows the spread of information from the initial source tweet to later users. Notice the differences in tree depth and branching in true vs false events. The distribution of node degrees is shown in Figure 6 indicate that the networks are following a skewed distribution, i.e., there are very few number of nodes with high connectivity. These highly connected nodes can heavily affect the mechanics of information diffusion.

Moreover, the degree centrality analysis in Figure 7 focuses structurally influential users in propagation network. Nodes with high centrality values often have strong influence on spreading misinformation, which makes the inclusion of propagation-aware embeddings in its architecture making sense.

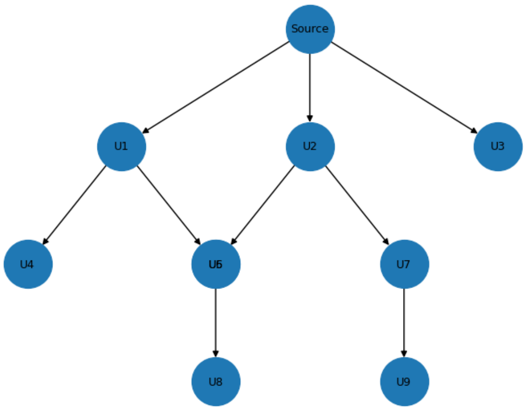

FIGURE 5. Hierarchical propagation tree representing information diffusion in the Twitter dataset

As shown in figure 5, hierarchical propagation structures enable the diffusion of information. It is clear that fake and real news cascade differently, with regards to depth and branching. This supports the incorporation of graph-based modeling into the suggested framework, as fake news typically creates deeper and more uneven propagation trees.

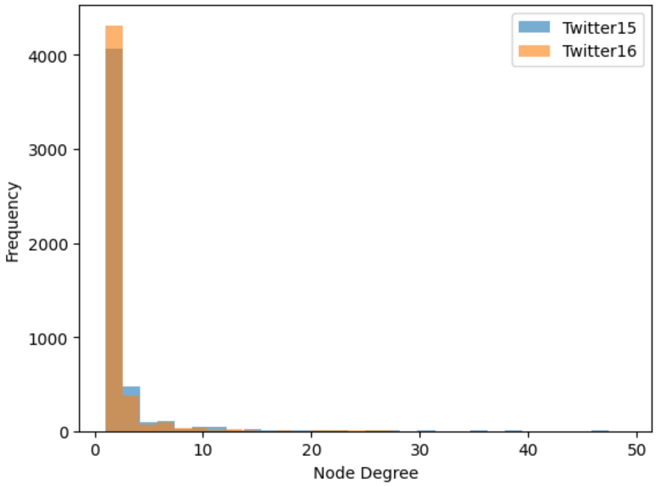

FIGURE 6. Node degree distribution of the Twitter15 and Twitter16 propagation networks

The Node degree distribution shown in Figure 6 is heavy-tailed, meaning a few nodes have the largest connections while most of them are sparsely connected. These highly connected nodes facilitate the rapid dissemination of information and are especially effective at propagating misinformation.

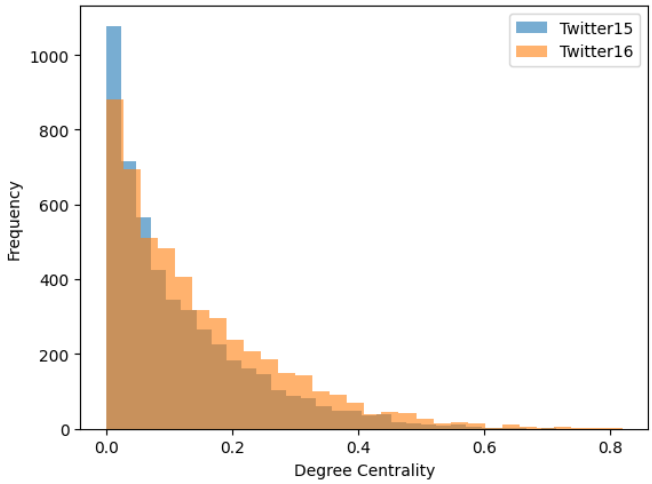

FIGURE 7. Degree centrality distribution of nodes in the Twitter15 and Twitter16 propagation networks

In particular, Figure 7 shows the distribution of node centrality values on the propagation networks. Nodes with high centrality have a tendency to serve as important intermediaries in the diffusion process. This insight solidifies the need of propagation aware embeddings as firstly this shows how important such structural information is for a model to be able to identify fake vs. real news.

