

Article

Integrating Message Content and Propagation Path for Enhanced False Information Detection Using Bidirectional Graph Convolutional Neural Networks

Jie Hu ^{1,*} , Mei Yang ¹, Bingbing Tang ¹ and Jianjun Hu ²

¹ College of Big Data Statistics, Guizhou University of Finance and Economics, Guiyang 550025, China; ymei_9600@163.com (M.Y.); tbb15838035836@163.com (B.T.)

² Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208, USA; jianjunh@cec.sc.edu

* Correspondence: jiehu@mail.gufe.edu.cn

Abstract: We investigate the impact of textual content and its structural characteristics on the detection of false information. We propose a Bidirectional Graph Convolutional Neural Network (ICP-BGCN) that integrates message content with its propagation paths for enhanced detection performance. Our approach leverages web propagation topology by transforming disconnected user posts into a bidirectional propagation graph, which integrates top-down and bottom-up pathways derived from post forwarding and commenting relationships. Using BERT embeddings, we extract contextual semantic features from both source texts and their propagated counterparts, which are embedded as node attributes within the propagation graph. The bidirectional graph convolutional neural network subsequently learns the feature representations of the event propagation network during information dissemination, merging these representations with the original text content features to achieve comprehensive disinformation detection. Experimental results demonstrate significant improvements over existing methods. On benchmark datasets Twitter15 and Twitter16, our model achieves accuracy rates of 89.7% and 91.7%, respectively, outperforming state-of-the-art baselines by 1.1% and 3.7%. The proposed ICP-BGCN exhibits strong cross-domain generalization, attaining 84.4% accuracy on the Pheme dataset and achieving improvements of 1.8% in accuracy and 3.8% in Macro-F1 score on SemEval-2017 Task 8.

Keywords: false information; bidirectional graph convolutional networks; propagation structure; feature fusion



Academic Editors: Gianluca Lax and Antonia Russo

Received: 24 January 2025

Revised: 19 March 2025

Accepted: 20 March 2025

Published: 21 March 2025

Citation: Hu, J.; Yang, M.; Tang, B.; Hu, J. Integrating Message Content and Propagation Path for Enhanced False Information Detection Using Bidirectional Graph Convolutional Neural Networks. *Appl. Sci.* **2025**, *15*, 3457. <https://doi.org/10.3390/app15073457>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Currently, the global number of Internet users is 5.16 billion, with 4.76 billion social media users, constituting nearly 60% of the global population [1]. With the accelerated development of social networking platforms, such as Twitter and Facebook, the process of information sharing and dissemination has significantly accelerated [2]. While social media provides a platform for free and diverse forms of expression, it inadvertently facilitates the spread of misinformation, which often fails to reflect the true nature of events. For instance, the recently circulating claim that “infection with *Mycoplasma pneumonia* will result in pneumonia and ‘white lung’” not only misleads public perception and induces psychological anxiety but also triggers social panic, disrupts daily life, and severely undermines social stability. Therefore, to mitigate the detrimental effects of misleading

content and to maintain social cohesion and stability, it is imperative to accurately identify false information.

False information is typically defined as unsubstantiated content that lacks supporting evidence, authoritative sources, or official statements [3]. In recent years, scholars globally have demonstrated significant interest in the detection of false information [4]. Previous studies primarily employed manual feature extraction techniques to derive significant attributes from various aspects, including text content [5], user information [6], and communication characteristics [7], for false information detection. However, these methods depend on manual feature extraction, which necessitates extensive domain expertise and manual annotation efforts. The vast scale and complexity of social media datasets further present substantial technical challenges. With the progress of deep learning, scholars are adopting these models to autonomously identify effective characteristics for disinformation detection. Ma et al. [8] were among the first to apply deep learning models to the field of false information detection. They input individual sentences into a recurrent neural network, using the hidden layer vectors to represent information features for false information detection. A key characteristic of false information is its rapid dissemination, often exhibiting a “viral” propagation pattern. Moreover, misinformation often spreads more extensively, quickly, and broadly than factual information, potentially causing significant confusion [3]. Consequently, researchers have sought to capture the temporal sequence characteristics of false information propagation by analyzing its propagation paths or networks on social media, thereby constructing time-series feature models [9,10]. With the continued advancement of graph representation learning, false information detection utilizing graph-based approaches has become a prominent research focus. Bian et al. [11] and Rahimi et al. [12] constructed graph structures based on reply or retweet relationships, analyzing false information through edge-based feature aggregation. However, most deep learning-based rumor detection methods rely on reliable propagation structures (which may not always be available) and large-scale datasets to enable such analysis. These methods not only fail to account for the instability in misinformation propagation structures but also lead to substantial resource consumption during data preprocessing and model training.

Our approach is based on the observation that the reinforcement of community structures plays a significant role in enhancing the efficacy of information dissemination, particularly in social networks, where rumor propagation structures demonstrate superior information representation capabilities [13]. However, existing deep learning-based rumor detection models often overlook the inherent uncertainty factors in real social networks when analyzing these propagation structures. Additionally, the effective utilization of original text features and their integration with other features remains an area requiring further exploration. In this study, we propose a multi-source information fusion approach for misinformation detection, which integrates global structural information with local semantic features of messages, leveraging both message content and propagation path features to improve detection accuracy and robustness. The discrete and fragmented nature of post texts is transformed into a propagation graph structure, while the BERT [14] (Bidirectional Encoder Representations from Transformers) model is employed to extract deep semantic features from both source texts and propagated texts. As an augmented node attribute within the dissemination graph framework, the bidirectional graph convolutional neural network is utilized to capture the topology of the event diffusion network during the analysis of information propagation patterns. Finally, the resulting propagation graph, enriched with reply semantic information, is combined with the semantic features of the source declaration and processed by the classifier. Based on the interplay between dissemination architecture and content, the global forwarding relationships are incorporated into the false information detection process. The key contributions of our study are as follows:

- We leverage the structural information of information dissemination to propose an effective disinformation detection method, termed ICP-BGCN. During the propagation of false information, interactive data, such as user comments and retweets, are embedded and represented using a bidirectional graph convolutional neural network. This network extracts global coupling features from both the nodes within the message propagation graph and the relationships among them, enabling feature extraction of disinformation dissemination data based on graph-based information representation.
- The proposed method is validated using two real-world datasets, Twitter15 and Twitter16. Experimental results show that the proposed model outperforms existing baseline methods in detecting false information. The rationality and effectiveness of the proposed ICP-BGCN method are further confirmed through ablation experiments.
- We investigate the characteristics of propagation paths for both disinformation and non-disinformation, as well as the differences in their dissemination dynamics, through visualization and statistical analysis of various topological metrics of the message propagation graphs, including degree distribution, diameter, and average path length of network nodes.

2. Related Work

In recent years, a variety of theories and techniques have been proposed for identifying false information. In this section, we summarize the prevailing methods for disinformation detection into two categories: content-based attributes and propagation structural elements.

