HGAPT: Heterogeneous Graph Augmented Prompt Tuning for Low-Resource Fake News Detection

Xingchen Ding, Fei Li, Chong Teng* , Donghong Ji
Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education,
School of Cyber Science and Engineering, Wuhan University, Wuhan, China
Email: {xingos,lifei_csnlp,tengchong,dhji}@whu.edu.cn

***Abstract*—The dissemination of disinformation on social media platforms has a significant impact on personal reputation and public trust. There has been a recent surge of interest in fake news detection. However, detecting low-resource fake news, particularly those pertaining to recent events that have not yet been disseminated by users and are typically in short text, remains challenging due to the lack of training data and prior knowledge. In this paper, we introduce a novel framework named the Heterogeneous Graph Augmented Prompt-based Tuning framework (HGAPT) that can leverage the metadata of news such as publisher and topic to construct a heterogeneous graph in same batch, which improve the performance of low-resource fake news detection. We have conducted extensive experiments on two low-resource fake news datasets that were collected from real-world sources. The results demonstrate that our proposed framework outperforms state-of-the-art methods, with superior detection performance at the zero-shot setting.**

***Index Terms*—Low-Resource, Fake News Detection, Social Media, Prompt Tuning, Heterogeneous Graph**

I. Introduction

With the advent of the Internet and social media platforms such as Twitter and Facebook, people now have convenient access to vast amounts of information. However, the widespread use of social media has also led to the proliferation of fake news, which can distort facts and spread misinformation, ultimately leading to negative consequences. For example, the dissemination of fake political news can erode public trust in both governments and journalism. To protect individuals, governments, and the news ecosystem from the negative impacts of fake news [1], it is crucial to develop automated techniques for detecting fake news on social media. Therefore, the field of fake news detection on social media has emerged as an area of great interest with the potential to provide significant benefits. Initial research on fake news detection focused on identifying effective features from various sources, such as textual content, user profiling data, and news diffusion patterns. Linguistic features such as writing styles and sensational headlines [2], and lexical and syntactic analysis [3], were explored to differentiate between fake news and real news. However, these feature-based methods were biased, time-consuming, labor-intensive, and susceptible to user manipulation. Recent studies [4], [5], [6] have attempted to tackle these challenges by utilizing various neural networks to acquire sophisticated representations for detecting fake news. [7], [8] also utilize the technique of fine-tuning Pre-trained Language Models (PLMs) for fake news detection. However, these methods exhibit suboptimal performance in low-resource settings without a sizable qualified training corpus.

For low-resource fake news, during the initial stage of a breaking event, only a limited amount of relevant short news with no comments and related knowledge is typically posted on social media. Therefore, it is appropriate to consider the metadata of the news, such as the publisher, topic, news source, and so on. People tend to believe news from a trusted and authoritative source, or news shared by publishers with a good reputation. In addition, certain topics, such as politics or controversial issues, are more likely to be targeted by fake news creators who seek to spread false information for their own agenda, which makes us more skeptical about news related to these topics. Drawing inspiration from the aforementioned observation, we propose a comprehensive solution to tackle the challenge of detection of fake news on social media platforms in low-resource environments. Our proposed solution is the Heterogeneous Graph Augmented Prompt-based Tuning framework (HGAPT). Specifically, we leverage news metadata to construct a batch-wise heterogeneous graph. In this graph, each labeled news item is represented by a node and edges between the nodes denote the similarity between the news items. HGAPT establishes edges in the heterogeneous graph based on whether two news items share common attributes such as speaker or topic, and further regulates the similarity of the representations learned by PLMs between each pair of news items to conform with the edges of the relation graph. The scalability of HGAPT is noteworthy since the heterogeneous graph is constructed based on a mini-batch of sampled data points per iteration.

To verify the effectiveness of our HGAPT framework, We developed two low-resource fake news datasets, respectively gathered from the Liar dataset [9] and www.politifact.com. Our experimental results demonstrate that our model outperforms other methods in the task of detecting fake news under low-resource settings and zero-shot settings.

II. Related Work

Fake News Detection. Initial research in the detection of fake news has primarily focused on developing effective features to distinguish between fake and true news. Researchers have

*Corresponding author
DOI reference number: 10.18293/SEKE2023-062

explored linguistic patterns, such as special characters and keywords [10], writing styles and sensational headlines [2], as well as lexical and syntactic features [3], and temporal-linguistic features [4] to identify fake news. In addition to linguistic features, some studies have also proposed a range of user-based features [10], [11], including factors such as the number of fans, registration age, and gender, as potential indicators of fake news. Nonetheless, the process of designing effective functions for fake news detection can be time-consuming and requires significant domain expertise.

