Neurocomputing 505 (2022) 362–374

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Dynamic graph neural network for fake news detection ☆

Chenguang Song, Yiyang Teng, Yangfu Zhu, Siqi Wei, Bin Wu *

Beijing Key Laboratory of Intelligence Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, PR China

ARTICLE INFO

Article history:
Received 26 October 2021
Revised 19 April 2022
Accepted 13 July 2022
Available online 18 July 2022
Communicated by Zidong Wang

Keywords:
Fake news detection
Dynamic propagation
Diffusion networks

ABSTRACT

The widespread of fake news on social media and other platforms can bring significant damage to the harmony and stability of our society. To defend against fake news, researchers have suggested various ways of dealing with fake news. In recent years, fake news detection has become the research focus in both academic and industrial communities. The majority of existing propagation-based fake news detection algorithms are based on *static* networks and they assume the whole information propagation network structure is readily available before performing fake news detection algorithms. However, real-world information diffusion networks are dynamic as new nodes joining the network and new edges being created. To address these shortcomings, we proposed a dynamic propagation graph-based fake news detection method to capture the missing dynamic propagation information in *static* networks and classify fake news. Specifically, the proposed method models each news propagation graph as a series of graph snapshots recorded at discrete time steps. We evaluate our approach on three real-world benchmark datasets, and the experimental results demonstrate the effectiveness of the proposed model.

1. Introduction

In recent years, we have witnessed a rise in the success of a number of online social media platforms such as Twitter[1], Facebook[2] and Sina Weibo[3]. It not only allows us to connect with people we never thought, but give us the opportunity to exchange of opinions and news propagation faster like never before. However, online social media platforms for news consumption is a double-edge sword because it drives the spread of fake news at the same time [2]. One possible explanation is that, compared to traditional news media that usually requires extensive research, fact checking and accurate coverage in order to be a reliable news resource, the absence of effective regulatory and fact-checking measures over each piece of news makes fake news can be easily created and published online for primary motivations of influencing opinions and seeking tempting rewards at low cost [3–6], which results in the fact that online social media platforms have become a primary source for spreading fake news [7].

Because the proliferation of fake news on social media may confuse and misguide public opinions, change the way people respond to real news and even disturb the social order [8,9], it has been listed by the World Economic Forum (WEF) as one of the main threats to our society [10]. Many works have shown that human beings are not good at distinguishing fake news from fake news [11]. To deal with fake news on social media, great efforts have been growing in fact-checking, however, largely centered on manual identification by a small group of highly credible fact-checkers. Unfortunately, manual fact-checking is labor-intensive and has difficulty in scaling with the volume of emerging fake news [5,11,12]. Consequently, it is essential to study computational fake news detection.

The inherent openness of social media platforms provides opportunities to trace and study the digital footprints of fake news [13]. By studying how users share and discuss fake news, we can develop effective detection and intervention techniques to automatically assess their veracity and avoid their dramatic effects [14]. A typical example is that a series of engineering features manually designed are fed into a machine learning-based algorithm to evaluate the authenticity of the given news [15]. However, the hand-craft feature extraction task is time consuming and poor in generalizability, and may result in biased features [16]. Recently, deep learning algorithms have been widely applied to many tasks such as sentiment analysis [17], fake news detection [18] and question answering [19], and have shown dramatic potential for capturing complex patterns automatically. Deep learning

☆An earlier version of this paper was presented at the 2021 7th IEEE International Conference on Cloud Computing and Intelligence Systems at Xian, China [1].

* Corresponding author.

E-mail addresses: scg@bupt.edu.cn (C. Song), tengyiyang@bupt.edu.cn (Y. Teng), zhuyangfu@bupt.edu.cn (Y. Zhu), Weisiqi@bupt.edu.cn (S. Wei), wubin@bupt.edu.cn (B. Wu).

[1] https://twitter.com/.
[2] https://www.facebook.com/.
[3] https://weibo.com/.

https://doi.org/10.1016/j.neucom.2022.07.057
0925-2312/

techniques can be fed with raw data, which means that deep learning-based fake news detection algorithms can bypass feature engineering [11]. Consequently, we has witnessed that a lot of deep learning-based fake news detection methods are proposed to mitigate the shortcomings of traditional machine learning methods [20].

Social media news spreads in the form of shares and re-shares of the source and shared posts, resulting in an information diffusion network [21,22]. Previous work has proven that fake and real news show difference in propagation patterns, which tells us that the news propagation network can be used to improve the performance of fake news detection algorithm [23–25]. Moreover, it is difficult for the individual users to control the spread patterns of news on social networks, which implies that propagation-based approaches may have better robustness [26,5]. Hence, much efforts have been devoted to investigated that how the news propagation network on social media can help to detect fake news. [27–31]. They have achieved remarkable success in fake news detection task.

Despite these advances, a major challenge now is that, most of existing news propagation network-based fake news detection algorithms overwhelmingly depend on *static* news propagation graphs, assuming the entire news propagation graph is readily available before performing fake news detection algorithms [5,32]. In fact, real-world information propagation networks are dynamic as new nodes joining the network and new edges being created. Fig. 1 illustrates the difference between *dynamic* and *static* news propagation networks. As shown in Fig. 1 (left), it shows a discrete-time dynamic news dissemination graph where each network represents a static graph snapshot and records user propagating behaviors occurs before time stamps $[t_1, t_2, \cdots, t_T]$, respectively. In contrast, Fig. 1 (right) depicts a *static* news propagation network which only presents the structural information of news propagation network without dynamic evolutionary patterns. Recently studies have been conducted showing the strong relationships of temporal engagement features of users with authenticity of social media news [33–37]. Fig. 2 displays the average number of tweets or posts [4] on three real-world and widely used public benchmark datasets [5]. The comparative analysis between fake and real news may help to understand why we need to build a dynamic propagation-based fake news detector. From Fig. 2, we can easily find that differences of temporal propagation patterns between fake and real news. To effectively utilize temporal information to improve the performance of fake news detection model, one essential task we confront is that build a time-aware fake news detection method to model temporal patterns of news propagation network.

Present work. To capture the missing dynamic propagation information, we propose a dynamic propagation network-based fake news detection architecture named **D**ynamic **G**raph **N**eural network for **F**ake news detection (DGNF). Specifically, we model the news propagation networks using the discrete-time dynamic graph (DTDG) in this work. We first aggregate temporal information of a news propagation graph into a sequence of static graph snapshots. Then, DGNF generates a dynamic representation for each snapshot using both structure-aware module and temporal-aware module. The structure-aware module is responsible for extracting features from local node neighborhoods and capturing local structural information in each static graph snapshot. The temporal-aware module is designed to capture temporal variations in the graph structure by flexibly weighting historical node representations over discrete time steps. Note that we extend the standard DGNF into two variants (i.e., DGNF-*tsn* and DGNF-*tcn*) with different temporal-aware modules (i.e., Temporal Self-Attention Networks (TSN) and Temporal Convolutional Networks (TCN) [38–40]). In summary, the contributions of this work include:

- In this paper, we study a novel problem of discrete-time dynamic news propagation network-based fake news detection task.
- We propose a discrete-time dynamic news propagation graph-based fake news detection framework named DGNF to capture dynamic evolution patterns and network structural information of social media news diffusion graph.
- We conduct extensive experiments on three real-world datasets to examine the performance of DGNF-*tsn* and DGNF-*tcn*, and the experimental results demonstrate the effectiveness of the proposed methods.

The rest of this paper is organized as follows. We briefly review the related work on fake news detection in Section 2. Section 3 introduces notations and formally defines the problem of fake news detection. Section 4 introduces the DGNF-*tsn* model and DGNF-*tcn* model. Section 5 describes the datasets and baselines used in our experiments, and then provide experimental settings and result analyses. Section 6 concludes the paper and discusses directions for future work.

2. Related Work

To date, a considerable amount of methods have been proposed to identify fake news using various features such as text, user and propagation. In this section, we review the existing work from two aspects: content-based and propagation-based fake news detection.

Content-based Fake News Detection. Normally, news content from online social media platforms is typically represented by tweets, replies to those tweets, and several attached images [8]. Because this news that needs to be verified mainly textual content, text are the most explicit features for identifying fake news [4,41,42], which is the main reason that some early researchers seek to directly evaluate news authenticity by quantify the difference of textual features between fake and real news [11,21]. A typical example is that a series of engineering features manually designed which can be topic, Bag of Words, or n-gram features, are then fed into one or more machine learning-based algorithms to evaluate the authenticity of the given news [43–45]. Although hand-crafted linguistic features-based methods showing promising results, these algorithms cannot fully capture fine-grained linguistic information and show poor performance in generalizing the information across disciplines [21].

To overcome the drawbacks, deep learning algorithms have been widely applied in fake news detection in part because they can automatically extract latent feature representation and capture complex contextual patterns from raw news text. A representative example is that, Ma et al. first designed a recurrent neural network (RNN)-based algorithm to better capture long term dependencies using most popular RNN architecture (i.e., Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)) [46]. Inspired by the promising results convolutional neural networks (CNN) have obtain on many text classification tasks, Wang et al. [47] later proposed a set of simplified CNN-based baseline algorithms that used only one layer of convolution on pre-trained word2vec embeddings [48]. By taking advantage of the attention mechanism can learn automatically importance weights that can be used to explain the contributions of words and sentences to a target claim, scholars have applied attention mechanism to interpret the identification of fake news detection algorithms [20,49]. Recently,

[4] The meaning of tweet is equal to post used in this paper.

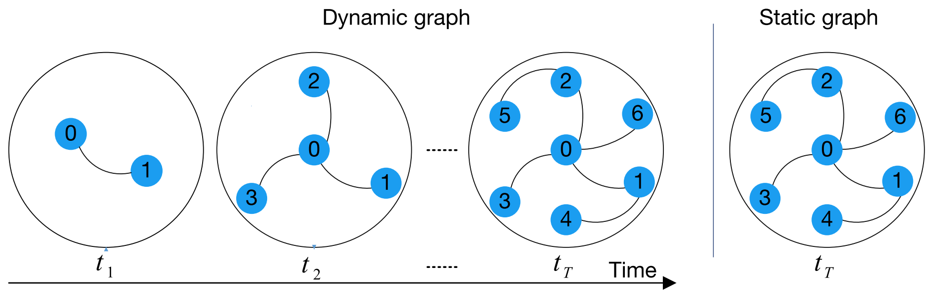


Fig. 1. Each network records users propagating behaviors occurs before time stamps $[t_1, t_2, \cdots, t_T]$, respectively. Each node denotes a tweet and each edge represents a share or retweet. Left: discrete-time dynamic propagation network of a piece of social media news. Right: static propagation network of a piece of social media news.