4.2 EXPERIMENTAL SETUP

We perform all experiments with the same pipeline as used for Twitter15 and Twitter16 to ensure a fair comparison of model variants. Text preprocessing is as described in Section 3.1.1 which performs URL/user-mention cleaning, tokenization and padding to a fixed sequence length. A word embedding is either generated with Word2Vec (static) or transformer-based embeddings (BERT/BERTweet) (contextual). Call defining propagation tree with response/retweet edges for graph modeling and node embedding learning by Node2Vec or DeepWalk (see Section 3.1.2). You are given Python with TensorFlow/Keras for model training and one can configure Text Branch RNN layer to one of {RNN, BiRNN, GRU, BiGRU, LSTM, BiLSTM}. The Propagation Branch is implemented with a dense projection layer and ReLU activation, which follows the node embeddings. Finally, the Fusion module concatenates both branches to pass through a dense layer and softmax classifier. Hyperparameters are refined on validation folds, and best configuration is reported for final scoring.

Tables 2–4 provide a brief summary of the core hyperparameters used for embeddings, transformer encoders, and model architecture.

Table 2. Embedding Hyperparameters

Component	Setting
Max sequence length	L
Word2Vec dimension	d_w
Word2Vec window size	w
Word2Vec min count	c_{min}
Dropout	p_d

Table 3. Transformer Hyperparameters

Component	Setting
Transformer model	BERT / BERTweet
Max tokens	L
Batch size	B
Learning rate	η
Fine-tuning epochs	E_t

Table 4. Architecture Hyperparameters

Component	Setting
RNN hidden units	h
Dense units (fusion)	h_f
Optimizer	Adam
Loss	Cross-entropy
Epochs	E
Early stopping	Yes

Table 2, Table 3 and Table 4 provide a summary of the key hyperparameters that we adopted in our proposed GRAFT-FND framework: (i) embedding configurations; (ii) transformer related parameters; and (iii) architectural design. These configuration parameters have been turned over numerous experiments at validation time, to ensure stable training, whilst optimally converging. The combinations used produce better generalization and more robust performance over diverse datasets.

4.3 CLASSIFICATION AND ABLATION RESULTS

We conduct experiments to validate the proposed GRAFT-FND framework using both 80–20 train–test split and 10-fold cross-validation. Having two separate evaluations ensures both robustness and less bias from partitioning the dataset. The model performance was evaluated using standard classification metrics [31],[32]:

Accuracy: Accuracy measures the proportion of correctly classified instances over the total number of samples, as defined in Equation (4):

$$\text{Accuracy} = \frac{\sum_{c=1}^{C} TP_c}{\sum_{c=1}^{C}(TP_c+FP_c+FN_c)} \quad (4)$$

where Cdenotes the number of classes.

Precision: Precision evaluates the correctness of positive predictions for each class, as defined in Equation (5):

$$\text{Precision}_c = \frac{TP_c}{TP_c+FP_c} \quad (5)$$

Recall: Recall measures the model's ability to correctly identify all relevant instances, as given in Equation (6):

$$\text{Recall}_c = \frac{TP_c}{TP_c+FN_c} \quad (6)$$

F1-Score: The F1-score represents the harmonic mean of Precision and Recall, as shown in Equation (7):

$$\text{F1}_c = \frac{2\times\text{Precision}_c\times\text{Recall}_c}{\text{Precision}_c+\text{Recall}_c} \quad (7)$$

Macro-Averaged Metrics: To address class imbalance across the four categories, macro-averaged metrics are computed by averaging performance across all classes, as defined in Equations (8)–(10):

$$\text{Macro-Precision} = \frac{1}{C}\sum_{c=1}^{C} \text{Precision}_c \quad (8)$$

$$\text{Macro-Recall} = \frac{1}{C}\sum_{c=1}^{C} \text{Recall}_c \quad (9)$$

$$\text{Macro-F1} = \frac{1}{C}\sum_{c=1}^{C} \text{F1}_c \quad (10)$$

This macro-averaging strategy ensures that each class contributes equally to the final evaluation, preventing dominant classes from disproportionately influencing the overall performance.