To tackle the aforementioned challenges, a variety of neural network models have been employed by researchers. For example, recurrent neural networks (RNNs) [4], convolutional neural networks (CNNs) [5], graph neural networks [12], multitask learning [13] have been utilized to learn representations from news content or diffusion graphs. Fine-tuning methods of pre-trained language models (PLMs) have thus far provided straightforward yet robust baselines in fake news detection [7], [14]. However, this approach is hampered by the gap between the pre-training and fine-tuning stages, leading to performance bottlenecks. In contrast to the aforementioned methods that either do not employ PLMs or use them insufficiently, we leverage an effective technique, i.e., prompt learning, to guide fake news detection utilizing a PLM.

Prompting for PLMs. Prompt-based learning has become increasingly popular for extracting knowledge from large language models. This trend is evident in recent research studies [15], [16] that have focused on prompt-tuning, which has gained significant attention over the past few years. With the advent of GPT-3, prompt-based learning has relied on handcrafted prompts to achieve impressive performance. More recently, AutoPrompt [17] and LM-BFF [18] have proposed automatic prompt construction through generating discrete tokens. In contrast to previous work on fake news detection [19], [20], our prompt-based framework primarily focuses on fake news detection by exploiting relations among a limited number of labeled news, which represents a novel exploration of this challenging task in a low resource setting.

Graph Structure Learning. Graph structure learning techniques aim to learn the graph structure and node embeddings of input samples simultaneously [21]. Typically, these methods consist of two iterative steps: (i) estimating the adjacency matrix that represents the graph structure using node embeddings, and (ii) employing graph neural networks (GNNs) to obtain new node embeddings based on the updated graph. Recently, graph structure learning has been utilized to estimate heterogeneous graphs among samples for effective propagation of supervised information [22], [23]. These methods estimate heterogeneous graphs that encode the similarity between sample embeddings. In contrast, our HGAPT model gauges similarity between samples using news prediction vectors without the need for additional parameters.

III. Problem Statement and Background

In this paper, we aim to address the problem of fake news detection by leveraging a pre-trained language model (PLM) and a few labeled news examples. In prompt-based tuning, each input sample consisting of a pair $(\mathbf{x}_i, y_i)$ is transformed into a pattern-verbalizer pair (PVP) [24], denoted by $(p(\mathbf{x}_i), v(y_i))$. The pattern mapping function $p(\cdot)$ takes $\mathbf{x}_i$ as input and produces a cloze question with masks. For instance, given a single sentence represented as '$\mathbf{x}_i = $ [CLS] *News.* [SEP], ' we can map it to a cloze question as follows: '$\mathrm{p}(\mathbf{x}_i) =$ [CLS] *News.* It was [MASK].[SEP],' where the tokens [CLS] and [SEP] serve as special start and end markers, respectively.

In the context of prompt-based tuning, the verbalizer function $v(\cdot)$ maps the label y_i to tokens that represent its semantic meaning. For instance, in this paper, labels such as "pants fire/false/mostly false/half true/mostly true/true" are mapped to tokens such as "fabricated/false/inaccurate/dubious/credible/authentic". Given a PVP, the representation of the input $\mathbf{x}_i$ is obtained by taking the token embedding $\mathbf{h}_i^{[\mathrm{MASK}]}$ corresponding to the [MASK] token. The class prediction $\hat{\mathbf{y}}_i$ is a probability distribution over all possible class labels, with the probability of the ground truth label y_i given $\mathbf{x}_i$ estimated as follows:

$$q(y_i \mid \mathbf{x}_i) = \frac{\exp\left(\mathbf{w}_{v(y_i)}^{\top} \cdot \mathbf{h}_i^{[\mathrm{MASK}]}\right)}{\sum_{y_j \in \mathcal{Y}} \exp\left(\mathbf{w}_{v(y_j)}^{\top} \cdot \mathbf{h}_i^{[\mathrm{MASK}]}\right)} \tag{1}$$

Here, $\mathbf{w}_v$ is the logit vector of token v in the vocabulary, and $\mathcal{Y}$ denotes the set of all possible class labels. Let $\mathbf{y}_i$ be a one-hot vector with all elements being 0 except for the one corresponding to the ground truth class label $y_i \in \{1, \ldots, C\}$. The model is trained by minimizing the cross-entropy loss $\mathcal{L}_{\mathrm{CE}}$ defined as:

$$\mathcal{L}_{\mathrm{CE}} = \sum_{i=1}^{N} -\log(\hat{\mathbf{y}}_i)^{\top} \mathbf{y}_i, \tag{2}$$

where $(\cdot)^{\top}$ denotes the transpose operation, and N is the total number of samples in the training set.