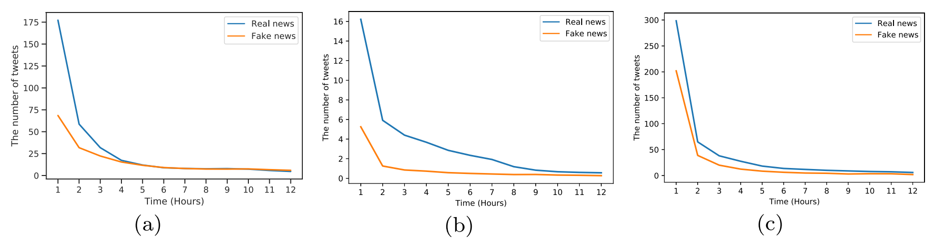


Fig. 2. (a) The average number of tweets for Weibo dataset at different timestamps; (b) The average number of tweets for FakeNewsNet dataset at different timestamps; (c) The average number of tweets for Twitter dataset at different timestamps [5].

introducing external knowledge has become an increasingly popular way to help models understand text in deep learning for Natural Language Processing (NLP). Researchers have explored external knowledge in spotting fake news detection because explaining the news content usually need enough background or professional knowledge [50,51,9]. In addition to the above work, academic researchers also have used generative and adversarial training techniques[52], domain adaption techniques [53,54], weak supervision techniques, Sentiment or emotion information [55–57], or visual cues [58] to detect fake news on social media.

Propagation-based Fake News Detection. In addition to the news content-based methods, more recent work highlighted the importance of social media news propagation patterns, also known as cascades, for identification of fake news [22,59–61]. The underlying assumption of propagation-based fake news detection is that the propagation patterns of fake news differ from true ones in some quantifiable way [59]. Propagation-based fake new detection methods often combine text-based data with structural information. To reflect the collective behavior of users engaging with spreading fake news, the propagation patterns of social media news are usually modeled as directed or undirected graphs [22]. According to the difference between nodes in interactions such as user-to-user or user-to-content, established work in propagation-based fake news detection generally falls in two categories: homogeneous and heterogeneous graph-based approaches [2,62]. We usually call a news propagation network with a single type of nodes and edges as homogeneous graph, and with two or more types of nodes and edges as heterogeneous graph. For instance, Ma et al. [63] proposed a RNN-based algorithm (i.e., RvNN) that infers fake news veracity based on the cascade structure of news propagation. Specifically, this algorithm models information cascade as top-down and bottom-up the news propagation tree structure. RvNN is a typical homogeneous graph-based approach because each node represents a tweet or reply and each edge denotes tweet-retweet relationship. Following the similar idea with [63], Bian et al. [64] proposed the first graph neural network (GNN)-based model, Bi-directional graph convolutional networks (BiGCN), for fake news detection by learning on both the top-down diffusion and bottom-up propagation of social media news. BiGCN outperformed RvNN and other competitive methods by a substantial margin on three benchmark datasets. Yang et al. [65] unified heterogeneous information network representation and graph adversarial learning in a multi-task learning framework to detect fake news. Nguyen et al. [66] proposed the Factual News Graph (FANG) model which is an inductive heterogeneous network representation architecture, explores relationship among news article, sources and users, and could predict the veracity of social media news spreading online by mining social structure and engagement patterns of individuals.

Whereas the aforementioned approaches have been gained some positive results, fewer efforts have been devoted to modeling dynamic news cascade. Here, we purposefully focus only on homogeneous DTDG. The studies closely related to this work are [5,67]. Because [5] models news propagation patterns using homogeneous continuous-time dynamic graphs (CTDG), it is difficult to make a fair comparison between [5] and this work directly. Choi et al. [67] models news propagation patterns using homogeneous DTDG. In [67], the authors also employed self-attention mechanism to capture temporal information. DGNF-*tsn* is different from [67] in following aspects: First, one of the important differences between [67] and DGNF-*tsn* is that we introduce a special visible matrix to prevent information leakage from the future into the past; Second, we use the positional embeddings to help temporal attention to capture ordering information, and additional residual connections to improve the stability of the model; Third, our proposed model learns the temporal propagation patterns from the node-level, which leads to a more fine-grained nodal relationship modeling for dynamic news propagation graphs; Fourth, the output of temporal attention in DGNF-*tsn* is the feature representations of last snapshots, while the output of attention mechanism in [67] is the average of feature embeddings of all snapshots.

3. Problem Formulation

Similar to [5], let $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ be an undirected unweighted static graph representing a news propagation network. $\mathscr{V} = \{v_1, \cdots, v_i, \cdots, v_{\mathscr{N}}\}$ is the set of nodes and $\mathscr{E}$ is the set of edges, where v_i represents a tweet, $\mathscr{N}$ represents the number of relevant tweets in $\mathscr{G}$, and $\mathscr{M}$ denotes the number of the observed interaction events. $\mathbf{h}_i \in \mathbb{R}^d$ is the feature representation of tweet v_i. Each edge $e_{ij} \in \mathscr{E}$ denotes node v_i has a response to v_j, and also can be formulated as an undirected unweighted adjacency matrix $\mathscr{A} = [a_{ij}]_{\mathscr{N}\times\mathscr{N}}$, where

$$a_{ij} = \begin{cases} 1 & if \;\; e_{ij} \in \mathscr{E} \\ 0 & otherwise \end{cases} \tag{1}$$

Thus, we can define an undirected unweighted discrete-time dynamic news propagation network as a series of static graph snapshots $\mathbb{G} = \{\mathscr{G}(1), \cdots, \mathscr{G}(t), \cdots, \mathscr{G}(T)\}$ where n is the number of sessions. Each snapshot $\mathscr{G}(t) = \{\mathscr{V}(t), \mathscr{E}(t)\}$ consists of the nodes set $\mathscr{V}(t) = \left\{v_1^{(t_1)}, \cdots, v_i^{(t_i)}, \cdots, v_{\mathscr{N}(t)}^{(t_{\mathscr{N}(t)})}\right\}$ and edges set $\mathscr{E}(t)$where $t_i \leqslant t$. Note that the nodes set $\mathscr{V}(t)$ and the edges set $\mathscr{E}(t)$could change over time. Each node $v_i^{(t_i)} \in \mathscr{V}(t)$indicates that the *i-th* tweet v_i is published at time point t_i. Each edge $e_{ij}^{(t_j)} \in \mathscr{E}(t)$ denotes that node v_j has a response to v_i at time point t_j where $t_i \leqslant t_j \leqslant t$. $\mathscr{N}(t) = |\mathscr{V}(t)|$ is the number of nodes in $\mathscr{G}(t)$. $\mathscr{M}(t) = |\mathscr{E}(t)|$ is the number of edges in $\mathscr{G}(t)$. $\mathbf{h}_i(t) \in \mathbb{R}^d$ is the feature representation of tweet $v_i^{(t_i)}$ at time stamp t. $\mathscr{A}(t) = [a_{ij}(t)]_{\mathscr{N}(t)\times\mathscr{N}(t)}$ is the undirected unweighted adjacency matrix of $\mathscr{G}(t)$, where

$$a_{ij}(t) = \begin{cases} 1 & if \;\; e_{ij}^{(t_j)} \in \mathscr{E}(t) \\ 0 & otherwise \end{cases} \tag{2}$$

When $t = T, \mathscr{G} = \mathscr{G}(T), \mathscr{V} = \mathscr{V}(T), \mathscr{A} = \mathscr{A}(T)$ and $\mathscr{N} = \mathscr{N}(T)$. $\mathbb{G}$ is associated with a ground-truth label $y \in \{0, 1\}$ describing its veracity, where y = 0 indicates $\mathbb{G}$ is true news, and $y = 1$means $\mathbb{G}$ is fake news. We formulate the fake news detection task in this paper as follows.

Problem Definition: Given a collection of static news propagation network snapshots $\mathbb{G} = \{\mathscr{G}(1), \cdots, \mathscr{G}(t), \cdots, \mathscr{G}(T)\}$ over discrete time stamps, this paper aims to learn a mapping function $\mathscr{F} : \mathscr{F}(\mathbb{G}) \rightarrow \hat{y}$ to give a predicted label for $\mathbb{G}$.

4. Model

4.1. Model Framework

Fig. 3 provides an overview of the proposed framework. The structure-aware module and temporal-aware module are fundamental modules of the DGNF. The input is a collection of static graph snapshots $\mathbb{G}$, and the output is a corresponding class label. For a static graph snapshot $\mathscr{G}(t)$ from $\mathbb{G}$, the model first produces the raw feature representation of each node in $\mathscr{G}(t)$ via input embedding layer. Second, the adjacency matrix $\mathscr{A}(t)$ and node feature representations of $\mathscr{G}(t)$ are feed as input to the structure-aware module to extracts features from local node neighborhoods. Third, the sequences of node representations output by structure-aware module is fed into the temporal-aware module to capture network dynamic evolutionary patterns. At last, we can use any dynamic embeddings of a static graph snapshot $\mathscr{G}(t)$ to predict the veracity of $\mathbb{G}$.

4.2. Input Embeddings

The input embedding layer is designed to create corresponding raw feature representation for each tweet. Each word of node $v_i^{(t_i)}$ is first mapped into a sequence of pretrained word embeddings:

$$[\omega_1, \cdots, \omega_j, \cdots] \leftarrow \text{WordEmbed}(v_i^{(t_i)}) \tag{3}$$

where $\omega_j \in \mathbb{R}^d$. The raw node $v_i^{(t_i)}$ feature representation of a tweet is the average of all of the word vectors (i.e., $\mathbf{x}_i(t) \in \mathbb{R}^d$).

