MSynFD: Multi-hop Syntax Aware Fake News Detection

Liang Xiao
Beijing Institute of Technology
School of Computer Science
Beijing, China
patrickxiao@bit.edu.cn

Qi Zhang
Tongji University
School of Computer Science
Shanghai, China
zhangqi_cs@tongji.edu.cn

Chongyang Shi*
Beijing Institute of Technology
School of Computer Science
Beijing, China
cy_shi@bit.edu.cn

Shoujin Wang
University of Technology Sydney
School of Computer Science
Sydney, Australia
shoujin.wang@uts.edu.au

Usman Naseem
Macquarie University
School of Computing
Sydney, Australia
usman.naseem@mq.edu.au

Liang Hu
Tongji University
School of Computer Science
Shanghai, China
lianghu@tongji.edu.cn

ABSTRACT

The proliferation of social media platforms has fueled the rapid dissemination of fake news, posing threats to our real-life society. Existing methods use multimodal data or contextual information to enhance the detection of fake news by analyzing news content and/or its social context. However, these methods often overlook essential textual news content (articles) and heavily rely on sequential modeling and global attention to extract semantic information. These existing methods fail to handle the complex, subtle twists[1] in news articles, such as syntax-semantics mismatches and prior biases, leading to lower performance and potential failure when modalities or social context are missing. To bridge these significant gaps, we propose a novel **m**ulti-hop **s**yntax aware **f**ake **n**ews **d**etection (MSynFD) method, which incorporates complementary syntax information to deal with subtle twists in fake news. Specifically, we introduce a syntactical dependency graph and design a multi-hop subgraph aggregation mechanism to capture multi-hop syntax. It extends the effect of word perception, leading to effective noise filtering and adjacent relation enhancement. Subsequently, a sequential relative position-aware Transformer is designed to capture the sequential information, together with an elaborate keyword debiasing module to mitigate the prior bias. Extensive experimental results on two public benchmark datasets verify the effectiveness and superior performance of our proposed MSynFD over state-of-the-art detection models.

CCS CONCEPTS

• **Computing methodologies → Artificial intelligence**; **Neural networks**.

---

*Corresponding author.

[1]A "subtle twist" refers to a slight, inconspicuous, or nuanced change or alteration that is unexpected and not immediately apparent.

KEYWORDS

Fake News Detection, Graph Neural Network, Debiasing

1 INTRODUCTION

The explosion of news consumption and sharing on social media platforms has created an unprecedented environment for the rapid dissemination of fake news. With the ease and speed at which information can be shared online, false narratives and misleading content can quickly gain attraction and reach a wide range of audiences. This proliferation of fake news poses a significant risk to society as it has the potential to manipulate public opinions, distort facts, and undermine trust in credible sources of information [13, 14]. Recognizing this issue, there is a growing recognition of the urgent need to address the challenge of detecting fake news [56].

With the impressive advancements in deep learning, deep neural networks have gained widespread adoption in fake news detection in recent years. Various advanced neural models have been explored for fake news detection, including Recurrent Neural Networks (RNN) [19], Convolutional Neural Networks (CNN) [43, 52], attention networks [26, 51], and Graph Neural Networks (GNN) [37, 56]. These models leverage news texts or visual content and contextual information to identify the distinguishing features of fake news, yielding impressive detection performance. While the integration of multimodal information and social context has proven beneficial for detecting fake news, approaches relying heavily on visual and contextual cues suffer from the absence of such modalities or context, thus limiting their practicality in real-life scenarios. Consequently, text-based approaches have attracted significant attention as they primarily rely on news text, serving as the most crucial source of information in various fake news detection models.

Prevalent text-based detection approaches primarily revolve around RNN-based [8, 36], CNN-based [22, 28], and attention-based methods [9, 36, 51], which are inclined to capture comprehensive semantic correlations. However, these existing methods often lead to the acquisition of irrelevant information or word associations,
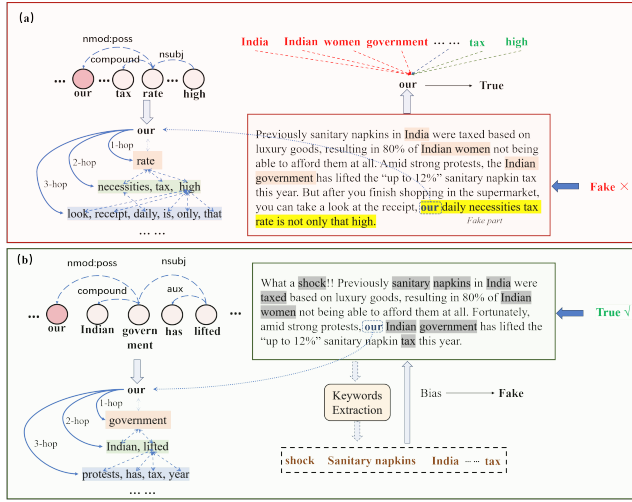
Figure 1: (a) A fake news example with misleading information is highlighted in yellow. The word correlations above show how irrelevant words affect the understanding of the center word 'our', then mislead the detection result; (b) A true news example including keywords marked in grey and words leading to potential prior bias list below. The left region of both (a) and (b) shows syntax-associated words towards the center word 'our' at the 3-hops case and the local structure of the syntactic dependency tree.

presenting limitations when detecting fake news with subtle twists. Such kind of fake news articles often contain mostly true information but introduce false details through slight reversals or comparisons. As illustrated in Figure 1(a), since most of the news content is about India, it is misleading that 'our' refers to 'India', which causes the misunderstanding of the entire news segment. Such syntax-semantics mismatch, e.g., referential transfer, easily deceives and degrades the aforementioned semantic-targeted models.

Additionally, it is crucial to address the presence of prior biases towards specific words, which has often been overlooked in previous methods. These biases arise from the statistical tendencies of neural models towards historical data and can result in an unfair viewpoint [10, 45, 55, 58], leading to misclassification of news articles, particularly those containing fake news [11]. Figure 1(b) illustrates this issue, where preconceived notions about the emotional word "shock" and the entity word "India" can easily influence interpretation and judgment, potentially leading to the misidentification of genuine news as fake news. Zhu et al. [58] first introduced causal learning to mitigate entity bias in fake news detection, explicitly improving the generalization ability of detectors to future news data. However, we recognize that these prior biases primarily originate not only from key entities in news articles but also from significant contextual indicators such as emotional words like "shocks" in Figure 1(b). Since fake news often exhibits distinctive writing styles [59], characterized by exaggeration or extreme stances, it becomes imperative to adaptively learn and mitigate biases towards specific words rather than focusing on entity words.

To tackle the aforementioned challenges, a practical solution is to incorporate a syntactical dependency graph as supplementary information to enhance semantic learning and facilitate debiasing.

However, modeling such syntactical dependency graphs presents three critical issues that need to be tackled: 1) Insufficient information from adjacent perception: The structure of adjacent perception may not provide enough contextual information. 2) Noisy information from imperfect parsing performance: Imperfect parsing can introduce noisy information into the syntactical dependency graph. 3) Lack of sequential information in syntactical dependency graphs [35]: Syntactical dependency graphs inherently lack sequential information. These issues pose significant challenges when it comes to effectively incorporating syntax analysis to address syntax-semantics mismatch and mitigate prior biases.