IV. Proposed Method

In this section, we introduce the proposed HGAPT, as illustrated in Figure 1. Our approach leverages supervised signals from training news samples by constructing and learning on batch-wise heterogeneous graphs. This method effectively enhances the detection of fake news.

A. MASK Representation Augment

A well-known challenge in prompting is the need for a fixed number of positions for the label, e.g., a single mask is needed for words present in the dictionary such as Yes/No; however, we need multiple positions to predict more complex ones with multiple tokens such as Half true. The label inventory commonly contains words tokenised into multiple tokens. [24] proposed a simple verbalisation technique where the original labels are replaced with words that can be represented with a single token from the vocabulary, e.g., Half true → Dubious. However, this will lead to the degradation of the model's performance for fake news detection, because of the loss of label information during the mapping process.

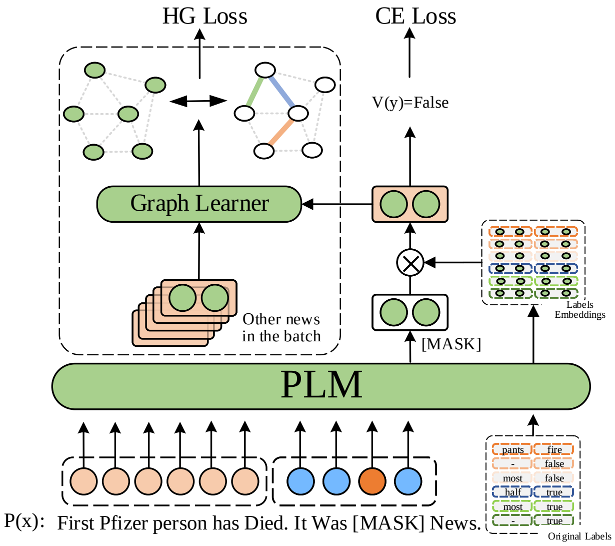

Fig. 1. A high-level illustration of the proposed HGAPT framework for fake news detection.

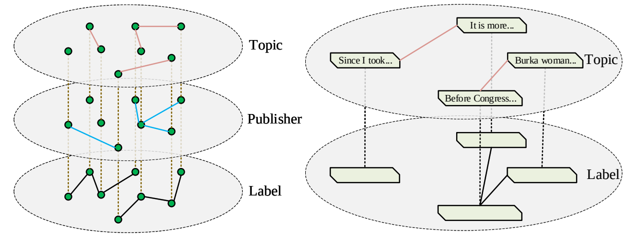

Fig. 2. A example of heterogeneous graph where each node represents a news and the three layers correspond to three types of relations.

Here, we propose a simple, yet effective, approach to overcome this problem. We take the original label inventory and tokenise all words, as shown in 1. In the original labels box, we see six labels for fake news detection tasks and their tokens – {'pants','fire'}, {'false'}, {'mostly', 'false'}, {'half', 'true'},{'mostly', 'true'},and {'true'}. For each token of a label, we extract the vector representation from the PLM's token embeddings $v_{TE}^{L_t} = \operatorname{Tok\,Emb}(L_t)$. Afterwards, we obtain the final label representation (LE_L) using an element-wise averaging for all $v_{TE}^{L_t}$ (see Eq. 3).

$$LE_L = \frac{1}{N}\sum_{t=0}^{N} \operatorname{TokEmb}(L_t)\,; \forall_L \in \{\text{ Labels }\} \tag{3}$$

Finally, to obtain the augmented MASK representation $\hat{\mathbf{y}}$ for each example, we take the dot product between the MASK representation for the masked token position, and each of the LE_L vectors.

B. Heterogeneous Graph Construction

Given a mini-batch $\mathcal{N} = (\mathbf{x}_i, y_i)\}_{i=1}^{N}$ comprising N randomly sampled sequence-label pairs, whose indices are stored in $\mathcal{I} = \{1, \ldots, N\}$. Our objective is to leverage additional supervised information by constructing a heterogeneous graph model.

