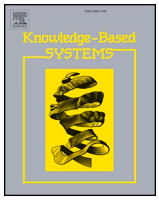
Knowledge-Based Systems 292 (2024) 111625

Contents lists available at ScienceDirect

Knowledge-Based Systems

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



A novel approach for rumor detection in social platforms: Memory-augmented transformer with graph convolutional networks

Qian Chang , Xia Li *, Zhao Duan

School of Information Management, Central China Normal University, Wuhan, PR China

ARTICLE INFO

Keywords:
Rumor detection
Social network
Graph classification
Graph convolutional networks
Transformer

ABSTRACT

Rumor detection in social media platforms is of critical importance owing to the widespread dissemination and impact of false information. Conventional approaches to rumor detection frequently rely on labor-intensive manual fact-checking or handcrafted features that may not adequately account for the complex nature of rumor propagation. To overcome these limitations, recent studies in deep learning, such as the recurrent neural network-based method and natural language processing techniques, have shown promise in capturing sequential patterns and analyzing textual content. However, these approaches often overlook the valuable information embedded in the global structural characteristics of rumor propagation. Herein, we propose a novel approach, named memory-augmented Transformer with graph convolutional networks (GCNs-MT), for rumor detection on social platforms. Our model integrates long short-term memory cells and the multi-head attention mechanism in Transformers to capture local dependencies and global dependencies in the propagation of rumors. By incorporating GCNs, a powerful deep learning framework for structured data, we aim to leverage the structural information of rumor propagation for improved detection performance. Additionally, we construct a Chinese dataset encoded and embedded by pretrained word embeddings (Word2Vec and bidirectional encoder representations from transformers [BERT]) based on real-world tweets from Weibo. Extensive evaluations on self-constructed Chinese and curated benchmark English datasets demonstrate the effectiveness of GCNs-MT in detecting and combating misinformation in social media platforms. The proposed GCNs-MT framework offers a comprehensive and efficient solution for rumor detection, addressing the challenges in social platforms posed by the rapid dissemination and complex nature of rumors.

1. Introduction

Rumors, characterized as unverified or false information circulating among users, give rise to grave concerns regarding the integrity and reliability of online discourse [1]. Rumors' propensity to swiftly disseminate and exert influence over public opinion necessitates the development of robust and efficient detection mechanisms to effectively address the challenges they pose [2]. Traditional methods without efficiency, such as manual fact-checking, are beset by time constraints, resource intensiveness, and an inherent inability to keep pace with the rapid dissemination of rumors [3,4]. Consequently, the emergence of advanced computational techniques and machine learning models has garnered considerable attention as a promising avenue for combating the pernicious effects of rumor propagation.

Conventional machine learning methods for rumor detection have relied on manual feature engineering, encompassing user characteristics, textual content, and propagation patterns. These features serve as inputs to train supervised classifiers, including decision trees, random forests (RF), and the support vector machine (SVM) [5,6]. Although these approaches have demonstrated some effectiveness, their reliance on labor-intensive and time-consuming handcrafted feature engineering remains a significant limitation. In addition, these handcrafted features may not capture the high-level representations necessary to accurately capture the complex nature of rumor propagation and dispersion [3,4].

To address these limitations, recent advancements in deep learning have been utilized to capture intricate representations from various sources, including rumor propagation paths and networks. Remarkably, recurrent neural network (RNN)-based models have demonstrated their efficacy in capturing sequential patterns within the nature of rumor propagation [7,8], enabling robust rumor screening [9–11]. Moreover, the integration of natural language processing (NLP) techniques,

* Corresponding author.
E-mail address: lixia@ccnu.edu.cn (X. Li).

https://doi.org/10.1016/j.knosys.2024.111625
Received 26 July 2023; Received in revised form 1 February 2024; Accepted 7 March 2024
Available online 8 March 2024

exemplified by pretrained word embeddings, such as Word to Vector (Word2vec) [12] and Bidirectional Encoder Representations from Transformers (BERT) [13], has been investigated to augment the model's capacity for analyzing and comprehending the semantic context of textual data in the context of rumor detection tasks [14,15]. Due to the powerful NLP capabilities of Transformers, several models based on Transformers have been applied to rumor detection tasks [16,17]. The multi-head attention mechanism in Transformers enables the models to capture long-range dependencies and effectively reason over the entire input sequence [13]. This capability enhances the models' ability to detect subtle patterns and linguistic cues that distinguish rumors from factual information [16]. By synergistically incorporating these techniques, the overall performance of rumor detection systems can be significantly enhanced by effectively harnessing the rich textual information inherent in social media posts.

However, recognizing that existing approaches often overlook the valuable information embedded in the structural characteristics of rumor propagation is essential [18–20]. The dynamics of social media events unfold through the interactions of users, who have the ability to engage with content by retweeting and commenting. These interactions give rise to a propagation network, wherein each user and their actions contribute to the evolving structure around a particular event. This network serves as a valuable source of information and insights into the characteristics of the event's propagation. The features embedded within this network structure are instrumental in understanding the patterns of information dissemination, user engagement, and the overall impact of the event in the online space. These structural features serve a crucial role in comprehending the underlying dynamics and patterns of rumor spread [21,22]. In response to this challenge, researchers have turned to the incorporation of graph convolutional networks (GCNs), a potent deep learning paradigm renowned for its ability to extract intricate high-level representations from structural data. This emerging approach presents a promising avenue in the domain of rumor detection. GCNs are adept at modeling intricate global structural relationships within graphs or trees, rendering them ideal for capturing the nuanced nature of rumor propagation. Several researchers have constructed the concept of propagation networks to capture the structural features of content propagation [3,22]. These networks exhibit a tree-structured network, encompassing the event itself, the user, the event–user interaction, and the user–user relationships. Such a network structure not only aligns with the application approach of GCNs but also reflects the exogenous and structural information associated with rumor propagation in real-world scenarios [22,23].

In this study, we present a novel approach, namely GCNs with memory-augmented Transformer (GCNs-MT), designed for rumor detection on social platforms. The proposed approach harnesses the strength of GCNs for node-level information convolution and integration. At the core of our model lies a unique design, the memory-augmented Transformer, which combines RNN cells and the multi-head attention mechanism in transformer architecture [24,25]. This fusion allows our model to capture both local and global dependencies from the rumor propagation pattern, which essentially transforms the current local representation into the global representation by the self-attention mechanism of Transformers. By integrating memory and Transformer-based architectures, our model excels at reasoning over the rich contextual information embedded in social media data. The memory component facilitates the retention and retrieval of important information from previous interactions while the Transformer component allows for effective information processing and aggregation. The main contributions of our study are as follows:

- A unique recurrent manner is designed to retain global graph-level information and generate global dependencies across all networks. The multi-head attention mechanism transforms local graph representation into a global graph representation.
- We introduce a pioneering approach for rumor detection, GCNs-MT, which learns structural information about the propagation network and captures local and global dependencies of event sequences.
- We conduct experiments on three real-world datasets in Chinese and English corpora to demonstrate the applicability and state-of-the-art performance of GCNs-MT and its effectiveness in identifying and combating rumors on social platforms.

2. Related work

The majority of prior studies on rumor detection have primarily focused on feature extraction from text content, user profiles, and propagation patterns, utilizing traditional machine learning algorithms such as SVM [5,19] and RF [26] classifiers, which rely on handcrafted features for low-level detection. However, these approaches suffer from low efficiency and limited accuracy in detecting rumors [5,6].

In recent years, the exceptional capabilities demonstrated by deep learning techniques in extracting high-level feature representations have ignited a surge in their ubiquitous application across rumor detection studies [2,11]. Ma et al. [27] employed the RNN to capture temporal content features for rumor detection. Building upon this work, Chen et al. [28] further improved the method by incorporating attention mechanisms to derive feature information from textual content using attention scores. Similarly, Yu et al. [29] proposed a convolutional neural network (CNN)-based approach to capture key features distributed throughout input sequences and aggregate them to form high-level representations. Wang et al. [30] introduced an RCNN model that synergistically combines RNN and CNN, enabling the capture of semantic text features while simultaneously learning sentiment features. Inspired by generative adversarial learning methods, Ma et al. [31] utilized discriminator models from generative adversarial networks as classifiers, while the corresponding generators were designed to generate uncertain or conflicting voices. GRU was employed as an encoder for time-series dialogue encoding, resulting in improved results in rumor detection. However, several of these methods rely on statistical feature extraction from raw text content, overlooking the linguistic aspects of the text. To address the influence of rumor semantics on detection accuracy, researchers have explored deep learning architectures from the field of NLP in rumor detection. Alkhodair et al. [32] jointly trained a Word2Vec model with unsupervised goals for learning word embeddings and RNN models with supervised goals for rumor detection. Kaliyar et al. [14] proposed FakeBERT, a deep learning method that integrates parallel blocks of a single-layer deep CNN framework with various kernel sizes and filters, along with BERT, effectively handling ambiguity and yielding remarkable results in disinformation detection. Furthermore, leveraging the remarkable capabilities exhibited by Transformers in language translation, sentiment analysis, and text classification within the field of NLP, several studies have incorporated Transformers into rumor detection methodologies. Lv et al. [17] introduced a novel approach called TMIF for automatic rumor detection. The approach integrates textual and image modalities through interactive fusion, using Transformers to capture the multilevel dependencies between different modalities while mitigating the impact of heterogeneous data. Taking advantage of the powerful natural language generation capabilities of Transformers, Ma et al. [16] employed Transformers architecture within the generator of a GAN network to enhance post-generation. Their innovative approach aims to create posts that closely resemble the source posts while preserving the authentic propagation structure and contextual information. Through adversarial training, the model successfully captures low-frequency yet crucial nontrivial patterns, leading to significant improvements in postgenerational quality. While these methods analyze content-level, user-level, media-level, and temporal-level information, they do not effectively capture structure-level features [2,6,33].

