DiHAN: A Novel Dynamic Hierarchical Graph Attention Network for Fake News Detection

Ya-Ting Chang
ya-ting.chang@student.unsw.edu.au
University of New South Wales
Sydney, NSW 2052, Australia

Zhibo Hu
zhibo.hu@student.unsw.edu.au
University of New South Wales
Sydney, NSW 2052, Australia

Xiaoyu Li
xli216@stevens.edu
Stevens Institute of Technology
Hoboken, NJ 07030, USA

Shuiqiao Yang
shuiqiao.yang@data61.csiro.au
Data61
Marsfield, NSW 2122, Australia

Jiaojiao Jiang*
jiaojiao.jiang@unsw.edu.au
University of New South Wales
Sydney, NSW 2052, Australia

Nan Sun
nan.sun@unsw.edu.au
University of New South Wales
Canberra, ACT 2600, Australia

ABSTRACT

The rapid spread of fake news on social media has caused great harm to society in recent years, which raises the detection of fake news as an urgent task. Recent methods utilize the interactions among different entities such as authors, subjects, and news articles to model news propagation as a static heterogeneous information network (HIN). However, this is suboptimal since fake news emerges *dynamically*, and the *latent chronological interactions between news* in HIN are essential signals for fake news detection. To this end, we model the dynamics of news and associated entities as a News-Driven Dynamic Heterogeneous Information Network (News-DyHIN), where the temporal relationships among news articles are well captured with meta-path based temporal neighbors. With the support of News-DyHIN, we propose a novel fake news detection framework, named **D**ynam**i**c **H**ierarchical **A**ttention **N**etwork (DiHAN), which learns news representations via a hierarchical attention mechanism to fuse temporal interactions among news articles. In particular, DiHAN first employs a temporal node level attention to learn the temporal information from meta-path based news neighbors through the modeled News-DyHIN. Then, a semantic attention layer is adopted to fuse different types of meta-path based temporal information for news representation learning. Extensive evaluations conducted on two public real-world datasets demonstrate that our proposed DiHAN achieves significant improvements over established baseline models.

KEYWORDS

Fake news detection, dynamic heterogeneous information network, representation learning

ACM Reference Format:
Ya-Ting Chang, Zhibo Hu, Xiaoyu Li, Shuiqiao Yang, Jiaojiao Jiang, and Nan Sun. 2024. DiHAN: A Novel Dynamic Hierarchical Graph Attention Network for Fake News Detection. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), October 21–25, 2024, Boise, ID, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3627673.3679950

*Corresponding author.

CIKM '24, October 21–25, 2024, Boise, ID, USA

ACM ISBN 979-8-4007-0436-9/24/10…$15.00
https://doi.org/10.1145/3627673.3679950

1 INTRODUCTION

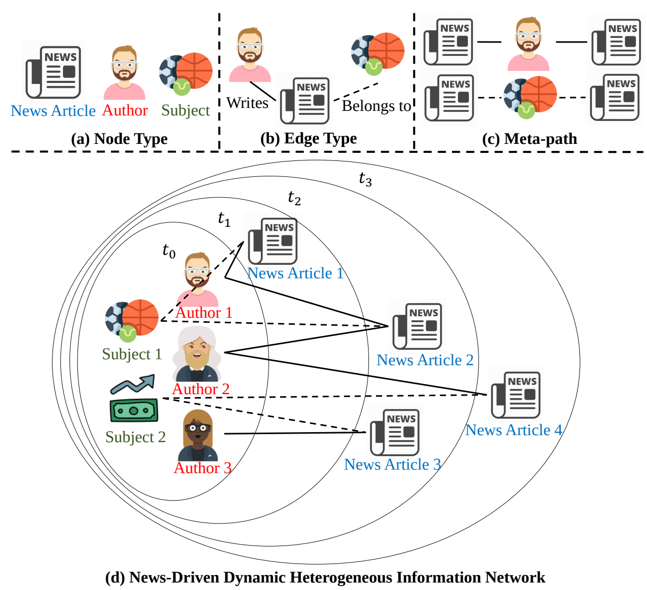

Figure 1: An illustrative example of a news-driven dynamic heterogeneous information network (News-DyHIN). (a) Three types of nodes (i.e., News article, Author, and Subject). (b) Two types of edges (i.e., Writes, Belongs to). (c) Two meta-paths (i.e., News-Author-News, News-Subject-News). (d) A News-DyHIN consists of three types of nodes and two types of edges and preserves the chronological order of different news articles, where $t_0 < t_1 < t_2 < t_3$ denoting the publishing time of news articles.

In recent years, the widespread use of various social media platforms—such as Twitter, Facebook, Instagram, LinkedIn, and TikTok—has significantly facilitated information dissemination. However, these easily accessible platforms act as double-edged swords, providing convenience while also exposing users to the pervasive issue of fake news. The rampant spread of fake news on social media has inflicted profound and lasting harm on society, affecting areas such as politics, health, and finance [21]. For instance, fake news was reported to have influenced events like Brexit [2] and the 2016 US presidential election [1]. Therefore, detecting the truth and falsehood of news and preventing fake news spreading is both essential and urgent.

Traditional methods of fake news detection often focus on the semantics of news articles, such as using the bag-of-words model to represent them [27]. However, in reality, news articles are not isolated entities. Instead, they are interconnected through various entities like news sources, authors, and subjects. Consequently, link relationships between news articles can be established based on their shared sources, creators, and subjects.

As a result, various heterogeneous information networks (HINs) have been constructed to incorporate different node entities and their relationships, thereby encompassing more comprehensive information [19, 24]. With the construction of HINs, recent studies have developed various graph neural network (GNN) models for fake news detection using graph convolution operations. These operations can aggregate information from neighboring news articles through different meta-paths [17]. To better aggregate heterogeneous information, models like the Heterogeneous Graph Attention Network (HAN) [29] and Hierarchical Graph Attention (HGAT) [11] have been proposed for processing heterogeneous information networks. Notably, HAN [29] was the first to apply a graph attention mechanism [26] to HINs. However, these methods primarily focus on either homogeneous or heterogeneous static networks and do not account for the temporal interactions between news articles.

Furthermore, the chronological order of news articles usually indicate their latent interactions, e.g., some newly published news referencing older news or similar news emerging within the same time span. Therefore, it is crucial to leverage the dynamic and temporal relationships among news articles for effective fake news detection. Although recent studies like the Heterogeneous Temporal Graph Neural Network (HTGNN) [9] can aggregate temporal information from time-slice-based temporal HINs, their temporal aggregation strategy fragments the continuous interactions among news articles, leading to suboptimal performance in fake news detection. Other recent models, such as TGNF [23] and Bi-GCN [3], have been proposed to capture news embeddings from dynamic homogeneous graphs. However, these models focus on the temporal propagation of news articles rather than the temporal relationships among the articles themselves.