A heterogeneous graph, denoted as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, F, R, \mu, \phi\}$. where $\mathcal{V}$ represents the set of nodes, $\mathcal{E}$ the set of edges, F the set of node types and R the set of edge types, where $|F|+|R|> 2$. Each node $v_i \in \mathcal{V}$ is associated with a node type mapping function $\mu : \mathcal{V} \longrightarrow F$, and each edge $e_i \in \mathcal{E}$ is associated with an edge type mapping function $\phi : \mathcal{E} \longrightarrow R$ [25] . Let $\mathcal{G}$ denotes the heterogeneous graph among the N training samples in $\mathcal{N}$. In particular, $\mathcal{V}$ is a set of nodes where each node $v_i \in \mathcal{V}$ corresponds to one training sample $\mathbf{x}_i$, $\mathcal{E} = \{e_{ij}\}$ is a set of edges between the N training samples, and R is a set of relation types of news. Hence, we establish the edge e_{ij} between a node v_i and another node v_j if these nodes have the same relation. Formally, e_{ij} is set to

$$e_{ij} = \begin{cases} x & \text{if } v_i \text{ is related to } v_j \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

Where x is selected from the set $\{0.2, 0.5, 1\}$ and its value depends on the type of edge. In this paper, our primary focus is on utilizing news metadata, including class, topic, publisher, etc., to construct a heterogeneous graph as depicted in Figure 2.

C. Heterogeneous Graph Learning

On the heterogeneous graph $\mathcal{G}$ of mini-batch $\mathcal{N}$, we decompose the original classification task into three sub-problems: (i) a node classification problem that aims to accurately classify each node into its corresponding class, (ii) an edge prediction problem that aims to establish connections between nodes belonging to the same class and disconnect nodes belonging to different classes, and (iii) a heterogeneous graph learning problem that aims to minimize the dissimilarity between the predicted graph and the original graph.

The problem of node classification is equivalent to the initial fake news classification task. As such, we can generate a class prediction $\hat{\mathbf{y}}_i$ for v_i, which corresponds to $\mathbf{x}_i$, by utilizing 1. We can then calculate the loss $\mathcal{L}_{\text{CE}}$ using 2.

In the edge prediction problem, we aim to predict the relationship between two nodes v_i and v_j by establishing $\hat{e}_{ij}$ based on the relevance between their corresponding representations $\hat{\mathbf{y}}_i$ and $\hat{\mathbf{y}}_j$:

$$\hat{e}_{ij} = g\left(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j\right), \tag{5}$$

where $\hat{\mathbf{y}}_i$ and $\hat{\mathbf{y}}_j$ are obtained using the MASK Representation Augment module. In this paper, we use cosine similarity to compute the function $g(\cdot,\cdot)$, although other choices are possible. One could leverage auxiliary heterogeneous graphs or calculate $\hat{e}_{ij}$ based on representation similarity, such as $g\left(\mathbf{h}_i^{[\text{CLS}]}, \mathbf{h}_j^{[\text{CLS}]}\right)$ and $g\left(\mathbf{h}_i^{[\text{MASK}]}, \mathbf{h}_j^{[\text{MASK}]}\right)$. However, we prefer to use $\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j$ as they carry more semantic information that is relevant to each news, yielding better empirical performance.

To measure the loss of edge prediction, we use the $\mathcal{L}_{\text{edge}}$ loss defined as follows:

$$-\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{A}(i)} e_{ij}\log\left(\hat{e}_{ij}\right) + \left(1 - e_{ij}\right)\log\left(1 - \hat{e}_{ij}\right), \tag{6}$$

where $\mathcal{A}(i) = j \in \mathcal{I}$ and $i \neq j$ is the set of all nodes except v_i, and e_{ij} is the true relationship between v_i and v_j. Note that we

Algorithm 1 Heterogeneous Graph Augmented Prompt-based Tuning

Input: A small set of news C.
Output: Assign news labels y to given unlabeled target data.

```
for each mini-batch N_i of the news C do:
    Pass N_i to the PVP and then PLM to obtain its [MASK] token representation mask_i.
    Pass mask_i to the representation augmented module to obtain its class-level feature ŷ_i.
    Compute the classification loss L_CE.
    for each relation edge_i of heterogeneous graph G do:
        for each event-level feature ŷ_i of mini-batch N_i do:
            for each event-level feature ŷ_j of mini-batch N_i do:
                Compute the loss of edge prediction L_{edge_i^n} as Eq.6.
        Compute the joint loss L_{edge_i} as Eq.6.
    Compute the joint loss L_HG.
    Jointly optimize all parameters of the model using the loss L = L_CE + αL_HG.
```

avoid introducing additional parameters to the function $g(\cdot,\cdot)$ to reduce the risk of overfitting, given the limited number of labeled samples.