With the emergence and development of GCNs, numerous researchers have started to leverage the GCN framework to capture

structural-level features. Dong et al. [34] introduced a sophisticated GCN-based model, referred to as GCNSI, which addresses the challenging task of identifying multiple rumor sources without relying on prior knowledge of the underlying propagation model. Bian et al. [35] introduced the Bi-GCN model, which captures the structural information of rumor propagation trees through both top-down and bottom-up information propagation. Lu and Li [36] put forward a novel model named GCAN, which leverages convolutional and recursive neural networks (RvNN) to acquire user-based representations during forwarding propagation based on user features. Through graph construction to simulate potential interactions among users, they effectively employ GCN to learn graph-aware representations of user interactions. Sun et al. [37] introduced a novel dual-dynamic GCN that employs two GCNs to capture structural information from two distinct time stages. Through the integration of these networks with a sophisticated temporal fusion unit, the model adeptly simulates the dynamic changes in message propagation and background knowledge within a unified and cohesive framework. These studies exemplify the utility of GCNs in capturing structural features for rumor detection. By leveraging the power of GCNs, researchers have been able to extract and model the underlying structural characteristics of rumor propagation networks.

Our proposed model is motivated by prior research in the field and leverages NLP techniques, specifically pretrained word embeddings (Word2Vec and BERT), for encoding textual content and obtaining high-level representations. In addition, we employ GCNs to capture structural information at a granular level. To enhance the ability to capture dependencies at a global structural level, we integrate a global structural memory module based on long short-term memory (LSTM) and Transformer. By ingeniously incorporating a multi-head attention mechanism inspired by a Transformer, our model facilitates seamless bidirectional interactions between each position in the sequence and the entire memory module, facilitating the effective capture of global contextual information. This attention mechanism empowers Transformers to capture both local and global dependencies, thereby promoting a comprehensive understanding of the data. Through the integration of these components, our model aims to leverage text content, user profiles, and structural information to improve rumor detection performance and provide a more comprehensive approach to data analysis.

3. Rumor detection approach: memory-augmented transformer with graph convolutional networks

In this section, we present our innovative approach, GCNs-MT, for rumor detection on social platforms. Leveraging the power of memory networks and the multi-head attention mechanism in Transformers, our model aims to capture local dependencies and global dependencies in the propagation of rumors. By integrating memory and Transformer-based architectures, we enhance the ability of our model to effectively reason over the rich contextual information embedded in social networks. Fig. 1 illustrates the representation of the dataset and the architecture of the GCNs-MT, which provides an overview of how the data are organized and processed within the model. Moreover, Fig. 2 provides a more detailed architecture of the two core modules in our model, namely the GCNs module and the memory-augmented Transformer. These modules play a crucial role in capturing and integrating graph-level information for enhanced rumor detection.

3.1. GCNs module for node-level embedding and graph-level embedding

As mentioned above, our approach involves preprocessing the raw text data and leveraging NLP techniques, specifically Word2Vec and BERT, to perform node-level embedding. This embedding process captures the semantic representation of individual nodes. Subsequently, we establish connections between nodes based on their relationships (comments, retweets, and other behaviors) to form a graph structure, as shown in Fig. 1. In the context of rumor detection, our dataset $\mathcal{G} = \{G_1,$

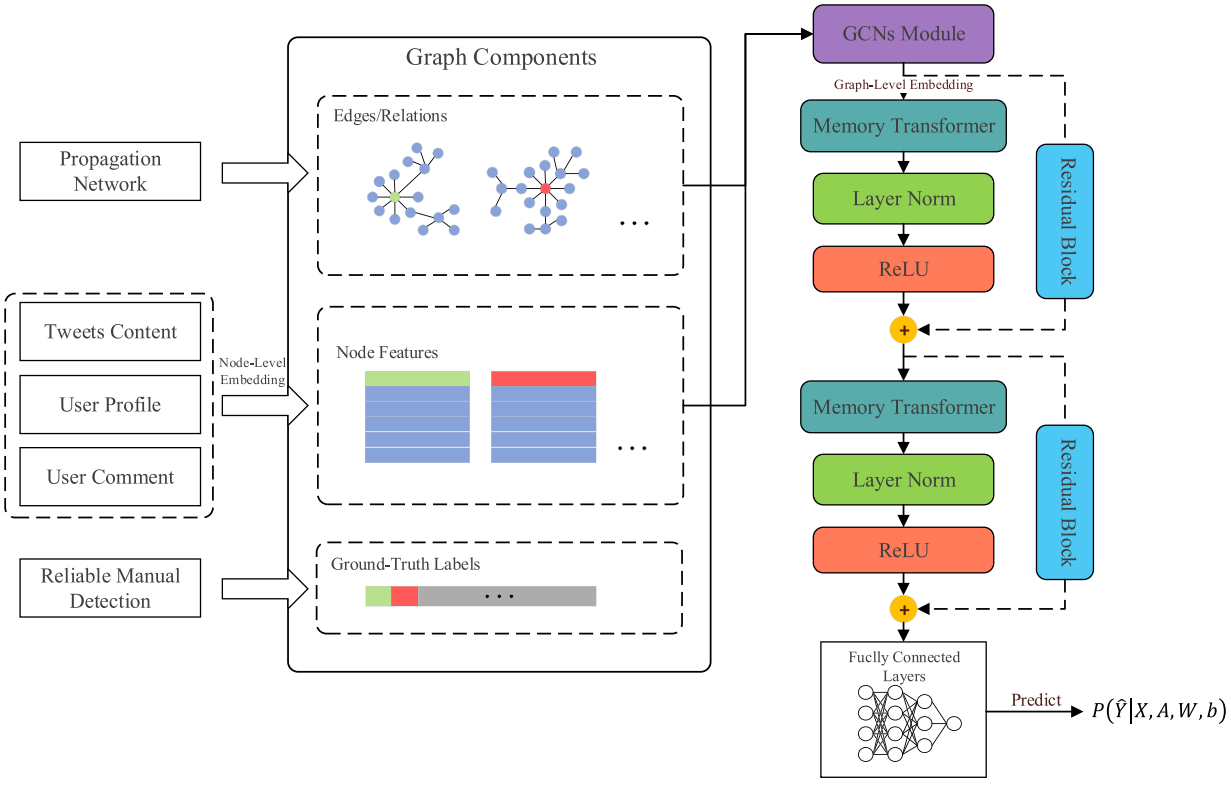


Fig. 1. The architecture of GCNs-MT. The left portion of the figure showcases the graph structures comprising root nodes (tweets) and the corresponding node feature tensor, accompanied by the corresponding ground-truth labels vector for each graph. On the right side, the process architecture of GNNs-MT is depicted, elucidating the sequential flow of operations.

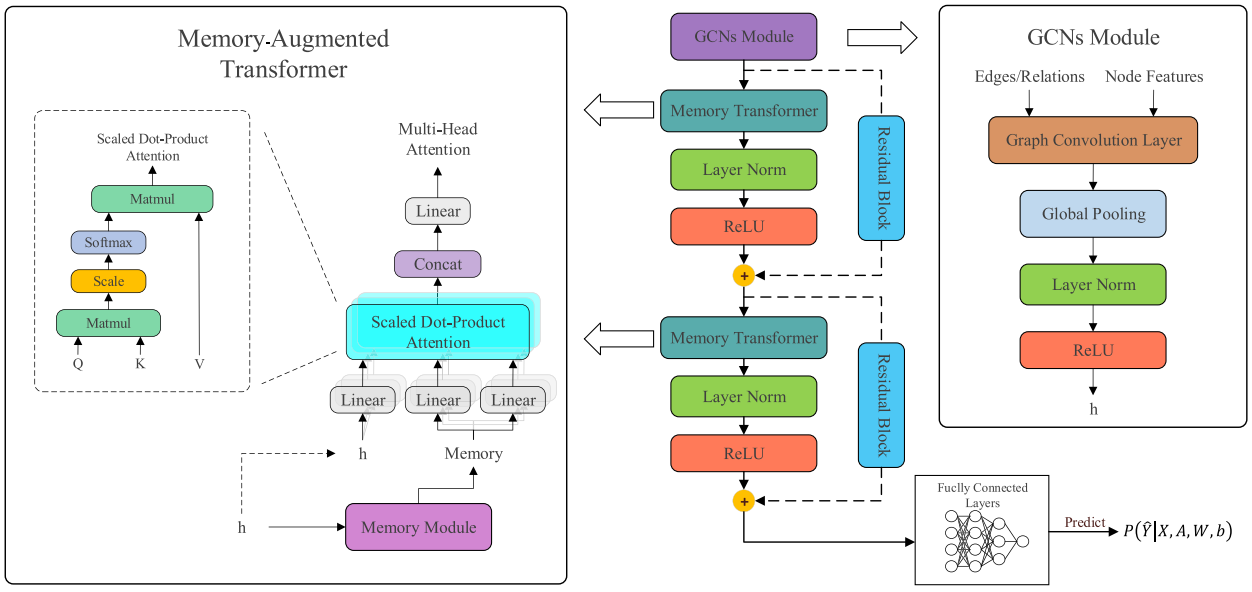


Fig. 2. The architecture of GCNs Module and Memory-Augmented Transformer.