In light of the above discussion, we present a novel approach called Multi-hop Syntax aware Fake News Detection (MSynFD) that leverages the information provided by a syntactical dependency graph among news pieces. To address the limited perception range, we introduce the Subgraph Aggregation Attention (SAA) module. The module employs a syntactical multi-hop subgraph aggregation mechanism to extend the perception range of words, enabling capturing more comprehensive information about hierarchical syntactic structures. To tackle noisy information, we incorporate an adaptive gating mechanism into the SAA module to filter out noisy structural information, maintaining more relevant and reliable information. Recognizing the reliability of direct relations, we further introduce a graph relative position bias mechanism that emphasizes the significance of low-hop relations. Furthermore, to tackle the lack of sequential information, we devise a sequential relative position-aware Transformer to capture sequential information for complementing the syntactical dependency graph. Our proposed Transformer seamlessly integrates with the SAA module, improving the interpretation and detection of fake news. Extensive experiments on public datasets verify the effectiveness and state-of-the-art performance of our detection method.

The main contributions of this paper are as follows:

- We propose a novel multi-hop syntax-aware fake news detection model, named MSynFD, to deal with fake news with subtle twists, effectively tackling syntax-semantics mismatch and mitigating prior biases in news articles.
- We design a multi-hop subgraph aggregation mechanism to capture comprehensive syntactic information, seamlessly integrating with a relative position-aware Transformer.
- We design a keywords-based debiasing to mitigate the preconceived notion within the news piece.

2 RELATED WORK

2.1 Fake News Detection

Fake news detection is conventionally framed as a binary classification task. This task can be broadly categorized into two main approaches: social-context-based and content-based [32].

Social-Context-Based Detection: Social-context-based methods revolve around the dynamics of news dissemination. Representative methods include 1) News dissemination-based approaches, which use GNN-based methods to model social interactions between users, news, and media sources [23, 33, 44, 56]; 2) User credibility-based approaches, which prioritize assessing the credibility of users and news sources in the context of fake news dissemination [1, 16]; 3) Feedback-based approaches, which rely on the user actions, e.g., comment [30, 57] and preference [5, 42].
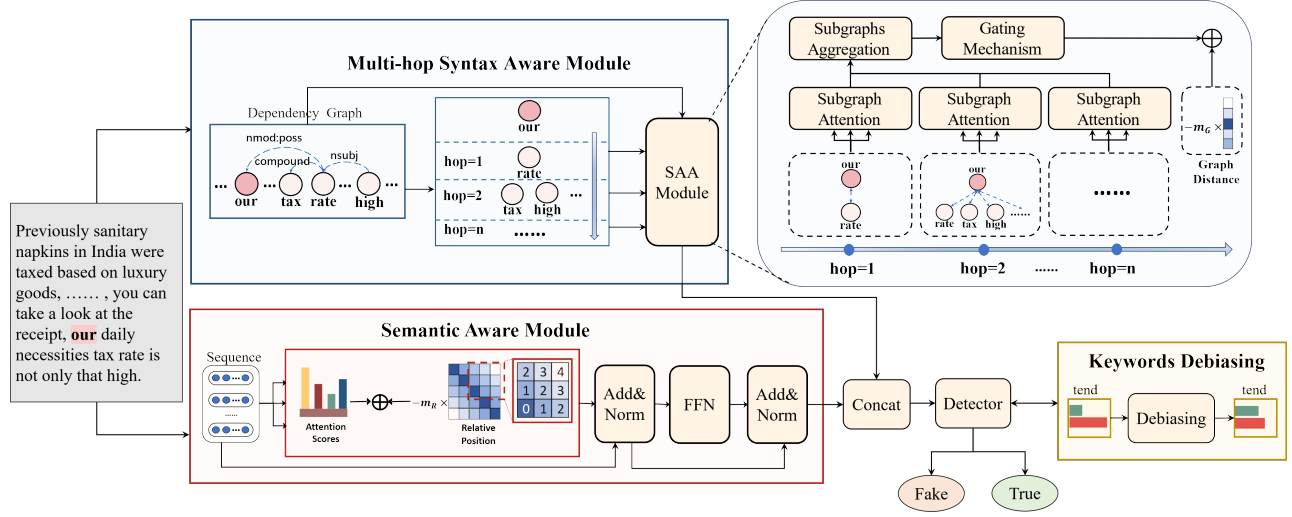
Figure 2: Overview of our MSynFD fake news detection method.

Content-Based Detection: Content-based methods are grounded in analyzing news content, incorporating text, visuals, and additional information to detect fake news. In the early stages, this analysis primarily relied on manual extraction of content, thematic elements, and user-related information, Detection techniques included machine learning models, including Decision Tree [2]and SVM [49]. More recently, deep learning models have achieved exceptional performance in the detection of fake news across various forms, including both unimodal text and textual-graphical multimodal data. For instance, RNN-based [8, 19, 20] methods leverage the sequential nature of textual data, while CNN-based [22, 43, 52] methods borrow from convolution concepts in computer vision to extract textual features. Attention-based [20, 26, 36, 41, 51] methods, which are particularly popular, utilizing attention mechanism [39] to capture relations within or between text from a global perspective. GNN-based methods focus on textual graph construction within documents[38] or the syntactical dependency relation between words [18, 34]. Additionally, methods using external factual verification [15, 48, 53] contribute to enhanced detection performance.

Both content-based and social-context-based approaches necessitate effective text content modeling for node encoding. Moreover, since irrelevant connections caused by RNN-based, CNN-based, and Attention-based methods could bring noisy information, syntactical dependency information should be considered introduced in text content modeling. While previous studies have leveraged syntactical dependency graphs, there remains a need for deeper exploration of these graphs to extract more syntactical relations and filter out noisy connections that may introduce irrelevant information. Besides, prior biases are another factor that needs to be considered, as they can impact the generalization capacity of fake news detection[58]. However, little research has been dedicated to understanding and mitigating such biases.

2.2 Graph Neural Networks

In the context of fake news detection, GNN-based methods are predominantly employed in social-context-based approaches for modeling news dissemination and interactions[23, 24, 44, 56]. Nevertheless, GNNs have also demonstrated success in modeling textual

content based on syntactical dependency graphs. These approaches typically entail using GNN-based methods, like GCN [12, 35] and GAT [7, 40], to encode the syntax graph predicted by off-the-shelf dependency parsers, subsequently generating textual graph embeddings tailored to specific tasks, and more recent research focuses on synergizing semantic and syntactical components to complement semantic information[17, 18, 34, 46]. However, GNN-based approaches face limitations. Traditional GNNs struggle with information exchange between non-local neighborhoods when twoword nodes are not in proximity. This challenge arises because the number of layers constrains the traditional approach to message passing, and extending this to larger values leads to overfitting and the loss of critical information [47, 54]. Although strategies like expanding the syntactical dependency graph to a global relation graph [47] and employing the graph spatial encoding [50] have shown promise, they introduce new issues, including an influx of irrelevant information and a lack of perception regarding sub-connected statements. In response, we propose aggregating subgraphs from a global syntactical dependency graph, attempting to enhance the scope of perceived word nodes while filtering out irrelevant information. To the best of our knowledge, this represents a novel contribution to fake news detection.

3 PROBLEM DEFINITION

With a news piece as input, our objective is to determine whether they are fake news based on its textual information. Specifically, each news piece $C= \{P, G, K, Y\}$ consists of the news text P containing n words $P= \{w_1, w_2, \cdots, w_n\}$. The syntactical dependency graph $G= (V, E)$ obtained by HanLP and Stanford CoreNLP tools[2] for Chinese and English news respectively, where V is the set of graph nodes corresponding to the words in P, and E is the set of edges representing the syntactical dependency relations between words. The keywords K are obtained by KeyBERT [6] containing m words $K= \{k_1, k_2, \cdots, k_m\}$, and the ground-truth label $Y \in \{0, 1\}$, where 1 and 0 denote the news piece is fake or true. The purpose of the fake news detection is to predict whether the label C is 1 or 0.