In order to address the problem of heterogeneous graph learning, we optimize the model to minimize the heterogeneous graph prediction loss $\mathcal{L}_{\mathrm{HG}}$ for each mini-batch $\mathcal{N}$ as a whole. Here, $f_i \in \{0.2, 0.5, 1\}$ represents a mapping of edge types to control the contribution of the corresponding $\mathcal{L}_{\mathrm{edge}_i}$, and $\mathcal{L}_{\mathrm{edge}_i}$ denotes the loss function for each edge type i in the heterogeneous graph. The expression for $\mathcal{L}_{\mathrm{HG}}$ is given by:

$$\mathcal{L}_{\mathrm{HG}} = \sum_{i \in F} f_i * \mathcal{L}_{\mathrm{edge}_i} \tag{7}$$

D. Model Training

We jointly train the model with the cross-entropy and supervised objectives:

$$\mathcal{L} = \mathcal{L}_{\mathrm{CE}} + \alpha \mathcal{L}_{\mathrm{HG}}$$

where α is a hyperparameter to control the contribution of this $\mathcal{L}_{\mathrm{HG}}$. Algorithm 1 presents the training process of our approach.

V. Experiments

In this section, we introduce the experiments to evaluate the effectiveness of HGAPT. Specifically, we aim to answer the following evaluation questions: **EQ1**: Can HGAPT improve low-resource fake news classification performance by exploiting the limited supervised information of news? **EQ2**: How effective are heterogeneous graph construction in improving the detection performance of HGAPT? **EQ3**: Can HGAPT improve the performance of zero-shot fake news detection task?

A. Datasets and Experiment Settings

Currently, there are no publicly available benchmarks for detecting fake news on social media in low-resource settings. In this paper, we utilize the original testing set from the LIAR dataset [9] for testing. Additionally, we randomly select 512 examples per class from the original training set and used them as the training set. Additionally, we create a new fake news

TABLE I
THE STATISTICS FOR THE TWO DATASETS.

General	# news	Avg. # of word per news	# train samples per class
LiarFewShot	7829	16	512
PolitifactFewShot	12904	20	512
Metadata	**# of topic**	**# of speaker**	**# of news_source**
LiarFewShot	3153	2420	/
PolitifactFewShot	/	3326	13

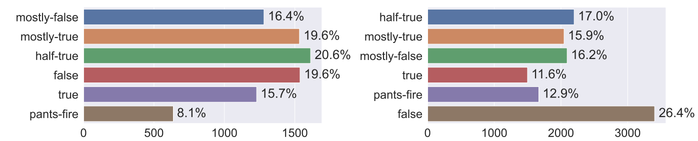

Fig. 3. Distribution of label from LiarFewShot and PolitifactFewShot datasets.

dataset, called the PolitiFactFewShot dataset, using authoritative sources from www.politifact.com. Table I and Figure 3 provide detailed statistics for both datasets. Our evaluation metrics, accuracy and F1 score. We also transformed the original six-class datasets into two classes to verify the performance of our model in a general setting for detecting fake news, for instance, the label ‘pants fire/false/mostly false” is categorized as ‘fake’, while ‘half true/mostly true/true’ is classified as ‘true’. We selected the hyperparameter α from [0 : 0.2 : 0.5 : 1], and the batch size from [4 : 8 : 16] for our HGAPT model.

B. Baseline Model

We compare our proposed model with several state-of-the-art baseline methods, which are described as follows: 1)**RNN** [4]: This model is based on a recurrent neural network (RNN) with gated recurrent units (GRU) for learning relevant post features over time in rumor detection; 2)**AttLSTM** [26]: This is a long short-term memory (LSTM) model that uses attention mechanism to consider the importance of words in the relevant posts; 3)**FNDML** [13]:This model employs Multitask Learning (ML) methodologies to train reliable classifiers to detect fake news; 4)**FT+ERINE** [27]: We use an existing fine-tuning technique based on the ERNIE pretrained language model; 5)**FakeBERT** [7]: This model combines different parallel blocks of the single-layer deep CNN having different kernel sizes and filters with the BERT; 6)**ParallelBERT** [14]: This model uses two parallel BERT networks to perform fake news detection. One of the BERT networks encodes news, and another encodes news-related knowledge; 7)**PT-*** [28]: We improve an existing prompt-based tuning technique on the ERNIE PLM for fake news detection and extend it for our task. The * in PT- represents different extensions, including knowledgeable prompt learning (KPL) [20], supervised contrastive learning (SCL) [29], and our proposed Heterogeneous Graph Augmented (HGA) framework.