$G_2, \cdots, G_N\}$ comprises N tree-like propagation networks of tweets. Each G_i corresponds to the i th event, where N represents the total number of events in the dataset. Specifically, for each G_i, we have $G_i = \{X_i, E_i\}$, where $X_i = \{r^i, x_1^i, x_2^i, \cdots, x_{n_i-1}^i\}$ denotes the node feature representations. In addition, $E_i = \{e_{sd}^i | s, d = 1, 2, \cdots, n_i\}$ represents the set of edges, capturing the interactions between posts in G_i. Here r^i signifies the root node representation, x_j^i is the j-th leaf node representation, and e_{sd}^i signifies the interaction between the s-th and d-th nodes in G_i. Moreover, for the d-th node responding to an s-th node, we observe a directed edge $e_s^i \rightarrow e_d^i$. The total number of posts in G_i is denoted by n_i. The set of edges E_i can be succinctly represented using an adjacency matrix, $A_i \in \{0,1\}^{n_i \times n_i}$, where each entry a_{sd}^i is defined as

$$a_{sd}^i = \begin{cases} 1, & if\ e_{ds}^i \in E_i \\ 0, & otherwise \end{cases}. \tag{1}$$

Moreover, each graph G_i has a corresponding ground-truth label $y_i \in (0,1)$, where $y_i = 1$ and $y_i = 0$ indicate that the event is classified as a rumor and a non-rumor, respectively. The principal objective in the field of rumor detection is to train a classifier f: $\mathcal{G} \rightarrow Y$, where $\mathcal{G}$ represents the set of events within the dataset and Y corresponds to the set of ground-truth labels. The fundamental task of the classifier is to make accurate predictions regarding the labels of individual events, taking into consideration influential factors such as textual contents, user profiles, and the intricate propagation patterns formed by the interrelated posts associated with each event. To capture the structural information and relationships within the event propagation network, we employ graph convolution operations. Specifically, we utilize the GCNs framework to perform node-level embeddings and aggregate information across the graph. The graph convolution operation can be summarized as follows:

$$H = GCNs(X, A) \tag{2}$$

where GCN($\cdot$) refers to a specific graph convolution operation function within the aforementioned GCN frameworks, X is the node embedding, and A denotes the adjacency matrix. In our study, we leverage three popular and efficient GCN frameworks: GCN [38], graph attention network (GAT) [39], and GraphSAGE (SAmple and aggreGatE, short for SAGE) [40]. In this study, we present a concise exposition of GCN and its utility for node-level convolution and embedding. The GCN operation is mathematically expressed as follows:

$$H^{(l+1)} = \sigma\left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^{(l)} W^{(l)}\right) \tag{3}$$

where $H^{(l)}$ denotes the node representations in the l-th layer. The activation function $\sigma(\cdot)$ (e.g., the ReLU function) is element-wise applied to the output. The degree matrix D is a diagonal matrix containing the degrees of individual nodes along the diagonal, and $W^{(l)}$ symbolizes the learnable weight matrix specific to the l-th layer of the GCN. Notably, $H^{(0)}$ corresponds to the initial node feature embedding denoted as X, which serves as the input to the first layer of the GCNs module. To capture graph-level embeddings, we leverage global pooling techniques to aggregate node-level features, enabling the generation of higher-level representations at the graph level. Global pooling operations play a pivotal role in summarizing information across all nodes within the graph, resulting in a condensed graph-level feature. Various pooling methods can be employed, such as max pooling, mean pooling, or graph-level attention mechanisms. In our model framework, we adopt the approach of average pooling, which facilitates the seamless integration of node-level information. Let $h_{ij}^{(l+1)} \in \mathbb{R}^d$ denote the j-th node feature vector of the i th graph in the l-th layer, where d is the hidden dimension. Then the global mean pooling operation can be expressed as follows:

$$h_i^{(l+1)} = \frac{1}{n_i} \sum_{j=1}^{n_i} h_{ij}^{(l)} \tag{4}$$

$$H^{(l+1)} = ReLU\left(\text{LayerNorm}\left(H^{(l)}\right)\right) \tag{5}$$

where $h_i^{(l+1)} \in \mathbb{R}^d$ is the pooled graph-level feature vector, and $H^{(l)}$ represents a set of hidden tensors in the l-th layer, which contains the representations of all nodes in certain graphs. The ReLU($\cdot$) function applies element-wise rectification to the aggregated features, while the LayerNorm($\cdot$) operation normalizes the activations within each layer to improve model stability and convergence.

3.2. Memory-Augmented transformer

The GCNs-MT framework not only leverages the GCNs module to capture the structural characteristics of rumor propagation but also integrates the memory-augmented transformer to facilitate seamless interactions between local and global dependencies of events. Through this integration, the model adeptly transforms local sequences into

global representations based on the memory vector, enabling a comprehensive and contextually rich understanding of rumor propagation. The memory-augmented transformer incorporates a memory module that utilizes multiple RNN units to capture global structural patterns. Specifically, we employ the LSTM [41] cells to retain and update a memory tensor $m \in \mathbb{R}^d$, which is designed to capture global dependencies and obtain the global representation of graph-level embeddings. The LSTM cell operates over multiple time steps, updating the memory tensor based on the current hidden state and the previous memory tensor.

Consider a batch of inputs with a sequence length T. Within this batch, the LSTM cell operates on the inputs as if they were sequential data, even though they are not temporally related. This cell allows the model to process the inputs effectively and capture relevant patterns within the dataset. For batch t, the LSTM cell takes as input the hidden state $h_{(t)} \in \mathbb{R}^{T \times d}$ and the previous memory tensor $m_{(t-1)}$, where $m_{(t-1)}$ can be considered as a global representation retained and updated by a recurrent manner. The LSTM cell equations can be written as follows:

$$i_{(t)} = \sigma\left(h_{(t)}W_{hi} + m_{(t-1)}W_{mi} + b_i\right) \tag{6}$$

$$f_{(t)} = \sigma\left(h_{(t)}W_{hf} + m_{(t-1)}W_{mf} + b_f\right) \tag{7}$$

$$o_{(t)} = \sigma\left(h_{(t)}W_{ho} + m_{(t-1)}W_{mo} + b_o\right) \tag{8}$$

$$g_{(t)} = tanh\left(h_{(t)}W_{hg} + m_{(t-1)}W_{mg} + b_g\right) \tag{9}$$

$$c_{(t)} = f_{(t)} \odot c_{(t-1)} + i_{(t)} \odot g_{(t)} \tag{10}$$

$$m_{(t)} = o_{(t)} \odot tanh\left(c_{(t)}\right) \tag{11}$$

where $\sigma(\cdot)$ represents the sigmoid activation function. The memory gate ($g_{(t)}$) is computed based on the current hidden state and the previous memory tensor. The memory cell ($c_{(t)}$) is updated using the input gate ($i_{(t)}$) and forget gate ($f_{(t)}$), and the memory tensor ($m_{(t)}$) is obtained by applying the output gate ($o_{(t)}$) to the hyperbolic tangent of the memory cell. Similar to the multi-head attention mechanism, we can stack multiple LSTM cells in a parallel paradigm in a multilabel classification task.

Next, we use the multi-head attention mechanism to transform the current local hidden state into a global hidden state. After obtaining the memory tensor, we utilize it as the key and value inputs for the Transformer module, while the hidden state h serves as the query input. This process allows us to construct a memory-augmented Transformer for enhanced rumor detection. The memory tensor is used as the key and value inputs for the Transformer module, while the hidden state h serves as the query input. The attention weights $\alpha_{c,i}^{(l)}$ are computed by taking the dot-product between the query and key vectors and applying a softmax function to normalize the scores:

$$q_{c,i}^{(l)} = h_i^{(l)}W_{c,q}^{(l)} + b_{c,q}^{(l)} \tag{12}$$

$$k_{c,i}^{(l)} = m_i^{(l)}W_{c,k}^{(l)} + b_{c,k}^{(l)} \tag{13}$$

$$\alpha_{c,i}^{(l)} = \frac{q_{c,i}^{(l)}, k_{c,i}^{(l)}}{\sum_{j=1}^{M} q_{c,ij}^{(l)}, k_{c,ij}^{(l)}} \tag{14}$$