4.3. Structure-Aware Module

The structure-aware component is composed of multiple stacked graph attention networks (GAT) [68] to extracts features from local node neighborhoods in each snapshot. For a multilayer GAT, given a node $v_i^{(t_i)}$ with a hidden state $\mathbf{h}^l \in \mathbb{R}^{1\times d^l}$ at layer l, GAT can update hidden state of node $v_i^{(t_i)}$ at layer $l+1$ with the following layer-wise propagation rule:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\sum_{j\in N_i} \alpha_{ij}^l \mathbf{h}_j^l \mathbf{W}^l\right) \in \mathbb{R}^{1\times d^{(l+1)}} \tag{4}$$

where N_i is the neighborhood of node $v_i^{(t_i)}$, α_{ij}^l is the attention value of node $v_i^{(t_i)}$ to $v_j^{(t_j)}$ in the l^{th} layer, $\mathbf{h}_j^l$ is feature representation of node $v_j^{(t_j)}$ and output by $(l-1)^{th}$ layer, $\mathbf{W}^l \in \mathbb{R}^{d^l\times d^{(l+1)}}$ is a trainable weight matrix, and $\sigma(\cdot)$ is an activation function. Note that, when $l = 0, \mathbf{h}_j^0 = \mathbf{x}_j(t)$ and $d^0 = d$. Specifically, the attention coefficient between node $v_i^{(t_i)}$ and $v_j^{(t_j)}$ can be expressed as:

$$\alpha_{ij}^l = \frac{\exp\left(\text{LeakReLU}\left(\left[\mathbf{h}_i^l\mathbf{w}^l \middle\| \mathbf{h}_j^l\mathbf{w}^l\right]\mathbf{w}\right)\right)}{\sum_{k\in N_i}\exp\left(\text{LeakReLU}\left(\left[\mathbf{h}_i^l\mathbf{w}^l \middle\| \mathbf{h}_k^l\mathbf{w}^l\right]\mathbf{w}\right)\right)} \tag{5}$$

where $\mathbf{w} \in \mathbb{R}^{2d^{(l+1)}\times 1}$ is weight vector. For a node $v_i^{(t_i)}$ in a static graph snapshot $\mathscr{G}(t)$, its feature representation can be denoted as $\mathbf{h}_i^{l+1}(t) \in \mathbb{R}^{1\times d^{(l+1)}}$. Thus, in the range $1 \leqslant t \leqslant T$, the feature representation sequence of node $v_i^{(t_i)}$ is defined as:

$$\mathbf{H}_i^{(l+1)} = \left[\mathbf{h}_i^{(l+1)}(1), \cdots, \mathbf{h}_i^{(l+1)}(t), \cdots, \mathbf{h}_i^{(l+1)}(T)\right] \in \mathbb{R}^{n\times d^{(l+1)}} \tag{6}$$

4.4. Temporal-Aware Module

We can get DGNF-*tsn* model and DGNF-*tcn* model by designing two variants of DGNF models based on TSN and TCN to encode the temporal information from the sequence of node representations output by structure-aware module. Specifically, we use temporal convolutional and temporal self-attention blocks to aggregate node features over time, respectively. *The output of temporal-aware module has the same dimension with its input.*

Temporal convolutional block. The temporal convolutional component is mainly composed of multiple stacked TCN layers. As shown in Fig. 4, we provide an illustration to help readers easily understand the TCN block, the key module of which is TCN. Formally, the TCN of the $(r+1)^{th}$ TCN layer can be written as:

$$\mathrm{F}^{(r+1)}(\mathbf{U}_i^r(t)) = \left(\mathbf{U}_i^r *_{\tau_{(r+1)}} \mathrm{f}\right)(t) = \sum_{j=0}^{z-1} \mathrm{f}_j^T \mathbf{U}_i^r(t - \tau_{(r+1)} j) \tag{7}$$

where $*$ is 1D convolution operator, $\mathrm{f} \in \mathbb{R}^{z\times d^{(l+1)}}$ is a convolutional filter with size z, $\tau_{(r+1)}$ is the dilation factor which can be set as $(z-1)^r$ to obtain exponential growth receptive field, and $\mathbf{U}_i^r(\leqslant 0) := 0$. When $r = 0$, for any node $v_i^{(t_i)}$, the input of 1^{th} TCN layer is $\mathbf{H}_i^{l+1} = \mathbf{U}_i^0$. We add weight norm, ReLU activation function and dropout operation to each TCN layer. Thus, the output of $(r+1)^{th}$ TCN layer is:

$$\mathbf{U}_i^{(r+1)} = \left[\mathrm{F}^{(r+1)}(\mathbf{U}_i^r(1)), \cdots, \mathrm{F}^{(r+1)}(\mathbf{U}_i^r(t)), \cdots, \mathrm{F}^{(r+1)}(\mathbf{U}_i^r(T))\right] \tag{8}$$

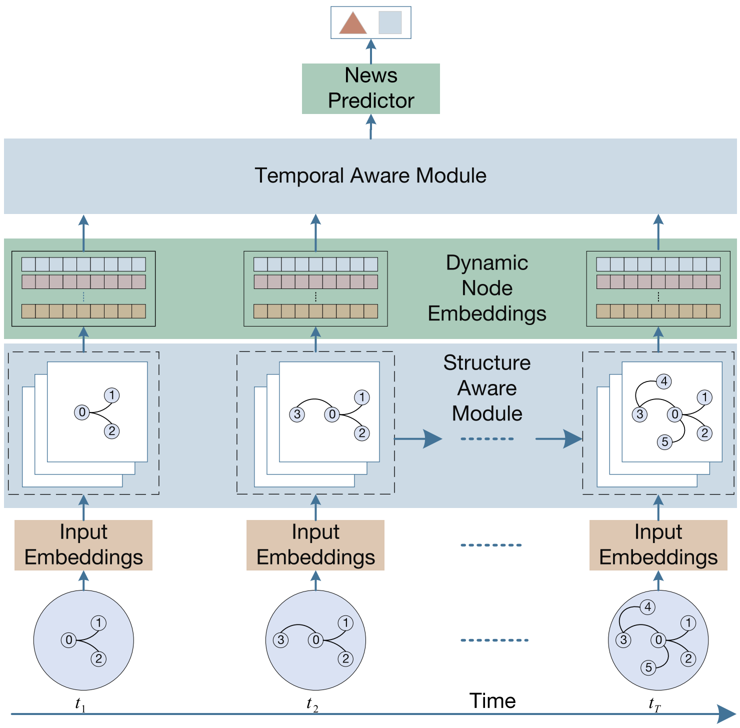


Fig. 3. The proposed model framework.

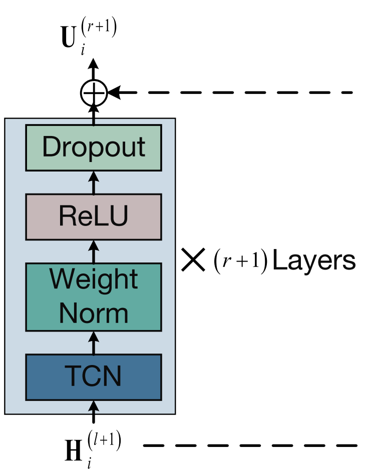


Fig. 4. An illustration for the TCN blocks.

To make $\mathbf{U}_i^{(r+1)}$ and $\mathbf{H}_i^{(l+1)}$ have the same dimension (i.e., $\mathbf{U}_i^{(r+1)} \in \mathbb{R}^{n \times d^{(l+1)}}$), the number of filters of each TCN layer is set as $d^{(l+1)}$. Specifically, the TCN is constructed based on causal convolution and dilated convolution. The causal convolution ensures the output $\mathbf{U}_i^{(r+1)}(t)$only depend on input $\mathbf{H}_i^{(l+1)}(\leqslant t)$, which means that there is no information leakage from the future to past. The dilated convolution makes receptive fields exponential to the number of TCN layers. By adjusting dilation factor τ, size of filter f and the number of TCN layers, the Temporal convolutional module can achieve flexible receptive fields, and explore the full temporal information.

Temporal self-attention block. In this subsection, we present another temporal-aware module (i.e., temporal self-attention block), the input of which is the series of representations for a particular node $v_i^{(t_i)}$ recorded at n different time steps (i.e., $\mathbf{H}_i^{(l+1)}$). The temporal self-attention block is a stack of multiple TSN layers. Same to TCN, TSN is also designed to capture graph evolution at different time steps. To enable each node representation in $\mathbf{H}_i^{(l+1)}$ carries unique position information, we used the positional encoding (PE) function introduced in [69] to embed temporal position of each snapshot. The function is defined as:

$$\mathrm{PE}(pos, 2j) = \sin\left(pos/10000^{2j/d^{(l+1)}}\right) \tag{9}$$

$$\mathrm{PE}(pos, 2j+1) = \cos\left(pos/10000^{2j/d^{(l+1)}}\right) \tag{10}$$

where $pos \in [0, \cdots, pos, \cdots, n-1]$ represents temporal position of each snapshot, $j \in \left[0, d^{(l+1)}/2\right)$ refers to the the dimension. For any node $v_i^{(t_i)}$, the position embeddings $\mathbf{P} \in \mathbb{R}^{d^{(l+1)}}$ would be combined with $\mathbf{H}_i^{(l+1)}$ to obtain the following sequence of representations across n time steps:

$$\begin{aligned}\mathbf{H}_i^{(l+1)} + \mathbf{P} = \Big[&\mathbf{h}_i^{(l+1)}(1) + \mathbf{p}(1), \cdots, \mathbf{h}_i^{(l+1)}(t) \\ &+\mathbf{p}(t), \cdots, \mathbf{h}_i^{(l+1)}(T) + \mathbf{p}(T)\Big] \in \mathbb{R}^{n \times d^{(l+1)}}\end{aligned} \tag{11}$$