We evaluate these models in the most challenging setting of detecting fake news in low-resource settings, where news on social media is typically shorter in length and less frequent.

C. Results

In order to answer EQ1, we conducted a comparison of HGAPT with the baselines outlined in Section V-B for low-resource fake news classification. The performance of our

TABLE II
TEST PERFORMANCE (%) MEASURED ON LOW-RESOUCE FAKE NEWS DATASETS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Model	LiarFewShot				PolitifactFewShot			
	6 classes		**2 classes**		**6 classes**		**2 classes**	
	Acc.	Mac-F_1	Acc.	Mac-F_1	Acc.	Mac-F_1	Acc.	Mac-F_1
RNN	0.254	0.237	0.569	0.508	0.208	0.192	0.534	0.389
AttLSTM	0.257	0.222	0.584	0.507	0.232	0.233	0.540	0.391
FNDML	0.273	0.282	0.594	0.587	0.247	0.244	0.577	0.482
FT+ERINE	0.332	0.304	0.624	0.602	0.248	0.254	0.618	0.591
FakeBERT	0.352	0.334	0.655	0.663	0.288	0.294	0.638	0.610
ParallelBERT	0.363	0.332	0.662	0.657	0.308	0.306	0.653	0.616
PT+KPL	0.383	0.379	0.692	0.712	0.332	0.454	0.719	0.674
PT+SCL	0.398	0.397	0.702	0.698	0.343	0.461	0.729	0.688
PT+HGA	**0.426**	**0.417**	**0.735**	**0.730**	**0.373**	**0.484**	**0.765**	**0.736**

proposed method, as well as all compared methods, on the LiarFewShot and PolitifactFewShot test sets is presented in Table II. It is observed that the first group of baselines exhibited poor performance due to the limitations of hand-crafted features in effectively encoding semantic information from short news content. Additionally, these methods were unable to perform deep feature interactions in low-resource settings. In the second group, ParallelBERT outperformed FakeBERT and FT+ERINE, which used limited labeled data for training. This can be attributed to ParallelBERT's use of additional news-related knowledge. The third group evaluated prompt-based tuning techniques, which showed improved performance over the fine-tuning learning baselines. This is because prompt-based tuning extracts rich semantic features from fewer samples using human-crafted prompts in conjunction with large pre-trained language models. Unlike KPL and SCL, we utilized more supervised information to enhance the performance of fake news detection.

By contrast, our proposed Heterogeneous Graph Augmented Prompt-based Tuning (HGAPT) approaches demonstrated superior performance compared to their counterparts, achieving improvements ranging from 2.8% (3%) to 17.2% (17%) in terms of accuracy scores on the LiarFewShot (PolitifactFewShot) datasets with 6 classes. These results highlight the strong discriminatory power of our approach, particularly for low-resource fake news scenarios. Furthermore, when applied to the LiarFewShot (PolitifactFewShot) datasets with 2 classes, our method outperformed fine-tuning learning baselines, achieving an accuracy improvement of approximately 8% (11.3%). These findings suggest that metadata associated with news articles plays a critical role in distinguishing between fake and true news, and our approach effectively leverages this information to improve classification accuracy.

D. Model Analysis

To address EQ2, we conducted additional ablation studies on the various modules of HGAPT. Figure 4 and Figure 5 depict the experimental results obtained on the LiarFewShot dataset. In our evaluation, we first assessed the impact of constructing heterogeneous graphs in different ways. As shown in Figure 4, we observed that using only label information to construct the

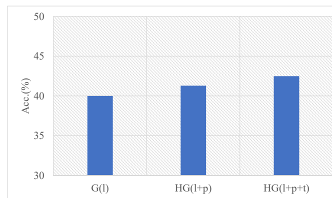

Fig. 4. Effect of different heterogeneou graphs, where "l, p, t" devotes "Label, Publisher, Topic".

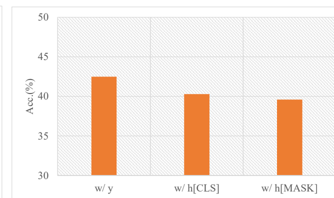

Fig. 5. Obtain edge representation in different ways.