2.1. False Information Detection Based on Content Features

Content-based detection methods focus on extracting text features, such as lexical features, text length, and grammatical features, to distinguish between false information and genuine information based on content. Early studies utilized classification algorithms, including Bayes [15], Support Vector Machine (SVM) [16], and random forest [17], to classify false information based on content representation. Choudhury et al. [18] used an optimized genetic algorithm to solve the problem of fake news detection and compared the performance of several common classifiers, including support vector machine, Naive Bayes, random forest, and logistic regression, to test their effectiveness in identifying fake news in different datasets.

Deep learning approaches facilitate the automated learning of high-level data characteristics and enable more accurate identification of patterns and correlations within the data, establishing themselves as the predominant detection methodology. Nasir et al. [19] proposed an innovative hybrid deep learning framework that integrates convolutional neural networks (CNNs) with recurrent neural networks (RNNs) for the classification of deceptive information. Shelke et al. [20] utilized foundational attributes from multiple categories, including user, content, and lexical features, and developed a hybrid deep learning model that combines Bidirectional LSTM (BiLSTM) with Multilayer Perceptron (MLP) frameworks to improve detection accuracy. Yang et al. [21] employed BERT and CNN to extract textual information from user comments, followed by LSTM to further extract sentiment features, which were then fused with content features to form false information representations. Li et al. [22] utilized the pre-trained BERT model to transform raw Weibo text into vector representations, which were then fed into a recurrent convolutional neural network for the identification of false information. Feng Lizhou et al. [23] proposed a detection method based on graph convolutional networks integrated with an attention mechanism, which incorporates the forwarding relationship features and semantic characteristics of texts within comments. Chen et al. [24] pioneered the integration of the attention mechanism with a residual network for rumor detection. First, the residual network, enhanced by the

attention mechanism, captured long-range dependencies within the text. Subsequently, CNN was used to select important components and local features, achieving notable results in false information detection and early detection tasks. Furthermore, to improve the effectiveness of disinformation detection, Weng et al. [25] and Lai et al. [26] employed dynamic feature vectors and large language modeling approaches, respectively, to capture semantic features. However, malicious actors may employ deepfake technology to mimic the writing techniques and expression styles of real events, thereby evading detection. This renders methods relying solely on content features ineffective in capturing relevant features for false information recognition [27].

2.2. False Information Detection Based on Propagation Structure and Content Features

With the rapid development of graph learning techniques, researchers have discovered significant differences in diffusion topology between false information and authentic content within social networks, prompting propagation path analysis to become one of the core foundations for online rumor identification. The propagation structure feature refers to a pattern feature of false information during its spread, encompassing the interactions among posts, the extent of the formed propagation route, and the duration of the transmission time. Investigators concentrate on discerning the disparities in the propagation traits of deceitful information versus veracious content for detection purposes. Certain academics employ deep learning techniques to extract the sequential attributes of misinformation propagation over time, drawing from the dissemination pathways or networks within social media platforms, intending to formulate a temporal feature model. Chen et al. [28] proposed a Recurrent Neural Network (RNN)-based model that learns deep representations of tweet sequences for false information detection. Some researchers also represent the non-sequential propagation structure features to realize false information detection. Ma et al. [8] use the non-sequential propagation structure of false information to learn the features that distinguish the content of tweets. The complete propagation structure can extract features more comprehensively than the partial structure. Bian et al. [11] used the graph convolutional network method to realize the construction of a false information detection model from two directions of propagation and diffusion. Feng et al. [29] also modeled the false information propagation structure as a bidirectional graph and designed three interpretable bidirectional graph data augmentation strategies. Using node-level and graph-level contrastive learning to capture the propagation characteristics of events. Hu Dou et al. [30] presented a disinformation detection method grounded in a multi-relation propagation tree, addressing the limitation whereby most current propagation characteristic-based approaches merely take into account the overt interaction dynamics during dissemination while neglecting the modeling of latent relationships. Qiang Zishan et al. [31] introduced a social media rumor detection model integrating dynamic propagation and community structure, which solved the problem of insufficient utilization of time information in existing rumor detection models. Verification using the community structure characteristics of false information propagation can improve the performance of the rumor detection model. Lin et al. [32] represented false information as different propagation threads and designed a hierarchical cue encoding mechanism to learn the context representation of false information from a language-independent perspective. They advocated an innovative zero-shot detection framework predicated on cue learning, which can identify false information from different domains or different languages.

False information detection methods based on propagation relationships rely on complete network topology data. When there are missing records or incomplete data in the propagation links, their detection accuracy significantly decreases. Furthermore, existing methods mainly focus on the structural features of the propagation path, while neglecting

the semantic information of the propagated content itself. This may lead to misjudgments about the nature of the information, such as its authenticity or emotional tendency.

2.3. Content-Based and Propagation Path-Based Methods

Recent studies have increasingly focused on combining content-based and propagation path-based methods to enhance the accuracy and robustness of misinformation detection. For instance, the GCAN model proposed by Lu and Li [33] integrates textual content with retweet sequences through a co-attention mechanism, achieving a 16% accuracy improvement while providing explainable predictions. Similarly, Heterogeneous Graph Attention Networks (Huang et al.) [34] construct tweet-word-user graphs to capture global semantic relationships and structural propagation patterns, enabling early-stage rumor detection. GARD (Tao et al.) [35] further emphasizes semantic evolvement during information diffusion by combining graph autoencoders with propagation structures, improving both robustness and early detection performance. Additionally, the GCNs-MT model (Chang et al.) [36] employs memory-augmented Transformers and Graph Convolutional Networks to jointly model textual dependencies and propagation dynamics, demonstrating strong cross-dataset generalization capabilities. These approaches highlight the complementary strengths of content and propagation features, advancing detection accuracy, explainability, and timeliness.

However, these methods overlook the feature representation of propagation nodes while focusing on the expression of propagation relations during the process. They also fail to consider the impact of using deep semantic features from the propagated text as initial node features for false information detection. Furthermore, during feature extraction, the original text features are underutilized, and the fusion of the original text features with other features is inadequate. To address these issues, this paper proposes a multi-source information fusion approach for false information detection based on graph convolutional networks. This method constructs a propagation graph by combining propagation users and structures, embedding both the source and propagated texts' semantic information as node features. Additionally, it fully leverages the source text information for multi-feature fusion, further improving detection performance.

3. Methodology

To address the aforementioned challenges, this study proposes ICP-BGCN (Integrated Content-Propagation Bidirectional Graph Convolutional Network), a novel framework for false information detection that integrates original text semantics, propagation context, and structural diffusion patterns. As illustrated in Figure 1, the framework comprises three interconnected modules:

Module 1. Propagation Structure Modeling: A hierarchical propagation graph is constructed using comment–retweet relationships. Semantic embeddings generated by Module 2 are mapped to graph nodes to initialize their representations. The DropEdge technique [37] is then applied to both top-down and bottom-up propagation graphs to mitigate over-smoothing and enhance model robustness. Finally, a bidirectional graph convolutional network (GCN) captures nuanced node interaction patterns within the propagation structure.

Module 2. Textual Feature Extraction: A pre-trained language representation model (BERT) encodes the source post and its associated comments/retweets to extract deep semantic features.