To address the aforementioned limitations for fake news detection, we propose to model the dynamics of news and associated entities as a News-Driven Dynamic Heterogeneous Information Network (News-DyHIN) where the temporal relationships among news could be captured through their common connections with other entities (i.e., sources, authors and subjects). As illustrated in Figure 1, News-DyHIN has three types of nodes, and two types of edges, and the temporal interactions among news articles are reserved through two types of meta-paths (i.e., News-Author-News, News-Subject-News). News-DyHIN captures the temporal relationships between news articles via continuously adding new articles and edges in a natural chronological order. For example, News Article 4 published at t_3 can have latent interactions with News Article 2 through the News-Author-News path, and News Article 3 through the News-Subject-News path. While News Article 2 can only have interactions with earlier published News Article 1 published at t_1.

To exploit News-DyHIN for fake news detection, we propose a Dynamic Hierarchical Attention Network (DiHAN). DiHAN employs a hierarchical attention process to capture the temporal relationships among news articles from different meta-paths in News-DyHIN. Specifically, DiHAN first employs a temporal node level attention mechanism to aggregate information from temporal neighboring news articles, where we adopt two meta-paths to obtain the temporal neighbors for news articles. Then, semantic level attention is adopted to fuse different types of meta-path based temporal information and learn node representations for fake news detection. Our contributions are summarized as follows:

- We model the dynamics of news articles and the associated entities as a news-driven dynamic heterogeneous information network (News-DyHIN) to preserve the temporal interactions among news articles via two types of meta-paths.
- We propose a novel model, namely Dynamic Hierarchical Attention Network (DiHAN) for fake news detection on News-DyHIN. DiHAN exploits a hierarchical attention process that includes a temporal node level attention mechanism for aggregating information from meta-path based temporal new articles neighbors and semantic attention mechanism to fuse the temporal information learned from different meta-paths.
- We conduct extensive experiments on two real-world fake news datasets. The experimental results show that our model outperforms the state-of-the-art baselines for fake news detection.

2 RELATED WORK

In this section, we review the state-of-the-art GNN-based methods for fake news detection. Based on the type of network structures used for fake news detection, we particularly divide existing methods into the following four categories and discuss the advantages of the proposed DiHAN model compared to them.

2.1 Graph Neural Networks and Network Embedding

Graph neural networks (GNNs) are a type of method for analyzing graph structures like social networks, academic networks, and other networks. The goal of network embedding is to encode network structural information and features and project the network into a low-dimensional latent space [4]. It has been demonstrated to be extremely effective in network analysis. In recent years, graph conventional networks (GCN) [13] was proposed to introduce the convolution operations from the computer vision domain to graph domain. Based on GCN, Hu et al. proposed a Multi-depth GCN [10] for fake news detection using the similarity of news to detect different levels of fakeness. Recent studies like Bi-Directional Graph Conventional Networks (Bi-GCN) [3] and SAFER [5] are proposed to apply GCN to capture the news embeddings from a homogeneous graph for Rumor Detection. Compared with GCN, Graph Attention Network (GAT) [26] is the first to employ attention mechanism into graph domain and aggregates features from the local neighbourhoods with self-learned weights.

2.2 Heterogeneous Information Network

However, the above studies only focus on analyzing homogeneous graphs until Heterogeneous Graph Attention Network (HAN) is proposed [29] to employ the attention mechanism in the heterogeneous graph. To capture the information from different types of nodes, Hierarchical Graph Attention (HGAT) [17] was proposed for fake detection, and it achieves good performance by aggregating the information from news creators and news subjects. Their work shows that there are biases towards truth or falsity in news that is released by specific authors or belongs to certain groups. A Recent study named Factual News Graph (FANG) [15] was proposed to capture the social context from sources like news, users, and sources into a high fidelity representation by dividing the fake news detection task into textual encoding and stance detection. Compared to other work on fake news detection, FANG is powerful for partial graph nodes.

2.3 Temporal Embedding Network

In the real world, networks are *dynamic* and vary over time in terms of newly appeared nodes and edges. Temporal graph attention (TGAT) [30] was proposed to learn the temporal interactions in the dynamic graphs. Rossi et al. improved TGAT by introducing a temporal memory module. They named this improved model Temporal Graph Network (TGN) [18]. News article networks are also dynamic. There are two types of temporal networks for news articles. The first one is called temporal propagation news networks like the tweets and retweets of news can form a dynamic propagation network. Recent studies like TGNF [23] capture the temporal propagation structure of news from the propagation networks formed by the users with tweet or retweet behaviors. The second type of temporal network are formed by the news with temporal events. News articles are published at different times, news article publishing events would create new nodes and new edges in such a network. None of the previous works attempt to build such a network and capture the temporal information of news neighbors.

2.4 Dynamic Heterogeneous Information Network

The heterogeneous information network like news article network can also be dynamic. Existing methods for analyzing dynamic heterogeneous information networks like heterogeneous temporal graph neural network (HTGNN) [9], heterogeneous dynamic graph attention network (HDGAN) [14] and DyHAN [32] utilize hierarchical aggregation to capture the heterogeneous information and employ cross-time aggregation to exchange the information across graphs of different time slices. Different from the above work, DyHAR [31] use a temporal attentive recurrent neural network (RNN) model to capture evolutionary patterns. However, the above models can't be applied to the heterogeneous network formed by news articles and its related components directly. Because these models aim to learn the temporal information from all the time slices, which would break down the continuous interactions among news articles into pieces. Detailed temporal information may be lost during this process.

3 DEFINITIONS AND PROBLEM FORMULATION

In this section, we introduce some definitions used in this paper and formulate the fake news detection problem.

Definition 1. **News Article**. News articles refer to news content published and propagated on social media platforms. All news articles published at time t and before can be represented by a set $\mathcal{N}(t) = \{n_1^{(t_1)}, n_2^{(t_2)}, \cdots\}$, where $n_i^{(t_i)}$ denotes the news article n_i published at time t_i. We denote the set $\mathcal{N} = \mathcal{N}(t)$ when time t reaches the maximum time (i.e., the time that all news in the dataset has been published).

Definition 2. **Author**. Authors refer people who write news articles. The set of authors can be denoted as $\mathcal{A} = \{a_1, a_2, \cdots\}$.

Definition 3. **Subject**. Subjects refer central ideas of news articles. The set of subjects can be denoted as $\mathcal{S} = \{s_1, s_2, \cdots\}$.

Definition 4. **News-DyHIN**. The news-driven dynamic heterogeneous information network (News-DyHIN) can be defined as $\mathcal{G}(t) = (\mathcal{V}(t), \mathcal{E}(t))$, which consists of the node set $\mathcal{V}(t) = \mathcal{N}(t) \cup \mathcal{A} \cup \mathcal{S}$ and the edge set $\mathcal{E}(t) = \mathcal{E}_a(t) \cup \mathcal{E}_s(t)$. Here, $\mathcal{E}_a(t) = \{(n_i^{(t_i)}, a_j) \in \mathcal{N}(t) \times \mathcal{A} \mid a_j \text{ writes } n_i^{(t_i)}\}$ and $\mathcal{E}_s(t) = \{(n_i^{(t_i)}, s_j) \in \mathcal{N}(t) \times \mathcal{S} \mid n_i^{(t_i)} \text{ belongs to } s_j\}$. In a News-DyHIN, the set of node types $\mathcal{V}_{\text{type}} = \{\text{News article, Author, Subject}\}$, and the set of edge types $\mathcal{E}_{\text{type}} = \{\text{Writes, Belongs to}\}$.