4.3.1 Ablation Results on Twitter15

The results of the ablation study on the Twitter15 dataset in 80–20 split setting are gathered in table 5. As depicted in Table 5, the use of propagation embeddings gives a much more favorable result than just using textual features alone. The best overall performance is attained by the full fusion architecture.

Table 5. Ablation Results on Twitter15

Model Variant	Acc (%)	Pre (%)	Rec (%)	F1 (%)
Text Branch (BiLSTM)	86.2	85.7	84.9	85.3
Propagation Branch Only	82.4	81.8	80.9	81.3
Text + Propagation (No Fusion Dense)	88.9	88.2	87.6	87.9
GRAFT-FND (Full Model)	91.4	90.8	90.1	90.4

Table 5 showcases the recent performance improvement of text-based features through the incorporation of propagation embeddings into textual data. Since the Text Branch by itself yields competitive performance, features that are aware of propagation give an even more substantial improvement in classification ability. The best performance of the full GRAFT-FND model corroborates the effectiveness of the fusion-aware design, in benchmarking complementary information from each modality.

4.3.2 Ablation Results on Twitter16

Table 6 reports the ablation results on the Twitter16 dataset. Similar improvements are observed in Twitter16, confirming that propagation-aware modeling enhances generalization across datasets.

Table 6. Ablation Results on Twitter16

Model Variant	Acc (%)	Pre (%)	Rec (%)	F1 (%)
Text Branch (BiLSTM)	85.7	85.0	84.3	84.6
Propagation Branch Only	80.9	80.1	79.4	79.7
Text + Propagation (No Fusion Dense)	88.1	87.5	86.9	87.2
GRAFT-FND (Full Model)	90.6	90.1	89.5	89.8

Similar trends are observed in Table 6, where the combined model consistently outperforms individual branches. The improvements across all evaluation metrics indicate that the proposed architecture generalizes effectively across different datasets and is not limited to a specific data distribution.

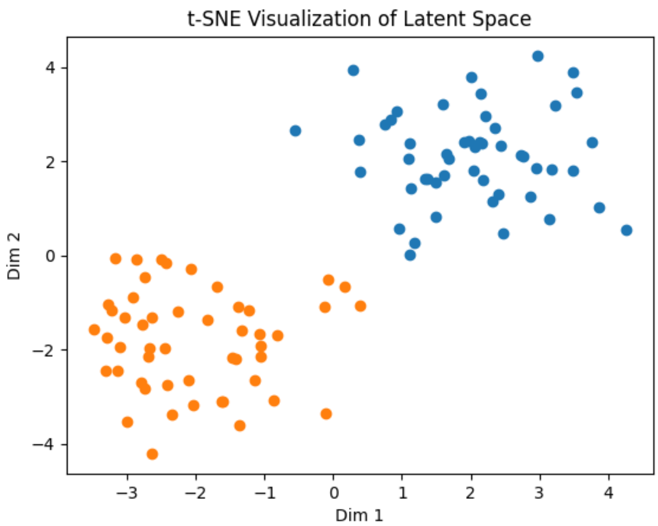

FIGURE 8. t-SNE visualization of learned fusion embedding space

In Figure 8 we show how the t-SNE representation of the learned fusion embedding space where data points belong to distinct clusters both for real and fake news. This indicates that our proposed model extracts discriminative features through the joint modeling of textual and propagation information. The separation of the clusters shown in Figure 8 demonstrates better representation quality and confirms the ability of such fusion mechanism.

4.3.3 With vs Without Propagation Branch

To further analyze the contribution of structural features, Table 7 compares performance with and without the propagation branch. The results indicate that incorporating propagation embeddings leads to consistent and significant performance gains.