where $\langle q, k\rangle = \exp\left(\frac{q^T k}{\sqrt{d}}\right)$ represents exponential scaled dot-product function, d is the hidden size of each head, and M is the number of LSTM cells. To the c-th head attention, the hidden state $h_i^{(l)}$ and the structural memory feature $m_i^{(l)}$ are transformed into query vector $q_{c,i}^{(l)} \in \mathbb{R}^d$ and $k_{c,i}^{(l)} \in \mathbb{R}^d$ respectively using different trainable parameters $W_{c,q}^{(l)}$, $W_{c,k}^{(l)}$, $b_{c,q}^{(l)}$, and $b_{c,k}^{(l)}$. After performing the multi-head attention mechanism, the message aggregation step follows. In this step, we transform the structural memory feature $m_i^{(l)}$ into a message vector $v_{c,i}^{(l)}$ for each head attention:

$$v_{c,i}^{(l)} = m_i^{(l)}W_{c,v} + b_{c,v}^{(l)} \tag{15}$$

$$\widehat{h}_i^{(l)} = \|_{c=1}^{C}\left[\alpha_{c,i}^{(l)}v_{c,i}^{(l)}\right] \tag{16}$$

where the $\|$ is the concatenation operation for C head attention. Furthermore, we incorporate a residual block [42] between layers to prevent model oversmoothing, as shown in Fig. 2. The gating mechanism $\varphi_i^{(l)}$ controls the contribution of the enriched representation $\widehat{h}_i^{(l+1)}$ relative to the original representation $h_i^{(l)}$ and is computed using the sigmoid activation function:

$$\varphi_i^{(l)} = sigmoid\left(h_i^{(l)}W_r + b_r^{(l)}\right) \tag{17}$$

$$h_i^{(l+1)} = ReLU\left(\text{LayerNorm}\left(\left[\left(1-\varphi_i^{(l)}\right)\widehat{h}_i^{(l)} + \varphi_i^{(l)} \,\|\, {}_{c=1}^{C} h_i^{(l)}\right]\right)\right). \tag{18}$$

Then the graph-level representation $h_i^{(l+1)}$ in the last layer of the memory-augmented transformer can be considered as the final basis for classification. By applying the gating mechanism and the residual connection, the residual block allows the model to selectively combine the original representation with the enriched information obtained from the attention mechanism, which helps prevent oversmoothing and ensures that important information is preserved throughout the layers of the memory-augmented transformer.

3.3. Graph classifier for rumor detection

Upon completing the memory-augmented transformer processing, the model's output traverses a series of fully connected layers for classification. The trainable parameters of the model undergo iterative updates via gradient descent, aiming to minimize the cross-entropy loss function. Inferring the predicted event label $\widehat{y}_i$ entails passing the output through a sequence of fully connected layers followed by a softmax activation function. Our proposed approach optimizes all model parameters by minimizing the cross-entropy divergence between the predicted probability distributions and the ground-truth distributions across the entire event dataset. The process can be formulated as follows:

$$\widehat{y}_i = softmax(\text{FC}(h_i)) \tag{19}$$

$$L_i = -\sum_{j=1}^{\mathscr{D}} y_{ij}\log\widehat{y}_{ij} \tag{20}$$

where $\widehat{y}_i \in \mathbb{R}^{\mathscr{D}}$ represents a probability vector containing the predicted probabilities for all classes used to predict the label of the i th event, and L_i represents the cross-entropy loss between the ground-truth label y_{ij} and the predicted probabilities $\widehat{y}_{ij}$. In addition, an L2 regularizer is incorporated into the loss function to manage the model's complexity and mitigate overfitting. This regularization term penalizes large values of the model parameters.

4. Experiments

In this section, we deploy our models to a Chinese dataset constructed from real Weibo data and two English corpora benchmark datasets built by other researchers. We thoroughly evaluate the accuracy and generalization capabilities of our model based on experimental results. We compare our model against other state-of-the-art baseline methods for rumor detection. Furthermore, we performed an ablation study to assess the individual contributions of each module in our model architecture. The thorough evaluation conducted in this study enables us to rigorously assess the efficiency and performance of our proposed

model in detecting rumors.

4.1. Experimental datasets

We conducted evaluations of our method on three real-world datasets, including a self-built Chinese corpus dataset and two benchmark English corpus datasets provided by other researchers. The raw data were crawled by Ma et al. [27]. In constructing our Chinese corpus dataset Weibo, we incorporated three distinct node features: 14-dimensional user profile features, 300-dimensional Word2Vec[1] features, and 768-dimensional BERT[2] features extracted from user comments, responses, and other textual content. Each graph in the dataset represents a tree structure that represents an event, with edges denoting retweeting or responding behavioral relationships. The ground-truth labels for events in our Chinese corpus were provided by the Sina Community Management Center (SCMC) ,[3] while the ground-truth labels for events from Twitter in the two English corpora, curated by other researchers, were annotated by well-known rumor detection systems Politifact[4] and Gossipcop.[5] The general statistics for the datasets are shown in Table 1.

In addition, we made histograms based on the counts of retweets and responses as well as the cumulative frequency of events across the three datasets, as shown in Fig. 3. The distribution of forwarding and reply counts exhibits significant variations among the datasets, which serves as a rigorous test to assess the comprehensiveness of our model's performance. The consistent patterns observed in the cumulative frequencies across the three datasets indicate that the extent of impact spread for all events may conform to a specific distribution.

4.2. Experimental setup

In our experiments, we evaluated the performance of well-known conventional machine learning algorithms, namely RF and SVM. In addition, we compared our proposed method with several state-of-the-art deep learning methods:

- Word2Vec-MLP and BERT-MLP: Two classification models based on multilayer perceptron (MLP) architecture. These models utilize features extracted from pretrained Word2Vec and BERT models, respectively, to encode the event representations for classification purposes.
- GCNFN[6] [43]: The GCNFN leverages deep geometric learning techniques to model the propagation network in conjunction with textual node embedding features for fake news detection. The architecture consists of two graph convolutional layers, followed by two fully connected layers, and a softmax layer for prediction.

Table 1
General statistics for the three datasets.

Identification Agency	Twitter		Weibo
Dataset	Politifact	Gossipcop	SCMC
#Rumors	157	2732	2313
#Non-rumors	157	2732	2351
#Graphs	314	5464	4664
#Total Nodes	41,054	314,262	2,856,741
#Total Edges	40,740	308,798	2,183,388
#Avg. Nodes per Graph	131	58	613
#Avg. Edges per Graph	130	57	612

- RvNN[7] [22]: The RvNN is a deep learning method suitable for information passing in tree-like rumor spreading networks, where rumor information is passed and aggregated in a recursive manner.
- Bi-GCN[8] [35]: According to the authors, the method represents the first application of GCN in a rumor detection model, incorporating a bidirectional propagation structure.
- UPFD[9] [23]: The UPFD is a deep learning model with strong performance in detecting fake news and rumors. This model leverages GNN-based techniques and an information fusion framework for effective analysis and classification.
- DCUK[10]: The DCUK is an innovative parallel stacking approach that combines the strengths of GAT, GAT with BERT, and transformer. This approach leverages these individual components, learns from them, and then seamlessly stitches them together to generate overall features for the classification task.

Our implementation of well-known conventional machine learning algorithms is based on the scikit-learn[11] library, while the implementation of deep learning baselines utilizes the PyTorch[12] library and the PyTorch Geometric (PyG)[13] library. We adopt the same train–validation–est split (70 %–10 %–20 %) for all models. Our models were trained with a consistent graph-level embedding size of 128, a batch size of 128, and an L2 regularization weight of 0.001. Learning rates were adjusted based on the specific graph convolution methods. For the models with graph convolution operations based on GAT and GCN, a learning rate of 0.001 was used, while the MT model under SAGE convolution used a learning rate of 0.005. To mitigate overfitting, we implemented early stopping with patience of 10 epochs during training. In addition, due to the specificity of the respective structures of the baseline models, we mostly use the same settings as in the open source repository of the original authors of the models during the training tests, while for the nonopen source models, we use settings similar to those used for our model training. In our study, we employ accuracy (Acc.) score, F1 score, precision (Prec.) score, and recall (Rec.) score as the evaluation metrics to assess the performance of our model on the datasets. The evaluation metrics can be calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{21}$$

$$Precision = \frac{TP}{TP + FP} \tag{22}$$

$$Recall = \frac{TP}{TP + FN} \tag{23}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{24}$$

where *TP* represents the true positive, *TN* denotes the true negative, *FP* represents the false positive, and *FN* represents the false negative. These metrics provide a comprehensive evaluation of the models' performance in terms of both correct classification and the trade-off between precision and recall.