Example. Figure 1 (d) is an example of a News-DyHIN. In the beginning, there are only three authors and two subjects in the News-DyHIN. Each time when a news article is published, it is added to the News-DyHIN, establishing an edge with the authors who write it and the subject to which it belongs. As time passes, more and more news articles will be involved in the News-DyHIN.

Definition 5. **Meta-path** [24]. A meta-path Φ is defined as a sequence in the form of $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \ldots \xrightarrow{R_n} A_{n+1}$, where $A_i \in \mathcal{V}_{\text{type}}$ and $R_i \in \mathcal{E}_{\text{type}}$. It defines a composite relation $R = R_1 \circ R_2 \circ \cdots \circ R_n$ among node types $A_1, A_2, \cdots, A_{n+1}$, where $\circ$ denotes the composition operator on relations.

Example. As shown in Figure 1 (c), two news articles can be connected via two meta-paths: News-Author-News (NAN) and News-Subject-News (NSN). Different meta-paths usually define different semantics. For example, if two news articles are connected via NAN, it means two articles share at least one same author.

Definition 6. **Meta-path based Temporal Neighbors**. Given a meta-path $\Phi \in \{NAN, NSN\}$, a news article $n_i^{(t_i)}$, and a time $t \leq t_i$, the meta-path based temporal neighbors of news article $n_i^{(t_i)}$ at time t are defined as the set $\text{Neighbor}(n_i^{(t_i)}, t) = \{n_j^{(t_j)} \in \mathcal{N}(t) \mid (n_i^{(t_i)}, n_j^{(t_j)}) \in R\}$, where R is the composite relation defined by the meta-path Φ.

Example. Taking Figure 1 (d) as an example, given the meta-path News-Author-News (NAN), the meta-path based temporal neighbors of News Article 2 at time t_1 are the 2-hop neighboring news articles sharing at least one same author with News Article 2 at time t_1 or before, i.e., News Article 1. Similarly, the meta-path based temporal neighbors of News Article 2 at time t_2 are News Article 1 and News Article 3.

Problem Formulation. Based on the concepts defined above, we can define the fake news detection problem formally as follows. Given a News-DyHIN $\mathcal{G}(t) = (\mathcal{V}(t), \mathcal{E}(t))$, the goal is to learn a mapping $\mathcal{F}; \mathcal{N} \rightarrow \tilde{\mathcal{Y}}$, to classify the veracity labels of the news articles in News-DyHIN.

4 THE PROPOSED MODEL

4.1 Framework Overview

In this section, we have a brief overview of our proposed model **D**ynam**i**c **H**ierarchical **A**ttention **N**etwork, namely DiHAN, for fake news detection with the support of News-DyHIN. Our model introduces two levels of attention mechanisms: (1) *temporal node level attention* and (2) *semantic level attention in a hierarchical manner*. Figure 2 (a) shows the entire framework of DiHAN. First, given a News-DyHIN at time t, a temporal node level attention module is introduced, which learns the neighboring news articles' temporal attention weights from the two meta-path based temporal news neighbours, and obtains the temporal node level embedding that aggregate neighboring new articles' temporal information. Then, the semantic level attention learns the weight of each meta-path and fuses representations of the node embeddings from the temporal node level attention layer. The embeddings of news articles that combine temporal information from different meta-paths are obtained and will be fed into a classifier for fake news detection.

4.2 Temporal Node Level Attention

Our temporal node level attention consists of two modules: temporal graph attention (TGAT) and temporal memory module (TMM) to extract comprehensive temporal features from News-DyHIN.

4.2.1 *Temporal Graph Attention.*

The temporal graph attention (TGAT) proposed by Xu et al. [30] is a method to aggregate temporal neighbors and the functional time encoding plays an important role in it. Same to [18, 30], we use the time encoding function, which is inspired by the classical harmonic analysis, and map time t to a feature vector $\phi(t) \in \mathbb{R}^{d_\phi}$ defined by

$$\phi(t) = \cos(\mathbf{w}_t \cdot t + \mathbf{b}_t), \tag{1}$$

where $\mathbf{w}_t, \mathbf{b}_t \in \mathbb{R}^{d_\phi}$ are learned parameters.

In order to apply TGAT on News-DyHIN to learn temporal information, the temporal neighbors used in TGAT are replaced by our meta-path based temporal neighbors. Note that TGAT layer only pays attention to the relative time difference, rather than the absolute time. Hence, given a news article $n_i^{(t)} \in \mathcal{N}$ and we denote its meta-path based temporal neighbors at time t by $\{n_1^{(t_1)}, n_2^{(t_2)}, \ldots, n_n^{(t_n)}\}$, where n in the subscript is the number of the target node's meta-path based temporal neighbors, we can encode the time differences between the target node and its meta-path based temporal neighbors as $\phi(t - t_1), \phi(t - t_2), \ldots, \phi(t - t_n)$, respectively, and the time encoding of the target node with respect to itself is $\phi(t - t) = \phi(0)$. Hence we can obtain the entity-temporal feature matrix which can be defined as follows:

$$\mathbf{Z}(t) = \begin{bmatrix} \mathbf{h}_i(t) \| \phi(0) \\ \mathbf{h}_1(t_1) \| \phi(t - t_1) \\ \vdots \\ \mathbf{h}_n(t_n) \| \phi(t - t_n) \end{bmatrix}. \tag{2}$$

where $\mathbf{h}_i(t), \mathbf{h}_1(t_1), \ldots, \mathbf{h}_n(t_n)$ are features of news articles $n_i^{(t)}$, $n_1^{(t_1)}, \ldots, n_n^{(t_n)}$, respectively, and $||$ denotes the concatenation operator. By forwarding it to three different linear projections, we obtain the Queries, Keys and Values as follows:

$$\begin{cases} \mathbf{Q}(t) = [\mathbf{Z}(t)]_0 \mathbf{W}_Q \\ \mathbf{K}(t) = [\mathbf{Z}(t)]_{1:n} \mathbf{W}_K \\ \mathbf{V}(t) = [\mathbf{Z}(t)]_{1:n} \mathbf{W}_V \end{cases} \tag{3}$$

Where the subscript of $\mathbf{Z}(t_i)$ denotes row slicing, and $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{(d+d_\phi) \times d_h}$ are trainable weight matrices which are used to capture the interaction between node features and time encodings. Self-attention layers [25] can be constructed by applying the scaled dot-product attention, which is defined as follows:

$$\mathrm{Attn}(\mathbf{Q}(t), \mathbf{K}(t), \mathbf{V}(t)) = \mathrm{softmax}\left(\frac{\mathbf{Q}(t)\mathbf{K}(t)^T}{\sqrt{d + d_\phi}}\right)\mathbf{V}(t). \tag{4}$$