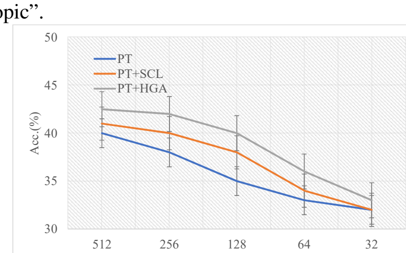

Fig. 6. Effect of samples per class.

graph resulted in the worst performance. On the other hand, the performance of HGAPT improved as we incrementally added more metadata of news to construct the heterogeneous graph. This demonstrates the advantages of using more supervised information and constructing a heterogeneous graph, allowing us to capture specific topology among news.

We further explore different approaches to obtaining $\hat{e}_{ij}$ in equation 6. Specifically, we consider three methods: (i) $\mathbf{w}/\mathbf{h}^{[\mathrm{CLS}]}$, which sets $\hat{e}_{ij} = \cos\left(\mathbf{h}_i^{[\mathrm{CLS}]}, \mathbf{h}_j^{[\mathrm{CLS}]}\right)$, where $\cos(\cdot,\cdot)$ denotes cosine similarity; (ii) $\mathbf{w}/\mathbf{h}^{[\mathrm{MASK}]}$, which sets $\hat{e}_{ij} = \cos\left(\mathbf{h}_i^{[\mathrm{MASK}]}, \mathbf{h}_j^{[\mathrm{MASK}]}\right)$; and (iii) $\mathbf{w}/\hat{\mathbf{y}}$, which is the approach adopted in HGA and sets $\hat{e}_{ij} = \cos\left(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j\right)$. Our experimental results, shown in Figure 5, demonstrate that HGA outperforms the other two methods. This validates that augmented class prediction carries more relevant information for discriminating news.

E. Zero-Shot Fake News Detection

Detecting fake news in minority domains has been shown to be difficult in previous studies [4], [8], [13] due to the lack of annotated resources. In order to address EQ3, we firstly compared various methods using different news samples, and

TABLE III

ZERO-SHOT FAKE NEWS DETECTION RESULTS (%). THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Target (Source)	Liar (Politifact)		Politifact (Liar)	
Model	Acc.	Mac-F_1	Acc.	Mac-F_1
FakeBERT	0.201	0.208	0.178	0.194
ParallelBERT	0.219	0.224	0.196	0.206
PT+KPL	0.244	0.272	0.221	0.206
PT+SCL	0.260	0.291	0.234	0.213
PT+HGA	**0.282**	**0.318**	**0.253**	**0.244**

evaluated their performance by measuring the accuracy on the LiarFewShot dataset obtained as we incrementally increased the number of samples. Figure 6 illustrates the impact of varying the number of labeled training samples on performance. It is evident that decreasing the number of training samples leads to reduce performance for all methods, with HGAPT consistently outperforming the other models.

Furthermore, we investigated the potential for low-resource fake news detection through domain transfer using HGAPT for zero-shot fake news detection. Specifically, these models were trained on a source training set and subsequently evaluated on the target test set. Table III presents the accuracy of several competitive models, highlighting the superiority of HGAPT over fine-tuning methods and state-of-the-art techniques. The results show that HGAPT achieves approximately 28% accuracy on the LiarFewShot dataset, and 25% accuracy on the PolitifactFewShot dataset, which is substantially better than most of the baseline models. Taken together, these experimental findings suggest that HGAPT not only enhances detection performance, but also exhibits superior zero-shot fake news detection capabilities.

VI. CONCLUSION

Our work introduces HGAPT, a novel prompt-based tuning framework that leverages heterogeneous graph augmentation to address the challenge of detecting low-resource fake news on social media. During the learning process, HGAPT constructs batch-wise heterogeneous graphs based on the metadata associated with each news item, and utilizes this information to fine-tune pretrained language models for solving both fake news classification and fake news relation prediction problems. By leveraging the limited supervised information available for news items, HGAPT is able to fully exploit the available data and achieve significant improvements over state-of-the-art models on both fake news classification and zero-shot detection tasks. Extensive experiments were conducted on two real-world datasets to demonstrate the effectiveness of our proposed model.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China No. 62176187, the Major Projects of the National Social Science Foundation of China No.11ZD189.

REFERENCES

[1] S. Volkova, K. Shaffer, J. Y. Jang, and N. O. Hodas, "Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter," *ACL*, 2017.

[2] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent features of rumor propagation in online social media," *International Conference on Data Mining*, 2013.

[3] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," *arXiv: Computation and Language*, 2017.

[4] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks," *International Joint Conference on Artificial Intelligence*, 2016.