---

[2]https://hanlp.hankcs.com/ and https://stanfordnlp.github.io/CoreNLP
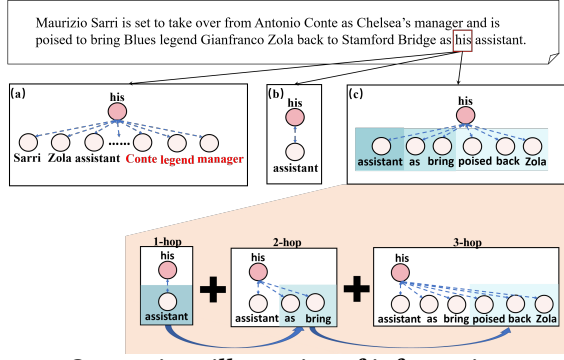
Figure 3: Comparison illustration of information propagation among (a) attention-based methods, (b) traditional GNN-based methods, and (c) our Multi-hop Syntax aware module.

4 METHOD

In this section, we discuss each component of our proposed MSynFD method in detail (as shown in Figure 2).

4.1 Input Encoding

For each news P with n words, i.e., $P = \{w_1, w_2, \cdots, w_n\}$, we feed it into BERT to obtain its representation $\widetilde{P} = \{\widetilde{w}_1, \widetilde{w}_2, \cdots, \widetilde{w}_n\}$. For each word w_i with m tokens $w_i = \{w_{sub1}, w_{sub2}, \cdots, w_{subm}\}$, we obtain its representation by summing the embeddings of its tokens.

4.2 Multi-hop Syntax Aware Module

We introduce the Subgraph Aggregation Attention (SAA) module. It consists of the syntactical multi-hop information aware mechanism and the adaptive gating mechanism and introduces the graph relative position bias. These components collectively capture information between words from the syntactical perspective and, importantly, prevent the formation of irrelevant connections.

When considering a central word, such as "his", as illustrated in Figure 3, the global connection of the attention-based method makes a lot of irrelevant connections, like "Conte" and "manager", brings noisy information to "his". Meanwhile, the information from adjacent words of the traditional GNN-based method often provides insufficient information. For instance, we could know little about "assistant". To address these limitations, multi-hop information becomes crucial for a more accurate understanding. For example, we can ascertain that the "assistant" refers to "Zola" and that he has been "poised" within 3-hop syntactical dependency relations. Accordingly, we have introduced a syntactical multi-hop information-aware mechanism, allowing us to perceive interactions within a range of m-hop. Firstly, we obtain m adjacent matrices, with $A^d \in R^{n \times n}$ representing the d-th hop subgraph from the syntactical dependency graph G. In these matrices, A_{ij}^d is set to 1 if word w_i can be reached from w_j within d words otherwise $A_{ij}^d = 0$. And we set $A_{ii}^d = 1$ for the self-connection, so the adjacency matrix $\widetilde{A}^d$ can be updated to $\widetilde{A}^d = A^d + I$. Note that the adjacent matrix indicates whether two words have a relation instead of the strength of the relation with specific values. Considering the varying word relations derived from different interaction scenarios within specific hop subgraphs, we introduce a hop-specific subgraph attention mechanism to determine the hop-based relation value.

Initially, we transform the news representation $\widetilde{P}$ into word node features $H = \{h_1, h_2, \cdots, h_n\}$ by the linear transformation with trainable parameters W_P, i.e., $H = W_P \widetilde{P}$. To account for the dynamics of word relations under different connections, we employ the hop-specific trainable weight matrix W_A^d, which is used to parameterize every word node. This enables the calculation of an edge weight matrix Z^d for the d-th hop subgraph, where the element z_{ij}^d signifies the relation value between word node i and word node j:

$$z_{ij}^d = LeakyReLU(W_A^d h_i, W_A^d h_j)\widetilde{A}_{ij}^d \tag{1}$$

As the perceived range expands, the potential for irrelevant and noisy information increases, diluting the special local information. To address this, we employ an adaptive adjustment mechanism to measure the importance of information from various subgraphs through a learnable parameter W_Z, allowing the model to balance the information from adjacent relation among subgraphs with varying hops. Denoting the set of multi-hop relation value $Z = [Z^1, Z^2, \cdots, Z^m]$, we have $S = \sigma(W_Z)Z$ where σ is the sigmoid function. To capture and filter the noise, we introduce a gating mechanism using another learnable parameter W_H. This is shared by the word nodes to discern and eliminate the noise, subsequently refreshing the value matrix as $S' = MS$:

$$M = \begin{cases} 1 & if \quad s_{ij} > t_i \\ 0 & else \end{cases} \tag{2}$$

where $T = W_H H$ and $T = [t_1, t_2, \cdots, t_n]$ is a set of adaptive thresholds to word nodes. Furthermore, shorter graph distances between two words indicate stronger relevance. Hence, we introduce a direct use of graph relative position from the global graph structure G, which is used as an attention bias added after the aggregation and filtering processes, enhancing adjacent attention between words within the syntactical structure during the softmax function-based attention calculation mechanism. The output graph representation $\widetilde{H} = \widetilde{S}H$, with $\widetilde{s}_{ij}$ in $\widetilde{S}$ can be defined as:

$$\widetilde{s}_{ij} = \frac{exp(s'_{ij} - m_G|d_{ij}|)}{\sum_{k=1}^{n} exp(s'_{ik} - m_G|d_{ik}|)} \tag{3}$$

where d_{ij} represents the graph's relative distance between word nodes i and j, m_G stands for a head-specific fixed slope. With h heads, the slopes are the geometric sequence: $\frac{1}{2^1}, \frac{1}{2^2}, \cdots, \frac{1}{2^h}$. We adjust the receptive field and filter the noise edges before using the graph relative position bias to ensure that only the relation between nodes' features is used to evaluate the reliability of information transmissions. To stabilize the learning process of the SAA module, the mechanism above is extended to the multi-head form with h heads. After concatenating the outputs from each head, the ultimate graph representation can be obtained after a normalization layer:

$$\widetilde{H} = Norm(concat(\widetilde{H}^{(1)}, \widetilde{H}^{(2)}, \cdots, \widetilde{H}^{(h)})) \tag{4}$$

4.3 Semantic Aware Module

Giving an input news representation $\widetilde{P}$, the information from syntactical structures may be limited, and potential syntactical errors might exist, so the transformer structure is employed to extract semantic information. The objective is to ensure that each word can obtain information from a global perspective while perceiving the

sequential structure. Inspired by the textual positional embedding researches in recent years[25, 27], we introduce a sequential relative position bias, which can be added after query-key dot product to promote higher attention scores between adjacent words in a sequence, leveraging the properties of softmax operator, to emphasize the stronger correlation among closer words. Specifically, for a transformer of multi-head design with h heads, we obtain $Q^{(l)}, K^{(l)}, V^{(l)}$ on the l-th head as the query matrix, key matrix, and value matrix through three distinct linear transformations, and utilize M_R as the sequential relative position matrix. As a result, the semantic representation on the l-th head $R^{(l)}$ can be defined:

$$Q = W_Q \widetilde{P}, \quad K = W_K \widetilde{P}, \quad V = W_V \widetilde{P} + b_V$$

$$R^{(l)} = softmax(\frac{Q^{(l)} K^{(l)T}}{\sqrt{d}} - m_R^{(l)} M_R) V^{(l)} \tag{5}$$

$$r_{ij}^{(l)} = \frac{exp(q_i^{(l)} k_j^{(l)} - m_R^{(l)} |i - j|)}{\sum_{k=1}^{n} exp(q_i^{(l)} k_k^{(l)} - m_R^{(l)} |i - k|)}$$

where W_Q, W_K, W_V, b_V are trainable parameters, $\sqrt{d}$ denotes the scaling factor. m_R is another head-specific fixed slope, equal to m_G in the experiments. We only introduce the trainable bias for $V^{(l)}$, which transforms the sequential relative position into a rigid bias, thereby encouraging the module to focus more on the sequential relation. After connecting the concatenated outputs from each head, a two-layer MLP is employed to extract higher-level semantic features, followed by two normalization layers and the residual structure. Thus, the final semantic representation is obtained:

$$\widetilde{R} = Norm(concat(R^{(1)}, R^{(2)}, \cdots, R^{(h)}) + \widetilde{P})$$
$$\widetilde{R} = Norm(\widetilde{R} + FFN(\widetilde{R})) \tag{6}$$

