Neural Networks 172 (2024) 106115

Contents lists available at ScienceDirect

Neural Networks

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

Graph global attention network with memory: A deep learning approach for fake news detection

Qian Chang, Xia Li *, Zhao Duan

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

ARTICLE INFO

Keywords:
Fake news detection
Graph convolutional networks
Social network
Graph classification

ABSTRACT

With the proliferation of social media, the detection of fake news has become a critical issue that poses a significant threat to society. The dissemination of fake information can lead to social harm and damage the credibility of information. To address this issue, deep learning has emerged as a promising approach, especially with the development of Natural Language Processing (NLP). This study introduces a novel approach called Graph Global Attention Network with Memory (GANM) for detecting fake news. This approach leverages NLP techniques to encode nodes with news context and user content. It employs three graph convolutional networks to extract informative features from the news propagation network and aggregates endogenous and exogenous user information. This methodology aims to address the challenge of identifying fake news within the context of social media. Innovatively, the GANM combines two strategies. First, a novel global attention mechanism with memory is employed in the GANM to learn the structural homogeneity of news propagation networks, which is the attention mechanism of a single graph with a history of all graphs. Second, we design a module for partial key information learning aggregation to emphasize the acquisition of partial key information in the graph and merge node-level embeddings with graph-level embeddings into fine-grained joint information. Our proposed method provides a new direction in news detection research with a combination of global and partial information and achieves promising performance on real-world datasets.

Introduction

A developing interest identified with online interactions on social media has drawn in numerous content participants, including the 1.09 billion mobile internet users in China, with 670 million photos and 100 million short videos shared daily in 2021 on Tencent WeChat (Institute, 2022), whereas misinformation is intentionally generated to agitate the public or manipulated several times in the process of widespread dissemination on social media platforms and other online channels (Gorrell et al., 2018; Gupta et al., 2012; Zhou & Zafarani, 2018). Malicious intent users use multimedia as a tool to proliferate fake news, which not only misleads public opinion but also puts social events at risk, such as incitement of violence or election interference, due to the low cost of maintaining social media as well as platform access, which results in significant challenges in the field of fake news detection (Bastick, 2021; Lazer et al., 2018; Vosoughi et al., 2018). Fake news can span a range of topics, styles, and platforms, and is characterized by various entities, such as the news article, the creators and spreaders of the news, and the surrounding social context. The endogenous and exogenous information conveyed by these aspects of news are crucial for accurate news characterization and play a pivotal role in the detection of false news (Wang et al., 2020; Zhang & Ghorbani, 2020).

One approach to mitigating the destructive impact caused by misinformation is expert-based fact-checking, which is labor-intensive and time-consuming. To break through the limitations of this manual approach, some previous researchers tried to program a range of handcrafted characteristics that were fed into a machine learning model to discern fake news attributed to technological advancement (Gupta et al., 2012; Kaur et al., 2020). However, fake news can be created and disseminated in various forms, including text, images, and videos, and can quickly evolve and adapt to new platforms and contexts (Khattar et al., 2019; Singhal et al., 2019). Furthermore, it can be difficult to distinguish fake news from genuine news, especially when it is well-crafted and strategically targeted (Zhang & Ghorbani, 2020). Detecting fake news has thus attracted widespread attention from researchers across various disciplines. Early research with conventional machine learning methods has yielded certain gains for the identification of fake news, but these methods often require hand-crafted features

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

https://doi.org/10.1016/j.neunet.2024.106115
Received 20 May 2023; Received in revised form 11 November 2023; Accepted 5 January 2024
Available online 8 January 2024

as the basis for detection, which makes it difficult to capture high-level representations of the news information, making the model's detection performance unable to achieve the desired results (Mridha et al., 2021; Wani et al., 2021). Deep learning, with its capacity to autonomously acquire meaningful high-level feature representations from data, has emerged as a promising approach for addressing this challenge (Monti et al., 2019; Wani et al., 2021).

Traditional deep learning techniques are commonly used for unimodal fake news detection, with a focus on the textual content of the news, such as the news headline and its corresponding subject content (Dhawan et al., 2022; Palani et al., 2022; Singh et al., 2021; Singhal et al., 2019). Content-based approaches for modeling the flow of information within fake news texts have predominantly relied on recurrent and convolutional neural networks (Goldani et al., 2021; Kaliyar et al., 2020; Yang et al., 2018). Nevertheless, it's important to acknowledge that these approaches might exhibit certain limitations in the domain of fake news detection. Recurrent and convolutional neural networks excel at handling sequential and spatial information, yet they might struggle to adeptly grapple with the intricacies of multi-source, multimodal data often encountered in fake news scenarios (Chen, 2015; Goldani et al., 2021). The intricacies of contextual and nuanced features, particularly within extended texts, and the complex non-linear traits of fake news could pose challenges that conventional feature extraction methodologies may inadvertently overlook. The development of Natural Language Processing (NLP) (Oshikawa et al., 2018; Z. Zhou et al., 2019) and Computer Vision (CV) (Kaur et al., 2020) has facilitated the use of multimodal features for the detection of fake news (Dhawan et al., 2022). These features comprise a range of information types, including news content, social media content, user profiles of fake news creators and propagators (Rangel et al., 2020), and audience demographics (Zhou & Zafarani, 2018). These diverse data types are encoded in various formats, such as textual and visual modalities, and are disseminated across social media networks (Khattar et al., 2019; Nakamura et al., 2019). The effective processing of these data and the accurate detection of fake news require sophisticated techniques and algorithms that can handle the complexity and scale of the data (Figueira & Oliveira, 2017; Gupta et al., 2012). While methods enhanced by NLP techniques may excel at extracting semantic information (Aggarwal et al., 2020; Altınel & Ganiz, 2018), they often fall short of capturing the structural information inherent in the dissemination of fake news across social networks. This limitation stems from their focus on text-based content analysis, leaving out critical insights from the network's topology and dynamics (Malhotra & Vishwakarma, 2020; Mridha et al., 2021).

Recent research has demonstrated the superior performance of Graph Convolutional Network Networks (GCNs) in fake news detection compared to traditional machine learning methods (Figueira & Oliveira, 2017; Varlamis et al., 2022) due to their ability to capture both the structural and semantic features of news articles and social media posts. And GCNs can be used to incorporate user information, such as user interactions, user profiles, and social network connections, to further improve the accuracy of fake news detection (Liu & Wu, 2018; Shu et al., 2020a). Previous research has shown that fake news tends to spread faster and wider than real news (Pierri & Ceri, 2019), and users who spread fake news are often more active and have higher centrality in the network (Gupta et al., 2022). Specifically, GCNs can incorporate such information into the model, providing a more fine-grained analysis of the dissemination dynamics of fake news by considering the interactions between users and their connections in the network. Furthermore, GCNs can capture the underlying factors that contribute to the spread of fake news, such as user beliefs, social influence, and network structure (Oshikawa et al., 2018; Varlamis et al., 2022; Zhou & Zafarani, 2018). However, these investigations have primarily concentrated on localized information within the distribution network of fake news, encompassing tweets, user profiles, and the underlying network structure (Vosoughi et al., 2018; Vziatysheva, 2020). The uniform global structural relationships that exist throughout holistic networks have been overlooked, and this broader contextual information holds significance for the extraction of structural features crucial in news detection.

In response to the limitations and obstacles observed in prior research, we have introduced a novel Global Attention Module with Memory within our framework. This module possesses the capability to identify the alignment between network structures through an attention mechanism and maintain this understanding through a memory mechanism. Consequently, it acquires an encompassing representation of the entire graph, facilitating enhanced fake news detection at a more profound semantic level. Simultaneously, the attention mechanism embedded in this module is adept at learning the global structural information of the fake news dissemination network, thereby bolstering the capacity to capture the overarching essence of fake news. Empirical studies from a sociological and psychological perspective have revealed that user preferences and online news consumption behaviors are not only influenced by endogenous preferences but also by exogenous contents present within social communication networks (Shu et al., 2019a). In addition, the veracity of news also has a significant impact on the structure of the social propagation network of news (Liu & Wu, 2018; Shu et al., 2020a), and aggregating localized information within a news communication network will also be an important component of fake news detection. In this study, we propose another Partial Key Message Learning strategy that focuses on aggregating the endogenous preferences of users within the social media communication network and the exogenous content that exists in the social communication network, aiming to better integrate the information within the network, and the news content is also an important piece of information for identifying fake news dissemination. The main contributions of this work are as follows:

- We employ pre-trained word embeddings, a feature engineering technique of NLP, to encode and embed news content, user profiles, and user-generated content in heterogeneous networks to construct a Chinese corpus dataset for fake news detection.
- We harness the capabilities of three advanced and widely recognized GCNs to extract pertinent features and amalgamate endogenous user preferences with exogenous contextual data. This strategic approach is aimed at enhancing the overall efficacy of fake news detection.
- We present a groundbreaking approach for fake news detection named the Graph Global Attention Network with Memory (GANM). The GANM approach combines two distinct strategies: the Global Attention Module with Memory for learning and retaining global structural information and the Partial Key Message Learning module for extracting essential messages. Our proposed approach was applied to real-world datasets, yielding promising results.
- We obtained quantitative analysis results on the significance of local and global information for fake news detection by comparing ablation experiments. These experiments revealed that local and global information play distinct roles in model learning due to their inherent differences.