4.3. Experimental results and performance

As shown in Tables 2–4, our meticulous analysis unequivocally reveals the pronounced supremacy of deep learning methodologies

[1] https://spacy.io/models/zh#zh_core_web_lg
[2] https://github.com/jina-ai/clip-as-service
[3] https://service.account.weibo.com/
[4] https://www.politifact.com/
[5] https://www.gossipcop.com/
[6] https://github.com/YingtongDou/GCNN
[7] https://github.com/majingCUHK/Rumor_RvNN/
[8] https://github.com/TianBian95/BiGCN
[9] https://github.com/safe-graph/GNN-FakeNews
[10] https://github.com/ltian678/DUCK-code/
[11] https://scikit-learn.org/stable/
[12] https://pytorch.org/
[13] https://github.com/pyg-team/pytorch_geometric

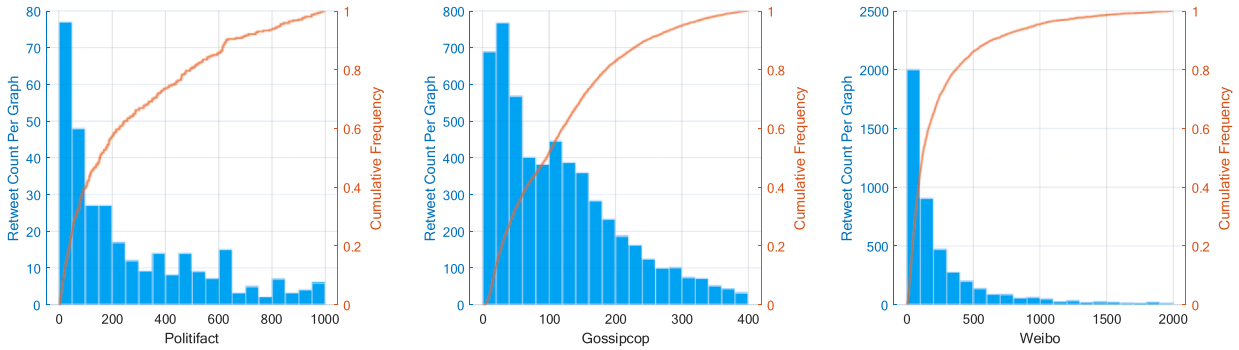


Fig. 3. Histogram of retweets and responses count and cumulative frequency for graphs of the three datasets.

Table 2
The results of comparative experiments on the Politifact dataset.

Model	Feature Source	Politifact			
		Acc.	F1	Prec.	Rec.
RF	Event Only	0.7557	0.7545	0.7615	0.7477
SVM	Event Only	0.7602	0.7580	0.7685	0.747
BERT-MLP	Event Only	0.7873	0.7983	0.7623	0.8378
Word2Vec-MLP	Event Only	0.7738	0.7685	0.7905	0.7477
GCNFN	Event+Social Context	0.8235	0.8267	0.8158	0.8378
RvNN	Event+User+Network	0.8145	0.8093	0.8365	0.7838
Bi-GCN	Event+User+Network	0.8281	0.8304	0.8230	0.8378
UPFD	Event+User+Network	0.8371	0.8378	0.8378	0.8378
DUCK	Event+User+Network	0.8416	0.8458	0.8421	0.8495
GCN-MT (ours)	Event+User+Network	0.8371	0.8235	0.7434	0.8922
GAT-MT (ours)	Event+User+Network	**0.8507**	**0.8465**	**0.8053**	**0.9231**
SAGE-MT (ours)	Event+User+Network	0.8462	0.8411	0.7965	0.8911

The best result is highlighted in bold, and the second-best result is highlighted in underline.

Table 3
The results of comparative experiments on the Gossipcop dataset.

Model	Feature Source	Gossipcop			
		Acc.	F1	Prec.	Rec.
RF	Event Only	0.8054	0.8037	0.8148	0.7928
SVM	Event Only	0.8190	0.8214	0.8142	0.8288
BERT-MLP	Event Only	0.8481	0.8478	0.8498	0.8458
Word2Vec-MLP	Event Only	0.8685	0.8675	0.8746	0.8604
GCNFN	Event+Social Context	0.9496	0.9495	0.9498	0.9493
RvNN	Event+User+Network	0.9451	0.9458	0.9348	0.9572
Bi-GCN	Event+User+Network	0.9593	0.9596	0.9554	0.9640
UPFD	Event+User+Network	0.9629	0.9632	0.9577	0.9687
DUCK	Event+User+Network	0.9726	0.9726	0.9703	0.9749
GCN-MT (ours)	Event+User+Network	0.9712	0.9713	0.9708	0.9718
GAT-MT (ours)	Event+User+Network	0.9773	0.9773	0.9796	0.9751
SAGE-MT (ours)	Event+User+Network	**0.9825**	**0.9825**	**1.0000**	**0.9655**

The best result is highlighted in bold, and the second-best result is highlighted in underline.

compared with conventional machine learning approaches in the domain of rumor detection. This marked advantage can be attributed to the inherent capacity of deep learning models to adeptly learn and harness intricate, nonlinear representations from the underlying data. By leveraging the hierarchical architecture of multiple layers, deep learning models proficiently capture and exploit complex patterns and latent associations present in rumor-related features. This astute

Table 4
The results of comparative experiments on the Weibo dataset.

Model	Feature Source	Weibo			
		Acc.	F1	Prec.	Rec.
RF	Event Only	0.8312	0.8324	0.8266	0.8382
SVM	Event Only	0.8529	0.8522	0.8564	0.8480
BERT-MLP	Event Only	0.8983	0.8980	0.9002	0.8958
Word2Vec-MLP	Event Only	0.9010	0.9009	0.9023	0.8995
GCNFN	Event+Social Context	0.9375	0.9376	0.9364	0.9387
RvNN	Event+User+Network	0.9593	0.9593	0.9578	0.9608
Bi-GCN	Event+User+Network	0.9721	0.9722	0.9695	0.9749
UPFD	Event+User+Network	0.9746	0.9746	0.9731	0.9761
DUCK	Event+User+Network	0.9825	0.9825	0.9834	0.9816
GCN-MT (ours)	Event+User+Network	0.9786	0.9787	0.9733	0.9841
GAT-MT (ours)	Event+User+Network	0.9831	0.9832	0.9890	0.9776
SAGE-MT (ours)	Event+User+Network	**0.9877**	**0.9878**	**0.9890**	**0.9866**

The best result is highlighted in bold, and the second-best result is highlighted in underline.

capability enables them to extract highly discriminative and information-rich representations, thus culminating in significantly improved and superior detection performance.

Furthermore, deep learning methods based on MLP architecture have significant shortcomings in capturing the complex structure inherent in rumor propagation networks. The shortcoming lies in their inability to explicitly model the graph structure on which information propagation depends, thus hindering their ability to fully understand the impact and propagation patterns of rumors. RvNN models, alternatively, although they can learn graph structure information representations, do so without being able to generate global dependencies, and their purely recurrent inference paradigm makes them cost-consuming to apply in practice. In contrast, GCN-based models are able to interpret graph structures in detail and discern how information spreads throughout the network through the learning process. As a result, GCN-based models are well able to encapsulate the complex dynamics of rumor propagation. The hybrid model, which combines BERT, GCNs, and the Transformers, gains a distinct advantage in extracting features from the networks of rumor propagation. The model's adeptness at learning pertinent high-level representations results in remarkable performance improvements in rumor detection.

Lastly, our proposed method, which combines GCNs and the memory-augmented Transformer, outperforms other GCN-based baseline approaches across all evaluation metrics. This performance can be attributed to the unique strengths of each component. The GCNs module captures graph-level dependencies and leverages the structural information in the rumor propagation network, enabling effective representation learning. The memory-augmented Transformer, consequently, enhances the ability of the model to capture global dependencies and

incorporate global contextual information, thereby improving its discrimination and generalization capabilities.

We conducted a comparative analysis of our three proposed models: GCN-MT, GAT-MT, and SAGE-MT. The performance comparisons are illustrated in Table 5, 6 and 7. In the realm of performance evaluation, the SAGE-MT model emerges as the standout performer. This supremacy is primarily attributed to SAGE's inductive learning capabilities, which excel in capturing higher-order dependencies within expansive rumor propagation network graphs and efficiently disseminating information throughout the intricate graph structures. The model's ability to encompass a broad array of relationships within the graph structure, even in the context of substantial networks, plays a pivotal role in its superior performance. Moreover, our evaluation of the GAT-MT model reveals that it excels when applied to the Politifact dataset, characterized by compact graphs and a scarcity of samples. This observation implies that these models exhibit remarkable resilience and capacity to perform well in scenarios with limited training data, rendering them particularly valuable for constrained data settings.

Moving on to training efficiency, we observed a notable trend wherein the SAGE-MT model consistently outperforms GCN-MT and GAT-MT, as visually depicted in Fig. 4. The superior training efficiency of SAGE-MT can be attributed to the astute aggregation strategy it employs through SAGE convolution. By incorporating random sampling during the training process, SAGE accelerates computation and convergence speed, significantly expediting the learning process. This strategic advantage not only streamlines model training but also enables the acquisition of high-level representations at an accelerated pace.

In our comprehensive evaluation, we assessed the models' sensitivity by analyzing the receiver operating characteristic (ROC) curve, as visually presented in Fig. 5. The ROC curve provides valuable insights into how the models' performance varies with changing binary discrimination thresholds, effectively illustrating the trade-off between false positive and true positive rates. The results demonstrate that, in the case of large datasets such as Gossipcop and Weibo, there is little discernible difference in sensitivity among the three models, and all of them exhibit exceptional performance. However, when scrutinizing the smaller Politifact dataset, a notable distinction emerges—the GCN-MT model exhibits higher sensitivity compared with GAT-MT and SAGE-MT. This observation underscores the GCN-MT model's exceptional discriminatory prowess in accurately discerning true from false rumors, thereby significantly diminishing the likelihood of false negatives. This model may excel at capturing subtle nuances in these smaller networks.