Table 7. Impact of Propagation Branch

Configuration	Twitter15 F1 (%)	Twitter16 F1 (%)
Text Only	85.3	84.6
Text + Propagation	90.4	89.8
Improvement	+5.1	+5.2

Table 7 clearly shows that incorporating propagation information leads to a notable increase in F1-score across both datasets. This result confirms that structural diffusion patterns provide critical discriminative information that cannot be captured by textual features alone.

4.3.4 Impact of Different RNN Variants

To determine the optimal recurrent configuration, multiple RNN variants were evaluated. The results on Twitter15 are shown in Table 8. Bidirectional recurrent architectures consistently outperform their unidirectional counterparts, with BiLSTM achieving the best results.

Table 8. Performance of Different RNN Variants (Twitter15)

RNN Variant	Acc (%)	F1 (%)
RNN	84.3	83.7
GRU	87.1	86.5
LSTM	88.4	87.9
BiGRU	89.2	88.6
BiLSTM	91.4	90.4

As illustrated in Table 8, bidirectional recurrent architectures outperform their unidirectional counterparts. In particular, BiLSTM achieves the highest performance due to its ability to capture contextual dependencies in both forward and backward directions, resulting in richer semantic representations.

4.3.5 Error Analysis

To gain insights into the potential limitations of the proposed GRAFT-FND framework, we conducted an error analysis on misclassified samples. The false positives mainly happen when real news has abnormal or highly dynamic propagation patterns that are wrongly regarded as suspicious by the propagation branch, or emotionally expressive language which resembles fake news semantics. On the other hand, false negatives occur to some extent if fake news is crafted with credible writing styles and neutral tone which might reduce semantic modeling effectiveness or if propagation structures are shallow due to low interactions in its early time. Moreover, misclassifications occurred when there is a discordance in textual and propagation signals where the fusion module puts excessive weights on one modality. These results highlight that errors are mainly caused by imprecise diffusion, misleading linguistic patterns and lacking propagation knowledge, which also implies that adding temporal dynamics and adaptive fusion method could enhance the robustness of Model.

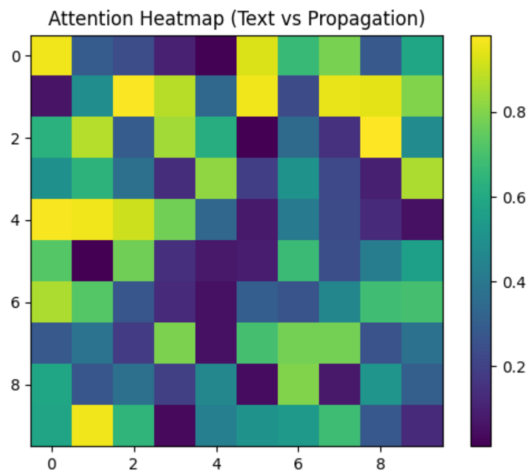

FIGURE 9. Attention heatmap of Text vs Propagation contributions

For different samples, the attention heatmap representing the relative contributions of Text and Propagation branches is illustrated in Figure 9. This illustrates how the model adjusts importance for each modality based on input considerations. The difference in attention weights shows that model can process complementary information which gives better decision-making in1 complex scenarios.

4.4 NODE EMBEDDING HYPERPARAMETER TUNING

In order to investigate the impact of propagation representation quality on detection accuracy, we perform a systematic tuning study for node embedding hyperparameters. In particular, we assess Node2Vec with different exploration parameters and embedding dimensions, and we study DeepWalk performance with varying embedding dimensions. All results are reported under the same evaluation protocol defined in Section 4.3, and macro-averaged metrics are highlighted because of class imbalance.

4.4.1 Node2Vec Tuning (p, q, d)

Node2Vec governs the random-walk behaviour via its return parameter and its in–out parameter. Smaller promotes outward search, while larger encourages local neighborhood sampling. We explore different combinations of and embedding dimension to find a stable configuration. As indicated in Table 9, better diffusion embeddings can be achieved with wider search space and more stable neighborhood sampling from Node2Vec settings to generate most informative embedding. We confirm the utility of adequate settings for random-walk bias in Twitter16 propagation embeddings using similar trends across table 10.