The hidden neighborhood representation of news article $n_i^{(t)}$ is the output of the scaled dot-product attention:

$$\hat{\mathbf{h}}_i(t) = \mathrm{Attn}(\mathbf{Q}(t), \mathbf{K}(t), \mathbf{V}(t)). \tag{5}$$

The neighbourhood representation concatenated with the target node features is fed into a feed forward neural network to obtain the node embedding of news article $n_i^{(t)}$ which is given by

$$\begin{aligned} \tilde{\mathbf{h}}_i(t) &= \mathrm{FFN}(\mathbf{h}_i(t) \| \hat{\mathbf{h}}_i(t)) \\ &= \mathrm{ReLU}([\mathbf{h}_i(t) \| \hat{\mathbf{h}}_i(t)]\mathbf{W}_0 + \mathbf{b}_0)\mathbf{W}_1 + \mathbf{b}_1 \end{aligned} \tag{6}$$

where $\mathbf{W}_0 \in \mathbb{R}^{(2d_h + d_\phi) \times d_f}, \mathbf{W}_1 \in \mathbb{R}^{d_f \times d_h}$ are weight matrices, $\mathbf{b}_0 \in \mathbb{R}^{d_f}, \mathbf{b}_1 \in \mathbb{R}^{d_h}$ are bias vectors.

4.2.2 *Temporal Memory Module.*

The temporal memory module (TMM) proposed by Ross et al. [18] enables the model to have the capability to memorize long term dependencies for each node in the graph. The memory of the model at time t consists of a memory vector $s_i(t)$ for each node i appeared so far. The memory of each node is initialized as a zero vector and updated after an interaction with other nodes. Its purpose is to store the historical interaction information of each node in a compressed format. An interaction between news article $n_i^{(t_i)}$ and news article $n_j^{(t_j)}$ through the meta-path can be represented by a message vector which is defined as follows:

$$\mathbf{m}_i(t) = [\mathbf{s}_i(t^-) \| \mathbf{s}_j(t^-) \| \phi(|t_i - t_j|)], \tag{7}$$

where $\mathbf{s}_i(t^-), \mathbf{s}_j(t^-)$ are memory vectors of node i and node j before time t, respectively, and $|t_i - t_j|$ denotes the timespan between news article $n_i^{(t_i)}$ and news article $n_j^{(t_j)}$.

For efficiency reason, TMM employs batch processing to simultaneously capture memory vectors for n previous interaction events,

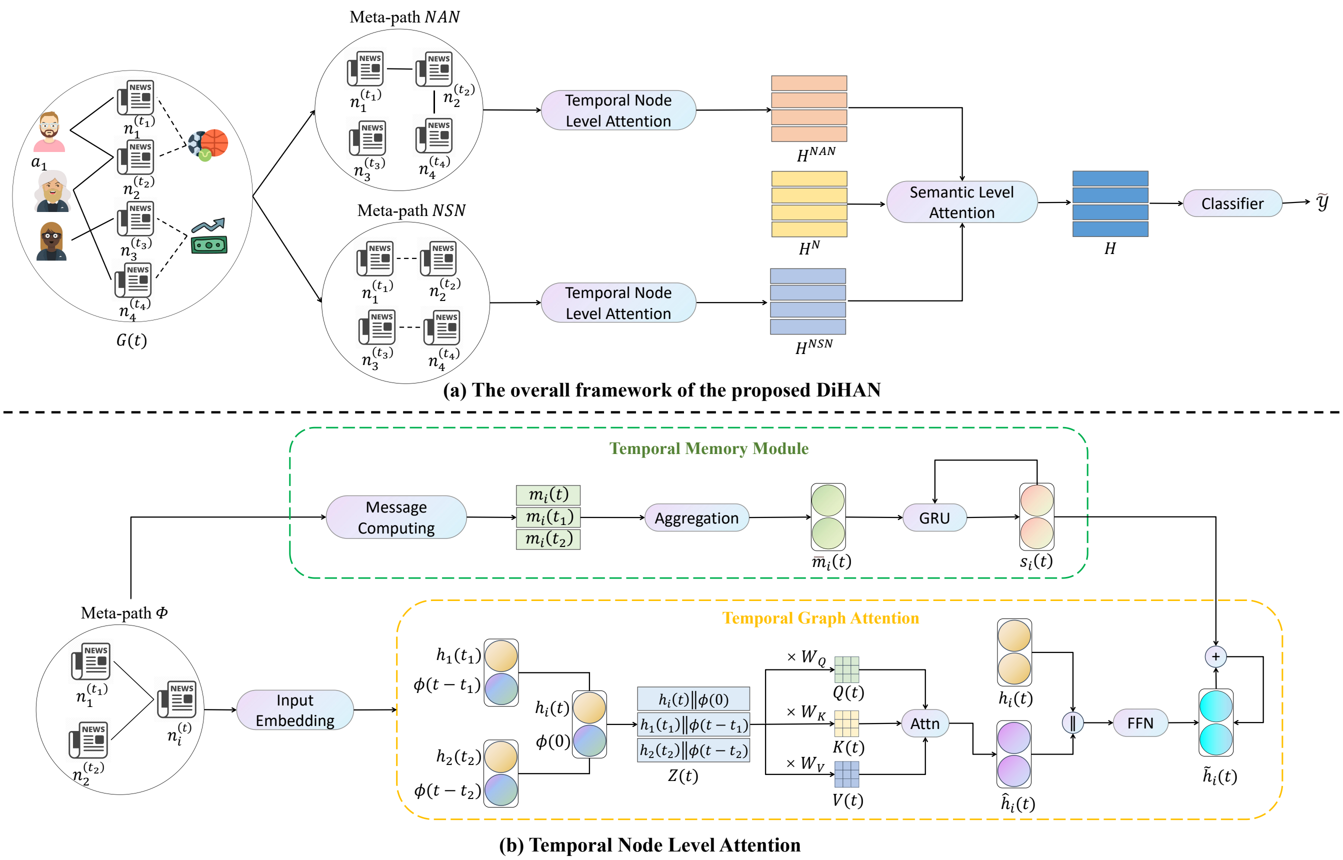

Figure 2: (a) Overall framework of the proposed DiHAN for fake news detection. DiHAN has a hierarchical attention process which includes a temporal node level attention mechanism and a semantic level attention mechanism. (b) The architecture of the temporal node level attention.