Intuitively, $\mathbf{h}_i^{(l+1)}(\leqslant t)$ not contribute equally to $\mathbf{h}_i^{(l+1)}(t)$ in a given dynamic graph. Thus, self-attention is naturally way to capture these relationships. In this paper, TSN adopts the scaled dot-product form of attention. Therefore, for a particular node $v_i^{(t_i)}$, its Queries, Keys and Values can be written as: $\mathbf{Q}_i = (\mathbf{H}_i^{(l+1)} + \mathbf{P}) \times \mathbf{W}_Q, \mathbf{K}_i = (\mathbf{H}_i^{(l+1)} + \mathbf{P}) \times \mathbf{W}_K$ and $\mathbf{V}_i = (\mathbf{H}_i^{(l+1)} + \mathbf{P}) \times \mathbf{W}_V$, where $\mathbf{W}_Q \in \mathbb{R}^{d^{(l+1)} \times d^{(l+1)}}, \mathbf{W}_K \in \mathbb{R}^{d^{(l+1)} \times d^{(l+1)}}$ and $\mathbf{W}_V \in \mathbb{R}^{d^{(l+1)} \times d^{(l+1)}}$. Formally, the self-attention can be formulated as:

$$\mathrm{Att}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) = \mathrm{softmax}(\mathbf{Q}_i \times \mathbf{K}_i^\top / \sqrt{\mathrm{d}^{(l+1)}}) \times \mathbf{V}_i \tag{12}$$

Same to causal convolution operation in TCN, TSN should only depend on the historical representations of each node. In other words, TSN also need to ensure that past information should not be visible to future. To prevent information leakage from the future into the past, a special visible matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ is defined as follows.

$$\mathbf{M}_{ij} = \begin{cases} 0 & t_i l225 \leqslant t_j \\ -\infty & t_i > t_j \end{cases} \tag{13}$$

Formally, the temporal self-attention can be defined as:

$$\mathrm{Att}_{mask}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) = \mathrm{softmax}[(\mathbf{Q}_i \times \mathbf{K}_i^\top + \mathbf{M})/\sqrt{\mathrm{d}^{(l+1)}}] \times \mathbf{V}_i \in \mathbb{R}^{n\times \mathrm{d}^{(l+1)}} \tag{14}$$

Similar to previous work [69], multihead attention mechanism is also adopted in this work. Thus, temporal multihead self-attention can be formulated as:

$$\mathrm{Att}_{mask}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) = \mathrm{Concat}(\mathrm{head}^1, \cdots, \mathrm{head}^j, \cdots, \mathrm{head}^H) \times \mathbf{W}_O \in \mathbb{R}^{n\times d^{(l+1)}} \tag{15}$$

$$\mathrm{head}_i^j = \mathrm{softmax}[(\mathbf{Q}_i^j \times \mathbf{K}_i^{j\top} + \mathbf{M})/\sqrt{d^{(l+1)}/H}] \times \mathbf{V}_i^j \tag{16}$$

where $\mathbf{Q}_i^j = (\mathbf{H}_i^{(l+1)} + \mathbf{P}) \times \mathbf{W}_Q^j, \mathbf{K}_i^j = (\mathbf{H}_i^{(l+1)} + \mathbf{P}) \times \mathbf{W}_K^j, \mathbf{V}_i^j = (\mathbf{H}_i^{(l+1)} + \mathbf{P}) \times \mathbf{W}_V^j, \mathbf{W}_Q^j \in \mathbb{R}^{d^{(l+1)}\times \frac{d^{(l+1)}}{H}}, \mathbf{W}_K^j \in \mathbb{R}^{d^{(l+1)}\times \frac{d^{(l+1)}}{H}}, \mathbf{W}_V^j \in \mathbb{R}^{d^{(l+1)}\times \frac{d^{(l+1)}}{H}}$ and $\mathbf{W}_O \in \mathbb{R}^{d^{(l+1)}\times d^{(l+1)}}$. Then, the output of temporal multihead self-attention is fed into an feed-forward neural network (FFN) followed by residual connections:

$$\mathbf{S}_i = [\mathrm{ReLU}(\mathrm{Att}_{mask}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i)\mathbf{W}_F + \mathbf{b}_F) + \mathrm{Att}_{mask}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i)] + (\mathbf{H}_i^{(l+1)} + \mathbf{P}) \tag{17}$$

where $\mathbf{W}_F \in \mathbb{R}^{d^{(l+1)}\times d^{(l+1)}}, \mathbf{b}_F$ is a bias term, and $\mathbf{S}_i = [\mathbf{S}_i(0), \cdots, \mathbf{S}_i(t), \cdots, \mathbf{S}_i(T)] \in \mathbb{R}^{n\times d^{(l+1)}}$ is the output of 1^{th} TSN layer. Although the temporal self-attention block can stack multiple TSN layers, in fact, it only slightly improve the model performance. Accordingly, this work adopts a single layer of TSN.

4.5. News Predictor

For the node $v_i^{(t_i)}$'s feature representations $\mathbf{U}_i^{(r+1)}$ and $\mathbf{S}_i$ output by the DGNF-*tcn* and the DGNF-*tsn* at different time steps, we take $\mathbf{U}_i^{(r+1)}(T)$ and $\mathbf{S}_i(T)$ as the final feature representations of node v_i, respectively. The representation of $\mathbb{G}$ is the average of all the node embeddings in $\mathscr{G}(T)$, which is then passed through a feed-forward neural network (FFN) layer and a softmax layer to predict the veracity of $\mathbb{G}$. Concretely,

$$\hat{y}_{tcn} = \mathrm{softmax}\left(\sigma\left[\left(\left(\sum_{i=1}^{N(T)} \mathbf{U}_i^{(r+1)}(T)\right)/N(T)\right)\mathbf{W}_{P1}\right] + \mathbf{b}_{P1}\right) \tag{18}$$

$$\hat{y}_{tsn} = \mathrm{softmax}\left(\sigma\left[\left(\left(\sum_{i=1}^{N(T)} \mathbf{S}_i(T)\right)/N(T)\right)\mathbf{W}_{P2}\right] + \mathbf{b}_{P2}\right) \tag{19}$$

where $\mathbf{U}_i^{(r+1)}(T) \in \mathbb{R}^{1\times d^{(l+1)}}, \mathbf{S}_i(T) \in \mathbb{R}^{1\times d^{(l+1)}}, \mathbf{W}_{P1} \in \mathbb{R}^{d^{(l+1)}\times 2}, \mathbf{W}_{P2} \in \mathbb{R}^{d^{(l+1)}\times 2}$, $\mathbf{b}_{P1} \in \mathbb{R}^{1\times 2}$ and $\mathbf{b}_{P2} \in \mathbb{R}^{1\times 2}$ are bias term, and $\sigma(\cdot)$ is an activation function. $\hat{y}_{tcn} = [\hat{y}_{tcn}^0, \hat{y}_{tcn}^1]$ and $\hat{y}_{tsn} = [\hat{y}_{tsn}^0, \hat{y}_{tsn}^1]$ denote that the probability of a given dynamic news propagation graph is real (i.e., $\hat{y}_{tcn}^0 = 0$ or $\hat{y}_{tsn}^0 = 0$) or fake (i.e., $\hat{y}_{tcn}^1 = 1$ or $\hat{y}_{tsn}^1 = 1$).

4.6. Objective Function

In this subsection, we present the objective function used to train DGNF-*tcn* and DGNF-*tsn* when being applied to fake news detection task. Because the fake detection task is viewed as a binary classification task in this work, for a given dynamic news propagation graph, we could define the objective functions of DGNF-*tcn* and DGNF-*tsn* as follows.

$$\mathscr{L}_{tcn}(\Theta_{tcn}) = -y\log(\hat{y}_{tcn}^1) - (1-y)\log(1-\hat{y}_{tcn}^0) \tag{20}$$

$$\mathscr{L}_{tsn}(\Theta_{tsn}) = -y\log(\hat{y}_{tsn}^1) - (1-y)\log(1-\hat{y}_{tsn}^0) \tag{21}$$

in which $\Theta_{tcn}$and Θ_{tsn}represent the parameters of DGNF-*tcn* and DGNF-*tsn*, respectively. The DGNF-*tcn* and DGNF-*tsn* models are trained by minimizing the corresponding objective function (i.e., Eq. 20 and Eq. 21).

5. Experiments

5.1. Datasets

In this work, we evaluate DGNF-*tcn* and DGNF-*tsn* over three datasets (i.e, Weibo, FakeNewsNet and Twitter), which are widely used public benchmark for detecting fake news. FakeNewsNet and Twitter datasets is constructed from Twitter social media platform, while Weibo dataset is constructed from Chinese Sina Weibo social media platform. Because all datasets contain time stamps, retweet or reply relationships, and textual information, we can build a discrete dynamic news propagation network for each piece of social media news. Table 1 summarizes more detailed statistics of three real-world and publicly available datasets.

- Weibo: This Weibo dataset is generated by Ma et al. [46], and collected from the Sina Weibo platform which is one of the most popular Chinese online social media platforms. After pre-processing this dataset, we present the number of items finally used in this paper and more details in Table 1.
- FakeNewsNet: The FakeNewsNet dataset is first presented in [24]. The news content is crawled from GossipCop[5] and PolitiFact[6]. The tweets, retweets and replies concerning a piece of news article are crawled from twitter social media platform via Twitter API. After pre-processing this dataset, we present the number of items finally used in this paper and more details in Table 1.
- Twitter: The Twitter dataset is generated by Ma et al. [70]. The Twitter dataset actually consists of Twitter15 and Twitter 16 datasets. In our work, the non-rumors and real rumors in these two Twitter datasets are regarded as real and fake news, which differs from rumor detection task in some previous research [70]. After pre-processing this dataset, Table 1 shows more details and the number of items finally used in this paper.

Same to previous work [64,71,5], the source tweet, retweets, and replies are regard as nodes. We treat the interactions between nodes such as retweet or reply behaviors as edges. The recorded time stamps of retweets or replies are regard as time the edges were created.

5.2. Experimental Setup

To facilitate the comparison with existing work [64,71], each fake news dataset is split into into training and test sets, containing 80% and 20% social media news, respectively. Moreover, we performed the experiments with fivefold cross-validation, which is consistent with existing research. For Weibo and FakeNewsNet datsets, we set the time points of each snapshot as [0.0, 0.5, 1.0,

[5] https://www.politifact.com/
[6] https://www.gossipcop.com/

Table 1
The statistics of three benchmark fake detection datasets [5].