4.4. Ablation study

In the ablation experiments, we conducted two distinct types of analyses, specifically, component ablation and information ablation, as illustrated in Table 8. Within the framework of component ablation, a series of systematic experiments were executed to elucidate the influence of various individual components or modules inherent in our proposed model. To this end, we rigorously removed or made deliberate modifications to designated elements of the model to evaluate their respective contributions to the overall performance. By contrasting the performance outcomes of these modified versions with the complete GCNs-MT model, we were able to glean valuable insights into the efficacy and relative significance of each constituent. In the information ablation study, we employed inputs, such as events bereft of structural information, user profiles enriched with propagated structural data, and user comment features, to gauge their impact on the comprehensive performance metrics. The component variant models include the following:

- GCNs-MT w/o GCNs: We remove the graph convolution module and feed the pooled features directly into the MT module set. This variant will consider the impact on model identification performance when graph structure information processing is removed.
- GCNs-MT w/o MT: This variant model removes the MT module, and the hidden embeddings output by the graph convolution layer will be directly pooled and fed into the fully connected layers as graph-level representations. This variant model serves the purpose of evaluating the significance of the proposed MT module in the context of rumor detection.
- GCNs-MT w/o Memory: The variant is a model modification that excludes the LSTM cell and reverts the original Transformer component. In this variant, the LSTM cell initially responsible for retaining global dependencies is removed. In addition, the multiple attention mechanism in the converter is replaced by a self-attention mechanism.
- GCNs-MT w/o Transformer: The variant is a model adaptation where the Transformer module is excluded from the original model. In this variant, the Transformer, which incorporates a multi-head attention mechanism, is omitted. This alteration is to assess the impact of applying the multi-head attention mechanism within the model.

The information variant models include the following:

Table 5
Comparison of the best performance of the proposed model on the Politifact dataset.

Model	Feature	Politifact			
		Acc.	F1	Prec.	Rec.
GCN-MT	Profile	0.8009	0.8053	0.8053	0.8053
	Word2Vec	0.8326	0.824	0.7699	0.8878
	BERT	0.8371	0.8235	0.7434	0.8922
GAT-MT	Profile	0.7602	0.7440	0.6814	0.8191
	Word2Vec	**0.8507**	**0.8465**	**0.8053**	**0.9231**
	BERT	0.8326	0.8279	0.7876	0.8725
SAGE-MT	Profile	0.7783	0.7742	0.7434	0.8077
	Word2Vec	0.8416	0.8341	0.7788	0.8980
	BERT	0.8462	0.8411	0.7965	0.8911

Table 6
Comparison of the best performance of the proposed model on the Gossipcop dataset.

Model	Feature	Gossipcop			
		Acc.	F1	Prec.	Rec.
GCN-MT	Profile	0.9354	0.9360	0.9421	0.9299
	Word2Vec	0.9686	0.9687	0.9703	0.9672
	BERT	0.9712	0.9713	0.9708	0.9718
GAT-MT	Profile	0.9365	0.9382	0.9629	0.9147
	Word2Vec	0.9712	0.9713	0.9729	0.9698
	BERT	0.9773	0.9773	0.9796	0.9751
SAGE-MT	Profile	0.9237	0.9253	0.9442	0.9072
	Word2Vec	0.9712	0.9715	0.9776	0.9655
	BERT	**0.9825**	**0.9825**	**1.0000**	**0.9655**

Table 7
Comparison of the best performance of the proposed model on the Weibo dataset.

Model	Feature	Weibo			
		Acc.	F1	Prec.	Rec.
GCN-MT	Profile	0.9259	0.9258	0.9269	0.9246
	Word2Vec	0.9743	0.9743	0.9749	0.9737
	BERT	0.9786	0.9787	0.9733	0.9841
GAT-MT	Profile	0.9157	0.9158	0.9167	0.9150
	Word2Vec	0.9789	0.9789	0.9804	0.9774
	BERT	0.9831	0.9832	0.9890	0.9776
SAGE-MT	Profile	0.9243	0.9243	0.9246	0.9241
	Word2Vec	0.9801	0.9801	0.9816	0.9786
	BERT	**0.9877**	**0.9878**	**0.9890**	**0.9866**

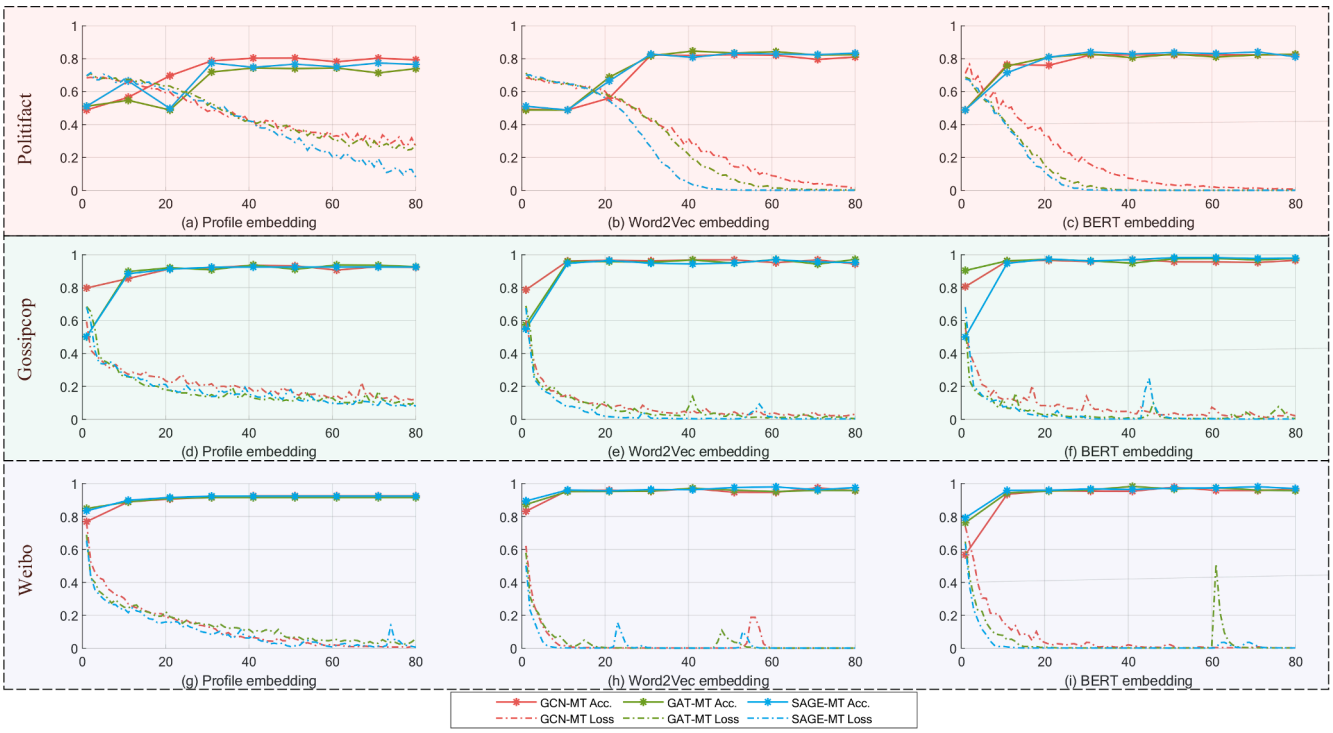


Fig. 4. The performance of the proposed models during training. The solid lines show the trend of model test set accuracy with the training epoch, and the dashed lines show the trend of cross-entropy loss with the training epoch.