Related work

Deep learning approach for fake news detection

Many conventional machine learning techniques have been extensively employed to identify false information propagation online, including Support Vector Machine (SVM) (Hussain et al., 2020), Naive Bayes (NB) (Aphiwongsophon & Chongstitvatana, 2018), and Random Forests (RF) (Basu et al., 2022). Nevertheless, these methods predominantly concentrate on feature extraction from news content, encompassing source, headline, body text, image, or video. Misinformation spreaders can exploit this reliance on textual news to undermine machine learning-based models (Mridha et al., 2021). Moreover, the majority of traditional machine learning methods necessitate manual

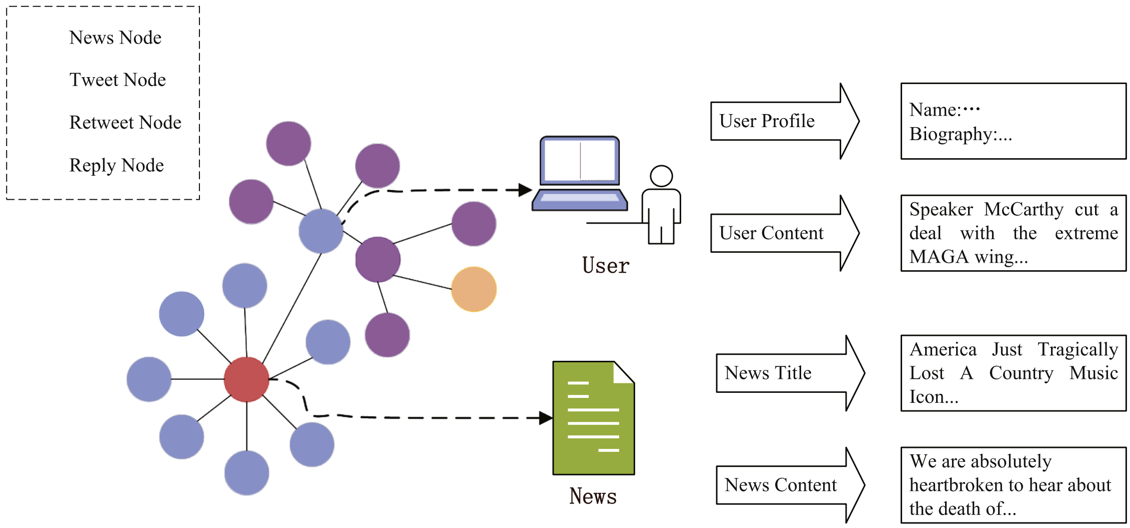

Fig. 1. Overview of the heterogeneous graphical representation of news propagation networks. The news propagation networks are represented as a tree graph, where the red node is the root news node of the tree graph, the other leaf nodes represent the tweeting users, the retweeting users, and the replying users, while the edges of the graph represent the reposting behaviors of the users towards the news, the commenting behaviors of the news, and the replying behaviors of the comments.

feature engineering, impeding the extraction of high-level feature representations and thereby leading to inefficiency in their approaches (Baarir & Djeffal, 2021; Zhang & Ghorbani, 2020).

With the recent advancements in deep learning technology, algorithms such as recurrent neural networks (RNN) and autoencoders have emerged as powerful tools for natural language embedding, memorizing essential semantic sequences, and capturing underlying semantic relationships, resulting in high-level feature representations (Choudhary & Arora, 2021). Consequently, deep learning algorithms have gained popularity in online fake news detection, particularly RNN and CNN-based deep learning methods. Bahad et al. (2019) proposed a fake news detection model based on bi-directional LSTM-recurrent neural networks, and their experimental results showed that the bi-directional LSTM-RNN model outperformed the uni-directional model. Shu et al. (2019b) applied a Bi-directional GRU (Bi-GRU) architecture and introduced a sentence-comment co-attentive sub-network model called dEFEND (interpretable fake news detection), which leverages news content and user comments for fake news detection. Kumar et al. (2020) introduced a Multi-level Convolutional Neural Network (MCNN) that incorporates local convolutional features and global semantic features to effectively capture semantic information in article text for classifying news as fake or genuine. The CNN and RNN-based methods do not rely on manually created text features, enabling them to capture news contextual information and semantic structure effectively, making them viable solutions for fake news representation and detection. Unlike RNN or CNN, the attention mechanism preserves word dependencies in a sentence regardless of their distance from each other. Lu and Li (2020) developed a novel model named GCAN for predicting fake news based on source tweets and their propagation-based users through a unique co-attention network mechanism that combines source-interaction co-attention representation, source-propagation co-attention representation, and GRU-based propagation representation. Although these deep learning-based methods yield high-level feature representations, their feature sources mainly focus on the news content, overlooking joint features related to users, dissemination networks, and social contexts. This vulnerability allows creators of fake news to exploit the system for targeted attacks, making it necessary to reduce the detection model's dependency on news text.

Graph Neural Networks (GNNs) are a novel technique that applies deep learning algorithms to graph structures. The unique structure of GNN holds the potential to unify content-based, communication-based, and social context-based approaches (Monti et al., 2019). GNN-based models can achieve comparable or higher performance than modern methods that do not rely on textual information (Han et al., 2020). Monti et al. (2019) proposed a geometric deep learning-based approach for fake news detection that utilizes a primarily GCN-based end-to-end network framework consisting of two graph convolutional layers and two fully connected layers to integrate news-related information, such as user profiles, user interactions, network structure, dissemination patterns, and content, through the construction of heterogeneous graph data. Bahad et al. (2019) introduced a novel bi-directional graph model called Bi-Directional Graph Convolutional Networks (Bi-GCN), which explores two aspects of features by manipulating the top-down and bottom-up propagation of rumors. It employs GCN with top-down rumor propagation-directed graphs to learn the pattern of rumor propagation and GCN with opposite rumor diffusion-directed graphs to capture the structure of rumor diffusion. On the other hand, Van-Hoang et al. (2020) utilized GraphSAGE as the core module of graph convolutional operations for inductive representation learning. Compared with transduction models (e.g., GCN, GAT, etc.), GraphSAGE does not require maintaining all nodes, making it scalable during training and efficient during inference without re-processing the entire graph. This approach enables more efficient news processing and detection, especially when dealing with larger spreading networks. Wei et al. (2021) proposed a novel edge-enhanced Bayesian graph convolutional network (EBGCN) to capture robust structural features by employing a Bayesian approach to adaptively think about the reliability of potential relationships between users in a propagation network. Furthermore, the advancement of NLP technology has introduced novel solutions to the realm of fake news detection. Many researchers have synergistically harnessed GNNs and NLP techniques, paving the way for cutting-edge advancements in this domain. Sun et al. (2022) leverage BERT (Devlin et al., 2018) to extract semantic information from textual content. They incorporate Graph Adversarial Contrastive Learning (GACL) into the loss function to discern disparities between dialogue threads in the same and distinct categories. Additionally, they introduce an Adversarial Feature Transformation (AFT) module to create conflicting samples, compelling the model to unearth event-invariant features. Tian et al. (2022) harnessed BERT to bolster semantic analysis, employed Transformer for capturing the significance of attentional connections among tweets, and integrated the Graph Attention Network (GAT) to gather graph structural insights. This amalgamation resulted in the development of a comprehensive

DUCK model. While these studies have made significant contributions and demonstrated commendable performance, they primarily concentrated on local information, specifically intra-graph information, when gathering insights about communication structures. However, they largely neglected the broader perspective of global structural information. It is this gap that our current study endeavors to bridge. Our research strives to incorporate global information spanning multiple news stories into the detection process. This ensures that the assessment of news authenticity isn't confined solely to the acquisition of local information.

News propagation networks

With the increasing prevalence of social media platforms as a means for news propagation, the traditional media is no longer the sole source for such information, leading to a transformation in the way news is disseminated and an increase in research on the news propagation network (Martens et al., 2018). This network is typically modeled as a heterogeneous graph consisting of nodes for news articles, users, and replies, where the relationships between these nodes are defined by the behavior of users in terms of tweeting and retweeting. When examining the dissemination of fake news, this network can be viewed as a heterogeneous graph, with a root node representing the news article and leaf nodes representing tweet nodes, which correspond to the users disseminating the news. The leaf nodes of the tweet nodes represent the retweet users. The news propagation network can be represented in the form of a heterogeneous graph, as shown in Fig. 1.

The news propagation network serves as a valuable source of information within the realm of fake news detection. The feature sources of this network can be categorized into two distinct groups: endogenous and exogenous information (Feng, 2022; Zhou & Zafarani, 2018; X. Zhou et al., 2019). Endogenous information is concerned with the interactions and relationships that occur between news nodes, tweeting nodes, and retweeting nodes, as well as aspects such as the title and content of the news article, user tweeting behavior, and comment content. On the other hand, exogenous information pertains to contextually relevant information from social media platforms that is not easily encoded in machine learning models, including details related to users' social network relationships, personal information, and geographic location. Together, these various sources of information provide a multifaceted framework of clues and support for the effective detection of fake news (Dou et al., 2021; Shu et al., 2020b). Therefore, by utilizing the characteristics of news propagation networks and advanced machine learning methods such as graph neural networks, the detection and identification of fake information can be effectively achieved. Analyzing and modeling news propagation networks can reveal the propagation pathways and influence of false information, which is of great significance for developing effective strategies for preventing the spread of fake information.

Graph convolutional networks

Graph convolutional networks (GCNs) are a class of neural networks designed to process data represented in graph structures. GNNs are capable of learning both node-level and graph-level representations by iteratively aggregating the information of the neighbors of each node in the graph. The key operation in GNNs is the graph convolution operation, which is used to propagate node-level information to higher-level graph representations. There are various types of GCNs, each with its own unique architecture and set of parameters. Many popular GCNs are widely employed in scientific research and practical applications, including Graph Convolutional Network (GCN) (Kipf & Welling, 2016), Graph Attention Network (GAT) (Velickovic et al., 2017), and GraphSAGE (SAGE) (Hamilton et al., 2017).

GCN is one of the most widely used types of GCNs. GCN employs a spectral graph convolution operation to learn node embeddings by leveraging the graph Laplacian matrix, which encodes the graph structure. The convolutional operation is performed by multiplying the node feature matrix with the graph Laplacian, followed by a non-linear activation function. For a node i, its $(l+1)$-th layer feature representation $h_i^{(l+1)}$ can be obtained by taking a weighted average of the l-th layer feature representations $h_j^{(l)}$ of its neighboring nodes j. Specifically, the convolution operation formula of GCN is:

$$h_i^{(l+1)} = \sigma\left(\sum_{j\epsilon N(i)} \frac{1}{\sqrt{d_i d_j}} h_j^{(l)} W^{(l)}\right) \tag{1}$$

where $N(i)$ denotes the set of first-order neighboring nodes of node i, d_i is the degree of node i, $W^{(l)}$ is the learnable weight matrix of the l-th layer, and σ is the activation function.