Statistic	Weibo	FakeNewsNet	Twitter
# of fake news	2,131	2,079	578
# of real news	2,207	2,089	569
# of users	1,309,645	45,109	29,858
Avg. time length	1577 Hours	1951 Hours	158 Hours
Avg. # of tweets	378	42	30
Max. # of tweets	1999	1315	323
Min. # of tweets	10	3	2

1.5, 2.0, 4.0, 8.0, 16.0, 32.0, 64.0, 128.0, 256.0, 512.0, 1024.0, 2048.0, max] hours (i.e., $n = H = 16$), while the Twitter dataset is set as [0.0, 0.5, 1.0, 1.5, 2.0, 4.0, 8.0, max] hours (i.e., $n = H = 8$). Therefore, for Weibo and FakeNewsNet datsets, the number of TCN layers could be set as 4 (i.e., $r = 3$), and the Twitter dataset only need set as 3 (i.e., $r = 2$). Because we used the pretrained Google BERT embedding to represent each token in a sentence [72], the dimension of word vectors is $d = 768$. For structure-aware component, we set the number of GAT layers as 2 (i.e., $l = 1$), and the dimension of output features of 2^{th} GAT layer is set as 64 (i.e., $\mathbf{H}_i^{(l+1)} = \mathbf{H}_i^{(2)} \in \mathbb{R}^{n\times d^{(l+1)}} = \mathbb{R}^{n\times 64}$). Specifically, we used the default setting of GAT in structure-aware component. We train DGNF-*tcn* and DGNF-*tsn* with a learning rate of $1e^{-5}$. Because the input of DGNF is a collection of static news propagation network snapshots $\mathbb{G} = \{\mathcal{G}(1), \cdots, \mathcal{G}(t), \cdots, \mathcal{G}(T)\}$, we set batchsize and the number of epochs as 1 and 200, respectively. Four standard evaluation metrics (i.e., accuracy, precision, recall, and the F_1 score) in fake news detection are adopted to help readers understand of the models' performance.

5.3. Baseline Approaches

The baseline methods consist of commonly used machine learning models (i.e., DTC [15], SVM-TS [73], and SVM-RBF [36]) trained on different hand-engineered features, and deep learning models (i.e., RvNN [63], StA-HiTPLAN [49], GAT [68], GCN [74], VAE-GCN [75], BiGCN [64], STS-NN [71] and DynGCN [67]) trained on network structure and content semantics features. DTC model is designed based on decision tree algorithm [15]. SVM-RBF and SVM-TS are support vector machine-based algorithms which utilize statistics and temporal features to predict the authenticity of news [73,36]. RvNN utilizes tree-structured RNN to encodes text and network features for fake news detection [63]. StA-HiTPLAN is a neural network algorithm that model long distance interactions between tweets through a multi-layer transformer network [49]. GAT [68] and graph convolutional networks (GCN) [74] are the most popular *static* graph-based representation learning algorithms. To make fair comparisons, we set the number of layers of GAT and GCN as 2. VAE-GCN [75] stands for Variational Graph Autoencoder-GCN, which utilizes a encoder and decoder architecture to identify fake news. BiGCN is a GCN-based fake news detection algorithm and represents news diffusion path through the way of top-down and bottom-up trees [64]. STS-NN models news propagation graph with deep spatial temporal neural network [71]. DynGCN [67] is a DTDG-based fake news detection model that models static snapshots using GCN and models temporal information using scaled dot-product and additive attention mechanism.

5.4. Results and Analysis

To demonstrate the effectiveness of the proposed method, in this subsection, we evaluate DGNF-*tsn* and DGNF-*tcn* using three real-world benchmark datasets. We respectively report the news classification results of the baselines and our methods on three real-world benchmark datasets in Table 2, Table 3 and Table 4 with the best model high lighted in bold font. From these tables, we can yield the following observations:

- Overall, we start by observing that DGNF-*tsn* and DGNF-*tcn* achieve better performance compared to all baseline algorithms in terms of accuracy and F_1 score across three real-world benchmark datasets, which suggests that temporal variations in graph structure preserved in our model provides effective information to help improve fake news detection.
- Moreover, another important observation is that most of the common machine learning based algorithms (e.g., DTC and SVM-TS) that highly depend on the hand-engineered features show significantly decreased F1 and accuracy, which is consistent with existing research, and partly explained by the fact that deep learning-based fake news detection algorithms can bypass feature engineering and capture complex patterns automatically.
- There is no clear evidence showing superiority of any *static* network-based approaches over others. Graph neural network-based models outperform RNN and transformer based methods in the identification of fake news in terms of accuracy and F_1 score. Although GAT, GCN and BiGCN fail to capture temporal information, we yet can observe that GAT, GCN and BiGCN achieve better performance than most baselines across various benchmark datasets. These results further verify the advantage of news propagation network-based models in identifying fake

Table 2
Performance comparisons of different methods on Weibo dataset.

Method	Accuracy	Fake News			Real News		
		Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
DTC	0.809	0.806	0.813	0.810	0.812	0.806	0.809
SVM-RBF	0.823	0.824	0.820	0.822	0.821	0.825	0.823
SVM-TS	0.859	0.825	0.891	0.850	0.871	0.818	0.836
RvNN	0.896	0.904	0.883	0.893	0.889	0.909	0.899
StA-HiTPLAN	0.870	0.869	0.866	0.867	0.871	0.874	0.872
GAT	0.931	0.924	0.937	0.931	0.939	0.926	0.932
GCN	0.932	0.923	0.940	0.931	0.941	0.924	0.933
VAE-GCN	0.906	0.907	0.902	0.904	0.906	0.911	0.908
BiGCN	0.933	0.928	0.939	0.930	0.940	0.929	0.930
STS-NN	0.912	0.912	0.908	0.910	0.911	0.915	0.913
DynGCN	0.937	0.934	0.937	0.936	0.939	0.937	0.938
DGNF-*tsn*	0.956	**0.960**[†]	0.951	0.955	0.953	**0.962**	0.957
DGNF-*tcn*	**0.957**	0.958	**0.954**	**0.957**	**0.957**	0.960	**0.958**

[†] *Bold text indicates the maximum value.*

Table 3
Performance comparisons of different methods on FakeNewsNet dataset.

Method	Accuracy	Fake News			Real News		
		Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
DTC	0.782	0.780	0.783	0.782	0.783	0.780	0.781
SVM-RBF	0.788	0.786	0.789	0.787	0.789	0.786	0.788
SVM-TS	0.811	0.808	0.796	0.791	0.828	0.820	0.809
RvNN	0.828	0.827	0.796	0.801	0.818	0.857	0.829
StA-HiTPLAN	0.800	0.802	0.794	0.798	0.797	0.805	0.801
GAT	0.885	0.886	0.883	0.884	0.884	0.887	0.885
GCN	0.873	0.872	0.874	0.873	0.874	0.873	0.873
VAE-GCN	0.865	0.865	0.863	0.864	0.864	0.866	0.865
BiGCN	0.889	0.890	0.888	0.889	0.888	0.891	0.890
STS-NN	0.858	0.867	0.847	0.857	0.848	0.868	0.858
DynGCN	0.896	0.897	0.894	0.895	0.895	0.898	0.896
DGNF-*tsn*	**0.922**†	**0.925**	0.918	**0.921**	**0.919**	**0.926**	**0.922**
DGNF-*tcn*	0.917	0.916	**0.919**	0.917	**0.919**	0.916	0.917

† *Bold text indicates the maximum value.*

Table 4
Performance comparisons of different methods on Twitter dataset.

Method	Accuracy	Fake News			Real News		
		Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
DTC	0.704	0.717	0.683	0.699	0.693	0.726	0.709
SVM-RBF	0.732	0.740	0.724	0.731	0.725	0.741	0.733
SVM-TS	0.707	0.715	0.698	0.706	0.700	0.717	0.709
RvNN	0.805	0.818	0.788	0.803	0.793	0.822	0.807
StA-HiTPLAN	0.780	0.777	0.783	0.780	0.782	0.776	0.779
GAT	0.865	0.879	0.849	0.864	0.852	0.882	0.866
GCN	0.858	0.860	0.855	0.858	0.857	0.861	0.859
VAE-GCN	0.841	0.847	0.836	0.841	0.836	0.847	0.841
BiGCN	0.864	0.867	0.862	0.865	0.861	0.866	0.863
STS-NN	0.834	0.838	0.829	0.834	0.829	0.838	0.833
DynGCN	0.873	0.884	0.861	0.872	0.862	0.885	0.873
DGNF-*tsn*	**0.899**†	**0.899**	**0.897**	**0.892**	**0.899**	**0.901**	**0.900**
DGNF-*tcn*	0.891	0.897	0.886	0.891	0.885	0.897	0.891

† *Bold text indicates the maximum value.*

Table 5
Ablation study results on Weibo dataset.

Method	Accuracy	Fake News			Real News		
		Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
DGNF-tsn^+	**0.959**†	**0.967**	0.952	**0.959**	0.951	**0.966**	**0.958**
DGNF-tcn^+	0.958	0.963	**0.955**	**0.959**	0.954	0.962	**0.958**
DGNF-tsn^-	0.949	0.956	0.943	0.950	0.942	0.955	0.949
DGNF-tcn^-	0.948	0.952	0.946	0.950	0.945	0.951	0.948
DGNF-tsn^*	0.955	0.957	**0.955**	0.956	0.953	0.955	0.954
DGNF-tcn^*	0.955	0.960	0.948	0.954	0.951	0.962	0.956
DGNF-*tsn*	0.956	0.960	0.951	0.955	0.953	0.962	0.957
DGNF-*tcn*	0.957	0.958	0.954	0.957	**0.957**	0.960	**0.958**

† *Bold text indicates the maximum value.*

Table 6
Ablation study results on FakeNewsNet dataset.