Module 3. False information Detection: Propagation structural features are fused with textual semantics through attention mechanisms. The enriched representations are fed into a classifier to predict whether the source post is false information.

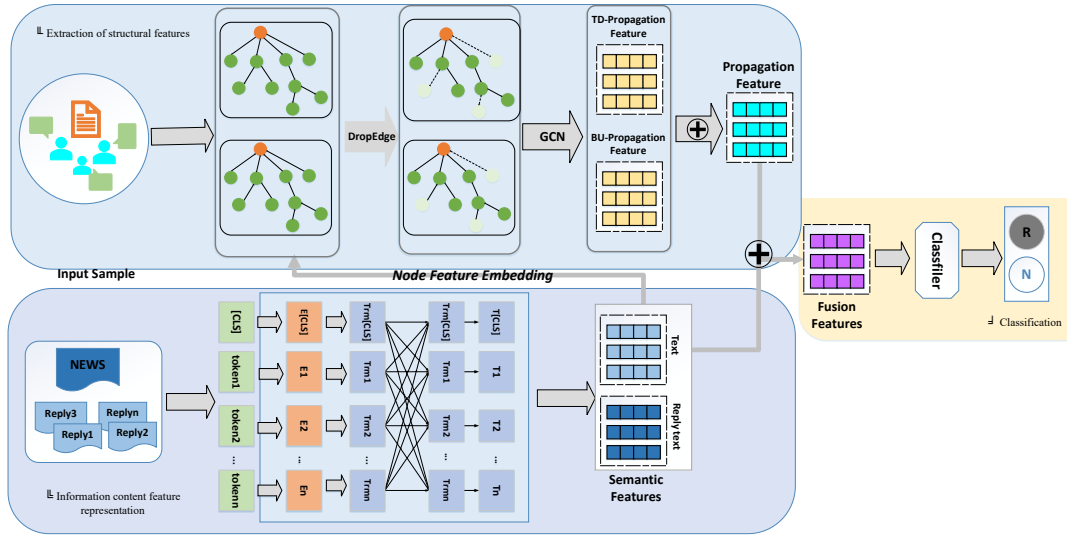


Figure 1. Framework of ICP-BGCN model.

3.1. Construction of Information Propagation Graphs

3.1.1. Building the Information Propagation Graph

In constructing the information dissemination graph, we treat each post as a node, while edges between posts are defined based on comment and retweet relationships. Specifically, if tweet A is a reply to tweet B, then there will be a directed edge between A and B in the graph to indicate the direction of information flow. Thus setting $C = \{c_1, c_2, \dots, c_m\}$ for the false information detection data sets, including c_i for the data set of the event i , m for the cumulative sum of events. $c_i = \{r_i, w_i^{(1)}, w_i^{(2)}, \dots, w_i^{(n_i-1)}, G_i\}$, where n_i is the total number of posts for event i , r_i represents the source post, $w_i^{(j)} \in V_i (j = 1, 2, \dots, n_i - 1)$ is the j -th related response posts, and $G_i = (V_i, E_i, A_i)$ on behalf of the propagation graph of event i . The undirected graph $G_i = (V_i, E_i, A_i)$ is constructed according to the comment or retweet relationship in the event, where $V_i = \{r_i, w_1^i, w_2^i, \dots, w_{n_i-1}^i\}$ represents the set of nodes in the graph, $E_i = \{e_{st}^i | s, t = 0, \dots, n_i - 1\}$ indicates the set of response behaviors in the process of information propagation, and the adjacency matrix $A_i \in \{0, 1\}^{n_i \times n_i}$ represents the response relationship of event propagation in the propagation graph G , which is defined as Equation (1).

$$A_{st}^i = \begin{cases} 1, & e_{st}^i \in E_i \\ 0, & \text{other} \end{cases}, \quad (1)$$

Based on inter-node relationships, we construct three distinct propagation graph structures with differentiated characteristics. First, the Undirected Graph (UD-Graph) employs binary relationships to represent node connectivity, preserving adjacency relationships while disregarding interaction directionality. Second, the Top-Down Propagation Graph (TD-Graph) adheres to hierarchical information flow patterns, where directional edges from parent posts (source posts r_i) to child posts (response posts $w_i^{(j)}$) explicitly delineate the tree-like diffusion pathways of information dissemination. In contrast, the Bottom-Up Propagation Graph (BU-Graph) establishes reverse edge configurations (child-to-parent orientation) to form information convergence networks, a topological design that systematically captures how diverse information sources aggregate toward specific nodes.

As can be seen from Figure 2, we construct the top-down propagation graph structure $G_i^{TD} = \langle V_i, E_i^{TD} \rangle$ and a bottom-up diffusion graph structure $G_i^{BU} = \langle V_i, E_i^{BU} \rangle$ for graph G_i , and use adjacency matrices A_i^{TD} and A_i^{BU} to represent the connection relationship of

nodes on the graph, respectively. In this paper, the semantic feature expression obtained by processing the original post and the response text content by BERT pre-training model is used as the feature vector of the post, i.e., the initial feature matrix of the propagation structure and the diffusion structure are the same. $X^{TD} = X^{BU} = [x_0, x_1, \dots, x_{n-1}] \in \mathbb{R}^{n_i \times d}$, where n_i represents the number of posts for event i , and d is the dimensionality of the feature space. $x_0 \in \mathbb{R}^d$ denotes the feature vector of the original post, while $x_j \in \mathbb{R}^d (j = 1, 2, \dots, n_i - 1)$ represents the feature vector of the response post. The adjacency matrices are transpositions of each other, i.e., $A_i^{TD} = (A_i^{BU})^T$.

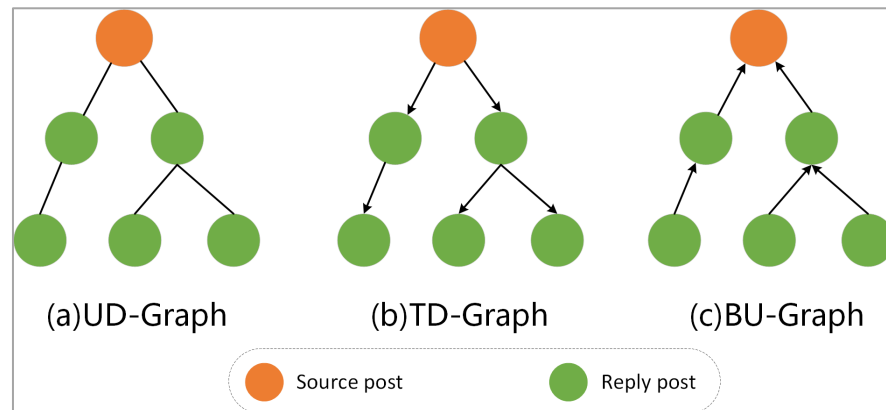


Figure 2. Information event propagation diagram.