GAT employs an attention mechanism to learn node embeddings. The attention mechanism allows the model to learn to assign different weights to the neighboring nodes when computing the final node representation, based on the similarity between the nodes. The GAT convolution operation can be written as:

$$h_i^{(l+1)} = \sigma\left(\sum_{j\epsilon N(i)} \alpha_{i,j}^{(l)} h_j^{(l)} W^{(l)}\right) \tag{2}$$

where $\alpha_{i,j}^{(l)}$ denotes the attention coefficient between nodes i and j in the l-th layer. The other parameter symbols have the same meaning as those expressed in the previous GCN convolution operation formula.

GraphSAGE is a GNN model that employs a graph-level aggregation function to compute node embeddings. SAGE operates by iteratively aggregating the feature information of neighboring nodes to compute a node's embedding. The key innovation in SAGE is the use of a learnable aggregator function that can be trained end-to-end with the rest of the model. The SAGE convolution operation can be written as:

$$h_i^{(l+1)} = \sigma\left(\left[h_i^{(l)} \| \frac{1}{|N(i)|} \sum_{j\epsilon N(i)} h_j^{(l)}\right] W^{(l)}\right) \tag{3}$$

where $\|$ denotes the vector concatenation operation. The other parameter symbols have the same meaning as those expressed in the previous GCN and GAT convolution operation formulas.

Problem statement

In the realm of fake mews detection, let $G = \{G_1, G_2, \cdots, G_N\}$ denotes the dataset, where Gᵢ corresponds to the i-th propagation network of news, and N represents the total number of events encompassed within the dataset. Specifically, $G_i = \{X_i, E_i\}$ signifies the node feature set $X_i = \{r^i, x_1^i, x_2^i, \cdots, x_{n_i-1}^i\}$, while a set of edges $E_i = \{e_{sd}^i | s, d = 1, 2, \cdots, n_i\}$ captures the relationships between responded posts and retweeted or responsive posts within G_i. Here, r^i signifies the embedding of the root node, x_j^i embodies the embedding of the j-th leaf node, and e_{sd}^i characterizes the interaction between the s-th and D-th nodes. In particular, an edge $e_s^i \rightarrow e_d^i$ is present if the D-th node responds to the s-th node. The parameter n_i denotes the total count of posts in G_i. Furthermore, the set of edges E_i is concisely portrayed using an adjacency matrix $A_i \in \{0,1\}^{n_i \times n_i}$, where

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

Moreover, each distinct graph G_i corresponds to an associated ground-truth label $y_i \in \{0,1\}$. Specifically, $y_i = 1$ when the related news is confirmed as true, while $y_i = 0$ if the news is ascertained to be false. The core aim of the news detection task is to formulate a classifier function $f(G) \rightarrow Y$. In this context, G signifies the ensemble of news

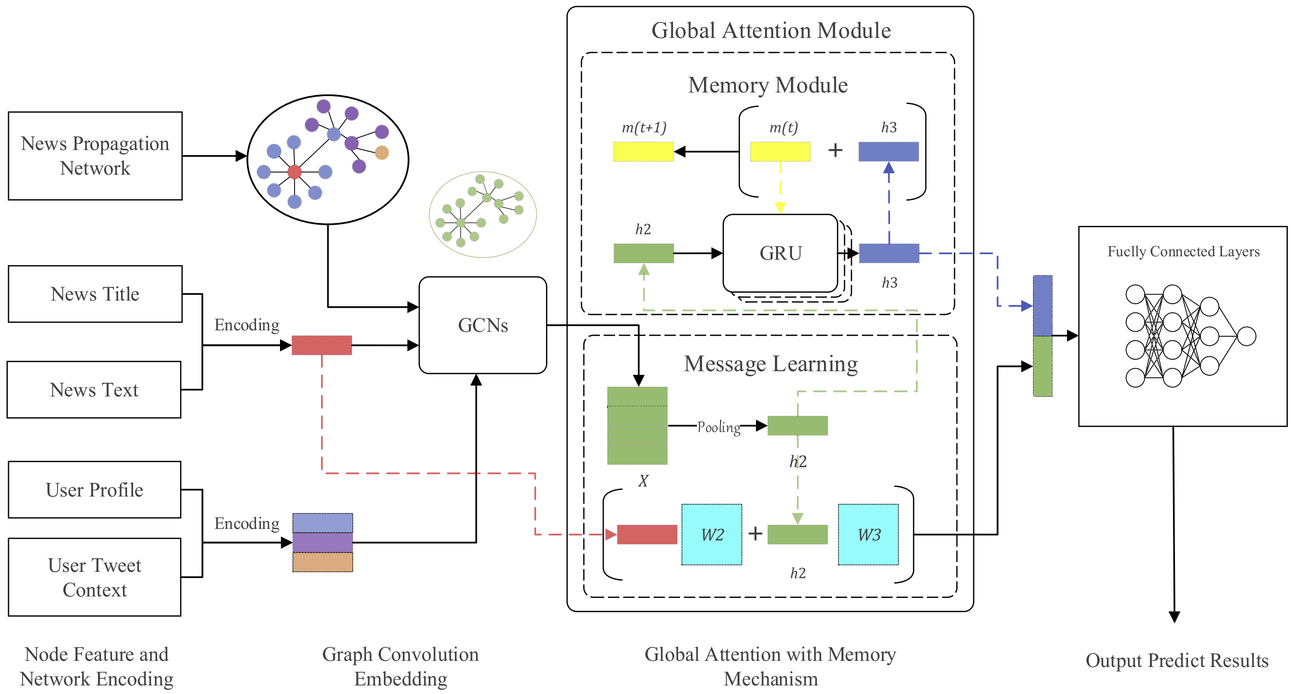

Fig. 2. An overview illustration of the GANM framework. The colored rectangles represent different tensors, while the solid black arrows depict the flow to the next network layer. The dashed arrows indicate tensor migration or duplication. The notes below show the main operations of the GANM framework.

within the dataset, while Y represents the set of corresponding ground-truth labels. The objective of this classifier function is to predict the label of a given event by considering a diverse array of factors, encompassing textual content, user profiles, and the structural propagation patterns formed by the interconnected posts associated with that particular event.

The proposed approach: GANM

Our proposed Graph Global Attention Network with Memory (GANM) is an end-to-end neural network aimed at enhancing the comprehension of global graph structure while capturing essential local information for precise graph classification tasks. An overview of the GANM is presented in Fig. 2. Initially, we embed features of different node types into the graph, which serves as input to a graph convolutional network responsible for feature convolutional embedding. Subsequently, the resultant convolutional node features are fed into our core modules, the Global Attention Module with Memory and the Partial Key Message Learning module. The Global Attention Module with Memory is responsible for capturing the global structural information dependency and storing it in the memory component, and the Partial Key Message Learning is responsible for learning the key partial messages. The outputs of the two modules are fused to produce a comprehensive feature vector. Lastly, the fully connected layers module reduces dimensionality and produces the output essential for graph classification.

Node-level embedding and graph-level embedding

As previously elucidated, the node features within the graph stem from the NLP encoding process. Through the establishment of edges interconnecting nodes, the graph $G(V,E)$ is constructed. This collection of graphs is denoted as $\{G_1, G_2, \cdots, G_n\}$. Three classical graph convolution methods GCN, SAGE and GAT are applied in node-level embedding, which can be simply expressed as:

$$X = GCNs(G(V,E)) \tag{4}$$

where $GCNs(\cdot)$ denotes a composite network function encompassing multiple layers of graph convolution, and The resultant feature matrix X encapsulates the outcome of multi-layer graph convolution. Further compression of the matrix data is performed below. Global pooling is a type of aggregation for graph-level embedding operation that combines information from all nodes in a graph into a single vector. This operation is typically used as the final layer of a GNN to produce a fixed-size feature vector that can be used for downstream tasks such as node classification or graph classification. Formally, let $X \in \mathbb{R}^{N \times F}$ be the matrix of node features for a graph with N nodes and F features per node. The goal of global pooling is to produce a fixed-size feature vector $h_2 \in \mathbb{R}^F$ that summarizes the information in X. The global mean pooling operation can be written as:

$$h_2^{agg} = \frac{1}{n_i} \sum_{j \in G_i} X_{ij} \tag{5}$$

where G_i signifies the graph to which node j belongs, X_{ij} represents the j-th embedded feature vector pertaining whole graph G_i after undergoing convolution with GCNs, and n_i denotes the total number of all nodes in the graph G_i.

Two strategies: global attention module with memory and partial key message learning

The process of embedding raw node information and constructing a graph is typically succeeded by the application of GCNs. These networks embed the node features of the graph and establish a hierarchical representation. After pooling, the resulting feature vector h_2 encapsulates fundamental characteristics of the graph. To engender a more comprehensive and valuable feature representation of the graph, we employ two distinct strategies with respect to the vector h_2.

The first strategy harnesses an attention mechanism to extract selective information from both the current and historical input graphs. We term this approach the "Global Attention Module with Memory," which can be conveniently abbreviated as the "Global Memory Module." In this approach, a fixed set of attention weights is computed and iteratively updated within the context of a Gated Recurrent Unit (GRU) cell.

These weights serve as a form of global attention that evaluates the corresponding historical graph features. This method facilitates the focused integration of the most informative aspects of each graph into the overall feature representation.

The second strategy involves amalgamating information derived from the attention mechanism with data pertaining to key nodes or core nodes within the present graph. Specifically, we concatenate attention-weighted historical graph features with the current graph features and employ fully connected layers to discern the importance of the features associated with key nodes. This fusion process not only captures the intrinsic attributes of the current graph but also captures the shared patterns and variances that exist across multiple graphs. The outcome is a more holistic and informative feature representation that is conducive to the demands of graph classification tasks.