4.4 Fake News Detector

For each news piece, we possess both the multi-hop graph representation $\widetilde{H}$ and the semantic representation $\widetilde{R}$. These two representations are then concatenated, yielding the fusion representation $\widetilde{F} = concat(\widetilde{R}, \widetilde{H})$. Next, we use a sequence attention mechanism to gather information from each word:

$$F = \sum_{i=1}^{n} softmax(W_{Fi} \widetilde{f_i} + b_{Fi}) \widetilde{f_i} \tag{7}$$

where W_F and b_F are trainable parameters. And in the end, we feed F into a two-layer MLP to get the prediction y':

$$y' = softmax(W_2(ReLU(W_1 F + b_1)) + b_2) \tag{8}$$

where W_1, W_2, b_1, b_2 are trainable parameters.

4.5 Keywords Debiasing

We introduce a keywords debiasing module to mitigate prior bias from keywords. First, we train a simple keyword encoder with a pre-trained BERT to obtain prior keyword representation $K = \{k_1, k_2, \cdots, k_m\}$. Then, we use the maximum pooling to capture the most salient features of each keyword. Next, we train another classification layer to obtain the prediction from keywords y_K:

$$\widetilde{K}_{max} = Maxpool(BERT(K))$$
$$y_K = softmax(W_4(ReLU(W_3 \widetilde{K}_{max} + b_3)) + b_4) \tag{9}$$

Table 1: Statistics of the datasets

Dataset	Weibo			GossipCop		
	Train	Val	Test	Train	Val	Test
Fake	2561	499	754	2024	604	601
Real	7660	1918	2957	5039	1774	1758
Total	10221	2417	3711	7063	2378	2359

where W_3, W_3, b_3, b_4 are trainable parameters.

For the training phase, the final prediction $\hat{y} = \alpha(y') + (1-\alpha)(y_K)$ fusion y' and y_K while α is a hyper-parameter to balance the two terms. We train the whole framework with the cross-entropy loss:

$$\mathcal{L} = \sum_{P, y \in \mathcal{D}} -y log(\hat{y}) - (1-y) log(1-\hat{y})$$
$$+ \beta \sum_{P, y \in \mathcal{D}} -y log(y_K) - (1-y) log(1-y_K) \tag{10}$$

where β is to balance the two loss functions of fusion prediction and keywords-based prediction, and both α and β are set as 0.1 in the experiments. This training procedure encourages the model to focus on and capture the prior keyword bias, allowing the fake news detector to learn less biased information. In the validation and test procedure, we only use y' as the prediction of the model.

5 EXPERIMENTS

5.1 Datasets

We evaluate our MSynFD on two real-world datasets. The Weibo dataset [29] ranging from 2010 to 2018 is used as the Chinese dataset, and the GossipCop data from FakeNewsNet [31] is used as the English dataset. Each news piece is labeled as fake or real in both datasets, and we only use the news content in the experiments. Besides, we keep the same dataset splitting as the organizers provide, where both datasets are segmented in chronological order to simulate real-world scenarios. Detailed statistics of both datasets used in our experiments are shown in Table 1.

5.2 Baselines

We choose nine content-based representative and/or state-of-the-art methods in fake news detection tasks for comparison, including RNN, CNN, GNN, attention, and debiasing models, and unimodal or multi-modal models. Since social-context-based methods focus on modeling information transmission and show high dependence on transmission structure, they are not included as baselines.

Bi-GRU [3] is an RNN-based model that uses a bidirectional GRU network to learn semantic associations within news.

EANN [43] is a multi-modal fake news detection model that uses TextCNN for text representation and uses an adversarial learning method to obtain the invariant features of news.

BERT [4] is a popular pre-training model used for fake news detection. We use the original BERT model for the GossipCop dataset and the Chinese version of BERT for the Weibo Dataset.

MDFEND [21] is a multi-domain-based fake news detection model integrating the Mixture of Experts (MOE).

HMCAN [26] is a multi-modal fake news detection model that designs a hierarchical encoding network to capture the rich hierarchical semantics text information of news.

BERT-Emo [57] is a fake news detection model that combines the emotional features of news content and social contexts.

Table 2: Fake news detection results on the Weibo dataset and the GossipCop dataset. The second best-performing methods are underlined, and ∗ indicates the statistically significant improvement (i.e., two-sided t-test with $p < 0.05$).

Method	Weibo						GossipCop					
	Acc	macF1	AUC	spAUC	F1$_{real}$	F1$_{fake}$	Acc	macF1	AUC	spAUC	F1$_{real}$	F1$_{fake}$
BiGRU	0.8214	0.7172	0.8354	0.6636	0.8887	0.5456	0.8379	0.7730	0.8634	0.7358	0.8943	0.6516
EANN	0.8197	0.7162	0.8276	0.6649	0.8875	0.5448	0.8517	0.7926	0.8765	0.7586	0.9033	0.6820
BERT	0.8474	0.7601	0.8754	0.7102	0.9048	0.6155	0.8439	0.7873	0.8781	0.7579	0.8968	0.6778
MDFEND	0.7786	0.7051	0.8301	0.6691	0.8519	0.5584	0.8518	0.7905	0.8712	0.7543	0.9037	0.6772
HMCAN	0.8289	0.7257	0.8300	0.6674	0.8939	0.5575	0.8490	0.7843	0.8479	0.7386	0.9025	0.6660
BERT-Emo	0.8438	0.7586	0.8743	0.7061	0.9019	0.6154	0.8455	0.7912	0.8800	0.7631	0.8974	0.6849
BERT-Emo-ENDEF	0.8584	0.7731	0.8838	0.7278	0.9121	0.6341	0.8520	0.8010	0.8855	0.7674	0.9020	0.6987
CMMTN	<u>0.8706</u>	<u>0.7812</u>	0.8723	<u>0.7438</u>	<u>0.9211</u>	<u>0.6412</u>	<u>0.8593</u>	<u>0.8117</u>	0.8889	0.7770	0.9064	<u>0.7170</u>
MGIN-AG	0.8666	0.7753	**0.8959**	0.7375	0.9185	0.6320	<u>0.8593</u>	0.8072	<u>0.8916</u>	<u>0.7788</u>	<u>0.9074</u>	0.7069
MSynFD	**0.8787**∗	**0.7889**∗	<u>0.8903</u>	**0.7656**∗	**0.9266**∗	**0.6512**∗	**0.8699**∗	**0.8164**∗	**0.8949**∗	**0.7904**∗	**0.9155**∗	**0.7173**

BERT-Emo-ENDEF [58] is a fake news detection method that introduces an entity debiasing framework (ENDEF) in the BERT-Emo model to mitigate the bias within news pieces.

CMMTN [41] is a multi-modal fake news detection model that uses a masked Transformer to filter the noise or irrelevant context.

MGIN-AG [34] is a multi-modal rumor detection model that uses GCN to generate augmented features from claims, and attention mechanisms to extract the embedded text from images.