which can be defined as follows:

$$\bar{\mathbf{m}}_i(t) = \text{agg}(\mathbf{m}_i(t_1), \mathbf{m}_i(t_2), \dots, \mathbf{m}_i(t_n)), \tag{8}$$

where agg is an aggregation function, and $t_1, t_2, \dots, t_n \leq t$. Here, for the sake of simplicity and efficiency, we keep only the most recent message as the result of aggregation. Then, the memory vector of a node is updated upon each batch of events, which can be defined as follows:

$$\mathbf{s}_i(t) = \text{GRU}(\bar{\mathbf{m}}_i(t), \mathbf{s}_i(t^-)), \tag{9}$$

where GRU [6] plays the role of memory updater here. When node i appears in the next batch at time t, the feature is:

$$\tilde{\mathbf{h}}_i(t) \leftarrow \tilde{\mathbf{h}}_i(t) + \mathbf{s}_i(t). \tag{10}$$

Then, for each meta-path Φ, the temporal node level attention layer will generate the embeddings for news article nodes from News-DyHIN. By stacking them together, we obtain the matrix consisting of news article node embeddings which is defined as:

$$\mathbf{H}^{\Phi} = [\mathbf{h}_i(t), i = 1, \dots, |\mathcal{N}|]. \tag{11}$$

Given the meta-path set $\{NAN, NSN\}$, we can obtain 2 groups of temporal node level embedding $\{\mathbf{H}^{NAN}, \mathbf{H}^{NSN}\}$ by feeding the news article subgraphs into the temporal node level attention layer as shown in Fig 2.

4.3 Semantic Level Attention

Through the temporal node level attention, we obtain different semantic-specific node embeddings of news articles by aggregating different meta-path based temporal neighbors. Since different meta-paths reveal different semantic information, each group of temporal node embeddings based on a given meta-path can only reflect node from one aspect. To fuse the information from different meta-paths, it is necessary to know how important each meta-path is by learning the weights for the meta-paths.

Let $M = \{NAN, NSN, N\}$, where NAN and NSN are meta-paths defined previously, and, to cooperate with the self-attention mechanism, we add a meta-path N which considers news article itself as a neighbor node and we denote the raw input features of news articles in $\mathcal{N}$ by $\mathbf{H}^N$. Then Given a meta-path $\Phi \in M$, the importance of it can be shown as follows:

$$w_{\Phi} = \frac{1}{|\mathcal{N}|} \sum_{n_i^{(t)} \in \mathcal{N}} \mathbf{q}_{\Phi}^T \cdot \tanh(\mathbf{W}_{\Phi} \tilde{\mathbf{h}}_i(t) + \mathbf{b}_{\Phi}), \tag{12}$$

where $\mathbf{W}_\Phi \in \mathbb{R}^{d_h \times d_h}$ is a weight matrix, $b_\Phi \in \mathbb{R}^{d_h}$ is a bias vector, and $\mathbf{q}_\Phi \in \mathbb{R}^{d_h}$ is trainable attention vector. By normalizing the above importance of all meta-paths, we have the importance of meta-path Φ:

$$\beta_\Phi = \mathrm{softmax}(w_\Phi) = \frac{\exp(w_\Phi)}{\sum_{\Phi' \in M} \exp(w_{\Phi'})}. \tag{13}$$

With the learned weights as coefficients, we obtain the final node representations for $\mathcal{N}(t)$ by fusing all semantic-specific embeddings as follows:

$$\mathbf{H} = \sum_{\Phi \in M} \beta_\Phi \cdot \mathbf{H}^\Phi. \tag{14}$$

4.4 Loss Function

After obtaining the final embedding, we use some labeled news article nodes to train an MLP classifier. The classifier consists of a feedforward neural network (FFN) layer and a softmax layer, which are defined as follows:

$$\hat{\mathcal{Y}} = \mathrm{softmax}[\tanh(\mathbf{H}\mathbf{W}_c + \mathbf{b}_c)] \tag{15}$$

where $\mathbf{W}_c \in \mathbb{R}^{d_h \times 2}$ is a weight matrix, $\mathbf{b}_c \in \mathbb{R}^h$ is a bias vector and $\hat{\mathcal{Y}} = (\hat{y}_{i,j}) \in \mathbb{R}^{|\mathcal{N}| \times 2}$. $\hat{y}_{i,0}$ is the predicted probability of news article $n_i^{(t)}$ being real news, whereas $\hat{y}_{i,1}$ is the predicted probability of news article $n_i^{(t)}$ being fake news.

Let $\mathcal{Y} = [y_1, y_2, \ldots, y_m]$ where y_i be the ground-truth label of the news articles. We minimize the binary cross-entropy loss for the labeled news article nodes through backpropagation. The loss function is given as follows:

$$L(\mathcal{Y}, \hat{\mathcal{Y}}) = -\sum_{i=1}^{|\mathcal{N}|} [y_i \log(\hat{y}_{i,1}) + (1 - y_i) \log(\hat{y}_{i,0})]. \tag{16}$$

The overall process of DiHAN is shown in Algorithm 1.

Algorithm 1: The overall process of DiHAN.

Input : The News-DyHIN $\mathcal{G}(t) = (\mathcal{V}(t), \mathcal{E}(t))$,
The initial news feature $\{\mathbf{h}_i(t) : \forall n_i^{(t)} \in \mathcal{N}\}$,
The meta-path set $M = \{NAN, NSN\}$,
The batch size of TMM.
Output: The final news embedding $\mathbf{H}(t)$,
The predicted probabilities $\hat{\mathbf{Y}}$.

for $\Phi \in M$ **do**
  for $n_i^{(t)} \in \mathcal{N}$ **do**
    Find the meta-path based temporal neighbors $\{h_1(t_1), h_2(t_2), \ldots, h_n(t_n)\}$;
    for $j = 1 \ldots n$ **do**
      Calculate time encoding $\phi(t - t_j)$;
    end
    $\mathbf{Z}(t) \leftarrow \begin{bmatrix} \mathbf{h}_i(t) \| \phi(0) \\ \mathbf{h}_1(t_1) \| \phi(t - t_1) \\ \vdots \\ \mathbf{h}_n(t_n) \| \phi(t - t_n) \end{bmatrix}$;
    $\mathbf{Q}(t) \leftarrow [\mathbf{Z}(t)]_0 \mathbf{W}_Q$;
    $\mathbf{K}(t) \leftarrow [\mathbf{Z}(t)]_{1:n} \mathbf{W}_K$,
    $\mathbf{V}(t) \leftarrow [\mathbf{Z}(t)]_{1:n} \mathbf{W}_V$;
    $\hat{\mathbf{h}}_i(t) \leftarrow \mathrm{Attn}(\mathbf{Q}(t), \mathbf{K}(t), \mathbf{V}(t))$;
    $\tilde{\mathbf{h}}_i(t) \leftarrow \mathrm{FFN}(\mathbf{h}_i(t) \| \hat{\mathbf{h}}_i(t))$;
    for *each batch of neighbors* **do**
      for $n_j^{(t_j)}$ *in each batch* **do**
        $\mathbf{m}_i(t_j) \leftarrow [\mathbf{s}_i(t^-) \| \mathbf{s}_j(t^-) \| \phi(t - t_j)]$;
      end
      $\bar{\mathbf{m}}_i(t) \leftarrow \mathrm{agg}(\mathbf{m}_i(t_1), \ldots, \mathbf{m}_i(t_j))$;
      $\mathbf{s}_i(t) \leftarrow \mathrm{GRU}(\bar{\mathbf{m}}_i(t), \mathbf{s}_i(t^-))$;
      $\tilde{\mathbf{h}}_i(t) \leftarrow \tilde{\mathbf{h}}_i(t) + \mathbf{s}_i(t)$;
    end
  end
  $\mathbf{H}^\Phi = [\mathbf{h}_i(t_i), i = 1, \ldots, |\mathcal{N}|]$;
end
for $\Phi \in M \cup \{N\}$ **do**
  Calculate the weight of meta-path w_Φ;
end
$\mathbf{H} = \sum_{\Phi \in M \cup \{N\}} w_\Phi H^\Phi$;
Calculate the prediction probabilities for news articles $\tilde{\mathbf{Y}}$ with Eq. 15;
Update parameters in DiHAN with Eq. 16;