Global attention module with memory

In our proposed approach, the inclusion of macroscopic information from all graphs, in conjunction with the intrinsic data contained within the graphs (i.e. node features and graph structure), constitutes a vital feature within our methodology. However, the direct simultaneous input of all graphs into the network is both impractical and unfeasible. To circumvent this, we adopt a global attention mechanism, augmented by a memory strategy reminiscent of those employed in time-series networks. After the node features are processed by the GCNs, the graph is assigned a new feature vector h_2, as shown in the equation. To learn and store historical global information in all graphs ever inputted, we use a GRU architecture. The current pooled feature vector h_2 and the previous memory vector $m^{(t-1)} \in \mathbb{R}^F$ are used as inputs to the GRU. Formally, let $r^{(t)}, z^{(t)}, n^{(t)}$ be the reset gate, update gate and output vector of the GRU for the t-th input stage of the memory module. The following equations summarize the proposed global attention mechanism with memory strategy:

$$r^{(t)} = \sigma\left(h_2 W_{rh} + b_{rh} + m^{(t-1)} W_{rh} + b_{rm}\right) \tag{6}$$

$$z^{(t)} = \sigma\left(h_2 W_{zh} + b_{zh} + m^{(t-1)} W_{zm} + b_{zm}\right) \tag{7}$$

$$n^{(t)} = tanh\left(h_2 W_{nh} + b_{nh} + r^{(t)} \odot m^{(t-1)} W_{nm} + b_{nm}\right) \tag{8}$$

$$h_3^{(t)} = \left(1 - z^{(t)}\right) \odot m^{(t-1)} + z^{(t)} \odot n^{(t)} \tag{9}$$

where $h_3^{(t)}$ is the global feature vector obtained by the memory module, $\odot$ denotes Hadamard product, all the W denote learnable weight parameters, and all the b denote learnable bias parameters, the sigmoid function $\sigma(x) = \frac{1}{1+exp(-x)}$ and the hyperbolic tangent function $\tanh(x) = \frac{exp(x)-exp(-x)}{exp(x)+exp(-x)}$ were employed as activation functions. The update formula for the memory vector is represented as:

$$\alpha_{i,j} = \frac{exp\left(LeakyReLU\left(\left[h_3^{(t)} i \| m_j^{(t)}\right] W_1\right)\right)}{\sum_{k=1}^{F} exp\left(LeakyReLU\left(\left[h_3^{(t)} i \| m_k^{(t)}\right] W_1\right)\right)} \tag{10}$$

$$m^{(t+1)} = \sigma\left(m^{(t)} + \alpha^{(t)} h_3^{(t)} W\right) \tag{11}$$

where $\alpha_{\mathrm{i,j}}$ denotes the attention score of the i-th feature of the global memory feature m to the j-th feature of the global feature vector $h_3^{(t)}$, $W_1 \in \mathbb{R}^{2F\times F}$ denotes the learnable weight parameter in the vector $h_3^{(t)}$ update process. The utilization of GRU for modeling the temporal dynamics of global information proves highly effective in encoding interdependencies within the input graphs. This approach adeptly captures the evolving dynamics of the input graphs over time. Specifically, the GRU operates by updating a memory vector, drawing from the current input feature vector as well as the preceding memory vector, and subsequently generates a fresh memory vector for the subsequent input step. This memory vector serves as the repository for global information derived from all input graphs, which is then seamlessly integrated into the ultimate feature representation essential for classification. In scenarios where tasks involve multiple distinct ground-truth label categories, a multi-head GRU mechanism may be employed. To accommodate n mutually exclusive ground-truth label types, a parallel stacked GRU architecture with $n-1$ GRU cells is introduced into the model. This design facilitates the modeling of various label categories, ensuring the model's capacity to effectively handle the diverse aspects of the data.

Partial key message learning

To mitigate the risk of over-smoothing, we incorporate a Partial Key Message Learning module into our model, akin to a residual block. We use two learnable weight matrices to aggregate the overall feature vector h_2 and the node feature vector $N \in \mathbb{R}^F$ to obtain a fused feature vector h_3. Importantly, the weight matrices in this layer are designed to learn attention mechanisms over the dimensions of the feature vectors, in order to emphasize certain dimensions that are deemed more informative for the classification task. Formally, let $h_p^{(l+1)}$ denotes the vector generated by the Partial Key Message Learning module. The process can be presented as:

$$h_p^{(l+1)} = ReLU\left(NW_2^{(l)} + h_2 W_3^{(l)}\right) \tag{12}$$

where $ReLU(\cdot)$ is a piece-wise activation function $ReLU(x) = max(0,x)$, $W_2^{(l)}$ is the partial key message attention weight matrix, and $W_3^{(l)}$ is the graph overall feature vector attention weight matrix. In strategy two, we aggregate the overall feature vector h_2 of the graph and the feature vectors of the key nodes in the graph by means of full connectivity, where the learnable weight matrices are applied to the learning of the attention of each dimension of the features, as a way to improve the attention to the information of the key nodes and obtain features with more usable values.

Message fusion

Finally, we aggregate the feature vector of the global attention mechanism obtained in the Global Attention with Memory module and the feature vector of the partial learning mechanism obtained in the Partial Key Message Learning module to obtain as comprehensive feature tensor for our method. Let C be the number of classes of ground-truth labels, $O^{(l+m)}$ be the output of the entire network:

$$O^{(l+m)} = \mathrm{softmax}\left(FC\left(ReLU\left(\left[h_3^{(l)} \| h_p^{(l)}\right]\right)\right)\right) \tag{13}$$

where $\|$ denotes the concatenation operation, $FC(\cdot)$ denotes multi-layer fully connected network function, m is the number of layers in the multi-layer fully connected layer network. In the end, while a set of training graph pairs was inputted for training, the final layer output tensor $O\epsilon\mathbb{R}^{|D| \times C} O \in \mathbb{R}^{|D| \times C}$ is compared against the ground-truth labels using the following cross-entropy loss function:

$$L = -\frac{1}{|D|} \sum_{i=1}^{D} \sum_{j=1}^{C} y_{ij} log \widehat{y}_{ij} \tag{14}$$

where D is the set of training graph pairs, $|D|$ denotes the total number of graphs contained in D, y_{ij} is the ground-truth label of the i-th graph for class j, and $\widehat{y}_{ij}$ is the predicted probability of the i-th graph belonging to class j. The cross-entropy loss measures the difference between the predicted probability distribution and the ground-truth distribution, and is commonly used as the objective function for training our model.

Experiments

In this section, we present a comprehensive application of our method on a real-world dataset, offering a detailed description of our experimental settings. We meticulously showcase the experimental results obtained from our three proposed models for detecting fake news. Furthermore, we conduct a comparative analysis with state-of-the-art baseline models, aiming to demonstrate the superior performance and effectiveness of our approach in tackling this critical challenge. Ablation studies are also used to validate the effectiveness of the two strategies in our approach. The robustness and reliability of our method are thoroughly examined, providing valuable insights and contributing to the advancement of fake news detection research.

Experimental dataset and experimental setup

The experimental datasets encompass the Weibo dataset, a self-built Chinese corpus dataset that underwent pre-trained coding and embedding, and two widely recognized benchmark English corpus datasets—the Politifact dataset and the Gossipcop dataset. The three datasets have similar structural designs. Edges within the news propagation networks are established based on user behaviors, encompassing actions like retweeting, commenting on news, and user interactions, and such relationships can be crawled from official websites. Within the Chinese dataset, there are networks of false and real news dissemination on the Weibo platform. The news ground-truth labels are meticulously crafted based on fact-checked data supplied by the Sina Community Management Center, an official organization within the Weibo ecosystem. The English corpus datasets[1] (Dou et al., 2021) include fake news and real news propagation networks on Twitter, built from fact-checked information from Politifact and Gossipcop. Notably, these datasets are distinct from one another and contain news items that are not shared between them. The news retweet graph was originally extracted from FakeNewsNet[2] (Shu et al., 2017). The dataset creators crawled nearly 20 million historical tweets from users involved in the propagation of fake news on Twitter to generate the node characteristics for the dataset. The statistical overview of datasets is shown below:

As previously mentioned in the concept of news propagation networks, each graph of the dataset is a hierarchical tree-structured graph that includes a root node representing news and leaf nodes representing social media users who retweeted the root news, and the edges of the graph represent the act of tweeting and retweeting. Each graph in the dataset represents a news event where there is no direct relationship between events. The raw news data includes news titles and news context, and the raw user data includes user profiles and user tweet context. In English datasets, the raw data were encoded using pre-trained word embeddings, BERT[3] and spaCy[4] word2vec (Mikolov et al., 2013). The 10-dimensional user profile feature was encoded using word2vec. The 310-dimensional content feature was composed of a 300-dimensional user comment word2vec embedding concatenated with the 10-dimensional profile feature. The 768-dimensional content feature was generated using the pre-trained BERT model to encode the user profiles and user comments. Unlike the English corpus datasets, the Weibo dataset uses the Chinese corpus pre-trained word embeddings, Chinese corpus BERT and Chinese corpus word2vec,[5] and the user profile feature in the Weibo dataset is 14-dimensional. Similarly, the 314-dimensional content feature was composed of a 300-dimensional user comment word2vec embedding concatenated with the 14-dimensional profile feature.

[1] https://github.com/safe-graph/GNN-FakeNews.
[2] https://github.com/KaiDMML/FakeNewsNet.
[3] https://github.com/jina-ai/clip-as-service.
[4] https://spacy.io/models/en#en_core_web_lg.
[5] https://spacy.io/models/zh#zh_core_web_lg.

Table 1
General statistics of dataset.

Dataset		Politifact	Gossipcop	Weibo
General Statistics	#Total Graphs	314	5464	4664
	#Fake News	157	2732	2313
	#Total Nodes	41,054	314,262	2,856,741
	#Total Edges	40,740	308,798	2,852,077
	#Avg. Nodes per Graph	131	58	613
	#Avg. Edges per Graph	130	57	612
Encoding Feature Dimension	Profile	10	10	14
	word2vec	310	310	314
	BERT	768	768	768

We have implemented all models using the PyTorch[6] deep learning framework in Python 3.10 and the PyTorch-Geometric (PyG)[7] package for implementing GNNs models. A unified graph embedding size of 128, a batch size of 128, and L2 regularization weight of 0.001 were used for all models. According to the different convolution methods of graph convolutional networks, we have used different learning rates. Specifically, we set the learning rate to 0.001 for GAT and GCN, while the learning rate for the GANM under GraphSAGE convolution is set to 0.005. To prevent overfitting, we applied early stopping with patience of 10 epochs during the training process. Further hyperparameters for each model can be found in the code repository, which is publicly available on request for replication and extension of our work.