Since this work focuses on the textual content of news, all the multi-modal models are kept with their text-only version. For a fair comparison, the labels for the auxiliary event classification task of EANN and the domain labels of MDFEND are derived by clustering according to the publication year; BERT-Emo is a simplified version without the emotion in comments, and MGIN-AG does not use the embedded text in images but use the claim text itself as the replacement. While the results of Bi-GRU, EANN, BERT, MDFEND, BERT-Emo, and BERT-Emo-ENDEF would come from the [58], the remaining models will all use the same training parameters setting, and their classification results will be obtained by the same design of MLP classifier as our proposed MSynFD method, in which the activation function is ReLU and the dimension of hidden layer is set as 384. The heads of any multi-head structure are set to 12, and we report the average testing results over five runs.

5.3 Experimental Settings

Since the Weibo and GossipCop datasets have different average lengths, the maximum sequence lengths of the Weibo and Gossip-Cop datasets are set to 150 and 350, respectively, and the batch size is 32. All models are implemented using PyTorch, and the Adam optimizer is used with a learning rate of 1e-5, and gradually decreases during training according to the decay rate of 1e-6. The hops of the syntactical dependency graph for the Weibo dataset and the GossipCop dataset are set as 4 and 3, respectively. We use an early stop strategy for the label accuracy of the validation set, with a patience of 5 epochs. We adopt six metrics, including accuracy (Acc), macro F1 score (macF1), Area Under ROC (AUC), standardized partial AUC (spAUC), and the F1 scores of fake and real class ($F1_{fake}$ and $F1_{real}$) to evaluate detection performance.

5.4 Performance Results

Table 2 shows the performance of all comparative methods on two public real-world datasets, where the best performance is marked in bold. Results show that our proposed MSynFD has achieved the best performance on five crucial metrics compared with the SOTA fake news detection models. On Weibo, MSynFD yields 0.81%, 0.77%, 2.18%, 0.55%, and 1.00% improvement, over Acc, macF1, spAUC, $F1_{fake}$ and $F1_{real}$, and over AUC is 0.56% lower than MGIN-AG model. Additionally, on GossipCop, MSynFD yields 1.06%, 0.47%, 0.33%, 1.16%, 0.81%, and 0.03% improvement, over Acc, macF1, AUC, spAUC, $F1_{fake}$ and $F1_{real}$. The results demonstrate that the proposed method can capture the local syntactical dependency structure information of news and mitigate the priori bias from keywords, which can help better understand and analyze the news piece.

The adversarial mechanism and the MOE may not be able to learn enough about fake news patterns in the short-text context, which causes EANN and MDFEND to perform well on the Gossip-Cop dataset but not on the Weibo dataset. Further, comparing the results between the HMCAN and CMMTN, the noisy irrelevant connections from the attention mechanism affect the model performance; with the help of the mask mechanism, CMMTN could perform better on both datasets.

Finally, the results of MGIN-AG show that the GNN model does play a role, making MGIN-AG perform better than BiGRU and HMCAN on both datasets. The results compared between BERT-Emo and BERT-Emo-ENDEF show that the debiasing framework does help improve model performance for fake news detection, providing a basis for rationalizing our design of MSynFD.

5.5 Ablation Study

To verify the effectiveness of the different modules of MSynFD, we compare them with the following variants:

**MSynFD ¬ Se** removes the Semantic Aware Module, which loses the ability to perceive the sequential position structure.

MSynFD ¬ MSA removes the Multi-hop Syntax Aware Module, which makes the model lose the ability to perceive the local syntactical dependency structure.

MSynFD ¬ KD removes the keywords debiasing, which makes the model lose the ability to mitigate the priori bias from keywords within the news piece.

MSynFD-MH-GAT replaces the Multi-hop Syntax Aware Module with GAT to validate its effectiveness in obtaining local syntactical dependency structure information. For a fair comparison, we adjust the traditional GAT to multi-hops(MH)-GAT, whose adjacency matrix is set to be the same hops adjacency case as the

Table 3: Results of ablation study on both datasets

Method	Weibo		GossipCop	
	Acc	macF1	Acc	macF1
MSynFD ¬ Se	0.8739	0.7826	0.8618	0.8067
MSynFD ¬ MSA	0.8709	0.7792	0.8453	0.7656
MSynFD ¬ KD	0.8758	**0.7938**	0.8661	0.8121
MSynFD-MH-GAT	0.8717	0.7783	0.8512	0.8022
MSynFD	**0.8787**	0.7889	**0.8699**	**0.8164**

(a) Different hops　　　**(b) Different max lengths**
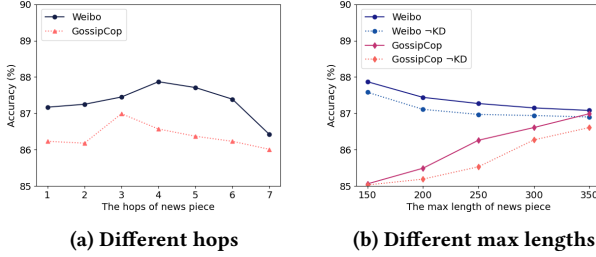
Figure 4: (a) Performance of the MSynFD model under different values of the parameter hops; (b) Performance of the MSynFD model and MSynFD ¬ KD under different values of the parameter max length.

original model, to ensure both models capture structural information at the same depth.

Table 3 shows that when comparing MSynFD with MSynFD ¬ Se reduces the accuracy of the proposed model by 0.48% and 0.81%, and macro F1 score by 0.63% and 0.97% on the Weibo and the GossipCop datasets, respectively. This means that the sequential representation module helps complete the global sequential information and improves the performance of fake news detection.

Further, for MSynFD ¬ MSA reduces the accuracy of the proposed model by 0.78% and 2.46%, and macro F1 score by 0.97% and 5.08% on the Weibo and the GossipCop datasets, respectively. It means that the local syntactical dependency structure information focused by the Multi-hop Syntax Aware Module can reduce the noisy information caused by irrelevant connections, confirming that model performance degrades much more in the long news dataset GossipCop than in the short text dataset Weibo.

For, MSynFD-MH-GAT reduces the accuracy by 0.70% and 1.87%, and macro F1 score by 1.06% and 1.42% on the Weibo and the GossipCop datasets, respectively. It means that though perceiving syntactical dependency structure at the same depth, the Multi-hop Syntax Aware Module is more effective than GAT due to the subgraph weighted aggregation mechanism.

Finally, MSynFD ¬ KD reduces the accuracy of the proposed model by 0.29% and 0.38%, and macro F1 score by -0.49% and 0.43% on the Weibo and the GossipCop datasets, respectively. The results show that keyword bias can improve performance in some situations (see section 5.6 - qualitative analysis for details).

5.6 Qualitative Analysis

To explore how the size of the perceived range and keyword bias affect the performance of fake news detection. We designed a series of experiments about the number of syntactical dependency graph hops and the max length of a news piece. The results shown in Figure 4 (a) indicate that the performances of both the Weibo dataset

and the GossipCop dataset increase and then decrease as hops increase, which means that the perceived range in the local syntactical dependency graph has a certain threshold. Before reaching it, the coverage of syntactical subgraphs is limited, leading to insufficient information. After that, the irrelevant noise will be brought and reduce the performance. The best hops for these two datasets are 4 and 3, respectively. This is because the informal phrase structure necessitates a broader range of word perception to gather sufficient information, causing the texts in the Weibo dataset with a more casual style to need larger hops than the texts in the Gossipcop dataset with a syntactically rigorous structure.

As shown in Figure 4 (b), the effect of the keywords debiasing presents a different scenario. For the Weibo dataset, as the max length of the news piece increases, the performance improvement from the keywords debiasing becomes more insignificant. We think that this may be due to the average length of the Weibo dataset being 120, so limited information makes the bias within keywords as important information for detection, and with the max length increasing, the percentage of padding in the news piece increases and reduces information density, creates further reliance on bias information, and alleviates the effect of keywords debiasing; On the other hand, for the GossipCop dataset, the performance improvement from the keywords debiasing is increasing first from insignificant and decreasing a little. Since the average length of the GossipCop dataset is 606, we think at first the length of 150 lacks information, causing the bias within keywords, which is important for detection too. As the max length increases, the informative patterns grow, which alleviates the reliance on biased information, making the debaising module more useful. With the max length increasing, more informative patterns are brought, and the effect of the keywords debiasing has been balanced.