Method	Accuracy	Fake News			Real News		
		Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
DGNF-tsn^+	**0.925**†	0.925	**0.925**	**0.925**	**0.926**	0.926	**0.926**
DGNF-tcn^+	0.919	0.921	0.919	0.920	0.918	0.920	0.919
DGNF-tsn^-	0.916	0.920	0.911	0.915	0.912	0.921	0.917
DGNF-tcn^-	0.911	0.905	0.918	0.912	0.917	0.904	0.911
DGNF-tsn^*	0.921	**0.927**	0.914	0.921	0.914	**0.928**	0.921
DGNF-tcn^*	0.918	0.920	0.916	0.918	0.917	0.920	0.919
DGNF-*tsn*	0.922	0.925	0.918	0.921	0.919	0.926	0.922
DGNF-*tcn*	0.917	0.916	0.919	0.917	0.919	0.916	0.917

† *Bold text indicates the maximum value.*

Table 7
Ablation study results on Twitter dataset.

Method	Accuracy	Fake News			Real News		
		Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
DGNF-*tsn*$^+$	**0.904**[†]	**0.912**	0.896	**0.904**	0.896	**0.912**	**0.904**
DGNF-*tcn*$^+$	0.895	0.907	0.883	0.895	0.885	0.908	0.896
DGNF-*tsn*$^-$	0.890	0.888	0.896	0.892	0.893	0.885	0.889
DGNF-*tcn*$^-$	0.885	0.891	0.879	0.885	0.879	0.890	0.885
DGNF-*tsn**	0.899	0.904	**0.899**	0.900	0.897	0.899	0.898
DGNF-*tcn**	0.890	0.888	0.896	0.892	0.893	0.885	0.889
DGNF-*tsn*	0.899	0.899	0.897	0.892	**0.899**	0.901	0.900
DGNF-*tcn*	0.891	0.897	0.886	0.891	0.885	0.897	0.891

[†] Bold text indicates the maximum value.

news, and the ability in modeling fine-grained structure information of news propagation network.

- For DTDG-based fake news detection methods, DGNF-*tsn* and DGNF-*tcn* outperform DynGCN in the identification of fake news w.r.t. all of the evaluation metrics. One possible reason is that DynGCN learns the temporal propagation patterns from the graph-level, while DGNF-*tsn* and DGNF-*tcn* from the node-level, which leads to a more fine-grained nodal relationship modeling for dynamic news propagation graphs.
- When comparing the proposed dynamic modeling methods (i.e., DGNF-*tsn* and DGNF-*tcn*), DGNF-*tsn* shows better overall performance. We conjecture that TSN can flexibly capture the interactions between nodes over multiple time steps, while TCN fails to extract the internal correlation information of node features over discrete time steps.

5.5. Ablation Study

In this subsection, we conduct experiments to further examine the effect of the time intervals of snapshots on DGNF-*tsn* and DGNF-*tcn*. Specifically, we make comparisons with the following variants of DGNF-*tsn* and DGNF-*tcn*:

- DGNF-*tsn*$^+$ and DGNF-*tcn*$^+$: For Weibo and FakeNewsNet datsets, we set the time point of each snapshot as [0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 3.0, 4.0, 6.0, 8.0, 12.0, 16.0, 24.0, 32.0, 48.0, 64.0, 96.0, 128.0, 192.0, 256.0, 384.0, 512.0, 768.0, 1024.0, 1536.0, 2048.0, max] hours, while the Twitter dataset is set as [0.0, 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 3.0, 4.0, 6.0, 8.0, max] hours.
- DGNF-*tsn*$^-$ and DGNF-*tcn*$^-$: For Weibo and FakeNewsNet datsets, we set the time point of each snapshot as [0.0, 1.0, 2.0, 8.0, 32.0, 128.0, 512.0, 2048.0, max] hours, while the Twitter dataset is set as [0.0, 1.0, 2.0, 8.0, max] hours.
- DGNF-*tsn** and DGNF-*tcn**: In structure-aware module, the GAT is replaced by GCN.

We present the experimental results of these variants in Table 5, Table 6 and Table 7. We start by observing that accuracy and F_1 score for DGNF-*tsn*$^+$ and DGNF-*tcn*$^+$ is larger than the accuracy and F_1 score for DGNF-*tsn* and DGNF-*tcn*, whereas the results are opposite for DGNF-*tsn*$^-$ and DGNF-*tcn*$^-$. This signals that the sampling frequency for news propagation network, to some extent, affects the experimental results. Another observation is that, compared to DGNF-*tsn* and DGNF-*tcn*, DGNF-*tsn** and DGNF-*tcn** that depend on GCN show slightly decreased accuracy and F_1 score.

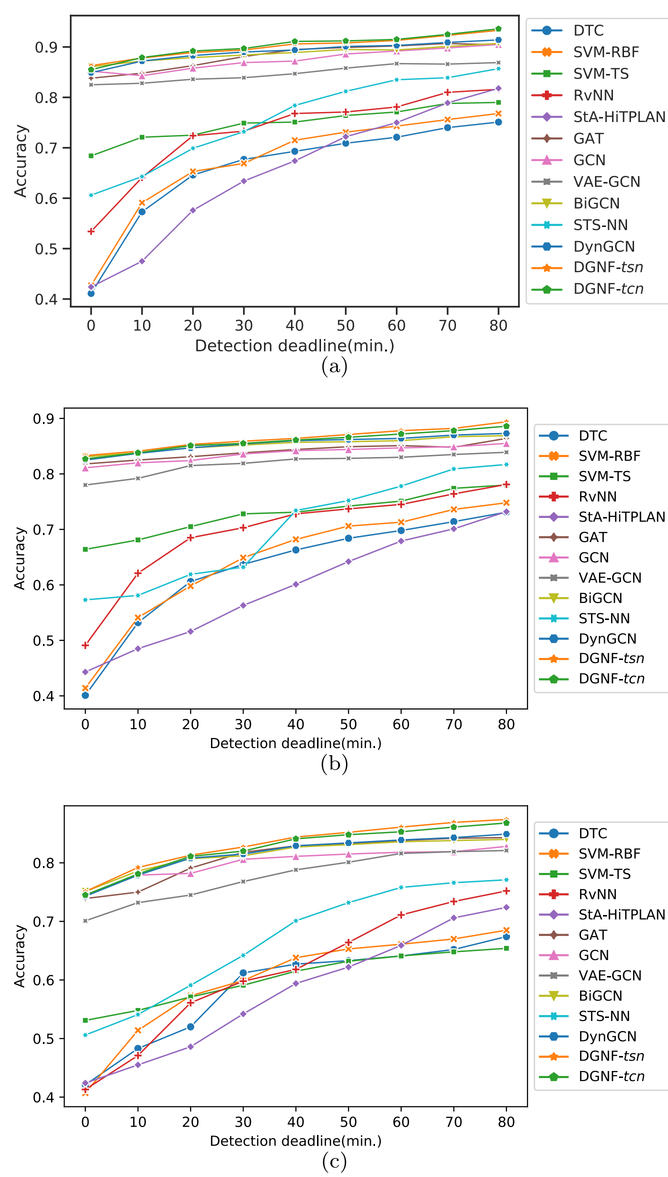


Fig. 5. (a) Impact of the check time stamps in detecting fake news on Weibo dataset; (b) Impact of the check time stamps in detecting fake news on FakeNewsNet dataset; (c) Impact of the check time stamps in detecting fake news on Twitter dataset.

5.6. Early Fake News Detection

A problem with fake news detection using real-world data is that there are the limited fake news instances observed for an algorithm to effectively identify fake news at the early stage of news propagation. Identifying fake news at the early stage is particularly crucial for taking early actions for fake news intervention before more individuals become exposed to fake news. To better

Table 8
Comparisons of average running time each epoch among some baselines (Minutes) [5].

Method	Weibo	FakeNewsNet	Twitter
GAT	1.1	0.7	0.3
GCN	0.7	0.6	0.2
BiGCN	2.9	1.5	0.8
DynGCN	5.1	2.9	1.0
TGNF	**54.2**[*]	**5.4**	**12.4**
DGNF-*tsn*	6.7	3.7	1.5
DGNF-*tcn*	6.6	3.4	1.4

[*] Bold text indicates the maximum value.

understand how the propagation time influence the performance of DGNF-*tsn*, DGNF-*tcn* and existing algorithms, we further evaluate their performance by accuracy on early fake news detection. We treat the news propagation timestamp as the detection deadline, which means that the tweets or replies published after the detection deadline are unavailable. By varying check time stamps in the range of $\{0, 10, 20, 30, 40, 50, 60, 70, 80\}$ minutes, the accuracy of several competitive models and the proposed method on three benchmark datasets is shown in Fig. 5. It can be observed from Fig. 5 that with the change of time delays, our model generally has a comparable performance compared to baselines in terms of early fake news detection accuracy across three benchmark datasets.

5.7. Comparison of the execution time

The static graph-based methods only need to process one completed news propagation graph each time. However, because DGNF model news propagation from the perspective of DTDG, and the input of DGNF is a collection of static news propagation network snapshots $\mathbb{G} = \{\mathcal{G}(1), \cdots, \mathcal{G}(t), \cdots, \mathcal{G}(T)\}$ over discrete time stamps, DGNF have to spend more execution time in the process of model learning. Table 8 compares the average running time of baseline methods and also the CTDG-based method TGNF [5] on three real world datasets in one epoch. The reason why we choose these methods is that they are all GNN-based methods and show better performance. All experiments are conducted on GeForce RTX 2080Ti GPU. Compared to the methods based on static news propagation graph, we observe that DGNF consistently shows a longer running time for each epoch across all the datasets. As shown in Table 9, Table 10 and Table 11, TGNF shows better performace than DGNF, the reason of which is that TGNF takes a more fine-grained way (i.e., CTDG) to model social media news dissemination network. However, DGNF shows better in learning efficiency [76] because it achieves greater accuracy improvement with less model learning time (actually, also with less computing resources).

6. Conclusion

In this paper, we study the problem of dynamic news propagation graph-based fake news detection task. Our contribution focuses on introducing a DTDG-based fake news detection method named DGNF, which utilizes structure-aware module and temporal-aware module to explicitly capture the temporal and network structure information, respectively. Specifically, we also provide, by designing two temporal-aware modules, two variants of DGNF, DGNF-*tsn* and DGNF-*tcn*. We conduct extensive experiments on three real-world benchmark datasets, and the identification results suggest that incorporating the temporal propagation and network structure information of online social media news can help fake news detection task to make more accurate predictions. Whereas the work presented in this paper focuses on the case of homogeneous discrete-time dynamic news propagation graph-based fake news detection task, in the future, complementary line of research could study and integrate social media user profiles using temporal heterogeneous news propagation graphs.