3.1.2. Extraction of Structural Features of Information Propagation

Given the extensive node count within the propagation graph G_i , the DropEdge technique is employed during model training to randomly eliminate edges in the original graph. This approach aims to circumvent over-fitting and mitigate information loss due to excessive smoothing. In the DropEdge algorithm, we first randomly sample a subset E_{drop} from the edge set E_i , with a size of $\lfloor p \cdot |E_i| \rfloor$. Subsequently, we generate a mask matrix $M \in \{0, 1\}^{n_i \times n_i}$, where M_{ij} is defined by the Equation (2). Finally, we use the Hadamard product ($\odot$) to modify the adjacency matrix A , resulting in $A' = A \odot M$.

$$M_{ij} = \begin{cases} 0, & e_{ij} \in E_{drop} \\ 1, & \text{others} \end{cases} \quad (2)$$

As shown in Figure 3, in this paper, the bidirectional Graph Convolutional Network (GCN) is used as a feature extractor to extract the propagation structure features of information. Since the original post contains the most abundant information, in the calculation process the characteristics of each node are combined with the characteristics of its root node to strengthen the information content of the source post. The procedure for extracting propagation characteristics is illustrated by Equations (3)–(6):

$$H_1^{TD} = \sigma(\hat{D}^{-\frac{1}{2}} \tilde{A}^{TD} \hat{D}^{-\frac{1}{2}} X W_0) \quad (3)$$

$$\tilde{H}_1^{TD} = \text{concat}(H_1^{TD}, X_{root}) \quad (4)$$

$$H_2^{TD} = \sigma(\hat{D}^{-\frac{1}{2}} \tilde{A}^{TD} \hat{D}^{-\frac{1}{2}} \tilde{H}_1^{TD} W_1) \quad (5)$$

$$H^{TD} = \text{concat}(H_2^{TD}, H_{1TD}^{root}) \quad (6)$$

where $\hat{D}$ is the degree matrix corresponding to the propagation graph, $\tilde{A} = A^{TD} + I_N$, W_0 , W_1 and are all learnable parameters, and the obtained H^{TD} is the top-down propagation characteristics of false information. Similarly, the bottom-up propagation characteristics

H^{BU} can be obtained, and the representation vector $h_p \in R^{1 \times d_p}$ of information propagation characteristics can also be obtained by average pooling after splicing the two. The pooling process is shown in Equation (7).

$$h_p = \text{Meanpool}(\text{concat}(H^{TD}, H^{BU})) \quad (7)$$

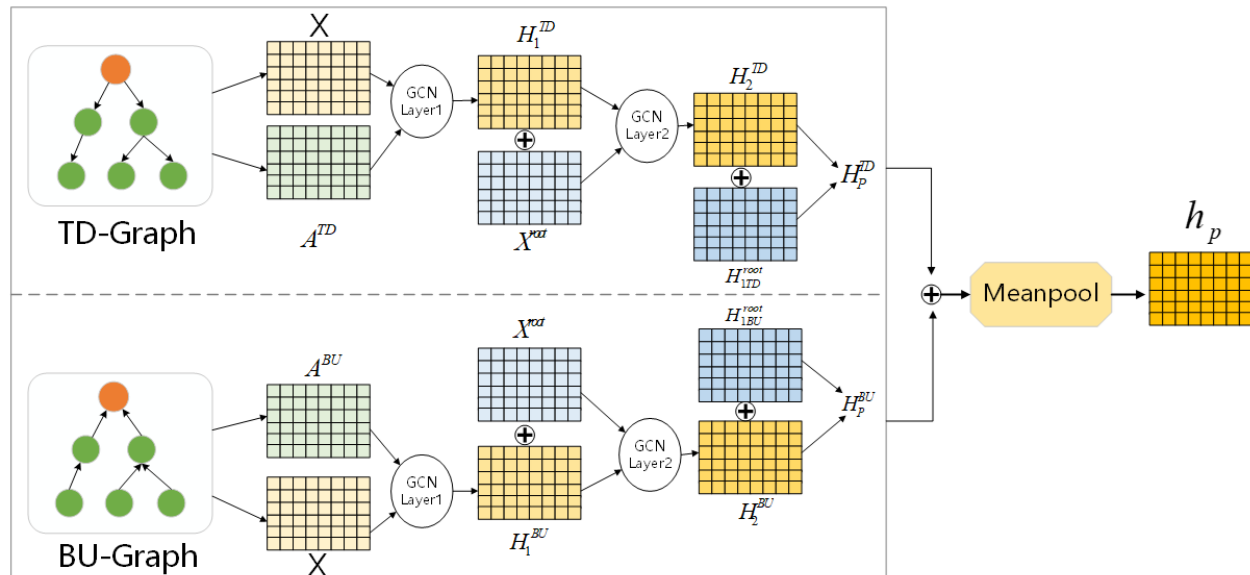


Figure 3. Extraction of propagation features.

3.2. Information Content Feature Representation

Both the original and propagated text content of false information contain critical semantic cues, and leveraging their semantic features comprehensively can enhance detection performance. Given that BERT captures bidirectional context-aware semantic representations to derive richer latent features, this study utilizes the pre-trained BERT model to generate contextual semantic embeddings from textual content, as illustrated in Figure 4.

The input to the pre-trained BERT consists of word embeddings (token embeddings), sentence embeddings (segment embeddings), and position embeddings (positing embeddings). The text information of length n is obtained $W = \{[CLS], token_1, token_2, \dots, token_n\}$ after word segmentation, where $[CLS]$ represents the token flag used for subsequent classification and $token_i$ represents the i -th word in W . The W is input into three embedding layers to obtain word vector E_{token} , text vector $E_{segment}$ and position vector $E_{position}$, respectively, and the three are superimposed to obtain a new vector $E = \{E_{[CLS]}, E_1, E_2, \dots, E_n\}$. The obtained vector E is used as the input for the model.

The overall architecture of BERT is composed of multiple stacked Transformer encoders. Each encoder layer contains a multi-head attention mechanism and a feedforward neural network. Through the multi-head attention mechanism of the encoder, there is a mutual attention edge between the encoded expression (Trm_i) of each token, which can achieve the same degree of attention between words with different distances. BERT internally stacks six encoders with the same structure in series to better learn the contextual semantic information of the input text. Among them, multi-head attention can be expressed as Equations (8) and (9):

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_n)\omega \quad (8)$$

$$\text{head}_i = \text{Attention}(Q\omega_i^Q, K\omega_i^K, V\omega_i^V) \quad (9)$$

where: n is the number of heads of the multi-head attention mechanism; $head_i$ is the output of the i -th head; and Q, K, V are obtained by linear transformation of the input feature matrix. $\omega^Q, \omega^K, \omega^V$ are the parameter matrices of Q, K and V learned after training, respectively.

Finally, the feature vector learned from the classification label $[CLS]$ is fed into the fully connected feedforward neural network layer to obtain the semantic expression S of the information text.