5.7 Case Study

To provide an intuitive demonstration of the functions of each part, we use test set data from two datasets to analyze the intermediate process. We first test the performance of the Multi-hop Syntax Aware Module and Semantic Aware Module. As shown in Figure 5, due to the use of sequential relative position bias, the focuses of sequential neighbors are significantly enhanced in Chinese news, especially in Figure 5 (a), while it does not work well in English news. This may be from the grammatical differences between Chinese and English. And, the distant irrelevant connection, like 'Iceland' to 'China' in Figure 5 (b) and 'we' to 'fashion' in Figure 5 (c), would still be built. The SAA Module does show the ability to avoid such irrelevant information while obtaining enough useful information. As shown in Figure 5 (c), the perception range is extended from the adjacent word "know" to the 3-hop adjacent word "Hadid". However, the hazard of information gaps still exists, as shown in Figure 5 (d); we cannot obtain how the photos are due to the limits of syntactical relations. So, the semantic complement is still necessary.

Then we analyze the distribution of prediction scores of our main model ablation keywords debiasing before and after, as Figure 6 shows, the keywords debiasing can mitigate the effect of words with prejudice (e,g, 'shock' in Figure 6 (a)) and words of authority (e.g. central bank in Figure 6 (a)). Although the keywords debiasing shows the ability to capture some non-entity keywords (e.g. 'shock', 'pay' in Figure 6 (a) and 'romantically' in Figure 6. (b)), it may ignore
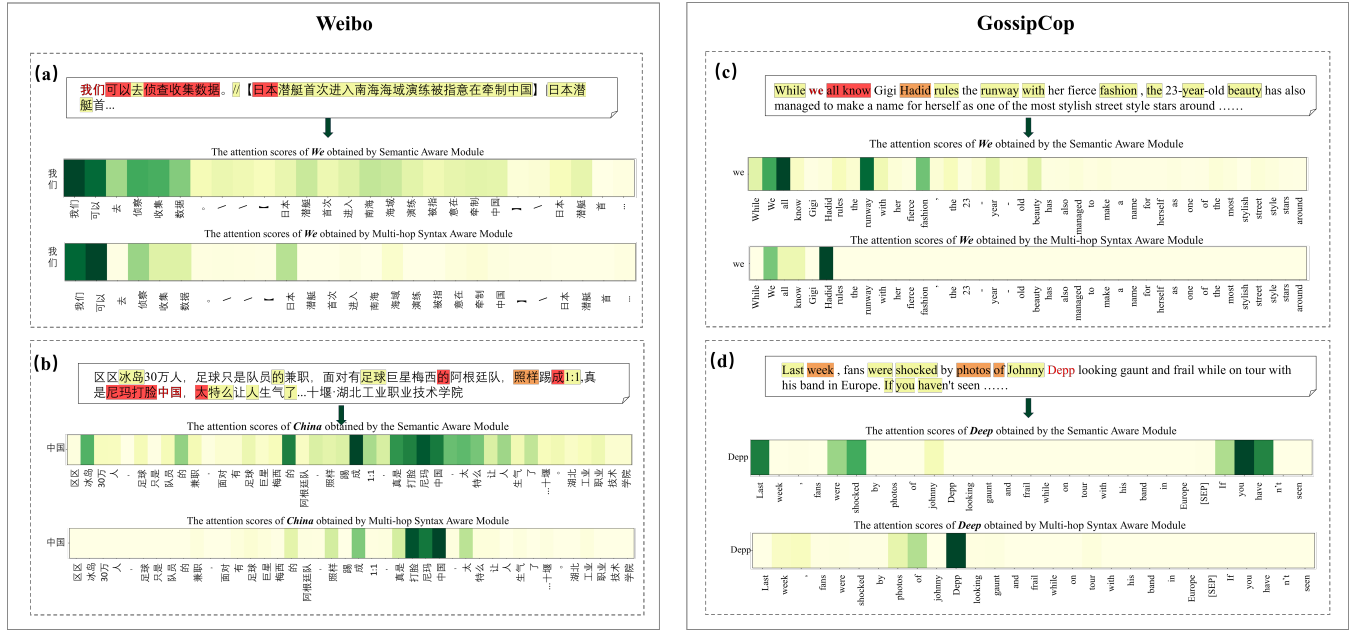
Figure 5: Case Study: Two pairs of cases from the Weibo and the GossipCop datasets, respectively. The center words focused by us are boldfaced, while the darker cell color indicates higher attention value, the yellow areas, orange areas, and red areas indicate the focus from the Semantic Aware Module, the Multi-hop Syntax Aware Module, or both focus.
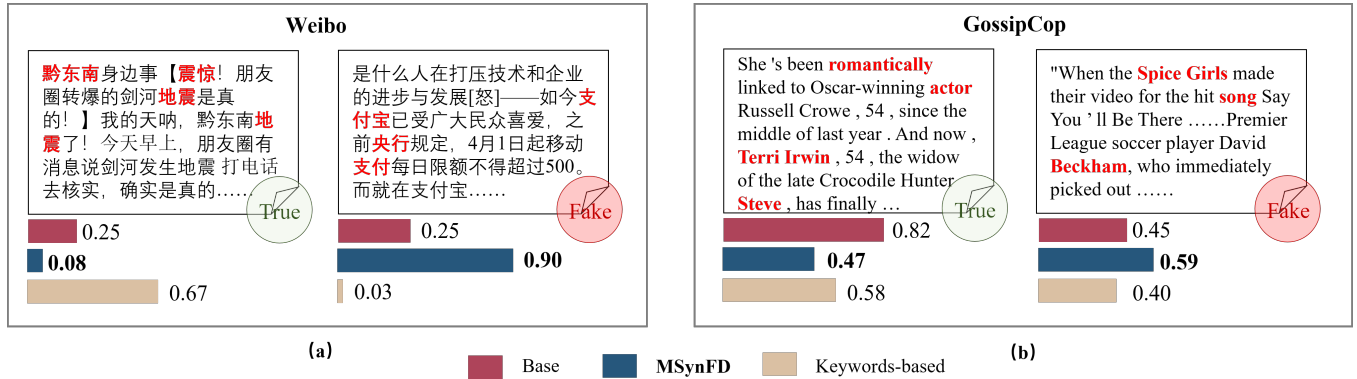

Figure 6: Case Study 2: Two pairs of cases from the Weibo and the GossipCop datasets respectively. For each dataset, one case is true while the other is fake. The keywords are boldfaced, and the lengths of bars represent the probability predictions of fake news of our base model (keywords debiasing ablation model), our MSynFD method, and the keywords-based model.

some important words that lead to misjudgment like 'Russell Crowe' due to the limits of Semantic-based keywords extraction method. Expanding the captured keywords is where our future research will focus on improvement.

6 CONCLUSION

In this paper, we propose a new fake news detection method, MSynFD, which uses a Multi-hop Syntax Aware Module to capture multi-hops syntactical dependency information within news pieces to extend the local syntax information of each word. Then, the Semantic Aware Module is used to obtain sequential aware semantic information. In the end, the Keywords Debiasing is mitigated into

the model to mitigate prior bias from keywords. The experimental results have shown that among the state-of-the-art methods, our proposed MSynFD method achieves the SOTA performance. Considering that the fake news detection task is one specific type of fine-grained semantic comprehension task, for future work, we plan to further explore the potential application of MSynFD on other fine-grained semantic comprehension tasks. Code is available at https://doi.org/10.5281/zenodo.10658674.