Experimental performance evaluation

According to our proposed approach, three models were generated by combining three classical graph neural convolutional network methods, named GAT-GANM, SAGE-GANM, and GCN-GANM. We evaluated the performance of our models on the dataset using two commonly used evaluation metrics: test accuracy score and F1 score. The test Accuracy (Acc.) score measures the proportion of correctly classified instances, while the F1 score considers both precision and recall. Specifically, we calculated the F1 score as the harmonic mean of precision and recall. Furthermore, we employed statistical tests to ascertain the significance of our model's outcomes. Our experimental results demonstrate the effectiveness of our proposed approach in detecting fake news, achieving prominent accuracy and F1 scores, as shown in Table 2.

The results shown in Table 2 indicate that the GCN-GANM model exhibits the most prominent overall performance across all datasets. Conversely, the SAGE-GANM model performs exceptionally well on the Weibo dataset, particularly when utilizing word2vec and BERT encoding methods. In contrast, the GAT-GANM model showcases relatively lower performance across all three datasets and encoding methods. Comparative analysis of the employed methods underscores the superiority of graph convolution operations utilizing GCN and SAGE over GAT across all experimental datasets. The observed outcomes can be explained by comparing the performance of the GCN, SAGE, and GAT models in fake news detection.

In terms of model architecture, the performance of the GAT model is relatively lower, potentially due to its emphasis on local inter-node dependencies. However, in this task, global inter-node dependencies among nodes might be crucial. GAT's sensitivity to the distances between nodes could limit its capability to capture long-distance dependencies, thereby affecting its performance. The outstanding performance of the SAGE-GANM model on the Weibo dataset might be attributed to its inductive learning ability, which allows it to capture

[6] https://pytorch.org/.
[7] https://pyg.org/.

Table 2
Performance of experimental results of our approaches for fake news detection.

Model	Encoding Method	Politifact		Gossipcop		Weibo	
		Acc.	F1	Acc.	F1	Acc.	F1
GAT-GANM	Profile	0.7849	0.7959	0.8509	0.8468	0.8713	0.8719
	word2vec	0.8387	0.8544	0.9474	0.9483	0.9617	0.9615
	BERT	0.8495	0.8511	0.9561	0.9550	0.9602	0.9604
SAGE-GANM	Profile	0.7527	0.7677	0.8772	0.8852	0.9225	0.9222
	word2vec	0.8602	0.8538	0.9737**	0.9739**	0.9779	0.9779
	BERT	0.8622***	0.8687**	0.9737	0.9735	**0.9835****	**0.9834*****
GCN-GANM	Profile	0.7849	0.8039	0.8684	0.8649	0.9216	0.9218
	word2vec	0.8602	0.8660	0.9649	0.9655	0.9715	0.9716
	BERT	**0.8602***	**0.8738****	**0.9825****	**0.9825****	0.9804**	0.9805**

The best results are highlighted in bold, and the second-best result is highlighted in underline. * denotes statistically significant under the *t*-test (* $p \leq 0.05$, * * $p \leq 0.01$, * * * $p \leq 0.001$).

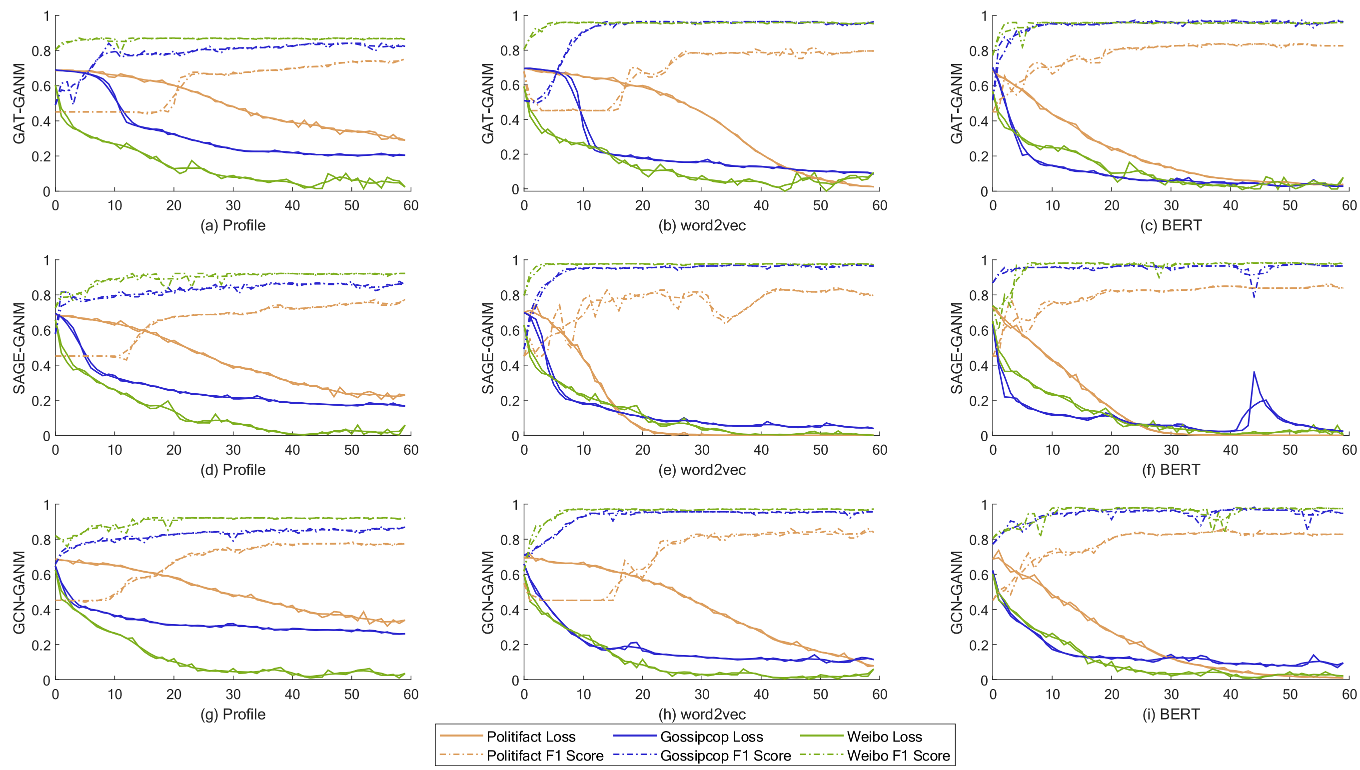

Fig. 3. Performance curves for nine unique training processes with smoothing trends. The colors represent different datasets, and the solid and dashed lines represent the loss and F1 scores of the training processes. The curves for each training process were smoothed, and the curves were transparent before smoothing (transparency $\alpha = 0.5$). The x-axis label for each subplot is the encoding method, and the y-axis label is the model name.

global information from local neighbors. Furthermore, SAGE's scalability advantage of not requiring the maintenance of the entire graph for operation proves beneficial, particularly for larger networks. The Weibo dataset features significantly larger and more intricate graphs compared to the Politifact and Gossipcop datasets, as illustrated in Table 1. The exceptional performance of the GCN model on the English corpus datasets could stem from its capacity to capture structural information within the graph. Proficiency in modeling propagation patterns and user relationships could enhance GCN's modeling capability. Moreover, GCN's flexibility in integrating propagation patterns and user information contribute to better distinguishing between true and false information. In summary, the varied nature of datasets and task requirements leads to the diverse performance of graph neural network models. GCN and SAGE excel on different datasets, while GAT may struggle to effectively capture global dependencies. Therefore, when selecting a model, careful consideration of the specific task context and performance metrics is essential. To visualize the training process of our three proposed models, we present a comprehensive figure containing nine subplots. Each subplot corresponds to a specific training epoch represented along the x-axis, showcasing the corresponding loss and F1 score, as shown in Fig. 3.

Fig. 3 provides a comprehensive depiction of the training process, highlighting the overall stability and efficiency of each model. In terms of convergence speed and F1 score for fake news detection, BERT encoding outperforms the other two encoding methods. Moreover, in terms of feature embedding effectiveness, training stability, and efficiency, the GCN-GANM model surpasses both the GAT-GANM and SAGE-GANM models. Notably, the GCN-GANM model exhibits superior performance, both in terms of training effectiveness and performance, on the Gossipcop dataset compared to the Politifact dataset. This discrepancy can be attributed to inherent differences in the dataset characteristics, resulting in variations in the training processes of the same model. To further showcase the effectiveness of our approach to feature extraction, we employed the GCN-GANM model with weight