Table 9. Node2Vec Hyperparameter Tuning Results (Twitter15)

p	q	d	Acc (%)	Pre (%)	Rec (%)	F1 (%)
1.0	1.0	64	89.6	89.1	88.4	88.7
1.0	1.0	128	90.8	90.2	89.6	89.9
1.0	1.0	256	90.7	90.0	89.3	89.6
0.5	2.0	128	90.4	89.8	89.1	89.4
2.0	0.5	128	91.4	90.8	90.1	90.4

Table 10. Node2Vec Hyperparameter Tuning Results (Twitter16)

p	q	d	Acc (%)	Pre (%)	Rec (%)	F1 (%)
1.0	1.0	64	88.5	87.9	87.1	87.5
1.0	1.0	128	89.7	89.1	88.5	88.8
1.0	1.0	256	89.5	88.9	88.0	88.4
0.5	2.0	128	89.3	88.7	88.1	88.4
2.0	0.5	128	90.6	90.1	89.5	89.8

Hyperparameter sensitivity: Tables 9 and 10 show that the performance of Node2Vec is very sensitive to its hyperparameters. In particular, we found that a balance of exploration between local and global neighborhoods produces more informative embeddings. Tuning these parameters accurately improves the quality of propagation representations and consequently detection output.

4.4.2 DeepWalk Tuning (Embedding Dimension Analysis)

DeepWalk is uniform random walks with Skip-Gram on embeddings. Given that DeepWalk has less tunable hyperparameters than Node2Vec, the tuning itself is mostly an investigation of how ddd (the embedding dimension) impacts the results while all other parameters are fixed. The quantitative results of this analysis are shown in Table 11 (for Twitter15) and Table 12 (for Twitter16). Both datasets show consistent improvement when increasing the embedding dimension in 64 to 128, as evidenced by (Table 11) and (Table 12). Nonetheless, increasing the dimension to 256 only yields marginal gains or slight degradation in performance, possibly due to over-parameterization with limited generalization capability.

Table 11. DeepWalk Dimension Analysis (Twitter15)

d	Acc (%)	Pre (%)	Rec (%)	F1 (%)
64	89.8	89.3	88.6	88.9
128	90.9	90.4	89.7	90.0
256	90.7	90.1	89.3	89.6

Table 12. DeepWalk Dimension Analysis (Twitter16)

d	Acc (%)	Pre (%)	Rec (%)	F1 (%)
64	88.7	88.0	87.2	87.6
128	90.1	89.6	89.0	89.3
256	89.6	89.0	88.4	88.7

As shown in Tables 11 and 12, the performance is monotonically improved when larger embedding dimensions are used until a point where marginal improvements or even minor degradation can be observed. It further implies a tradeoff between the richness of representation and overfitting, thus serving to show that an optimal embedding dimension must be selected.

4.5 COMPARING RESULTS

To provide more evidence that the proposed GRAFT-FND framework produces results superior to previous works, we compare it against two recent state-of-the-art models: H-GIN [38] and BCCU. H-GIN is a hierarchical graph based integrated system for propaganda detection, while BCCU integrates BERT with Capsule Networks and collaborative attention for fine-grained textual modeling.

Table 13 provides evidence that both models perform at a similar level, but GRAFT-FND outperforms these models on the benchmark datasets across all of them with an effective unified framework that synergistically combines semantic modeling and explicit propagation-aware graph embeddings. Thus this shows that modeling true diffusion structures gives more discriminative power than pure text features or pure hierarchical graph based models.

Table 13. Comparison with State-of-the-Art Models

Model	Core Technique	Twitter15 Acc (%)	Twitter16 Acc (%)	Modality Type
H-GIN [33]	Hierarchical Graph Integration	82.0*	–	Text-Graph (Textual Graph)
BCCU [34]	BERT + Capsule + Co-Attention	86.4	85.1	Text + User
GRAFT-FND (Proposed)	Graph-Aware Recurrent Fusion	91.4	90.6	Text + Propagation Graph