5 EXPERIMENTAL SETUP

This section details our experimental setup, including our research questions, the adopted datasets, and the compared baselines.

5.1 Research Questions

In our experiments, we aim to test the effectiveness of DiHAN and answer the following research questions:

(**RQ1**) Can DiHAN improve the performance of fake news detection and outperform its baselines?

(**RQ2**) Can existing dynamic heterogeneous graph neural networks that use a cross-time aggregation strategy to capture the temporal information from network time slices achieve an expected performance?

(**RQ3**) Can the proposed temporal node-level attention and semantic level attention modules in DiHAN learn the importance of meta-path-based temporal neighbors and semantic relationships, compared to general graph attention?

5.2 Datasets

We adopt two real-world fake news detection datasets: *GossipCop* and *PolitiFact* from the *FakeNewsNet*[1] developed by Shu et al. [20–22]. The news content in two datasets are crawled from two fact-checking platforms: GossipCop[2] and PolitiFact[3]. The most recent version of the datasets contain published time, author and subject information, which means they are suitable for our task. For the construction of News-DyHIN on *GossipCop*, the published time, authors, and subjects of news articles are required, so we removed any news articles that miss these pieces of information.

[1]https://github.com/KaiDMML/FakeNewsNet

[2]https://www.politifact.com/

[3]https://www.gossipcop.com/

Table 1: Statistics of the adopted datasets.

	GossipCop		PolitiFact	
# nodes	fake news	933	fake news	106
	real news	3395	real news	79
	authors	2876	authors	388
	subjects	341	sources	68
# edges	news-author	9859	news-author	467
	news-subject	4328	news-source	185
	NAN	526273	*NAN*	76
	NSN	188931	*NSN*	1920

For the *PolitiFact* dataset, as many news articles lack the information of subjects, so we replace the subject information with the source of the news articles to construct the News-DyHIN. After pre-processing the datasets, the statistics of News-DyHINs for two datasets are shown in Table 1.

5.3 Baselines

- **SVM** [7]: SVM is a classical supervised machine learning method for classification. We use the content of news articles as input to SVM for news classification.
- **DeepWalk** [16]: DeepWalk is a random walk based network embedding method. A logistic regression model is trained to classify the learned news article embeddings.
- **GCN** [13]: GCN is a semi-supervised method based on a variant of convolutional neural networks which apply directly to static homogeneous graphs.
- **GAT** [26]: GAT is an attention-based method used for node classification on static homogeneous graphs, and cannot utilize temporal information.
- **HAN** [29]: HAN is a heterogeneous graph neural network using node-level and semantic-level attentions to capture heterogeneity information but cannot utilize temporal information.
- **HGAT** [17]: HGAT is a heterogeneous graph neural network using a novel hierarchical attention mechanism to learn node embeddings but cannot utilize temporal information.
- **TGN** [18]: TGN is a method on dynamic graphs represented as sequences of temporal events, and it is mainly used for processing homogeneous or bipartite graphs. The News-DyHIN is treated as a dynamic homogeneous graph when testing TGN.
- **HTGNN** [9]: HTGNN is a dynamic heterogeneous graph neural network using both temporal and heterogeneous information to learn node representations. It has three layers: intra-relation aggregation, inter-relation aggregation, and across-time aggregation. Unlike authors or subjects, news articles do not have across-time properties, so in order to apply HTGNN on News-DyHIN, we add an attention layer at the end of HTGNN, which aggregates the information of authors and subjects into news articles.

As DeepWalk, GCN, GAT, HAN, and HGAT can not process dynamic information networks. Hence, the News-DyHIN is treated as a static heterogeneous graph.

5.4 Experimental Settings

The news article nodes are the target nodes to conduct the fake news detection task. We divide the news article nodes into 10 folds, of which 8 folds are used as the training set, 1 fold as the validation set, and 1 fold as the test set. In order to test the performance of the model under different sizes of training sets, we set up our experiments based on different train ratios; we use 4 and 8 among the 8 folds as the training set, respectively. We use *word2vec* in *Gensim* package to produce raw input feature representations of news articles, and the dimension is set up to 128. In the binary classification, since the goal is to detect fake news, we treat fake news articles as the positive class and the real news articles as the negative class.

Classical metrics for classification such as accuracy, precision, recall, and f1-score are used as evaluation metrics in the classification task. Moreover, for the *Gossipcop* dataset, the true and false classes are imbalanced, so the method proposed by Cui et al. [8], which is based on the effective number of samples, is used to handle the imbalance. The weights for each class in the loss function are calculated with $(1-\beta)/(1-\beta^n)$ where $\beta = 0.999$ and n is the size of samples in the class.

The proposed DiHAN and other GNNs are implemented via *Deep Graph Library* [28]. For DiHAN, parameters are randomly initialized and the model is optimized using Adam optimizer [12] with learning rate 0.005 and regularization parameter 0.0005. The dimension of the message vector is set as 64, the dimension of the semantic attention vector is set as 128, the dropout of the temporal node level attention is set as 0.1, the batch size in TMM is set as 5 and the number of attention head is set as 1. For other GNNs, Adam optimizer is adopted and their hyper-parameters are optimized using the validation set. For DeepWalk, the window size is set as 5, length of the random walk is set as 30, the number of walks per node is set as 10. For fair comparison, the embedding dimension of news articles is set as 128 for all the algorithms. We ran each algorithm 5 times and report the average results. The experiments are tested on a Tesla V100-SXM2 GPU with 32GB memory.

6 RESULTS AND ANALYSIS

In this section, we report our experimental results and answer our **RQ1** and **RQ2**.

We first address **RQ1** by analyzing Tables 2 and 3, we can observe that DiHAN achieves the best performance on accuracy and f1-score. Although it does not outperform HGAT and TGN on precision and recall when we use 40% or 80% of the news articles as training data, but it still outperforms all the other baselines at most times. Based on our experiment results, we find that when comparing with DiHAN, HGAT and TGN tend to judge news as "real" which lead to a higher precision and lower recall. In other words, although the precision of HGAT and TGN are higher than DiHAN in some cases, their ability to correctly identify fake news is weaker than DiHAN. Therefore, its high precision is not practical in our experiments.