REFERENCES

[1] Parisa Bazmi, Masoud Asadpour, and Azadeh Shakery. 2023. Multi-view co-attention network for fake news detection by modeling topic-specific user and news source credibility. *Information Processing and Management* 60, 1 (2023), 103146. https://doi.org/10.1016/j.ipm.2022.103146

[2] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information Credibility on Twitter *(WWW '11)*. Association for Computing Machinery, New York, NY, USA, 675–684. https://doi.org/10.1145/1963405.1963500

[3] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 1724–1734. https://doi.org/10.3115/v1/D14-1179

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR* abs/1810.04805 (2018). arXiv:1810.04805 http://arxiv.org/abs/1810.04805

[5] Yingtong Dou, Kai Shu, Congying Xia, Philip S. Yu, and Lichao Sun. 2021. User Preference-Aware Fake News Detection *(SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 2051–2055. https://doi.org/10.1145/3404835.3462990

[6] Maarten Grootendorst. 2020. KeyBERT: Minimal keyword extraction with BERT. https://doi.org/10.5281/zenodo.4461265

[7] Binxuan Huang and Kathleen Carley. 2019. Syntax-Aware Aspect Level Sentiment Classification with Graph Attention Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 5469–5477. https://doi.org/10.18653/v1/D19-1549

[8] Celestine Iwendi, Senthilkumar Mohan, Suleman khan, Ebuka Ibeke, Ali Ahmadian, and Tiziana Ciano. 2022. Covid-19 fake news sentiment analysis. *Computers and Electrical Engineering* 101 (2022), 107967. https://doi.org/10.1016/j.compeleceng.2022.107967

[9] Joonwon Jang, Yoon-Sik Cho, Minju Kim, and Misuk Kim. 2022. Detecting incongruent news headlines with auxiliary textual information. *Expert Systems with Applications* 199 (2022), 116866. https://doi.org/10.1016/j.eswa.2022.116866

[10] Xinyu Jiang, Qi Zhang, and Chongyang Shi. 2022. Hierarchical Neural Network with Bidirectional Selection Mechanism for Sentiment Analysis. In *IJCNN*. IEEE, 1–8.

[11] Shingo Kato, Linshuo Yang, and Daisuke Ikeda. 2022. Domain Bias in Fake News Datasets Consisting of Fake and Real News Pairs. In *2022 12th International Congress on Advanced Applied Informatics (IIAI-AAI)*. 101–106. https://doi.org/10.1109/IIAIAAI55812.2022.00029

[12] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*. https://openreview.net/forum?id=SJU4ayYgl

[13] An Lao, Chongyang Shi, and Yayi Yang. 2021. Rumor Detection with Field of Linear and Non-Linear Propagation. In *WWW*. 3178–3187.

[14] An Lao, Qi Zhang, Chongyang Shi, Longbing Cao, Kun Yi, Liang Hu, and Duoqian Miao. 2023. Frequency Spectrum is More Effective for Multimodal Representation and Fusion: A Multimodal Spectrum Rumor Detector. *CoRR* abs/2312.11023 (2023).

[15] Jiawen Li, Shiwen Ni, and Hung-Yu Kao. 2021. Meet The Truth: Leverage Objective Facts and Subjective Views for Interpretable Rumor Detection. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. Association for Computational Linguistics, Online, 705–715. https://doi.org/10.18653/v1/2021.findings-acl.63

[16] Quanzhi Li, Qiong Zhang, and Luo Si. 2019. Rumor Detection by Exploiting User Credibility Information, Attention and Multi-task Learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 1173–1179. https://doi.org/10.18653/v1/P19-1113

[17] Ruifan Li, Hao Chen, Fangxiang Feng, Zhanyu Ma, Xiaojie Wang, and Eduard Hovy. 2021. Dual Graph Convolutional Networks for Aspect-based Sentiment Analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online, 6319–6329. https://doi.org/10.18653/v1/2021.acl-long.494

[18] Tong Liu, Ke Yu, Lu Wang, Xuanyu Zhang, Hao Zhou, and Xiaofei Wu. 2022. Clickbait Detection on WeChat: A Deep Model Integrating Semantic and Syntactic Information. *Know.-Based Syst.* 245, C (jun 2022), 11 pages. https://doi.org/10.1016/j.knosys.2022.108605

[19] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting Rumors from Microblogs with Recurrent Neural Networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, New York, USA) *(IJCAI'16)*. AAAI Press, 3818–3824.

[20] Asutosh Mohapatra, Nithin Thota, and P. Prakasam. 2022. Fake News Detection and Classification Using Hybrid BiLSTM and Self-Attention Model. *Multimedia Tools Appl.* 81, 13 (may 2022), 18503–18519. https://doi.org/10.1007/s11042-022-12764-9

[21] Qiong Nan, Juan Cao, Yongchun Zhu, Yanyan Wang, and Jintao Li. 2021. MD-FEND: Multi-Domain Fake News Detection *(CIKM '21)*. Association for Computing Machinery, New York, NY, USA, 3343–3347. https://doi.org/10.1145/3459637.3482139

[22] Jamal Abdul Nasir, Osama Subhani Khan, and Iraklis Varlamis. 2021. Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights* 1, 1 (2021), 100007. https://doi.org/10.1016/j.jjimei.2020.100007

[23] Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. 2022. FANG: Leveraging Social Context for Fake News Detection Using Graph Representation. *Commun. ACM* 65, 4 (mar 2022), 124–132. https://doi.org/10.1145/3517214

[24] Huyen Trang Phan, Ngoc Thanh Nguyen, and Dosam Hwang. 2023. Fake news detection: A survey of graph neural network methods. *Applied Soft Computing* 139 (2023), 110235. https://doi.org/10.1016/j.asoc.2023.110235

[25] Ofir Press, Noah Smith, and Mike Lewis. 2022. Train Short, Test Long: Attention with Linear Biases Enables Input Length Extrapolation. In *International Conference on Learning Representations*. https://openreview.net/forum?id=R8sQPPpGCv0

[26] Shengsheng Qian, Jinguang Wang, Jun Hu, Quan Fang, and Changsheng Xu. 2021. Hierarchical Multi-Modal Contextual Attention Network for Fake News Detection *(SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 153–162. https://doi.org/10.1145/3404835.3462871

[27] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* 21, 1, Article 140 (jan 2020), 67 pages.

[28] I. Kadek Sastrawan, I.P.A. Bayupati, and Dewa Made Sri Arsa. 2022. Detection of fake news using deep learning CNN–RNN based methods. *ICT Express* 8, 3 (2022), 396–408. https://doi.org/10.1016/j.icte.2021.10.003

[29] Qiang Sheng, Juan Cao, Xueyao Zhang, Rundong Li, Danding Wang, and Yongchun Zhu. 2022. Zoom Out and Observe: News Environment Perception for Fake News Detection. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Dublin, Ireland, 4543–4556. https://doi.org/10.18653/v1/2022.acl-long.311

[30] Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu. 2019. DEFEND: Explainable Fake News Detection. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Anchorage, AK, USA) *(KDD '19)*. Association for Computing Machinery, New York, NY, USA, 395–405. https://doi.org/10.1145/3292500.3330935

[31] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. *Big Data* 8, 3 (2020), 171–188. https://doi.org/10.1089/big.2020.0062 arXiv:https://doi.org/10.1089/big.2020.0062 PMID: 32491943.