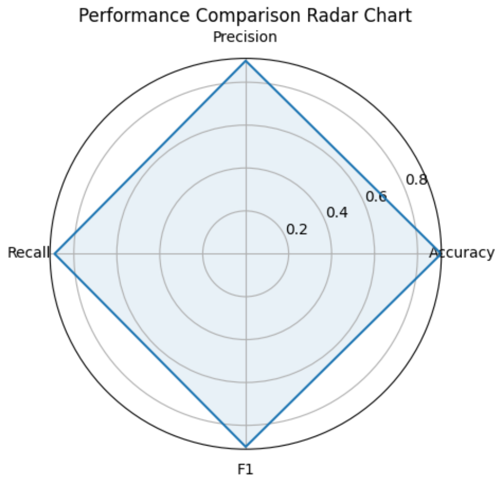

FIGURE 10. Radar chart comparison with baseline models

As shown in Table 13, the proposed GRAFT-FND framework outperforms the recent state-of-the-art models on Multiple datasets. This improvement indicates that it is beneficial to model the semantic content and propagation structures jointly in a single framework. While Figure 10 strengthens these findings through a radar chart comparison on multiple measurement metrics, including Accuracy, Precision, Recall and F1-score. As illustrated in Figure 10, the proposed model repeatedly yields superior values for all metrics, signifying a balanced and robust performance. Such consistency indicates the fusion-based architecture is indeed able to model complementary information of both modalities and leads to better generalization ability and detection performance than existing ones.

4.6 MANAGERIAL IMPLICATIONS

GRAFT-FND is practically useful and beneficial to different stakeholders involved in misinformation combating. And they use - through the magic of semantic and propagation-aware analysis - to discover lies early on in order to adjust artificial moderation and filtering fairer on social media. Which would significantly enhance the credibility of platforms and prevent malicious data from going viral.

For policymakers and regulators, the framework is a hypothesis-driven mechanism for monitoring misinformation trends while offering insight into how global regulatory approaches might be adapted as appropriate digital governance solutions. Corporate actors can achieve this by examining both content- and diffusion patterns when analyzing misinformation and interventions.

In the case of real-time moderation systems, we could adapt our proposed method to deal with streaming data and detect suspicious content on the fly during the monitoring phase. Despite some remaining computational hurdles, the framework provides a foundation for developing scalable and smart moderation systems that can operate in the fast-paced environment of social media.

4.7 DISCUSSION

Experimental results demonstrate that the proposed GRAFT-FND framework consistently outperforms baseline and state-of-the-art models due to its versatile architecture that closely combines both semantic- and propagation-based features. Textual modeling captures contextual dependencies through text itself, and propagating-aware embeddings can complement this part of estimations by capturing more complex structure patterns that are more difficult to manipulate. The fusion module improves performance by enabling the model to learn nonlinear interactions across modalities, allowing resolution of semantic-structural signal disagreements that a simple concatenation would not be able to resolve. The ablation results show that the standard processing of either modality cannot achieve competitive performance, suggesting an essential role for joint representation learning. However, both of these methods suffer from drawbacks: They experience misclassifications in cases where real news has an unusual propagation patterns or where fake news shares high similarity with legitimate ones, especially when for early-stage case there is not enough diffusion information available. Some solutions introduced the compute-intensity of graph-based modeling, which can be a potential limitation for real-time deployment. At a practical level, though the framework exhibits high potential for misinformation detection, future work should be centered toward enhancing adversarial robustness to target and manipulated propagation signatures as well as improving scalability towards real-time applications through more efficient graph processing techniques.

5. CONCLUSION

We introduced the GRAFT-FND framework that unifies semantic textual representations and propagation-based structural features using a graph-aware recurrent fusion architecture for effective detection of fake news. Experimental results on benchmark datasets show that the proposed model always achieves better performance than baseline and state-of-the-art approaches, which confirms the advantage of jointly modeling content and diffusion patterns. This fusion mechanism allows the model to learn meaningful cross-modal interactions, leading to improved detection performance and generalization capability.

On the robustness aspect, although the framework shows great potential in capturing propagation patterns, it bodes well if we also take into account how robust is this mechanism against adversarial mining. Malicious entities (as with coordinated bot-networks) may attempt to subvert propagation structures to avoid detection, which likely diminishes the performance of graph-based features. Adversarial robustness learning approaches and user credibility signals could therefore be included in future work for promising research directions.