Table 2: The Results of News Article Classification on *Gossipcop* Dataset

Train	Metric	SVM	DeepWalk	GAT	GCN	HAN	HGAT	TGN	HTGNN	DiHAN
40%	Accuracy	0.5162	0.5823	0.8241	0.8171	0.8169	**0.8335**	0.8217	0.8310	**0.8351**
	Precision	0.1813	0.2403	0.6393	0.6061	0.5932	**0.6791**	0.6066	0.6449	**0.6760**
	Recall	0.3548	0.4762	0.4194	0.4301	0.4968	0.4907	**0.5097**	0.4839	**0.5135**
	F1-score	0.2400	0.3195	0.5065	0.5031	0.5407	**0.5697**	0.5539	0.5529	**0.5836**
80%	Accuracy	0.4560	0.5782	0.8264	0.8209	0.8291	**0.8406**	0.8382	0.8398	**0.8572**
	Precision	0.1966	0.2366	0.6216	0.6021	0.6305	**0.6892**	0.6630	0.6750	**0.7362**
	Recall	0.4946	0.4697	0.4946	0.5090	0.5160	**0.5258**	0.5201	0.4989	**0.5283**
	F1-score	0.2813	0.3147	0.5509	0.5516	0.5675	**0.5965**	0.5829	0.5737	**0.6151**

† The top two are highlighted by **First**, **Second**.

Table 3: The Results of News Article Classification on *PolitiFact* Dataset

Train	Metric	SVM	DeepWalk	GAT	GCN	HAN	HGAT	TGN	HTGNN	DiHAN
40%	Accuracy	0.5395	0.5054	0.6947	0.6421	0.7095	**0.7895**	0.7579	0.7358	**0.7919**
	Precision	0.6283	0.6409	0.7273	0.6476	0.7058	**0.8484**	**0.8077**	0.6917	0.7938
	Recall	0.5815	0.5868	0.8098	0.8546	**0.8551**	0.7855	0.8182	0.8472	**0.8719**
	F1-score	0.6042	0.6127	0.7663	0.7368	0.7734	**0.8157**	0.8129	0.7616	**0.8310**
80%	Accuracy	0.5132	0.4194	0.7158	0.7264	0.7926	**0.8094**	0.7977	0.8074	**0.8210**
	Precision	0.5882	0.5710	0.7130	0.7560	0.7915	0.7949	**0.8440**	**0.8265**	0.8147
	Recall	0.5227	0.4358	**0.8727**	0.8535	0.8537	0.8541	0.8309	0.8165	**0.9273**
	F1-score	0.5535	0.4943	0.7848	0.8018	0.8214	0.8234	**0.8374**	0.8215	**0.8593**

† The top two are highlighted by **First**, **Second**.

From Table 3, we can observe that the evaluation results on the smaller dataset *PolitiFact* are better than on the bigger dataset *GossipCop*. Through careful analysis, we find that this is because the subjects of news articles in *PolitiFact* are replaced with the news sources, which leads to fewer subjects based meta-paths, and the model can easily learn the differences between fake news and real news. Also, the number of real and fake news in the *PolitiFact* dataset is balanced.

In Table 2 and 3, DiHAN shows better performances than HTGNN. However, since HTGNN is designed to process dynamic heterogeneous networks, but HTGNN does not show advantages over other baselines which are unable to process dynamic heterogeneous networks. The heterogeneous temporal network formed for HTGNN divides the timeline of the network into slides and may lose temporal information. For this reason, HTGNN cannot capture the temporal information of news nodes well. Thus, it indicates that existing dynamic heterogeneous graph neural networks that use cross-time aggregation to capture the temporal information from divided time slices cannot achieve good performance for fake news detection, which leads to our answer for the **RQ2**.

Compared DiHAN with HGAT and TGN, we can see that the overall performances of HGAT and TGN are close to DiHAN. They can outperform DiHAN on precision score in some cases. But HGAT exploits more information by aggregating comprehensive textual information for the related components like author and subject, and DiHAN does not incorporate these textual information but only uses the news-to-news mate-paths to build the network. Therefore, we can conclude that the textual information of news related components (i.e., authors and subjects) can benefit the effectiveness of fake news detection to some extent. Even though DiHAN does not employ the textual information of authors and subjects but its accuracy, recall and f1-score are higher than HGAT. These comparison results also show that the temporal graph structure of news is powerful for fake news detection. On the other hand, DiHAN works better on bigger dataset GossipCop than smaller PolitiFact, it is because the semantic and temporal information in of GossipCop is much more than PolitiFact. Thus DiHAN can learn the patterns much easier.

7 ABLATION STUDY

To address **RQ3**, we conducted four different tests to evaluate the impact and effectiveness of the temporal node-level attention module and the semantic level attention module. We compared three variants to assess the following perspectives: (1) the impact of the temporal node-level attention module, (2) the effect of temporal memory information, and (3) the effect of the temporal graph attention module (4) the effect of the semantic level attention module. The variants are designed as follows:

- DiHAN¬N: A variant of DiHAN with both temporal node-level attention modules removed and replaced with fixed equal weights for the temporal neighbors of news nodes from each meta-path.
- DiHAN¬M: A variant of DiHAN with the temporal memory module removed.
- DiHAN¬G: A variant of DiHAN with the temporal graph attention module removed and all the attention weights are set to one.
- DiHAN¬S: A variant of DiHAN with the semantic level attention module removed and all the attention weights are set to one.

Table 4: The comparison of news article classification between DiHAN and the variants on GossipCop dataset.

Model	Accuracy	Precision	Recall	F1-score
DiHAN	0.8572	0.7362	0.5283	0.6151
DiHAN¬N	0.8468	0.7226	0.4772	0.5763
DiHAN¬M	0.8472	0.7547	0.4301	0.5749
DiHAN¬G	0.8519	0.7302	0.4946	0.5897
DiHAN¬S	0.8426	0.6984	0.4731	0.5641

The ablation study was conducted on GossipCop dataset and the results are shown in table 4.

7.1 Effect of the temporal node-level attention

We compare the performance of DiHAN¬N and DiHAN, and as we can see, DiHAN achieves better results in all criteria. This demonstrates that incorporating temporal information of news can improve the performance of fake news detection. Additionally, the performance of DiHAN¬N is more similar to that of HAN. HAN adopts node-level attention to set the weight of neighboring nodes, which indicates that HAN does not learn useful weights from the neighboring nodes. In contrast, DiHAN learns the weights of neighboring nodes using temporal attention, which is more effective for fake news detection.