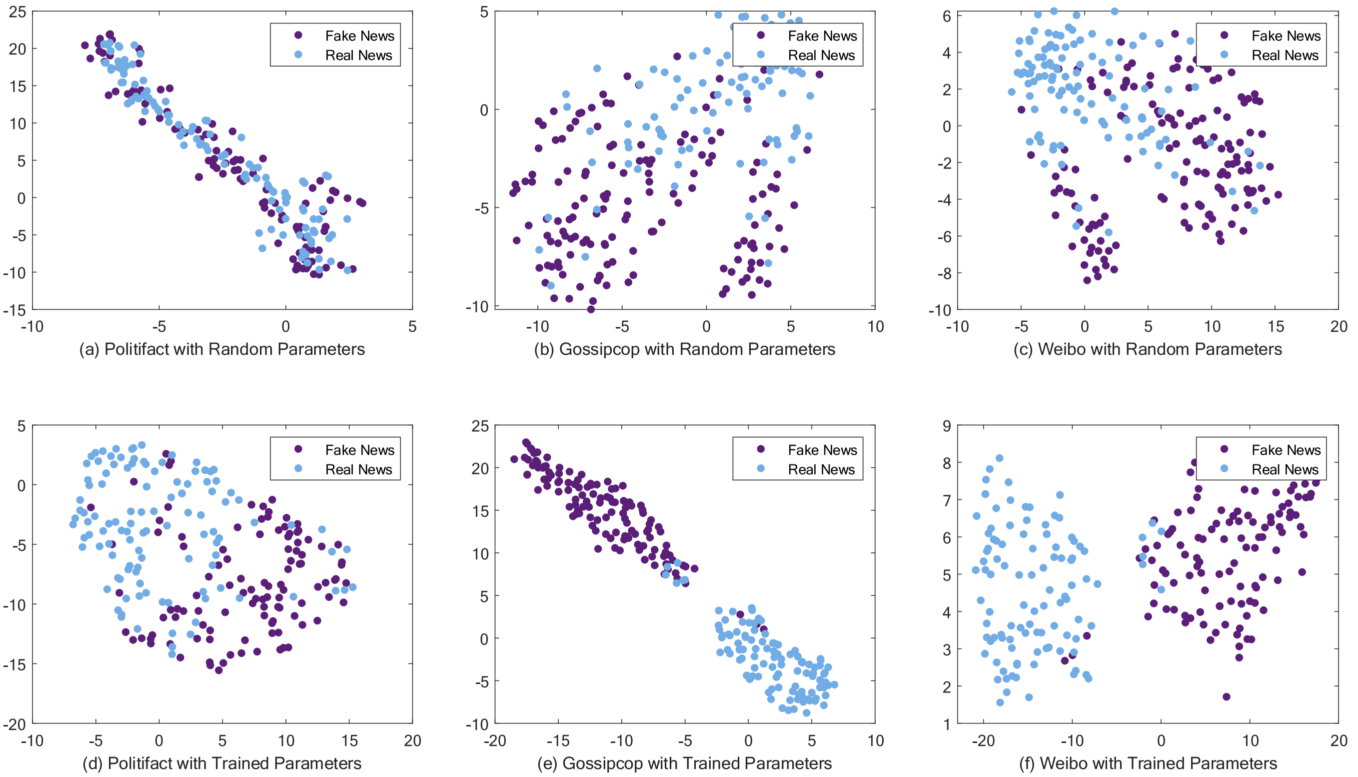

Fig. 4. t-SNE visualization of the datasets with random parameters and trained parameters by using GCN-GANM with BERT encoding. Subfigures (a), (b) and (c) show the t-SNE visualizations of some samples in the three datasets before model training. Subfigures (d), (e) and (f) show the t-SNE visualizations of some samples in the three datasets after model training.

parameters initialized randomly without undergoing any gradient descent optimization. Subsequently, the distribution of graphs was visualized using t-SNE visualization within one test batch, as presented in Fig. 4.

In Fig. 4, subplots (a), (b), and (c) illustrate the classification efficacy of real and fake news detection under random parameters. In this scenario, the original features are generated via pre-trained word embeddings. The t-SNE visualizations substantiate that discernible distinctions exist in the feature information integrated across various datasets. Particularly noteworthy is the observation that the Politifact dataset poses a relatively greater challenge in classification performance compared to both the Gossipcop and Weibo datasets, even when utilizing an exceptional pre-trained word embedding such as BERT. This challenge arises due to the inherent difficulty in extracting sufficiently informative data from the relatively constrained sample size of the Politifact dataset. These findings empirically underscore the inherent diversity inherent in the three datasets.

In contradistinction, subplots (d), (e), and (f) exhibit marked enhancements in the classification performance across all three datasets when contrasted with the outcomes under random parameters during the model parameter training phase. Notably, the Weibo dataset achieves impressive classification performance with outstanding classification quality for both true and false news after model parameter training, as shown in subplot (f). This pronounced effect can be attributed to the fact that the Weibo dataset has a higher average number of relevant users per news item and a higher correlation between the users and the news items compared to the other datasets. This augmented user-news interaction allows the model to comprehensively assimilate exogenous information during the training process.

Comparison with baseline methods

The comparison with baseline methods encompasses two primary categories: traditional machine learning algorithms for clustering and deep learning algorithms based on neural networks. The first type of baseline methods includes four traditional machine learning approaches for classification tasks: Naive Bayes (NB), Random Forest (RF), Support Vector Machine (SVM), and Passive Aggressive (PA). We employ the scikit-learn[8] library in Python to implement these approaches. For the textual raw data encoding of news, we use the NLTK[9] library and the tokenizer module of the sikit-learn library for the encoding process.

The second type of baseline method includes neural network-based deep learning approaches:

- word2vec-MLP: The word2vec-MLP method employs word2vec encoding and a Multi-Layer Perceptron (MLP) model for feature embedding to construct a classification task framework for fake news detection. The MLP is a fundamental deep learning method that has shown excellent performance in various fields. However, due to the structural limitations of the model, MLP cannot aggregate social network structure information.
- Text-CNN: The Text-CNN (Kim, 2014) uses CNN for text classification, which applies a novel CNN architecture called dynamic k-max pooling and has demonstrated outstanding performance in text classification tasks.
- HPFN[10]: The HPFN (Shu et al., 2020b) involves two stages of propagation: global and local propagation, which are performed hierarchically to capture the global and local semantic features of the news articles. The model also employs a gating mechanism to selectively propagate relevant information during the global and local propagation stages.

[8] https://scikit-learn.org/.
[9] https://www.nltk.org/.
[10] https://github.com/mdepak/fake-news-propagation.

Table 3
Comparison with baseline methods of experimental results.

Model	Feature source	Politifact		Gossipcop		Weibo	
		Acc.	F1	Acc.	F1	Acc.	F1
NB	News Only	0.7240	0.6980	0.8314	0.8227	0.8447	0.8456
RF	News only	0.7460	0.7427	0.8176	0.8333	0.8545	0.8550
SVM	News only	0.7495	0.7815	0.8379	0.8621	0.8474	0.8467
PA	News only	0.7682	0.7636	0.8571	0.8576	0.8778	0.8770
BERT-MLP	News only	0.7647	0.7636	0.8469	0.8711	0.9210	0.9212
text-CNN	News only	0.7759	0.8038	0.8530	0.8738	0.9087	0.9093
HPFN	Social context	0.8430	0.8430	0.8690	0.8710	0.9170	0.9172
dEFEND	News+user+network	0.8080	0.7532	0.9041	0.9283	0.9558	0.9560
GCNFN	News+user+graph	0.8316	0.8356	0.9638	0.9509	0.9528	0.9530
UPFD	News+user+graph	0.8462	0.8465	0.9723	0.9722	0.9755	0.9754
Bi-GCN	News+user+graph	0.8190	0.8113	0.9589	0.9590	0.9614	0.9613
EBGCN	News+user+graph	0.8281	0.8304	0.9412	0.9406	0.9638	0.9633
GAT-GANM (ours)	News+user+graph	0.8387	0.8544	0.9561	0.9550	0.9617	0.9615
SAGE-GANM (ours)	News+user+graph	<u>0.8622</u>	<u>0.8687</u>	<u>0.9737</u>	<u>0.9739</u>	**0.9835**	**0.9834**
GCN-GANM (ours)	News+user+graph	**0.8602**	**0.8738**	**0.9825**	**0.9825**	<u>0.9804</u>	<u>0.9805</u>

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

- dEFEND[11]: The dEFEND (Shu et al., 2019b) approach employs a BERT encoder and attention mechanism to extract features from article and headline text, which are inputted to an interpretable classifier to predict article authenticity. The model not only outputs predictions but also provides explanations for its decisions to help users better understand the decision-making process.
- GCNFN[12]: The GCNFN (Monti et al., 2019) method is a novel geometric deep learning-based approach for fake news detection. It converts the fake news detection problem on social media to a graph classification problem and extracts user profiles on social media using GCN, which establishes a graph representation with multiple information sources.
- UPFD[13]: The UPFD (Dou et al., 2021) approach utilizes NLP techniques to encode the news context and user context and obtain feature tensors for both news and users. GCNs are applied to integrate both endogenous and exogenous user information and learn joint user engagement embedding.
- Bi-GCN[14]: The Bi-GCN (Bian et al., 2020) uses a parallel bi-directional GCN structure to bi-directionally learn about the propagation and dispersion of information in the community, achieving excellent performance that has been recognized in many studies and used as a baseline method comparison.
- EBGCN: EBGCN (Wei et al., 2021) adopts the bidirectional graph convolutional structure introduced by Bi-GCN as its foundational framework. However, it further innovates by integrating Bayesian probabilistic content to infer edge weights between nodes.

Our approach achieves promising results on the three datasets. The comparison with baseline methods of experimental results is shown in Table 3.

Table 3 presents a comprehensive performance evaluation of our proposed method in comparison to various comparative methods across the Politifact, Gossipcop, and Weibo datasets. Our method clearly surpasses the performance of other baseline methods, establishing itself as the leading approach in fake news detection. Notably, GANM, when integrated with the GCN framework for convolution operation, emerges as the top-performing configuration, demonstrating exceptional overall efficacy.

First, the baseline algorithm underscores the superiority of deep learning over traditional machine learning approaches that rely on manually crafted features. Deep learning methods exhibit significantly better performance, underscoring their capacity to acquire high-level representations of news, enabling the capture of effective features. This reiterates the critical importance of deep learning in the realm of fake news detection.

Second, the GNNs-based approaches, tailored to leverage graph-structured data, outperform the more generalized approach. This outcome emphasizes the inherent advantages of employing graph-structured data in extracting and comprehending the propagation patterns of false news. It accentuates the significance of incorporating network structure considerations when addressing the task of fake news detection.

Finally, the proposed method consistently outperforms other deep learning techniques across all performance metrics. This observation underscores the effectiveness of incorporating the proposed Global Attention module with Memory and Partial Key Message Learning module into the fake news detection process. Previous state-of-the-art methods often neglect the global information that permeates the entire graph, thereby overlooking a vital structural feature of fake news dispersion. The integration of these modules empowers our approach to capture a comprehensive high-level representation of fake news, resulting in significantly improved fake news detection performance.