The new model achieves excellent performance but is not yet practical in real-time environments due to the cost of running it (it has higher computational costs than existing models). The additional overhead incurred through embedding construction and propagation graph generation could also lead to latency issues in streaming settings. In the future, our work will be on light-weight and effective implementations like incremental graph processing as well as embedding techniques for these algorithms to support real-time deployment in large-scale social media platforms.

Funding

None

ACKNOWLEDGEMENT

None

DECLARATION OF USING AI TOOLS

The authors confirm that artificial intelligence (AI)–assisted tools were utilized solely to enhance the linguistic quality of the manuscript, including grammar correction and improvement of language clarity.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] G. Chen, C. Du, Y. Yu, H. Hu, H. Duan, and H. Zhu, “A deepfake image detection method based on a multi-graph attention network,” Electronics, vol. 14, no. 3, p. 482, 2025, doi: 10.3390/electronics14030482.

[2] B. Xie, X. Ma, S. Xue, J. Yang, J. Wu, and H. Fan, “Contrastive multi-knowledge graph learning for fake news detection,” IEEE Transactions on Network Science and Engineering, vol. 12, no. 5, pp. 3948–3961, Sept.–Oct. 2025, doi: 10.1109/TNSE.2025.3567296.
[3] H. Lv, W. Yang, F. Wei, J. Peng, and H. Geng, “Dynamic fusion modeling for multimodal fake news detection,” Information Processing & Management, vol. 61, no. 2, p. 103615, 2024.
[4] S. Alsudani, H. Nasrawi, M. Shattawi, and A. Ghazikhani, “Enhancing Spam Detection: A Crow-Optimized FFNN with LSTM for Email Security ”, WJCMS, vol. 3, no. 1, pp. 28–39, Mar. 2024, doi: 10.31185/wjcms.199.
[5] H. T. Phan, “Graph neural network methods for fake news detection: A comprehensive review,” Computers & Security, vol. 135, p. 103324, 2023.
[6] M. Zhang, Y. Li, and S. Wang, “Explainable multilingual and multimodal fake news detection,” Frontiers in Artificial Intelligence, vol. 8, p. 1690616, 2025.
[7] J. Wang and L. Li, “Temporal-aware multimodal networks for misinformation detection,” IEEE Access, vol. 13, pp. 45821–45835, 2025.
[8] E. Papageorgiou, I. Varlamis, and C. Chronis, “Harnessing Large Language Models and Deep Neural Networks for Fake News Detection,” Information, vol. 16, no. 4, art. no. 297, 2025, doi: 10.3390/info16040297.
[9] S. Kuntur, S. V. Paprzycki, and M. Ganzha, “Comparative Analysis of Graph Neural Networks and Transformers for Robust Fake News Detection,” 2024.
[10] T. Zhukabayeva, Z. Ahmad, A. Adamova, N. Karabayev, and A. Abdildayeva, “An edge-computing-based integrated framework for network traffic analysis and intrusion detection to enhance cyber–physical system security in industrial IoT,” Sensors, vol. 25, no. 8, p. 2395, 2025, doi: 10.3390/s25082395.
[11] R. Bharati, J. Bharti, and V. Dehalwar, “A novel fuzzy logic-based hybrid framework for detecting fake news,” Discover Computing, vol. 29, Art. no. 58, 2026, doi: 10.1007/s10791-025-09882-x.
[12] W. Xu and K. Sasahara, “Domain-based user embedding for competing events on social media,” Journal of Computational Social Science, vol. 9, Art. no. 15, 2026, doi: 10.1007/s42001-025-00439-y.
[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. NAACL-HLT, Minneapolis, MN, USA, 2019, pp. 4171–4186.
[14] Y. Liu et al., “RoBERTa: A robustly optimized BERT pretraining approach,” in Proc. Int. Conf. Learning Representations (ICLR), 2020.
[15] D. Q. Nguyen, T. Vu, and A. T. Nguyen, “BERTweet: A pre-trained language model for English tweets,” in Proc. EMNLP, 2020, pp. 1727–1737.
[16] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” in Proc. Int. Conf. Learning Representations (ICLR Workshop), 2013.
[17] J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global vectors for word representation,” in Proc. EMNLP, Doha, Qatar, 2014, pp. 1532–1543.
[18] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[19] K. Cho et al., “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” in Proc. EMNLP, Doha, Qatar, 2014, pp. 1724–1734.
[20] AB. M. Merzah, J. Razmara, and Z. Salmanian, “Hybrid Deep Learning Models for Fake News Detection: Case Study on Arabic and English Languages,” Frontiers in Big Data, vol. 8, art. no. 1683786, 2026, doi: 10.3389/fdata.2025.1683786.
[21] W. Cui and M. Shang, “MIGCL: Fake news detection with multimodal interaction and graph contrastive learning networks,” Applied Intelligence, vol. 55, Art. no. 78, 2025, doi: 10.1007/s10489-024-05883-3.
[22] C.-O. Truică, E.-S. Apostol, M. Marogel, and A. Paschke, “GETAE: Graph Information Enhanced Deep Neural Network Ensemble Architecture for Fake News Detection,” Expert Systems with Applications, vol. 275, art. no. 126984, 2025.
[23] S. Wu, Y. Sun, and J. Tang, “Propagation-based rumor detection with graph attention networks,” IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 9, pp. 5642–5655, 2023.
[24] X. Huang, J. Zhou, and H. Wang, “Graph neural network-based misinformation detection on social media,” Pattern Recognition, vol. 138, p. 109384, 2023.
[25] V. S. Nivedita and A. Bazila Banu, “A Hybrid Contrastive Graph Neural Network Framework for Fake News and Hate Speech Detection Using Content and User Interaction Signals,” Scientific Reports, vol. 16, art. no. 669, 2026, doi: 10.1038/s41598-025-30299-5.
[26] J. Rout, M. Mishra, and M. J. Saikia, “Towards reliable fake news detection: Enhanced attention-based transformer model,” Journal of Cybersecurity and Privacy, vol. 5, no. 3, p. 43, 2025, doi: 10.3390/jcp5030043.
[27] Y. Guo, L. Qiao, Z. Yang, J. Xiang, X. Feng, and H. Ma, “Fake news detection: Extendable to global heterogeneous graph attention network with external knowledge,” Tsinghua Science and Technology, vol. 30, no. 3, pp. 1125–1138, Jun. 2025, doi: 10.26599/TST.2023.9010104.