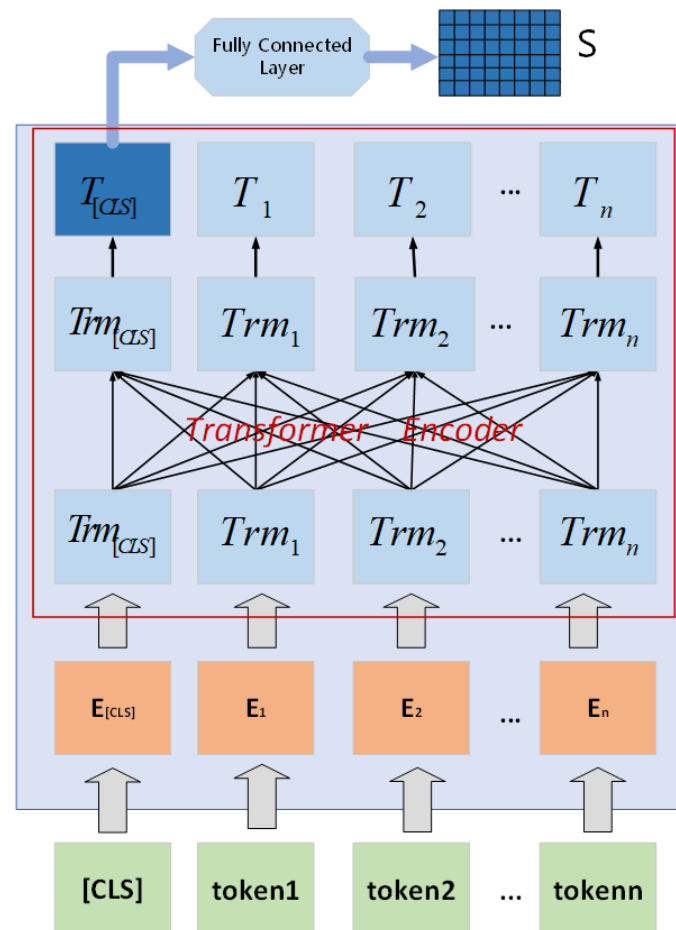


Figure 4. Extraction process for information content features.

3.3. Classification of False Information

After obtaining the semantic features (S) of the original text and the propagation structure features (h_p) enhanced by the root node, these features are concatenated to obtain the fusion feature. The process is shown in Equation (10):

$$F = \text{Concate}(h_p, S) \quad (10)$$

Ultimately, the amalgamated features are fed into a fully connected neural layer followed by a softmax layer to discern the event labels, with model refinement achieved through minimizing the cross-entropy loss function. The process is shown in Equations (11) and (12):

$$\hat{y} = \text{softmax}(W_c F + b_c) \quad (11)$$

$$\text{Loss} = -\sum_{i=1}^C y_i \log \hat{y}_i \quad (12)$$

where $\hat{y} \in \mathbb{R}^{1 \times C}$ is the probability vector of all classes predicting the event label, $y_i \in \{0, 1\}$ is the true label value, C is the number of classification classes, and both W_c and b_c are parameters that can be learned through training.

4. Experiments

4.1. Datasets

To effectively evaluate the performance of the model, we conducted experiments on four publicly available datasets: Twitter15, Twitter16, PHEME, and SemEval-17-task 8. Twitter15 and Twitter16, crawled and collated from the Twitter platform by Ma et al. [38], contain 1490 and 818 rumor-related events, respectively.

Each event is labeled based on the authenticity annotations from rumor verification websites, categorized into four classes: Non-rumors (NRs), verified False rumors (FRs), verified True rumors (TRs), and Unverified rumors (URs). PHEME [39], a benchmark dataset for rumor detection in social media, comprises 5746 tweets collected during breaking news events (e.g., Ferguson protests, Charlie Hebdo shooting). Labels categorize posts into Rumor or Non-Rumor based on fact-checking annotations. The SemEval-17-task 8 dataset [40], a widely adopted benchmark for rumor analysis, contains 325 real-world events with 3854 Twitter conversational threads. It focuses on fine-grained rumor verification with three labels: “True rumor” (TR), “False rumor” (FR), and “Unverified rumor” (UR), while also providing stance classification subtasks (Support, Deny, Query). In the dataset, nodes signify individual posts, edges delineate retweet or comment connections, and nodal characteristics are pre-trained utilizing BERT. The summarized metrics of the dataset are presented in the following Table 1.

Table 1. Statistics of the dataset.

Statistics	Twitter15	Twitter16	PHEME	SemEval-17
Number of events	1490	818	5473	325
Non-rumor	374	205	-	-
False rumors	370	205	3477	74
Unsubstantiated rumor	374	203	-	106
True rumor	372	205	1996	145
Number of users	276,663	173,487	36,443	-
Average number of posts	223	251	20	-
Maximum number of posts	1768	2765	346	-
Minimum number of posts	55	81	3	-

4.2. Experiment Settings

To confirm the efficacy of the model posited in this paper for the task of disinformation detection, this study contrasts it with other foundational techniques within the experimental framework. The methodological specifics are elucidated as follows:

DTC [41]: A classification model based on decision trees, which manually extracts features to obtain information credibility for false information detection.

SVM-RBF [6]: The model is based on a support vector machine with an RBF kernel, which utilizes the aggregate statistics of the postings for disinformation detection features by manually constructing features.

GRU [8]: A deep learning model based on RNN that detects false information by learning the propagation sequence of messages, i.e., the temporal structure characteristics of events.

RvNN [42]: A false information detection method based on a tree-structured recurrent neural network with GRU units.

PPC_RNN + CNN [43]: The model combines RNN and CNN to learn the representation of events through the user information on the message propagation path, and then identifies false information.

Bi-GCN [11]: This is a graph model using a bidirectional graph convolutional neural network. Features are extracted from top-down (top-down) and bottom-up (bottom-up) propagation directions of rumors for detection.

GCN-Bert [44]: The rumor detection method not only considers the features of the message itself, but also utilizes the rumor features of all relevant texts and words.

HAGNN [45]: A graph neural network-based disinformation detection model that captures high-level representations of textual content at different granularities, and fuses propagation structures for disinformation detection.

The experiment detailed in this paper is executed on the Ubuntu 22.04 platform, with the experimental environment consisting of Python 3.10 and PyTorch 2.1.0. Table 2 presents the precise specifications of the experimental setup. To ensure impartiality in comparison, the dataset was randomly partitioned into five segments and a five-fold cross-validation experiment was conducted on these segments. During training, the configurations were set as follows: hidden layer dimension to 64, iterations (epoch) to 200, batch size (batch_size) to 128, learning rate to 0.0005, and dropout rate to 0.2. The Adam optimization algorithm facilitated model parameter updates, while the early stopping technique was implemented to prematurely cease training if the validation set's loss value remained stagnant for 10 consecutive trials. Model efficacy was gauged using the evaluation metrics: Accuracy (Acc) and F1 score.

Table 2. Configuration of the experimental environment.

Experimental Environment	Experimental Setup
OS	Ubuntu22.04
Development environment	VScode 1.98.2
processor	Intel(R)Xeon [®] Silver 4116 CPU @ 2.10 GHz
Graphics card model	NVIDIA GeForce RTX 3090, RTX (24 GB)
programming language	Python 3.10
Deep Learning Framework	Pytorch 2.5.1+cu124

4.3. Results and Analysis

On Twitter15 and Twitter16 datasets, the proposed ICP-BGCN method is analyzed with seven baseline models such as the classical DTC, and the experimental results are shown in Tables 3 and 4.