Ablation study

In the ablation study, a series of experiments were conducted to comprehensively analyze the influence of various components or modules within our proposed model. The primary objective was to systematically eliminate or adjust specific segments of the model architecture to gauge their individual contributions to the overall performance. Through a comparative analysis of the outcomes obtained from these altered model versions against the performance of the complete GANM model, valuable insights regarding the efficacy and significance of each component were obtained. The ablation experiments encompassed the following variations:

- Variant 1: We established a variant of our GANM model by excluding two key components: The Global Attention Module with Memory and the Partial Key Message Learning component. This stripped-down variant consists solely of the graph convolution operation and the fully connected layers, serving as a control group for our study. By adopting this approach, we aim to dissect and evaluate the individual impacts of the aforementioned strategies that we have incorporated.
- Variant 2: The Variant 2 model introduces a modification by excluding the Partial key message Learning component within the

[11] https://github.com/heeyjunaid/dEFEND-Pytorch.
[12] https://github.com/YingtongDou/GCNN.
[13] https://github.com/safe-graph/GNN-FakeNews.
[14] https://github.com/TianBian95/BiGCN.

Table 4
Comparison of results of ablation experiments.

Encoding	Variant	Architecture module			Politifact		Gossipcop		Weibo	
		GCNs Module	Global Memory Module	Partial Message Learning Module	Acc.	F1	Acc.	F1	Acc.	F1
Profile	Variant1	✔	×	×	0.7419	0.7736	0.8158	0.8235	0.8866	0.8791
	Variant2	✔	✔	×	0.7957	0.8119	0.8509	0.8350	0.9175	0.9167
	Variant3	✔	×	✔	0.7634	0.7843	0.8421	0.8269	0.8969	0.8936
	GANM	✔	✔	✔	0.7849	0.8039	0.8772	0.8852	0.9225	0.9222
word2vec	Variant1	✔	×	✔	0.7634	0.7708	0.9035	0.9027	0.9175	0.9130
	Variant2	✔	✔	×	0.8280	0.8400	0.9474	0.9474	0.9545	0.9545
	Variant3	×	✔	×	0.7849	0.7778	0.9298	0.9298	0.9320	0.9322
	GANM	✔	✔	✔	0.8602	0.8660	0.9737	0.9739	0.9779	0.9779
BERT	Variant1	✔	×	✔	0.8280	0.8367	0.9123	0.9123	0.9278	0.9231
	Variant2	✔	✔	×	0.8495	0.8542	0.9474	0.9474	0.9620	0.9621
	Variant3	×	✔	×	0.8387	0.8421	0.9211	0.9204	0.9393	0.9394
	GANM	✔	✔	✔	0.8602	0.8738	0.9825	0.9825	0.9835	0.9834

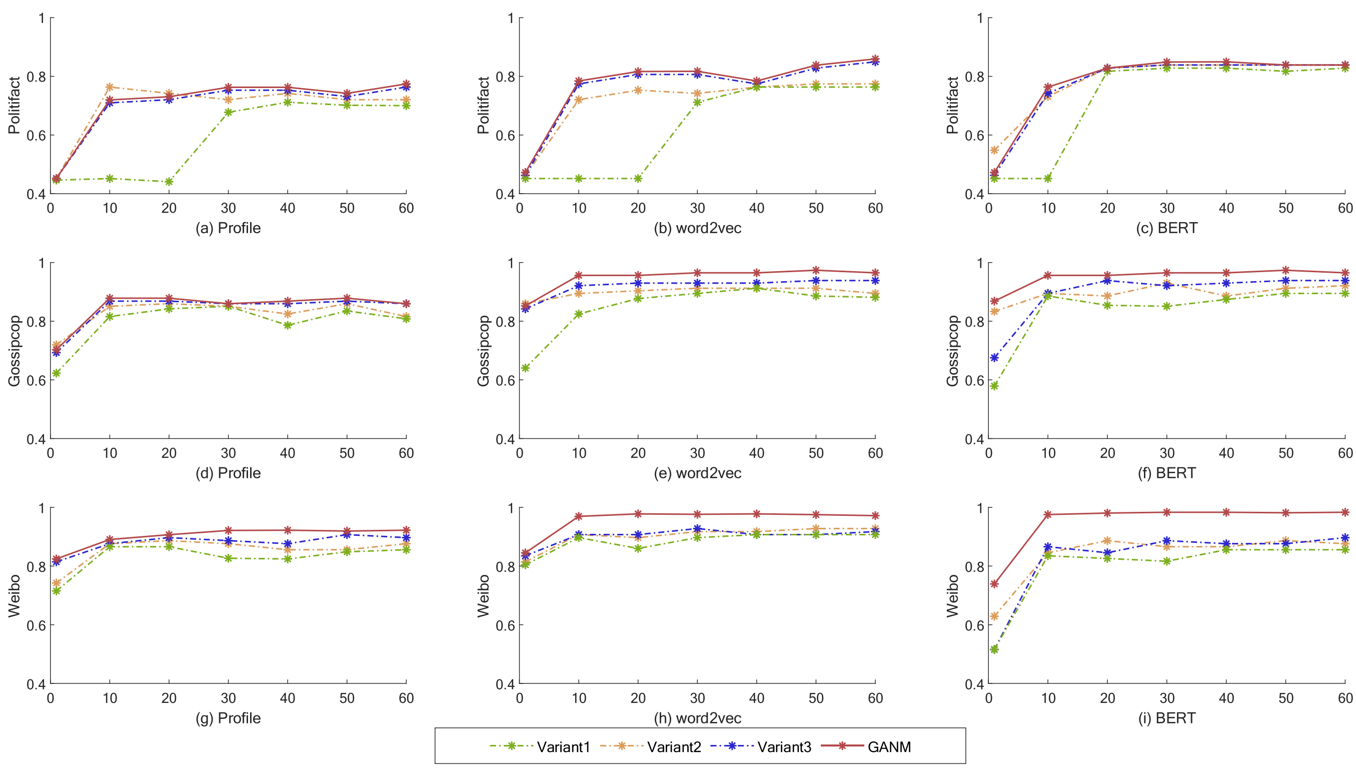

Fig. 5. Performance of the ablation experiment training process. The dashed curve corresponds to the variant model, while the solid curve represents the complete GANM model. The y-axis tick value denotes the F1 score, while the x-axis represents the epoch. The x-label specifies the coding method used, and the y-axis label indicates the dataset employed in the experiment.

GANM model. In this configuration, the embedding proceeds through the Global Memory Module before progressing to the subsequent stage of the fully connected layers. This sequence of operations takes place following the execution of the graph convolution operation and global pooling.

- Variant 3: In this specific variant model, we take an alternative approach by removing the Global Attention Module with Memory from the architectural design. In this scenario, the hidden state tensor output generated by the GCNs module directly interfaces with the Partial Key Message Learning component. This occurs subsequent to the processes of global pooling.

All the outcomes derived from the ablation experiments exhibit statistical significance, with a significance level of $p \leq 0.05$, as determined by the *t*-test. Table 4 provides a comprehensive overview of the outcomes obtained from the ablation experiments, presenting a detailed account of how the variants perform across diverse encoding methods (Profile, word2vec, and BERT). Notably, Variant 2 emerges as the frontrunner among the three variants in terms of overall performance. This outcome substantiates the pivotal role played by the Global Memory Module (referred to as the Global Attention Module with Memory) within the GANM architecture. The Global Memory Module is evidently instrumental in facilitating the grasp of structural information and enhancing the learning capabilities of the model. Significantly, the inclusion of the Global Memory Module ensures that the model avoids significant overfitting issues at later training stages. This prevents the occurrence of oscillating or declining accuracy and F1 score curves shown in Fig. 5, a phenomenon observed in comparison to variant 1 and variant 3, where this module is absent. The superior performance of variant 2 underscores the module's effectiveness in maintaining model stability throughout the learning process. Turning to Variant 3, which incorporates the Partial Key Message Learning module, notable trends emerge. This module exhibits commendable performance in the initial and early stages of the learning process. Its robust key message learning capacity enables rapid parameter updates, resulting in relatively prompt convergence to a favorable solution domain. However, this advantage begins to wane as the training progresses. The model encounters limitations in pushing performance further beyond a certain threshold, eventually leading to some degree of overfitting. Consequently, the test set accuracy of the model is compromised, showcasing its inability to maintain high accuracy levels.

The outcomes stemming from the ablation experiments unequivocally affirm the pivotal significance of the Global Memory Module within the GANM framework. This component distinctly amplifies the model's grasp of structural intricacies and its cognitive acuity in acquiring knowledge. Concurrently, the experiments shed light on the nuanced influence of the Partial Key Message Learning module, elucidating its dual impact on initial performance and susceptibility to overfitting tendencies. It becomes evident that both of these components constitute indispensable pillars underpinning the efficacy of GANM. Their symbiotic synergy manifests as a formidable amalgamation, endowing the model with the prowess to capture intricate graph dynamics and bolster its performance capabilities. In this intricate tapestry of interrelated modules, both the global memory module and the partial key information learning module emerge as indispensable assets, coalescing to empower GANM with multifaceted capabilities.

Conclusion

The proposed approach is designed to detect fake news in complex graph-structured data through the application of various techniques, such as deep learning, graph neural networks, and temporal modeling. The approach is capable of efficiently processing graphs with multiple modalities, attributes, and features while also identifying key nodes and subgroups that can improve classification accuracy. Therefore, the approach demonstrates remarkable robustness and superior performance on real-world datasets by utilizing these techniques. In addition, the approach has broad applications beyond fake news detection, including classification and prediction tasks for multiple graphs with critical root nodes or subgroups as well as graph sequences with time-varying properties. By combining powerful computational methods with detailed graph analysis, this approach offers an effective tool for addressing complex data analysis challenges. In the future, a promising direction is the ongoing refinement of the global memory module, potentially integrating adaptive mechanisms to accommodate a diverse array of graph structures. Furthermore, the pursuit of improving the method's scalability to effectively manage larger and more dynamic datasets remains an enticing endeavor.

CRediT authorship contribution statement

Qian Chang: Conceptualization, 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

Aggarwal, A., Chauhan, A., Kumar, D., Mittal, M., & Verma, S. (2020). Classification of fake news by fine-tuning deep bidirectional transformers based language model. *Eai Endorsed Transactions on Scalable Information Systems, 7*(27). https://doi.org/10.4108/eai.13-7-2018.163973. Article 10.