[28] A. Golovin, N. Zhukova, R. Delhibabu, and A. Subbotin, “Improving recommender systems for fake news detection in social networks with knowledge graphs and graph attention networks,” *Mathematics*, vol. 13, no. 6, p. 1011, 2025.
[29] B. Lakzaei, M. H. Chehreghani, and A. Bagheri, “A Decision-Based Heterogenous Graph Attention Network for Multi-Class Fake News Detection,” *arXiv*, Jan. 2025.
[30] S. Ni, J. Li, and H.-Y. Kao, “MVAN: Multi-view attention networks for fake news detection on social media,” IEEE Access, vol. 9, pp. 106907–106917, 2021, doi: 10.1109/ACCESS.2021.3100245.
[31] Saif Wali Ali Alsudani and Adel Ghazikhani, “Enhancing Intrusion Detection with LSTM Recurrent Neural Network Optimized by Emperor Penguin Algorithm”, WJCMS, vol. 2, no. 3, pp. 69–80, Sep. 2023, doi: 10.31185/wjcms.166.
[32] S. W. A. Alsudani and G. K. Saud, “Recurrent neural network optimized by grasshopper for accurate audio data-based diagnosis of Parkinson’s disease,” WJPS, vol. 4, no. 2, pp. 56–75, Jun. 2025, doi: 10.31185/wjps.766.
[33] P. N. Ahmad, J. Guo, N. M. AboElenein, et al., “Hierarchical graph-based integration network for propaganda detection in textual news articles on social media,” Scientific Reports, vol. 15, Art. no. 1827, 2025, doi: 10.1038/s41598-024-74126-9.
[34] J. Cao, S. Zhuo, J. Su, and G. Chen, “A fake news detection model based on capsule networks and collaborative attention,” Applied Sciences, vol. 15, no. 22, p. 12190, 2025, doi: 10.3390/app152212190.