7.2 Effect of the temporal memory module

We compare the performance of DiHAN¬M and DiHAN and observe that DiHAN outperforms DiHAN¬M. This performance gap arises because the removal of TMM reduces the model's ability to handle temporal dynamics and sequential event information. TMM is crucial for storing and aggregating past information and interactions, which enables the model to remember and use historical data. Without it, the model cannot understand historical context and patterns of events overtime, which impacts the model's ability to make accurate prediction and shows that temporal memory can improve the performance of fake news detection.

7.3 Effect of the temporal graph attention (TGA)

We compare the performance of DiHAN¬G and DiHAN, and we observe that DiHAN outperforms DiHAN¬G. This improvement can be attributed to TGA, which helps the model capture temporal dependencies and interactions within a dynamic network. TGA allows the model to focus on relevant temporal neighbors, and considers the importance of interactions over time. Without TGA, the model lacks temporal context and cannot effectively distinguish between recent and older interactions. This inability to differentiate the temporal relevance of interactions leads to worse performance and shows that temporal information is important in fake news detection.

7.4 Effect of the semantic-level attention (SLA)

We compare the performance of DiHAN¬S and DiHAN and observe that DiHAN outperforms DiHAN¬S. This is because SLA aggregates semantic information from different meta-paths and captures the relationships between various types of nodes in the network. It allows the model to distinguish between important and less important meta-paths, and focuses on the most relevant information. Without SLA, it treats all meta-paths equally and does not prioritise the crucial ones, which shows SLA is important to enhance fake news detection model.

8 CONCLUSION AND FUTURE WORK

In this paper, we proposed to model the dynamics of news articles and associated entities as a News-Driven Dynamic Heterogeneous Information Network (News-DyHIN) which preserves the temporal interactions among news articles. Based on News-DyHIN, we introduced a novel fake news detection model named Dynamic Hierarchical Attention Network (DiHAN). DiHAN employs a hierarchical attention process which includes a temporal node attention mechanism to learn the neighboring nodes' temporal attention weights and a semantic attention mechanism to learn the importance of different meta-paths. We conducted experiments on two real-world datasets and performed an ablation study for our model. The results demonstrate DiHAN can outperform SOTA baselines at most times. Our ablation study also shows that utilizing the temporal information of news and the dynamic structure of different news can improve the performance of fake news detection.

While DiHAN demonstrates effective performance, its scalability could be impacted by the significant growth in dataset size. Future work will focus on optimizing computational efficiency and incorporating distributed processing to handle large-scale, real-time data more effectively. In addition, news content may also be updated on social media platforms. Hence, improving the temporal node level attention to capturing the evolution of news content will be another future work that can extend DiHAN to more applicable scenarios.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives* 31, 2 (2017), 211–36.
[2] Marco T Bastos and Dan Mercea. 2019. The Brexit botnet and user-generated hyperpartisan news. *Social science computer review* 37, 1 (2019), 38–54.
[3] Tian Bian, Xi Xiao, Tingyang Xu, Peilin Zhao, Wenbing Huang, Yu Rong, and Junzhou Huang. 2020. Rumor detection on social media with bi-directional graph convolutional networks. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 549–556.
[4] Hongyun Cai, Vincent W Zheng, and Kevin Chen-Chuan Chang. 2018. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering* 30, 9 (2018), 1616–1637.
[5] Shantanu Chandra, Pushkar Mishra, Helen Yannakoudakis, Madhav Nimishakavi, Marzieh Saeidi, and Ekaterina Shutova. 2020. Graph-based modeling of online communities for fake news detection. *arXiv preprint arXiv:2008.06274* (2020).

[6] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).

[7] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning* 20, 3 (1995), 273–297.

[8] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. 2019. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9268–9277.

[9] Yujie Fan, Mingxuan Ju, Chuxu Zhang, and Yanfang Ye. 2022. Heterogeneous Temporal Graph Neural Network. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*. SIAM, 657–665.

[10] Guoyong Hu, Ye Ding, Shuhan Qi, Xuan Wang, and Qing Liao. 2019. Multi-depth graph convolutional networks for fake news detection. In *CCF International conference on natural language processing and chinese computing*. Springer, 698–710.

[11] Haeseung Jeong. 2021. *Hierarchical Attention Networks for Fake News Detection*. Ph. D. Dissertation. The Florida State University.

[12] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[13] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[14] Qiuyan Li, Yanlei Shang, Xiuquan Qiao, and Wei Dai. 2020. Heterogeneous dynamic graph attention network. In *2020 IEEE International Conference on Knowledge Graph (ICKG)*. IEEE, 404–411.

[15] Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. 2020. Fang: Leveraging social context for fake news detection using graph representation. In *Proceedings of the 29th ACM international conference on information & knowledge management*. 1165–1174.

[16] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 701–710.

[17] Yuxiang Ren and Jiawei Zhang. 2021. Fake news detection on news-oriented heterogeneous information networks through hierarchical graph attention. In *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[18] Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael Bronstein. 2020. Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637* (2020).

[19] Chuan Shi, Yitong Li, Jiawei Zhang, Yizhou Sun, and S Yu Philip. 2016. A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering* 29, 1 (2016), 17–37.

[20] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2018. FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media. *arXiv preprint arXiv:1809.01286* (2018).

[21] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter* 19, 1 (2017), 22–36.

[22] Kai Shu, Suhang Wang, and Huan Liu. 2017. Exploiting Tri-Relationship for Fake News Detection. *arXiv preprint arXiv:1712.07709* (2017).

[23] Chenguang Song, Kai Shu, and Bin Wu. 2021. Temporally evolving graph neural network for fake news detection. *Information Processing & Management* 58, 6 (2021), 102712.

[24] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. 2011. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *Proceedings of the VLDB Endowment* 4, 11 (2011), 992–1003.

[25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[26] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[27] Hanna M Wallach. 2006. Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd international conference on Machine learning*. 977–984.

[28] Minjie Wang, Lingfan Yu, Da Zheng, Quan Gan, Yu Gai, Zihao Ye, Mufei Li, Jinjing Zhou, Qi Huang, Chao Ma, Ziyue Huang, Qipeng Guo, Hao Zhang, Haibin Lin, Junbo Zhao, Jinyang Li, Alexander J. Smola, and Zheng Zhang. 2019. Deep Graph Library: Towards Efficient and Scalable Deep Learning on Graphs. *CoRR* abs/1909.01315 (2019). arXiv:1909.01315 http://arxiv.org/abs/1909.01315

[29] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous graph attention network. In *The world wide web conference*. 2022–2032.

[30] Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. 2020. Inductive representation learning on temporal graphs. *arXiv preprint arXiv:2002.07962* (2020).

[31] Hansheng Xue, Luwei Yang, Wen Jiang, Yi Wei, Yi Hu, and Yu Lin. 2020. Modeling dynamic heterogeneous network for link prediction using hierarchical attention with temporal rnn. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 282–298.

[32] Luwei Yang, Zhibo Xiao, Wen Jiang, Yi Wei, Yi Hu, and Hao Wang. 2020. Dynamic heterogeneous graph embedding using hierarchical attentions. In *European Conference on Information Retrieval*. Springer, 425–432.