Table 3. Comparative experimental results on Twitter15 dataset.

Methods	ACC	F1			
		NR	FR	TR	UR
DTC	0.454	0.733	0.355	0.317	0.415
SVM-RBF	0.318	0.318	0.318	0.318	0.318
GRU	0.646	0.792	0.574	0.608	0.592
RvNN	0.723	0.723	0.723	0.723	0.723
PPC_RNN + CNN	0.667	0.667	0.667	0.667	0.667
Bi-GCN	0.886	0.891	0.860	0.930	0.864
GCN-Bert	0.872	0.853	0.892	0.823	0.911
HAGNN	0.865	0.813	0.870	0.905	0.896
ICP-BGCN	0.897	0.904	0.959	0.850	0.869

Bold indicates optimal results.

Table 4. Comparative experimental results on Twitter16 dataset.

Methods	ACC	F1			
		NR	FR	TR	UR
DTC	0.465	0.643	0.393	0.419	0.403
SVM-RBF	0.553	0.553	0.553	0.553	0.553
GRU	0.633	0.772	0.489	0.686	0.593
RvNN	0.737	0.737	0.737	0.737	0.737
PPC_RNN + CNN	0.690	0.690	0.690	0.690	0.690
Bi-GCN	0.880	0.847	0.869	0.937	0.865
GCN-Bert	0.877	0.915	0.827	0.886	0.930
HAGNN	0.874	0.815	0.809	0.880	0.865
ICP-BGCN	0.917	0.931	0.891	0.962	0.883

Bold indicates optimal results.

As evidenced by the experimental results in Tables 3 and 4, the ICP-BGCN model achieves superior classification performance with accuracies of 89.7% and 91.7% on the Twitter15 and Twitter16 datasets respectively, demonstrating marked superiority over baseline models. Additionally, it exhibits exceptional performance across precision and recall metrics: achieving precision and recall rates of 90.9% and 89.3% on Twitter15, and 90.1% and 91.7% on Twitter16. This performance profile confirms its dual capability in maintaining classification accuracy while ensuring comprehensive sample coverage. Furthermore, ICP-BGCN performs well on the NR, FR, TR and UR criteria on both datasets. The model is able to maintain a high level of performance when dealing with different categories of samples, achieving more than 85% on all criteria. The combined advantages of the multidimensional metrics validate the robustness and stability of the model in handling different categories of rumor detection tasks.

Through experiments, it can be found that the detection method based on deep learning is superior to the detection method based on machine learning. On the Twitter15 dataset, the accuracy of ICP-BGCN is 44.3% higher than that of DTC and 57.9% higher than that of SVM-RBF. The main reason is that machine learning relies on manual extraction of features, which needs to rely on the experience and judgment of workers, while deep learning-based models can automatically capture deeper features and the correlation between features, thus better identifying false information.

Among the seven deep learning-based misinformation detection models, ICP-BGCN, HAGNN, GCN-Bert, Bi-GCN utilize graph neural networks to extract false information propagation structure features, and demonstrate better performance than the other three models. This shows that GNNs are effective to model the propagation process of information using propagation graphs and extract propagation structure features. Our ICP-BGCN model fuses the propagation structure and the semantic features of the message text. Compared with Bi-GCN, which only considers the structure of information dissemination, GCN-Bert, which utilizes text information features at different granularities, and HAGNN, which captures multi-level semantic information of text content and combines the structural features of dissemination networks, the detection accuracies on the Twitter15 dataset are improved by 1.1%, 2.5%, and 3.2%, respectively. It is also better than other models in various indicators, which shows that it is reasonable and effective to fully fuse the original text features, propagation text features, and propagation structure features to improve the accuracy of false information detection. Overall, the ICP-BGCN model outperforms the other eight models, which include traditional machine learning and deep learning approaches, in terms of detection accuracy and F1 score for each category to varying extents.

To comprehensively assess the cross-scenario generalization capability of our proposed ICP-BGCN model, we select the Pheme dataset with significant scenario variations as

our validation benchmark. As a misinformation detection benchmark dataset in breaking news scenarios, PHEME encompasses multiple crisis domains, including social, political, and health-related events [46], with its cross-domain characteristics providing an ideal experimental environment for evaluating model adaptability across different scenarios. Compared to conventional datasets, like Twitter15 and Twitter16, PHEME exhibits three distinctive characteristics. Firstly, in terms of data composition, the dataset contains high-risk events with heightened emotional intensity (e.g., public health crises, political scandals) and user interactions exhibiting more pronounced emotional signals [47], presenting multidimensional challenges for semantic modeling. Secondly, the PHEME dataset exhibits significant variations in event scales, where much of the information labeled as rumors actually originates from misclassifications of real events [46]. This label distribution characteristic constitutes a rigorous test of the model's discriminative power. Thirdly, regarding data scale, the limited rumor samples (with non-rumors constituting 63.5% of instances) intensify training challenges under class imbalance conditions [46].

These characteristics closely align with the complex scenarios of misinformation propagation in real-world settings. To validate ICP-BGCN's performance on cross-scenario datasets, we maintain identical parameter configurations to those used in Twitter15 and Twitter16 experiments, conducting comparative analyses with four baseline models. The experimental results for the PHEME dataset are shown in Table 5.

Table 5. Comparative experimental results on PHEME dataset.

Methods	Class	Accuracy	Precision	Recall	F1
GCAN [33]	R	0.834	0.769	0.758	0.761
	N		0.871	0.874	0.872
Bi-GCN [11]	R	0.824	0.753	0.734	0.741
	N		0.861	0.872	0.865
GACL-CADA [48]	R	0.808	0.523	0.722	0.599
	N		0.921	0.831	0.872
GAN [49]	R	0.823	0.765	0.760	0.760
	N		0.858	0.858	0.857
ICP-BGCN	R	0.844	0.770	0.794	0.775
	N		0.889	0.870	0.775

The four baseline models employ distinct technical approaches. GCAN [33] utilizes graph neural networks with dual co-attention mechanisms to achieve multimodal dynamic feature fusion across source text, user attributes, and propagation pathways. Bi-GCN [11] leverages bidirectional graph convolutional networks to concurrently model forward diffusion (source-to-retweeters) and reverse traceability (leaf-to-source) patterns in information dissemination. GACL-CADA [48] implements a class-aware adversarial domain adaptation framework to address cross-domain distribution alignment between historical data and emerging events. GAN [49] enhances detection robustness for hybrid true/false content through adversarial sample generation (e.g., semantically ambiguous text variants) coupled with discriminator-based decision boundary optimization.

Experimental results demonstrate that ICP-BGCN achieves an accuracy of 84.4%, representing a 1% improvement over the best baseline model (GCAN: 83.4%), while also delivering competitive performance in precision, recall, and F1-score. This finding illustrates that the model not only effectively identifies routine rumor patterns in Twitter15 and Twitter16 datasets but also exhibits robust performance in handling PHEME's high-emotion, semantically ambiguous crisis-related misinformation. This cross-domain adaptability highlights its robustness and potential for generalization across diverse propagation scenarios.