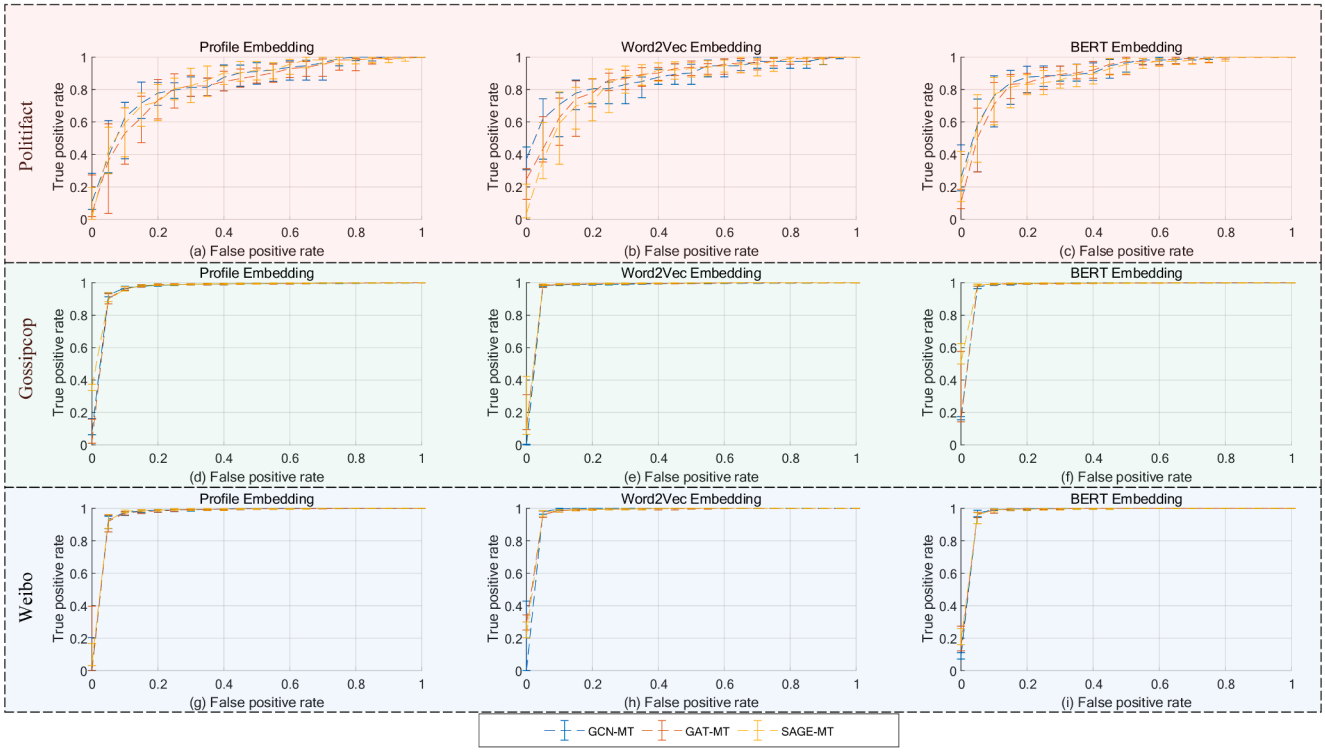


Fig. 5. ROC curves with pointwise confidence bounds. The true positive rate and the false positive rate indicate the proportion of samples in which the event is actually a rumor and a non-rumor, respectively, that the model is able to correctly identify.

- GCNs-MT w/ event: The variant is designed to focus exclusively on the event itself as the sole source of information input into the model. In contrast to the original model that incorporates various user characteristics, this variant intentionally excludes those factors.
- GCNs-MT w/ profiles: The GCNs-MT w/profiles variant is designed to concentrate exclusively on user profile information within networks featuring rumor propagation. In this model variant, the emphasis is placed on the structure of the propagation network,

Table 8
Variant models in ablation experiments.

Component Variant	Model Components		
	GCNs	Memory	Transformer
GCNs-MT w/o GCNs	×	✔	✔
GCNs-MT w/o MT	✔	×	×
GCNs-MT w/o Memory	×	✔	✔
GCNs-MT w/o Transformer	✔	✔	×
GCNs-MT	✔	✔	✔

Information Variant	w/o Propagation Structures	w/ Propagation Structures	
	Event	User Profiles	User Comments
GCNs-MT w/ event	✔	×	×
GCNs-MT w/ profiles	✔	✔	×
GCNs-MT w/ comments	✔	×	✔
GCNs-MT	✔	✔	✔

where the transmission of user profiles serves as a critical feature medium. This variant primarily considers the role of user profiles and their influence on rumor propagation within the network.

- GCNs-MT w/ comments: The variant is designed to exclusively consider user comments within network information on social media platforms. In this model variant, the primary focus is on the structure of the communication network, with user comments serving as the essential medium for delivering characteristics and information. This model variant places an emphasis on the influence of user comments and their subjective opinions on the overall performance of the model.

As depicted in Fig. 6, a series of experiments conducted with the building block variants of our model reveal several crucial insights. First, the full-fledged GCNs-MT model consistently outperforms the experiments conducted with variants lacking one of the model's key components. This performance underscores the comprehensive nature of our proposed method and underscores the indispensability of each individual construct. Second, when the model is devoid of the MT module, a significant drop in performance is observed. In contrast, the absence of either a single memory or Transformer module results in a relatively less pronounced decline in performance. This finding indicates that both the memory and Transformer modules contribute positively to the model's efficacy. Notably, the omission of the memory module has a more substantial impact, causing precision scores to drop significantly and leading to an increased misclassification of non-rumors as rumors. This outcome underscores the crucial role played by the generated global dependency representations in shaping the model's classification results. Furthermore, the absence of graph convolution operations, which results in the model lacking access to crucial structural information from the network, also exerts a significant adverse effect on the model's performance. This result highlights the pivotal role of propagating network structural information in the context of rumor detection. In summary, the experiments underscore the comprehensive and interdependent nature of the model's components, with the MT module, memory module, and graph convolution operations each playing a unique and vital role in enhancing the model's performance in rumor detection.

As for the information variant experiments, we can find from the results presented in Fig. 7 that utilizing the complete and unaltered content of the information source yields more significant and notable effects across all three experimental datasets. This observation underscores the importance of comprehensive information when it comes to enhancing the model's performance in rumor screening. Moreover, experimenting with using only the event itself as the source of information for rumor screening proves to be challenging, with comparatively less favorable results. In contrast, methods that leverage network structural information perform better, which suggests that, while the event itself is a critical component, the model's ability to harness additional contextual and structural information significantly contributes to its efficacy in rumor detection. It is worth noting that user profiles and the content of user comment tweets within the social media communication network emerge as vital sources of features. Among these, the content of users' comment tweets regarding the event stands out as the second most effective source of information, following the complete information source. This result highlights the substantial influence of user-generated content and commentary on the model's ability to detect rumors effectively.

5. Conclusion

The propagation of rumors on social media platforms can have unpredictable consequences, as the continuity of human civilization relies on the dissemination of accurate knowledge and attitudes. Incorrect information needs to be detected and corrected. Our proposed approach combines the graph convolution operation of GCNs, the structural memory of LSTM cells, and the multi-head attention mechanism of the Transformer. Through extensive evaluations of Chinese and English datasets, GCNs-MT consistently outperforms existing methods,

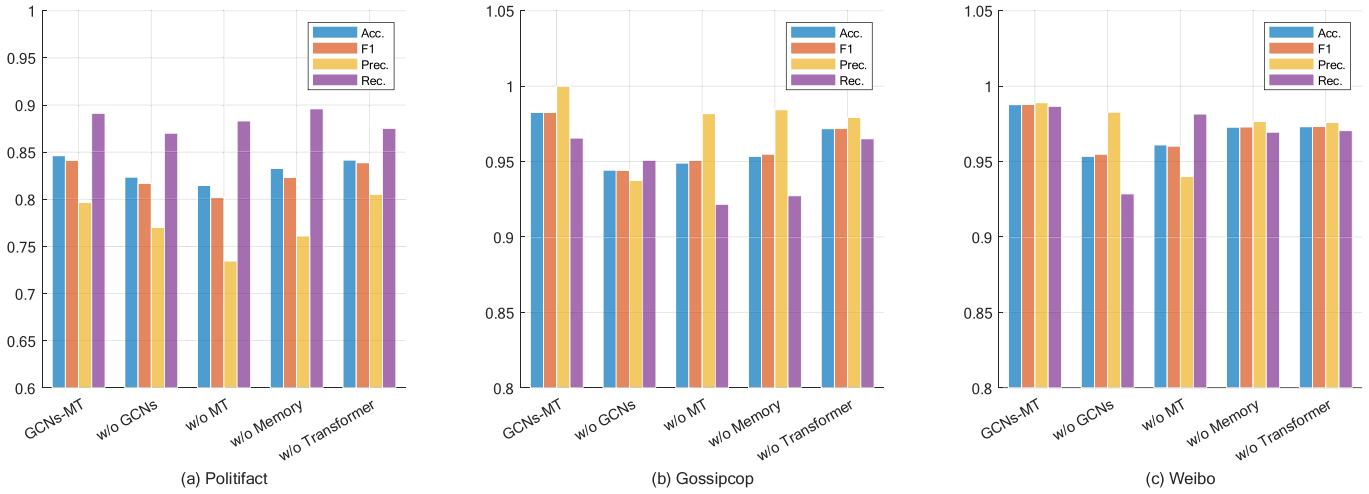


Fig. 6. Experimental performance of component variant models. Compact groups of bars indicate individual variant model results, with x-axis labels indicating which specific model they belong to. Different colored bars indicate different evaluation metrics.