[32] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *SIGKDD Explor. Newsl.* 19, 1 (sep 2017), 22–36. https://doi.org/10.1145/3137597.3137600

[33] Amila Silva, Yi Han, Ling Luo, Shanika Karunasekera, and Christopher Leckie. 2021. Propagation2Vec: Embedding Partial Propagation Networks for Explainable Fake News Early Detection. *Inf. Process. Manage.* 58, 5 (sep 2021), 17 pages. https://doi.org/10.1016/j.ipm.2021.102618

[34] Tiening Sun, Zhong Qian, Peifeng Li, and Qiaoming Zhu. 2023. Graph Interactive Network with Adaptive Gradient for Multi-Modal Rumor Detection *(ICMR '23)*. Association for Computing Machinery, New York, NY, USA, 316–324. https://doi.org/10.1145/3591106.3592250

[35] Hao Tang, Donghong Ji, Chenliang Li, and Qiji Zhou. 2020. Dependency Graph Enhanced Dual-transformer Structure for Aspect-based Sentiment Classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 6578–6588. https://doi.org/10.18653/v1/2020.acl-main.588

[36] Tina Esther Trueman, Ashok Kumar J., Narayanasamy P., and Vidya J. 2021. Attention-Based C-BiLSTM for Fake News Detection. *Appl. Soft Comput.* 110, C (oct 2021), 8 pages. https://doi.org/10.1016/j.asoc.2021.107600

[37] Vaibhav Vaibhav, Raghuram Mandyam, and Eduard Hovy. 2019. Do Sentence Interactions Matter? Leveraging Sentence Level Representations for Fake News Classification. In *Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13)*. Association for Computational Linguistics, Hong Kong, 134–139. https://doi.org/10.18653/v1/D19-5316

[38] Vaibhav Vaibhav, Raghuram Mandyam, and Eduard Hovy. 2019. Do Sentence Interactions Matter? Leveraging Sentence Level Representations for Fake News Classification. In *Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13)*. Association for Computational Linguistics, Hong Kong, 134–139. https://doi.org/10.18653/v1/D19-5316

[39] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You

Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 6000–6010.

[40] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *International Conference on Learning Representations*. https://openreview.net/forum?id=rJXMpikCZ

[41] Jinguang Wang, Shengsheng Qian, Jun Hu, and Richang Hong. 2023. Positive Unlabeled Fake News Detection Via Multi-Modal Masked Transformer Network. *IEEE Transactions on Multimedia* (2023), 1–11. https://doi.org/10.1109/TMM.2023.3263552

[42] Shoujin Wang, Xiaofei Xu, Xiuzhen Zhang, Yan Wang, and Wenzhuo Song. 2022. Veracity-Aware and Event-Driven Personalized News Recommendation for Fake News Mitigation. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 3673–3684. https://doi.org/10.1145/3485447.3512263

[43] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (London, United Kingdom) *(KDD '18)*. Association for Computing Machinery, New York, NY, USA, 849–857. https://doi.org/10.1145/3219819.3219903

[44] Jiaying Wu and Bryan Hooi. 2023. DECOR: Degree-Corrected Social Graph Refinement for Fake News Detection. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Long Beach, CA, USA) *(KDD '23)*. Association for Computing Machinery, New York, NY, USA, 2582–2593. https://doi.org/10.1145/3580305.3599298

[45] Junfei Wu, Qiang Liu, Weizhi Xu, and Shu Wu. 2022. Bias Mitigation for Evidence-Aware Fake News Detection by Causal Intervention *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 2308–2313. https://doi.org/10.1145/3477495.3531850

[46] Zeguan Xiao, Jiarun Wu, Qingliang Chen, and Congjian Deng. 2021. BERT4GCN: Using BERT Intermediate Layers to Augment GCN for Aspect-based Sentiment Classification. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 9193–9200. https://doi.org/10.18653/v1/2021.emnlp-main.724

[47] Bowen Xing and Ivor Tsang. 2022. DigNet: Digging Clues from Local-Global Interactive Graph for Aspect-level Sentiment Classification. *arXiv e-prints*, Article arXiv:2201.00989 (Jan. 2022), arXiv:2201.00989 pages. https://doi.org/10.48550/arXiv.2201.00989 arXiv:2201.00989 [cs.CL]

[48] Weizhi Xu, Junfei Wu, Qiang Liu, Shu Wu, and Liang Wang. 2022. Evidence-aware Fake News Detection with Graph Neural Networks. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 2501–2510. https://doi.org/10.1145/3485447.3512122

[49] Fan Yang, Yang Liu, Xiaohui Yu, and Min Yang. 2012. Automatic Detection of Rumor on Sina Weibo. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics* (Beijing, China) *(MDS '12)*. Association for Computing Machinery, New York, NY, USA, Article 13, 7 pages. https://doi.org/10.1145/2350190.2350203

[50] Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. 2021. Do Transformers Really Perform Badly for Graph Representation?. In *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (Eds.), Vol. 34. Curran Associates, Inc., 28877–28888. https://proceedings.neurips.cc/paper_files/paper/2021/file/f1c1592588411002af340cbaedd6fc33-Paper.pdf

[51] Seunghyun Yoon, Kunwoo Park, Joongbo Shin, Hongjun Lim, Seungpil Won, Meeyoung Cha, and Kyomin Jung. 2019. Detecting Incongruity between News Headline and Body Text via a Deep Hierarchical Encoder. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence* (Honolulu, Hawaii, USA) *(AAAI'19/IAAI'19/EAAI'19)*. AAAI Press, Article 98, 10 pages. https://doi.org/10.1609/aaai.v33i01.3301791

[52] Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2017. A Convolutional Approach for Misinformation Identification. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence* (Melbourne, Australia) *(IJCAI'17)*. AAAI Press, 3901–3907.

[53] Huaiwen Zhang, Quan Fang, Shengsheng Qian, and Changsheng Xu. 2019. Multi-Modal Knowledge-Aware Event Memory Network for Social Media Rumor Detection. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 1942–1951. https://doi.org/10.1145/3343031.3350850

[54] Mi Zhang and Tieyun Qian. 2020. Convolution over Hierarchical Syntactic and Lexical Graphs for Aspect Level Sentiment Analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Online, 3540–3549. https://doi.org/10.18653/v1/2020.emnlp-main.286

[55] Qi Zhang, Longbing Cao, Chongyang Shi, and Liang Hu. 2021. Tripartite Collaborative Filtering with Observability and Selection for Debiasing Rating Estimation on Missing-Not-at-Random Data. In *AAAI*. AAAI Press, 4671–4678.

[56] Qi Zhang, Yayi Yang, Chongyang Shi, An Lao, Liang Hu, Shoujin Wang, and Usman Naseem. 2023. Rumor Detection With Hierarchical Representation on Bipartite Ad Hoc Event Trees. *IEEE Transactions on Neural Networks and Learning Systems* (2023), 1–13. https://doi.org/10.1109/TNNLS.2023.3274694

[57] Xueyao Zhang, Juan Cao, Xirong Li, Qiang Sheng, Lei Zhong, and Kai Shu. 2021. Mining Dual Emotion for Fake News Detection. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) *(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 3465–3476. https://doi.org/10.1145/3442381.3450004

[58] Yongchun Zhu, Qiang Sheng, Juan Cao, Shuokai Li, Danding Wang, and Fuzhen Zhuang. 2022. Generalizing to the Future: Mitigating Entity Bias in Fake News Detection. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Madrid, Spain) *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 2120–2125. https://doi.org/10.1145/3477495.3531816

[59] Yongchun Zhu, Qiang Sheng, Juan Cao, Qiong Nan, Kai Shu, Minghui Wu, Jindong Wang, and Fuzhen Zhuang. 2023. Memory-Guided Multi-View Multi-Domain Fake News Detection. *IEEE Transactions on Knowledge and Data Engineering* 35, 7 (2023), 7178–7191. https://doi.org/10.1109/TKDE.2022.3185151