To further ensure the accuracy of our results and assess the generalization capability of our proposed ICP-BGCN model, we incorporate the SemEval-17-task 8 dataset as an additional benchmark in our experiments. The SemEval-17-task 8 dataset is widely adopted for rumor analysis and provides a rich set of Twitter conversational threads with fine-grained labels—“True rumor” (TR), “False rumor” (FR), and “Unverified rumor” (UR)—as well as stance classification tasks. The experimental results are shown in Table 6. The technical specifications of the four baseline models are as follows.

Table 6. Comparative experimental results on SemEval-17 dataset.

Methods	Accuracy	Macro-F1
HiTPLAN [10]	0.571	0.581
MTL2-hierarchical transformer [50]	0.643	0.657
Coupled hierarchical transformer [50]	0.678	0.680
HCD-MGN [51]	0.767	0.754
ICP-BGCN	0.785	0.792

HiTPLAN [10] employs a multi-level Transformer architecture to capture nuanced contextual representations from social media posts for deceptive content detection. MTL2-Hierarchical Transformer hierarchically [50] segments conversational threads into sub-threads, encodes contextual features using BERT embeddings, and aggregates cross-sub-thread semantics via Transformer fusion to enable multi-granular representation learning. Coupled Hierarchical Transformer extends MTL2 [50] by integrating multi-task learning through a hybrid attention mechanism that aligns BERT-derived semantics with stance-aware propagation patterns, jointly optimizing rumor verification and stance detection. Hierarchical Contrastive Disentangled Multi-task Graph Network (HCD-MGN) [51] enhances multi-task performance through (1) a feature decoupling module (PFN) separating shared/task-specific features, (2) dual graph encoders modeling propagation structures and semantic relationships, and (3) stance-aware contrastive learning for representation optimization.

On the SemEval-17 Task 8 dataset, the proposed ICP-BGCN model achieves state-of-the-art performance with 78.5% accuracy and 79.2% Macro-F1 score, demonstrating a 1.8% absolute performance improvement over the best baseline model HCD-MGN (76.7% accuracy). These results, together with those obtained from the Twitter15, Twitter16, and PHEME datasets, confirm that our model consistently generalizes across multiple social data sources, effectively capturing the unique propagation structures inherent in different social media scenarios.

4.4. Ablation Study

To validate the functionality of individual components within the ICP-BGCN model, this paper devises three distinct ablation studies as follows:

- (1) w/o BiPS: The bidirectional propagation graph module is removed. An undirected propagation graph is used to represent the propagation path, which only considers the structure of event propagation and does not consider its direction.
- (2) w/o Ps: Removed the information propagation graph. Without considering the influence of tweet propagation structure on detection, only the content semantic features of source tweets and response posts are used for false information detection.
- (3) w/o Text: Removed textual semantic information. The semantic features of the source tweet and the response post were removed, and the model only contained the bidirectional graph structure of the propagation relationship.

The experimental results are shown in Tables 7 and 8. Analyzing these, it can be found that each module plays a unique role, and removing or replacing any of the components will

affect the overall performance of the ICP-BGCN method. When the undirected propagation graph is used to replace the bidirectional propagation graph model, the accuracy of the model on the two datasets is decreased by 2.8% and 2.8%, respectively, indicating that the topology structure combining propagation and diffusion can more effectively capture and model the propagation characteristics of events and enhance the event detection rate. Furthermore, only considering the propagation structure or relying on semantic information alone for false information detection will lead to a significant decrease in the accuracy of the model, which indicates that the propagation structure and semantic information have significant effects on false information detection respectively. Concurrently, the integration of both elements significantly enhances the precision of disinformation detection.

Table 7. Comparison of ablation experimental results on Twitter15 dataset.

Methods	ACC	F1			
		NR	FR	TR	UR
w/o Bi-Ps	0.869	0.821	0.945	0.857	0.833
w/o Ps	0.838	0.830	0.941	0.777	0.751
w/o Text	0.842	0.735	0.653	0.956	0.764
ICP-BGCN	0.897	0.904	0.959	0.850	0.869

Table 8. Comparison of ablation experimental results on Twitter16 dataset.

Methods	ACC	F1			
		NR	FR	TR	UR
w/o Bi-Ps	0.889	0.769	0.671	0.957	0.903
w/o Ps	0.855	0.802	0.867	0.944	0.477
w/o Text	0.843	0.791	0.665	0.933	0.805
ICP-BGCN	0.917	0.931	0.891	0.962	0.883

4.5. Propagation Graph Analysis

In order to explore the impact of propagation paths on disinformation detection, we statistically analyze the structure of the propagation networks of disinformation and non-disinformation, aiming to reveal the differences in propagation patterns between the two. We integrated the labels of the Twitter15 and Twitter16 datasets, classified “verified true rumors (TR)”, “verified false rumors (FR)”, and “unverified rumors (UR)” into the category of “rumors”, and compared them with the data of “Non-rumors (NR)”. The visualization of information dissemination is shown in Figure 5 below. In an information dissemination relationship graph, each node represents a separate unit of information, which includes the original tweet, its related comments, or retweets. These nodes are connected by edges, which represent interactive behaviors between them, such as retweets or comments. We define the original tweet as the root node of the relationship graph, and all posts that directly reply to that tweet become its children. Following this logic, if a post $-i$ receives a reply from another post $-j$, then according to the order of information dissemination, post $-j$ becomes a child node of post $-i$, which is represented in the graph as node $-j$ is a subordinate node of node $-i$.

As shown in Figure 5, the information dissemination tree exhibits a broad structure in which most nodes belong to the shallow first-level responses [52]. In the disinformation dissemination network, there are obvious clusters of nodes, and the nodes within the cluster are more densely connected, while the nodes connected to nodes outside the cluster are relatively sparse, which reflects the high degree of aggregation in the disinformation dissemination network. This discrepancy may be explained by the fact that well-crafted rumors usually carry

inherent information content features that can trigger the replies and reposts of multiple internet celebrity individuals with high social influence. In contrast, naturally occurring true events are not well crafted to maximize their social impact, making it non-trivial to trigger reposts or replies of multiple internet celebrity individuals simultaneously.