Altınel, B., & Ganiz, M. C. (2018). Semantic text classification: A survey of past and recent advances. *Information Processing & Management, 54*(6), 1129–1153.

Aphiwongsophon, S., & Chongstitvatana, P. (2018). Detecting fake news with machine learning method. In *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* (pp. 528–531).

Baarir, N. F., & Djeffal, A. (2021). Fake news detection using machine learning. In *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being* (pp. 125–130).

Bahad, P., Saxena, P., & Kamal, R. (2019). Fake news detection using bi-directional LSTM-recurrent neural network. *Procedia Computer Science, 165*, 74–82.

Bastick, Z. (2021). Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. *Computers in Human Behavior, 116*, Article 106633.

Basu, A., Kiliroor, C. C., Basu, R., & Nag, R. (2022). Fake news predictor: A random forest-based web application for the prediction of fake news on social media. In *Electronic Systems and Intelligent Computing: Proceedings of ESIC 2021. Lecture Notes in Electrical Engineering* (pp. 163–173).

Bian, T., Xiao, X., Xu, T., Zhao, P., Huang, W., Rong, Y., et al. (2020). Rumor detection on social media with bi-directional graph convolutional networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 549–556).

Chen, Y. (2015). Convolutional neural network for sentence classification. *arXiv preprint arXiv:1408.5882.*.

Choudhary, A., & Arora, A. (2021). Linguistic feature based learning model for fake news detection and classification. *Expert Systems with Applications, 169*. https://doi.org/10.1016/j.eswa.2020.114171. Article 114171.

Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dhawan, M., Sharma, S., Kadam, A., Sharma, R., & Kumaraguru, P. (2022). Game-on: Graph attention network based multimodal fusion for fake news detection. *arXiv preprint arXiv:2202.12478*.

Dou, Y., Shu, K., Xia, C., Yu, P. S., & Sun, L. (2021). User preference-aware fake news detection. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 2051–2055).

Feng, Y. (2022). Misreporting and fake news detection techniques on the social media platform. *Highlights in Science, Engineering and Technology, 12*, 142–152.

Figueira, Â.I., & Oliveira, L. (2017). The current state of fake news: Challenges and opportunities. *Procedia Computer Science, 121*, 817–825.

Goldani, M. H., Safabakhsh, R., & Momtazi, S. (2021). Convolutional neural network with margin loss for fake news detection. *Information Processing & Management, 58* (1), Article 102418.

Gorrell, G., Bontcheva, K., Derczynski, L., Kochkina, E., Liakata, M., & Zubiaga, A. (2018). Rumoureval 2019: Determining rumour veracity and support for rumours. *arXiv preprint arXiv:1809.06683*.

Gupta, A., Li, H., Farnoush, A., & Jiang, W. (2022). Understanding patterns of COVID infodemic: A systematic and pragmatic approach to curb fake news. *Journal of Business Research, 140*, 670–683.

Gupta, M., Zhao, P., & Han, J. (2012). Evaluating event credibility on twitter. In *Proceedings of the 2012 SIAM International Conference on Data Mining* (pp. 153–164).

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems, 30*.

Han, Y., Karunasekera, S., & Leckie, C. (2020). Graph neural networks with continual learning for fake news detection from social media. *arXiv preprint arXiv:2007.03316*.

Hussain, M. G., Hasan, M. R., Rahman, M., Protim, J., & Al Hasan, S. (2020). Detection of bangla fake news using MNB and SVM classifier. In *2020 International Conference on Computing, Electronics & Communications Engineering (ICCECE* (pp. 81–85).

Institute, P. s. N. R. (2022). *China mobile internet development report*. http://finance.people.com.cn/n1/2022/0629/c1004-32460664.html.

Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet-a deep convolutional neural network for fake news detection. *Cognitive Systems Research, 61*, 32–44.

Kaur, S., Kumar, P., & Kumaraguru, P. (2020). Automating fake news detection system using multi-level voting model. *Soft Computing, 24*(12), 9049–9069.

Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). Mvae: Multimodal variational autoencoder for fake news detection. In *The World Wide Web Conference* (pp. 2915–2921).

Kim, Y. (2014). Convolutional neural networks for sentence classification. In *2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1746–1751).

Kipf, T.N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Kumar, S., Asthana, R., Upadhyay, S., Upreti, N., & Akbar, M. (2020). Fake news detection using deep learning models: A novel approach. *Transactions on Emerging Telecommunications Technologies, 31*(2). https://doi.org/10.1002/ett.3767. Article e3767.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., et al. (2018). The science of fake news. *Science (New York, N.Y.), 359*(6380), 1094–1096.

Liu, Y., & Wu, Y. F. (2018). Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In *, 32. Proceedings of the AAAI Conference on Artificial Intelligence*.

Lu, Y.J., & Li, C.T. (2020). GCAN: Graph-aware co-attention networks for explainable fake news detection on social media. *arXiv preprint arXiv:2004.11648*.

Malhotra, B., & Vishwakarma, D. K. (2020). Classification of propagation path and tweets for rumor detection using graphical convolutional networks and transformer based encodings. In *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)* (pp. 183–190).

Martens, B., Aguiar, L., Gomez-Herrera, E., & Mueller-Langer, F. (2018). The digital transformation of news media and the rise of disinformation and fake news.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M.M. (2019). Fake news detection on social media using geometric deep learning. *arXiv preprint arXiv:1902.06673*.

Mridha, M. F., Keya, A. J., Hamid, M. A., Monowar, M. M., & Rahman, M. S. (2021). A Comprehensive review on fake news detection with deep learning. *IEEE Access : Practical Innovations, Open Solutions, 9*, 156151–156170. https://doi.org/10.1109/access.2021.3129329

Nakamura, K., Levy, S., & Wang, W.Y. (2019). r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection. *arXiv preprint arXiv:1911.03854*.

Oshikawa, R., Qian, J., & Wang, W.Y. (2018). A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*.

Palani, B., Elango, S., & Viswanathan, K. V. (2022). CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT. *Multimedia Tools and Applications, 81*(4), 5587–5620.

Pierri, F., & Ceri, S. (2019). False news on social media: A data-driven survey. *ACM Sigmod Record, 48*(2), 18–27.

Rangel, F., Giachanou, A., Ghanem, B. H. H., & Rosso, P. (2020). Overview of the 8th author profiling task at pan 2020: Profiling fake news spreaders on twitter. In *CEUR Workshop Proceedings* (pp. 1–18).

Shu, K., Bernard, H. R., & Liu, H. (2019a). Studying fake news via network analysis: Detection and mitigation. *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, 43–65.

Shu, K., Cui, L., Wang, S., Lee, D., & Liu, H. (2019b). dEFEND: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD International Conference On Knowledge Discovery & Data Mining* (pp. 395–405).

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020a). Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data, 8*(3), 171–188.

Shu, K., Mahudeswaran, D., Wang, S., & Liu, H. (2020b). Hierarchical propagation networks for fake news detection: Investigation and exploitation. In *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 626–637).

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter, 19*(1), 22–36.

Singh, V. K., Ghosh, I., & Sonagara, D. (2021). Detecting fake news stories via multimodal analysis. *Journal of the Association for Information Science and Technology, 72*(1), 3–17.

Singhal, S., Shah, R. R., Chakraborty, T., Kumaraguru, P., & Satoh, S.i. (2019). Spotfake: A multi-modal framework for fake news detection. In *2019 IEEE fifth International Conference on Multimedia Big Data (BigMM)* (pp. 39–47).

Sun, T., Qian, Z., Dong, S., Li, P., & Zhu, Q. (2022). Rumor detection on social media with graph adversarial contrastive learning. In *WWW '22: Proceedings of the ACM Web Conference 2022* (pp. 2789–2797).

Tian, L., Zhang, X. J., & Lau, J. H. (2022). Duck: rumour detection on social media by modelling user and comment propagation networks. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4939–4949).

Van-Hoang, N., Sugiyama, K., Nakov, P., & Min-Yen, K. (2020). FANG: leveraging social context for fake news detection using graph representation. In *CIKM '20: Proceedings of the 29th International Conference on Information & Knowledge Management* (pp. 1165–1174).

Varlamis, I., Michail, D., Glykou, F., & Tsantilas, P. (2022). A survey on the use of graph convolutional networks for combating fake news. *Future Internet, 14*(3), 70.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *Stat, 1050*(20), 10.48550.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science (New York, N.Y.), 359*(6380), 1146–1151.

Vziatysheva, V. (2020). How fake news spreads online? *International Journal of Media & Information Literacy, 5*(2).

Wang, Y., Qian, S., Hu, J., Fang, Q., & Xu, C. (2020). Fake news detection via knowledge-driven multimodal graph convolutional networks. In *Proceedings of the 2020 International Conference on Multimedia Retrieval* (pp. 540–547).

Wani, A., Joshi, I., Khandve, S., Wagh, V., & Joshi, R. (2021). Evaluating deep learning approaches for covid19 fake news detection. In *Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1* (pp. 153–163).

Wei, L., Hu, D., Zhou, W., Yue, Z., & Hu, S. (2021). Towards propagation uncertainty: edge-enhanced Bayesian graph convolutional networks for rumor detection. *arXiv preprint arXiv:2107.11934*. arXiv:2107.11934.

Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P.S. (2018). TI-CNN: Convolutional neural networks for fake news detection. *arXiv preprint arXiv:1806.00749*.

Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management, 57*(2), Article 102025.

Zhou, X., & Zafarani, R. (2018). Fake news: A survey of research, detection methods, and opportunities. *arXiv preprint arXiv:1812.00315*, *2*.

Zhou, X., Zafarani, R., Shu, K., & Liu, H. (2019a). Fake news: Fundamental theories, detection strategies and challenges. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining* (pp. 836–837).

Zhou, Z., Guan, H., Bhat, M.M., & Hsu, J. (2019). Fake news detection via NLP is vulnerable to adversarial attacks. *arXiv preprint arXiv:1901.09657*.