[5] F. Yu, Q. Liu, S. Wu, L. Wang, and T. Tan, "A convolutional approach for misinformation identification," *International Joint Conference on Artificial Intelligence*, 2017.

[6] H. Guo, J. Cao, Y. Zhang, J. Guo, and J. Li, "Rumor detection with hierarchical social attention network," 2018.

[7] R. K. Kaliyar, A. Goswami, and P. Narang, "Fakebert: Fake news detection in social media with a bert-based deep learning approach," *Multimedia tools and applications*, vol. 80, pp. 11765–11788, 2021.

[8] N. Rai, D. Kumar, N. Kaushik, C. Raj, and A. Ali, "Fake news classification using transformer based enhanced lstm and bert," *International Journal of Cognitive Computing in Engineering*, vol. 3, pp. 98–105, 2022.

[9] W. Y. Wang, "" liar, liar pants on fire": A new benchmark dataset for fake news detection," *arXiv preprint arXiv:1705.00648*, 2017.

[10] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," *The Web Conference*, 2011.

[11] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on sina weibo," *Knowledge Discovery and Data Mining*, 2012.

[12] C. Yuan, Q. Ma, W. Zhou, J. Han, and S. Hu, "Jointly embedding the local and global relations of heterogeneous graph for rumor detection," *Cornell University - arXiv*, 2019.

[13] J. C. B. Cruz, J. A. Tan, and C. Cheng, "Localization of fake news detection via multitask transfer learning," *arXiv:1910.09295*, 2019.

[14] M. Farokhian, V. Rafe, and H. Veisi, "Fake news detection using parallel bert deep neural networks," *arXiv preprint arXiv:2204.04793*, 2022.

[15] T. L. Scao and A. M. Rush, "How many data points is a prompt worth," *North American Chapter of the Association for Computational Linguistics*, 2021.

[16] K. Zhou, J. Yang, C. Loy, Z. Liu, and J. L. D. Z. M. S. D. Z. Z. L. M. D, "Conditional prompt learning for vision-language models," 2023.

[17] T. Shin, Y. Razeghi, R. L. Logan, E. Wallace, and S. Singh, "Autoprompt: Eliciting knowledge from language models with automatically generated prompts," *Cornell University - arXiv*, 2020.

[18] T. Gao, A. Fisch, and D. Chen, "Making pre-trained language models better few-shot learners.," *Cornell University - arXiv*, 2020.

[19] C. B. El Vaigh, T. Girault, C. Mallart, and D. H. Nguyen, "Detecting fake news conspiracies with multitask and prompt-based learning," in *MediaEval 2021-MediaEval Multimedia Evaluation benchmark. Workshop*, pp. 1–3, 2021.

[20] G. Jiang, S. Liu, Y. Zhao, Y. Sun, and M. Zhang, "Fake news detection via knowledgeable prompt learning," *Information Processing & Management*, vol. 59, no. 5, p. 103029, 2022.

[21] Y. Zhu, W. Xu, J. Zhang, Q. Liu, S. Wu, and L. Wang, "Deep graph structure learning for robust representations: A survey," *arXiv: Learning*, 2021.

[22] V. G. Satorras and J. B. Estrach, "Few-shot learning with graph neural networks," *International Conference on Learning Representations*, 2018.

[23] P. Rodríguez, I. H. Laradji, A. Drouin, and A. Lacoste, "Embedding propagation: Smoother manifold for few-shot classification.," *arXiv: Computer Vision and Pattern Recognition*, 2020.

[24] T. Schick and H. Schütze, "Exploiting cloze-questions for few-shot text classification and natural language inference," *ECACL*, 2021.

[25] D. Jin, C. Huo, C. Liang, and L. Yang, "Heterogeneous graph neural network via attribute completion," *The Web Conference*, 2021.

[26] Y. Long, Q. Lu, R. Xiang, M. Li, and C. Huang, "Fake news detection through multi-perspective speaker profiles," in *IJCNLP 2017, Volume 2: Short Papers*, pp. 252–256, 2017.

[27] A. Malakhov, A. Patruno, and S. Bocconi, "Fake news classification with bert.," in *MediaEval*, 2020.

[28] J. Weng, Y. Hu, J. Qiu, and H. Huan, "Stprompt: Semantic-guided and task-driven prompts for effective few-shot classification," *arXiv preprint arXiv:2210.16489*, 2022.

[29] B. Gunel, J. Du, A. Conneau, and V. Stoyanov, "Supervised contrastive learning for pre-trained language model fine-tuning," *arXiv: Computation and Language*, 2020.