Table 9
Performance comparisons of DGNF-*tsn*, DGNF-*tcn* and TGNF on Weibo dataset.

Method	Accuracy	Fake News			Real News		
		Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
TGNF	**0.968**[†]	**0.962**	**0.975**	**0.969**	**0.974**	0.960	**0.967**
DGNF-*tsn*	0.956	0.960	0.951	0.955	0.953	**0.962**	0.957
DGNF-*tcn*	0.957	0.958	0.954	0.957	0.957	0.960	0.958

[†] *Bold text indicates the maximum value.*

Table 10
Performance comparisons of DGNF-*tsn*, DGNF-*tcn* and TGNF on FakeNewsNet dataset.

Method	Accuracy	Fake News			Real News		
		Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
TGNF	**0.935**[†]	**0.937**	**0.932**	**0.935**	**0.933**	**0.928**	**0.931**
DGNF-*tsn*	0.922	0.925	0.918	0.921	0.919	0.926	0.922
DGNF-*tcn*	0.917	0.916	0.919	0.917	0.919	0.916	0.917

[†] *Bold text indicates the maximum value.*

Table 11
Performance comparisons of DGNF-*tsn*, DGNF-*tcn* and TGNF on Twitter dataset.

Method	Accuracy	Fake News			Real News		
		Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
TGNF	**0.923**[†]	**0.932**	**0.914**	**0.923**	**0.914**	**0.932**	**0.923**
DGNF-*tsn*	0.899	0.899	0.897	0.892	0.899	0.901	0.900
DGNF-*tcn*	0.891	0.897	0.886	0.891	0.885	0.897	0.891

[†] *Bold text indicates the maximum value.*

CRediT authorship contribution statement

Chenguang Song: Conceptualization, Methodology, Software, Validation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Yiyang Teng:** Visualization, Data curation, Writing - review & editing. **Yangfu Zhu:** Software, Data curation, Writing - review & editing. **Siqi Wei:** Data curation, Writing - review & editing. **Bin Wu:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

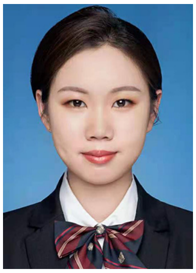
This work is supported by the NSFC-General Technology Basic Research Joint Funds (Grant No. U1936220) and the National Natural Science Foundation of China (Grant No. 61972047).

References

[1] C. Song, Y. Teng, B. Wu, Dynamic graph neural network for fake news detection, in: 2021 IEEE 7th International Conference on Cloud Computing and Intelligent Systems (CCIS), 2021, pp. 27–31, https://doi.org/10.1109/CCIS53392.2021.9754681.
[2] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: a data mining perspective, ACM SIGKDD Explorations Newsletter 19 (1) (2017) 22–36, https://doi.org/10.1145/3137597.3137600.
[3] S. Kumar, N. Shah, False information on web and social media: A survey (2018). arXiv:1804.08559.
[4] A. Bondielli, F. Marcelloni, A survey on fake news and rumour detection techniques, Information Sciences 497 (2019) 38–55, https://doi.org/10.1016/j.ins.2019.05.035.
[5] C. Song, K. Shu, B. Wu, Temporally evolving graph neural network for fake news detection, Information Processing & Management 58 (6) (2021) , https://doi.org/10.1016/j.ipm.2021.102712 102712.
[6] M. Guderlei, M. Aßenmacher, Evaluating unsupervised representation learning for detecting stances of fake news, in: Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, 2020, pp. 6339–6349, https://doi.org/10.18653/v1/2020.coling-main.558.
[7] X. Zhang, A.A. Ghorbani, An overview of online fake news: Characterization, detection, and discussion, Information Processing & Management, 57, 2020, p. 102025, 10.1016/j.ipm.2019.03.004.
[8] B. Guo, Y. Ding, L. Yao, Y. Liang, Z. Yu, The future of false information detection on social media: New perspectives and trends, ACM Comput. Surv. 53 (4). doi:10.1145/3393880.
[9] L. Hu, T. Yang, L. Zhang, W. Zhong, D. Tang, C. Shi, N. Duan, M. Zhou, Compare to the knowledge: Graph neural fake news detection with external knowledge, 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) ACL, 2021, pp. 754–763, https://doi.org/10.18653/v1/2021.acl-long.62.
[10] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H.E. Stanley, W. Quattrociocchi, The spreading of misinformation online, Proceedings of the National Academy of Sciences 113 (3) (2016) 554–559, https://doi.org/10.1073/pnas.1517441113.
[11] X. Zhou, R. Zafarani, K. Shu, H. Liu, Fake news: fundamental theories, detection strategies and challenges, in: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, ACM, 2019, pp. 836–837, https://doi.org/10.1145/3289600.3291382.
[12] Y. Zhao, J. Da, J. Yan, Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches, Information Processing & Management 58 (1) (2021) , https://doi.org/10.1016/j.ipm.2020.102390 102390.
[13] L. Wu, H. Liu, Tracing fake-news footprints: Characterizing social media messages by how they propagate, in: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM '18, ACM, 2018, pp. 637–645, https://doi.org/10.1145/3159652.3159677.
[14] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, R. Procter, Detection and resolution of rumours in social media: A survey, ACM Comput. Surv. 51 (2) (2018) 36, https://doi.org/10.1145/3161603.
[15] C. Castillo, M. Mendoza, B. Poblete, Information credibility on twitter, in: Proceedings of the 20th International Conference on World Wide Web, ACM, 2011, pp. 675–684, https://doi.org/10.1145/1963405.1963500.
[16] C. Song, N. Ning, Y. Zhang, B. Wu, A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks, Information Processing & Management 58 (1) (2021) , https://doi.org/10.1016/j.ipm.2020.102437 102437.
[17] R. Feldman, Techniques and applications for sentiment analysis, Commun. ACM 56 (4) (2013) 82–89, https://doi.org/10.1145/2436256.2436274.
[18] Y. Chen, J. Sui, L. Hu, W. Gong, Attention-residual network with cnn for rumor detection, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, ACM, 2019, pp. 1121–1130, https://doi.org/10.1145/3357384.3357950.
[19] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C.L. Zitnick, D. Parikh, Vqa: Visual question answering, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015.
[20] K. Shu, L. Cui, S. Wang, D. Lee, H. Liu, Defend: explainable fake news detection, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 395–405, https://doi.org/10.1145/3292500.3330935.
[21] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, Y. Liu, Combating fake news: a survey on identification and mitigation techniques, ACM Transactions on Intelligent Systems and Technology 10 (3) (2019) 21:1–21:42. doi:10.1145/3305260.
[22] N. Rosenfeld, A. Szanto, D.C. Parkes, A kernel of truth: Determining rumor veracity on twitter by diffusion pattern alone, in: Proceedings of The Web Conference 2020, WWW '20, ACM, 2020, pp. 1018–1028, https://doi.org/10.1145/3366423.3380180.
[23] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, Science 359 (6380) (2018) 1146–1151, https://doi.org/10.1126/science.aap9559.
[24] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, 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.
[25] M. Si, L. Cui, W. Guo, Q. Li, L. Liu, X. Lu, X. Lu, A comparative analysis for spatio-temporal spreading patterns of emergency news, Scientific Reports 10 (1) (2020) 1–13.
[26] F. Monti, F. Frasca, D. Eynard, D. Mannion, M.M. Bronstein, Fake news detection on social media using geometric deep learning (2019). arXiv:1902.06673.
[27] P. Wei, N. Xu, W. Mao, Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity, 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) ACL, 2019, pp. 4787–4798, https://doi.org/10.18653/v1/D19-1485.
[28] Y.-J. Lu, C.-T. Li, GCAN: Graph-aware co-attention networks for explainable fake news detection on social media, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics ACL, Online, 2020, pp. 505–514, https://doi.org/10.18653/v1/2020.acl-main.48.
[29] D.T. Vu, J.J. Jung, Rumor detection by propagation embedding based on graph convolutional network, International Journal of Computational Intelligence Systems 14 (2021) 1053–1065, https://doi.org/10.2991/ijcis.d.210304.002.
[30] K. Tu, C. Chen, C. Hou, J. Yuan, J. Li, X. Yuan, Rumor2vec: A rumor detection framework with joint text and propagation structure representation learning, Information Sciences 560 (2021) 137–151, https://doi.org/10.1016/j.ins.2020.12.080.
[31] Z. Wu, D. Pi, J. Chen, M. Xie, J. Cao, Rumor detection based on propagation graph neural network with attention mechanism, Expert Systems with Applications 158 (2020), 113595, https://doi.org/10.1016/j.eswa.2020.113595.
[32] L. Jian, J. Li, H. Liu, Toward online node classification on streaming networks, Data Min. Knowl. Discov. 32 (1) (2018) 231–257, https://doi.org/10.1007/s10618-017-0533-y.
[33] N. Ruchansky, S. Seo, Y. Liu, Csi: A hybrid deep model for fake news detection, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017, pp. 797–806, https://doi.org/10.1145/3132847.3132877.
[34] M. Lukasik, T. Cohn, K. Bontcheva, Point process modelling of rumour dynamics in social media, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing ACL, 2015, pp. 518–523, https://doi.org/10.3115/v1/P15-2085.
[35] K. Zhou, C. Shu, B. Li, J.H. Lau, Early rumour detection, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 1614–1623. doi:10.18653/v1/N19-1163.
[36] J. Ma, W. Gao, Z. Wei, Y. Lu, K.-F. Wong, Detect rumors using time series of social context information on microblogging websites, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15 ACL, 2015, pp. 1751–1754, https://doi.org/10.1145/2806416.2806607.
[37] Q. Zhang, J. Cook, E. Yilmaz, Detecting and forecasting misinformation via temporal and geometric propagation patterns, in: D. Hiemstra, M. Moens, J. Mothe, R. Perego, M. Potthast, F. Sebastiani (Eds.), Advances in Information Retrieval - 43rd European Conference on IR Research, ECIR 2021, Vol. 12657 of Lecture Notes in Computer Science, Springer, 2021, pp. 455–462. doi:10.1007/978-3-030-72240-1_48.
[38] S. Bai, J.Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling (2018). arXiv:1803.01271.