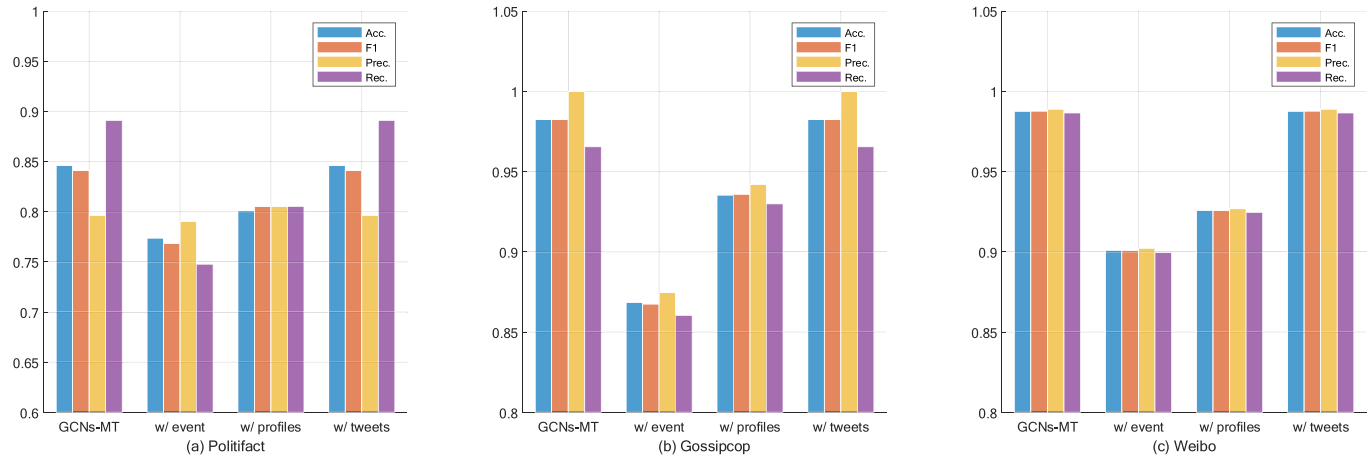


Fig. 7. Experimental performance of information variant models. Compact groups of bars indicate individual variant model results, with x-axis labels indicating which specific model they belong to. Different colored bars indicate different evaluation metrics.

demonstrating its effectiveness in identifying and combating misinformation. Our research opens up new avenues for addressing the challenges posed by false information in online environments. Future studies can further explore the potential of GCNs-MT and extend its application to other domains related to information dissemination and social media analysis.

CRediT authorship contribution statement

Qian Chang: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Xia Li:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision. **Zhao Duan:** Methodology, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Xia Li reports financial support was provided by The Ministry of Education in China of Humanities and Social Science Project. Xia Li reports financial support was provided by The National Natural Science Foundation of China.

Data availability

Data will be made available on request.

Acknowledgements

This research is partially supported by the National Natural Science Foundation of China (No.71503099, No.71974069), the Ministry of Education in China of Humanities and Social Science Project (No. 20YJC630067).

References

[1] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, R. Procter, Detection and resolution of rumours in social media: a survey, ACM Comput. Surv. (CSUR) 51 (2) (2018) 1–36.
[2] P. Meel, D.K. Vishwakarma, Fake news, rumor, information pollution in social media and web: a contemporary survey of state-of-the-arts, challenges and opportunities, Expert. Syst. Appl. 153 (2020) 112986.
[3] A.L. Opdahl, T. Al-Moslmi, D.T. Dang-Nguyen, M. Gallofré Ocaña, B. Tessem, C. Veres, Semantic knowledge graphs for the news: A review, 55, ACM Computing Surveys, 2022, pp. 1–38.
[4] J.A. Tucker, A. Guess, P. Barberá, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal B. Nyhan, Social media, political polarization, and political disinformation: a review of the scientific literature. Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018), 2018.
[5] H. Ahmed, I. Traore, S. Saad, Detection of online fake news using n-gram analysis and machine learning techniques, in: in Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments: First International Conference, ISDDC 2017, Vancouver, BC, Canada, October 26-28, 2017, Proceedings 1, Springer, 2017.
[6] V.U. Gongane, M.V. Munot A. Anuse, Machine learning approaches for rumor detection on social media platforms: a comprehensive survey. Adv. Mach. Intellig. Signal Proces., 2022: p. 649–663.
[7] J. Ma, W. Gao, Z. Wei, Y. Lu, K.F. Wong, Detect rumors using time series of social context information on microblogging websites, in: in Proceedings of the 24th ACM international on conference on information and knowledge management, 2015.
[8] S. Kwon, M. Cha, K. Jung, Rumor detection over varying time windows, PLoS. One 12 (1) (2017) e0168344.
[9] Y. Liu, Y.F. Wu, Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks, in: in Proceedings of the AAAI conference on artificial intelligence, 2018.
[10] A. Roy, K. Basak, A. Ekbal P. Bhattacharyya, A deep ensemble framework for fake news detection and classification. arXiv preprint arXiv:1811.04670, 2018.
[11] M.R. Islam, S. Liu, X. Wang, G. Xu, Deep learning for misinformation detection on online social networks: a survey and new perspectives, Soc. Netw. Anal. Min. 10 (2020) 1–20.
[12] T. Mikolov, K. Chen, G. Corrado J. Dean, Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.
[13] J. Devlin, M.W. Chang, K. Lee K. Toutanova, Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv: 1810.04805, 2018.
[14] R.K. Kaliyar, A. Goswami, P. Narang, FakeBERT: fake news detection in social media with a BERT-based deep learning approach, Multimed. Tools. Appl. 80 (8) (2021) 11765–11788.
[15] N.R. de Oliveira, P.S. Pisa, M.A. Lopez, D.S.V. de Medeiros, D.M. Mattos, Identifying fake news on social networks based on natural language processing: trends and challenges, Information 12 (1) (2021) 38.
[16] J. Ma, J. Li, W. Gao, Y. Yang, K.F. Wong, Improving rumor detection by promoting information campaigns with transformer-based generative adversarial learning, IEEe Trans. Knowl. Data Eng. 35 (3) (2023) 2657–2670.
[17] J.D. Lv, X.G. Wang, C.L. Shao, TMIF: transformer-based multi-modal interactive fusion for automatic rumor detection, Multimedia Systems, 2022.
[18] J. Ma, W. Gao, K.F. Wong, Detect rumors in microblog posts using propagation structure via kernel learning, Association for Computational Linguistics, 2017.
[19] K. Wu, S. Yang, K.Q. Zhu, False rumors detection on sina weibo by propagation structures, in: in 2015 IEEE 31st international conference on data engineering, IEEE, 2015.
[20] A. Bondielli, F. Marcelloni, A survey on fake news and rumour detection techniques, Inf Sci (Ny) 497 (2019) 38–55.
[21] X. Zhang, A.A. Ghorbani, An overview of online fake news: characterization, detection, and discussion, Inf. Process. Manage 57 (2) (2020) 102025.
[22] J. Ma, W. Gao, K.F. Wong, Rumor detection on twitter with tree-structured recursive neural networks, Association for Computational Linguistics, 2018.
[23] Y. Dou, K. Shu, C. Xia, P.S. Yu, L. Sun, User preference-aware fake news detection, in: in Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021.
[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Å.u. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. (2017) 30.
[25] Q. Chang, X. Li, Z. Duan, Graph global attention network with memory: a deep learning approach for fake news detection, Neural Network. (2024) 106115.

[26] L. Zeng, K. Starbird, E. Spiro, # unconfirmed: classifying rumor stance in crisis-related social media messages, in: in Proceedings of the International AAAI Conference on Web and Social Media, 2016.
[27] J. Ma, W. Gao, P. Mitra, S. Kwon, B.J. Jansen, K.F. Wong M. Cha, Detecting rumors from microblogs with recurrent neural networks. 2016.
[28] T. Chen, X. Li, H. Yin, J. Zhang, Call attention to rumors: deep attention based recurrent neural networks for early rumor detection. in Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2018 Workshops, BDASC, BDM, ML4Cyber, PAISI, DaMEMO, Springer, Melbourne, VIC, Australia, 2018. June 3, 2018, Revised Selected Papers 22.
[29] F. Yu, Q. Liu, S. Wu, L. Wang, T. Tan, A convolutional approach for misinformation identification, IJCAI, 2017.
[30] R. Wang, Z. Li, J. Cao, T. Chen, L. Wang, Convolutional recurrent neural networks for text classification, in: in 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 2019.
[31] J. Ma, W. Gao, K.F. Wong, Detect rumors on twitter by promoting information campaigns with generative adversarial learning, in: The world wide web conference, 2019.
[32] S.A. Alkhodair, S.H. Ding, B.C. Fung, J. Liu, Detecting breaking news rumors of emerging topics in social media, Inf. Process. Manage 57 (2) (2020) 102018.
[33] V. Vziatysheva, How fake news spreads online? Internat. J. Media Informat. Literacy 5 (2) (2020).
[34] M. Dong, B. Zheng, N. Quoc Viet Hung, H. Su, G. Li, Multiple rumor source detection with graph convolutional networks, in: in Proceedings of the 28th ACM international conference on information and knowledge management, 2019.
[35] 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: in Proceedings of the AAAI conference on artificial intelligence, 2020.
[36] Y.J. Lu C.T. Li, GCAN: graph-aware co-attention networks for explainable fake news detection on social media. arXiv preprint arXiv:2004.11648, 2020.
[37] M. Sun, X. Zhang, J. Zheng, G. Ma, Ddgcn: dual dynamic graph convolutional networks for rumor detection on social media, in: in Proceedings of the AAAI conference on artificial intelligence, 2022.
[38] T.N. Kipf M. Welling, Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907, 2016.
[39] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio Y. Bengio, Graph attention networks. stat, 2017. 1050(20): p. 10.48550.
[40] W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, Adv. Neural Inf. Process. Syst. 30 (2017).
[41] S. Hochreiter, J.r. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[42] G. Li, M. Muller, A. Thabet, B. Ghanem, Deepgcns: can gcns go as deep as cnns?, in: in Proceedings of the IEEE/CVF international conference on computer vision, 2019.
[43] F. Monti, F. Frasca, D. Eynard, D. Mannion M.M. Bronstein, Fake news detection on social media using geometric deep learning. arXiv preprint arXiv:1902.06673, 2019.