[39] Y. Yuan, Y. Wang, K. Liu, Perceiving more truth: A dilated-block-based convolutional network for rumor identification, Information Sciences 569 (2021) 746–765, https://doi.org/10.1016/j.ins.2021.05.014.
[40] A. Sankar, Y. Wu, L. Gou, W. Zhang, H. Yang, Dysat: Deep neural representation learning on dynamic graphs via self-attention networks, in: Proceedings of the 13th International Conference on Web Search and Data Mining, WSDM '20, ACM, New York, NY, USA, 2020, pp. 519–527, https://doi.org/10.1145/3336191.3371845.
[41] E. Agichtein, C. Castillo, D. Donato, A. Gionis, G. Mishne, Finding high-quality content in social media, in: Proceedings of the 2008 International Conference on Web Search and Data Mining, WSDM '08, ACM, 2008, pp. 183–194, https://doi.org/10.1145/1341531.1341557.
[42] C. Song, C. Yang, H. Chen, C. Tu, Z. Liu, M. Sun, Ced: Credible early detection of social media rumors, IEEE Transactions on Knowledge and Data Engineering (2019), https://doi.org/10.1109/TKDE.2019.2961675, 1–1.
[43] V.L. Rubin, Y. Chen, N.J. Conroy, Deception detection for news: three types of fakes, in: Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, American Society for Information Science, 2015, pp. 83:1–83:4..
[44] K. Sejeong, C. Meeyoung, J. Kyomin, C. Wei, W. Yajun, Prominent features of rumor propagation in online social media, in: IEEE 13th International Conference on Data Mining, IEEE, 2013, pp. 1103–1108, https://doi.org/10.1109/ICDM.2013.61.
[45] C. Castillo, M. Mendoza, B. Poblete, Predicting information credibility in time-sensitive social media, Internet Research 23 (5) (2013) 560–588, https://doi.org/10.1108/IntR-05-2012-0095.
[46] J. Ma, W. Gao, P. Mitra, S. Kwon, B.J. Jansen, K.-F. Wong, M. Cha, Detecting rumors from microblogs with recurrent neural networks, in: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, AAAI Press, 2016, pp. 3818–3824.
[47] W.Y. Wang, liar liar pants on fire: A new benchmark dataset for fake news detection, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2017, pp. 422–426, https://doi.org/10.18653/v1/P17-2067.
[48] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Proceedings of the 26th International Conference on Neural Information Processing Systems, Curran Associates Inc., 2013, pp. 3111–3119. doi:10.5555/2999792.2999959.
[49] L.M.S. Khoo, H.L. Chieu, Z. Qian, J. Jiang, Interpretable rumor detection in microblogs by attending to user interactions, in: The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI Press, 2020, pp. 8783–8790.
[50] Y. Dun, K. Tu, C. Chen, C. Hou, X. Yuan, Kan: Knowledge-aware attention network for fake news detection, Proceedings of the AAAI Conference on Artificial Intelligence 35 (1) (2021) 81–89.
[51] L. Cui, H. Seo, M. Tabar, F. Ma, S. Wang, D. Lee, Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '20, ACM, 2020, pp. 492–502, https://doi.org/10.1145/3394486.3403092.
[52] J. Ma, W. Gao, K.-F. Wong, Detect rumors on twitter by promoting information campaigns with generative adversarial learning, in: Proceedings of the 28th international conference on World Wide Web, ACM, 2019, pp. 3049–3055, https://doi.org/10.1145/3308558.3313741.
[53] A. Silva, L. Luo, S. Karunasekera, C. Leckie, Embracing domain differences in fake news: Cross-domain fake news detection using multimodal data, in: Proceedings of the 35th International Joint Conference on Artificial Intelligence, AAAI Press, 2021.
[54] C. Song, N. Ning, Y. Zhang, B. Wu, Knowledge augmented transformer for adversarial multidomain multiclassification multimodal fake news detection, Neurocomputing 462 (2021) 88–100, https://doi.org/10.1016/j.neucom.2021.07.077.
[55] X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, K. Shu, Mining dual emotion for fake news detection, in: Proceedings of the Web Conference 2021, WWW '21, ACM, 2021, pp. 3465–3476, https://doi.org/10.1145/3442381.3450004.
[56] K. Shu, G. Zheng, Y. Li, S. Mukherjee, A.H. Awadallah, S. Ruston, H. Liu, Early detection of fake news with multi-source weak social supervision, in: Proceedings of The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases 2020(ECML-PKDD 2020), 2020.
[57] Y. Liu, Y.-F.B. Wu, Fned: A deep network for fake news early detection on social media, ACM Trans. Inf. Syst. 38 (3). doi:10.1145/3386253.
[58] P. Qi, J. Cao, X. Li, H. Liu, Q. Sheng, X. Mi, Q. He, Y. Lv, C. Guo, Y. Yu, Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues, in: Proceedings of the 29th ACM International Conference on Multimedia, MM '21, ACM, 2021, pp. 1212–1220, https://doi.org/10.1145/3474085.3481548.
[59] L. Cheng, R. Guo, K. Shu, H. Liu, Causal understanding of fake news dissemination on social media, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, KDD '21, ACM, 2021, pp. 148–157, https://doi.org/10.1145/3447548.3467321.
[60] Z. He, C. Li, F. Zhou, Y. Yang, Rumor detection on social media with event augmentations, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '21, ACM, 2021, pp. 2020–2024, https://doi.org/10.1145/3404835.3463001.
[61] A. Lao, C. Shi, Y. Yang, Rumor detection with field of linear and non-linear propagation, in: Proceedings of the Web Conference 2021, WWW '21, ACM, 2021, pp. 3178–3187, https://doi.org/10.1145/3442381.3450016.
[62] X. Zhou, R. Zafarani, Network-based fake news detection: A pattern-driven approach, SIGKDD Explor. Newsl. 21 (2) (2019) 48–60, https://doi.org/10.1145/3373464.3373473.
[63] J. Ma, W. Gao, K.-F. Wong, Rumor detection on twitter with tree-structured recursive neural networks, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL, Melbourne, Australia, 2018, pp. 1980–1989, https://doi.org/10.18653/v1/P18-1184.
[64] T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, J. Huang, Rumor detection on social media with bi-directional graph convolutional networks, in: The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI Press, 2020, pp. 549–556.
[65] X. Yang, Y. Lyu, T. Tian, Y. Liu, Y. Liu, X. Zhang, Rumor Detection on Social Media with Graph Structured Adversarial Learning, in: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20, International Joint Conferences on Artificial Intelligence Organization, 2020, pp. 1417–1423.
[66] V.H. Nguyen, K. Sugiyama, P. Nakov, M.Y. Kan, Fang: Leveraging social context for fake news detection using graph representation, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM '20, ACM, 2020, pp. 1165–1174, https://doi.org/10.1145/3340531.3412046.
[67] J. Choi, T. Ko, Y. Choi, H. Byun, C.-K. Kim, Dynamic graph convolutional networks with attention mechanism for rumor detection on social media, PLOS ONE 16 (8) (2021) 1–17, https://doi.org/10.1371/journal.pone.0256039.
[68] P. Velikovi, G. Cucurull, A. Casanova, A. Romero, P. Li, Y. Bengio, Graph attention networks, in: International Conference on Learning Representations, 2018.
[69] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. u. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, Curran Associates Inc, 2017, pp. 5998–6008.
[70] J. Ma, W. Gao, K.-F. Wong, Detect rumors in microblog posts using propagation structure via kernel learning, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics ACL, 2017, pp. 708–717, https://doi.org/10.18653/v1/P17-1066.
[71] Q. Huang, C. Zhou, J. Wu, L. Liu, B. Wang, Deep spatial–temporal structure learning for rumor detection on twitter, Neural Computing and Applications (2020) 1–11.
[72] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT; Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics), Association for Computational Linguistics, 2019, pp. 4171–4186.
[73] F. Yang, Y. Liu, X. Yu, M. Yang, Automatic detection of rumor on sina weibo, in: Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, ACM, 2012, pp. 13:1–13:7. doi:10.1145/2350190.2350203.
[74] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: International Conference on Learning Representations (ICLR), 2017.
[75] H. Lin, X. Zhang, X. Fu, A graph convolutional encoder and decoder model for rumor detection, in: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 2020, pp. 300–306, https://doi.org/10.1109/DSAA49011.2020.00043.
[76] G. Menghani, Efficient deep learning: A survey on making deep learning models smaller, faster, and better, CoRR abs/2106.08962. arXiv:2106.08962.

Chenguang Song received the B.E. and M.E. degrees in electronic and information engineering, and electronic and communication engineering from Xuchang University and Chongqing University of Posts and Telecommunications, in 2014 and 2018, respectively. He is currently pursuing the Ph.D. degree in computer science and technology with the Beijing University of Posts and Telecommunications, since 2018. His research interests include data mining, fake news detection, complex network, and social network analysis.



Yiyang Teng is currently pursuing the Ph.D. degree in computer science and technology at Beijing University of Posts and Telecommunications, since 2018. Her research interests include computer vision, social relation recognition, and social network analysis.

Yangfu Zhu is currently pursuing the Ph.D. degree in computer science and technology at Beijing University of Posts and Telecommunications, since 2020. His research interests include natural language processing, personality computing, and social network analysis.

Siqi Wei received the B.E. and M.E. degrees from Nanchang University and Liaoning University, in 2018 and 2021, respectively. He is currently pursuing the Ph.D. degree in computer science and technology with the Beijing University of Posts and Telecommunications, since 2021. His research interests include deep learning, fake news detection, and social network analysis.

Bin Wu received the B.E. degree from the Beijing University of Posts and Telecommunications, in 1991, and the M.E. and Ph.D. degrees from the Institute of Computing Technology of Chinese Academic of Sciences, in 1998 and 2002, respectively. He joined the Beijing University of Posts and Telecommunications as a Lecturer, in 2002, where he is currently a Professor. He has published more than 100 papers in refereed journals and conferences. His current research interests include social computing, data mining, and complex network.