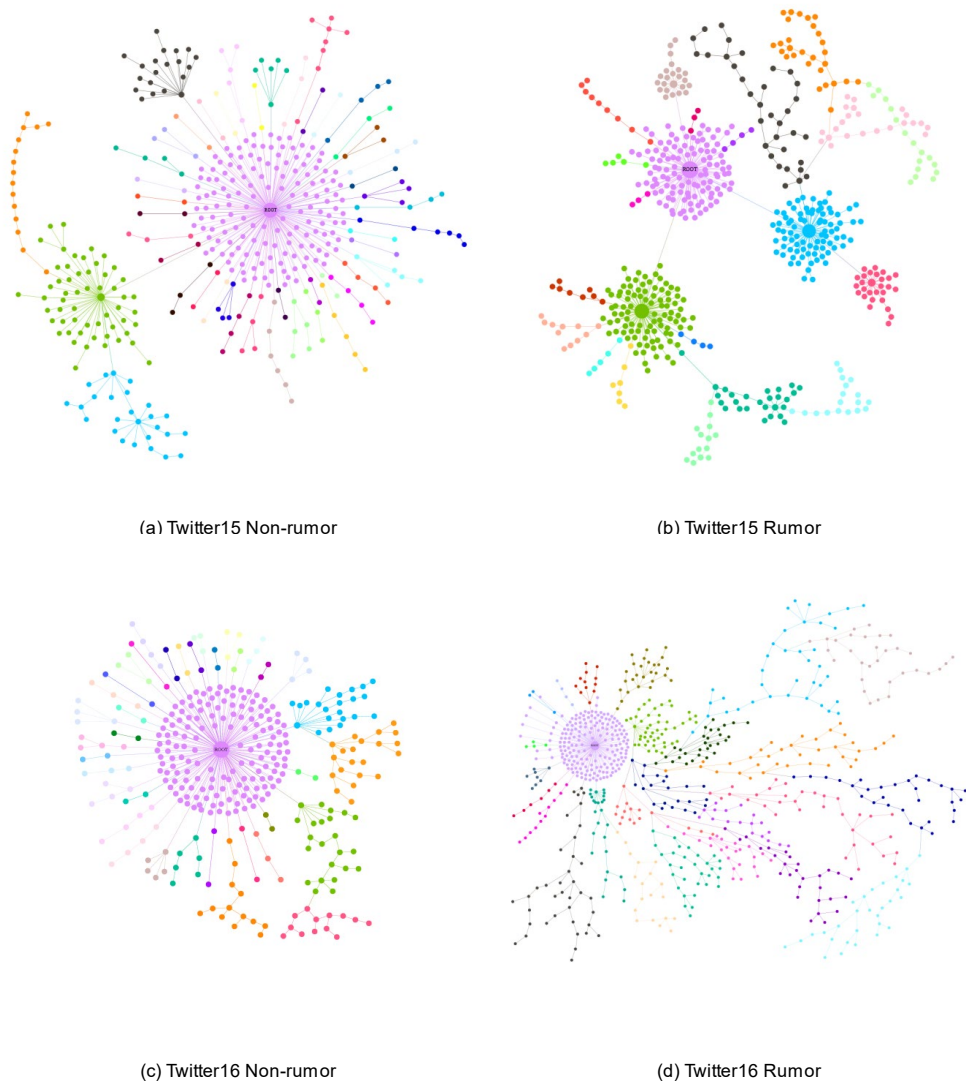


Figure 5. Comparison of information dissemination network diagrams. (a,c) represent the dissemination networks of typical non-disinformation events in the two datasets, respectively; and (b,d) represent the dissemination maps of typical disinformation in the two datasets, respectively.

In addition, the structure of the disinformation dissemination network is more complex, the dissemination path is usually longer, and the connections between nodes are closer, which suggests that the rumor information not only spreads quickly in the process of dissemination but also gets strengthened and consolidated in specific groups. In contrast, a lower aggregation trend is presented in the dissemination network of non-false information. Its path distribution is more uniform and the connection between nodes is sparse, all these features indicating that the breadth of information dissemination is limited and decentralized.

Further, we computed and compared the topological metrics of the information dissemination graphs, such as the number of nodes, number of edges, average path length, degree distribution, etc. We used the average values of the metrics for all graphs in both categories, non-spurious information and spurious information, as the final comparison

data. The calculation results of the network structure metrics are detailed in Table 9; see Figure 6 for the visualization of the degree distribution.

Table 9. Calculated network metrics for datasets.

Indicators	Twitter15		Twitter16	
	Non-Rumor	Rumor	Non-Rumor	Rumor
Nodes	52.786	51.369	43.458	50.829
Edges	103.572	100.738	84.915	99.657
Diameter	6.875	6.838	6.592	7.081
Average Path Length	2.676	2.715	2.612	2.791
Assortativity	−0.740	−0.763	−0.743	−0.748
Density	0.114	0.113	0.154	0.127

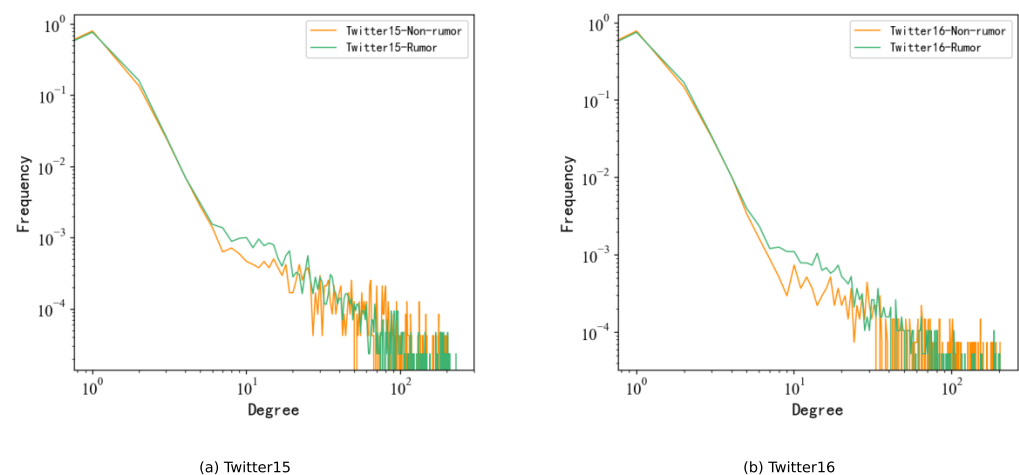


Figure 6. Plot of the degree distribution of the dataset.

In the network metrics calculation results shown in Table 9, the results of the two datasets show opposite trends. Specifically, in the Twitter15 dataset, network metrics such as the number of nodes and the number of edges present higher values in the non-rumor category; however, in the Twitter16 dataset, the values of these same network metrics in the rumor category exceed those in the non-rumor category. According to the relevant literature discussion, false information spreads farther, faster, deeper, and wider in terms of speed, scope, depth, and breadth compared to non-false information [3,53,54]. In addition, the experimental results in Section 4.3 show that disinformation detection based on propagation paths performs significantly better in the Twitter16 dataset than in the Twitter15 dataset. Therefore, we believe that the network characteristics revealed by the Twitter16 dataset are more in line with the characteristics displayed by disinformation and non-disinformation during the propagation process, where key indicators, such as the number of nodes, the number of edges, the graph diameter, and the average path length of disinformation, are larger than those of non-disinformation. That is, there is more user participation, more frequent forwarding of information, a wider dissemination range, and more complex dissemination paths in the dissemination process of disinformation. Due to its unrepresentative sample failing to encompass a broader or more balanced range of user groups and message types, the Twitter15 dataset may not exhibit network characteristics consistent with theory. In summary, the Twitter16 dataset is more consistent with the characteristics of real rumor propagation, and our method is more valid for this type of data.

In addition, by analyzing the coefficient of congruence and network density values, the analysis reveals that the information dissemination network is a low-density heterogeneous network. In such a network structure, the connections between nodes are not tightly connected, but show a tendency to connect between highly connected nodes and lowly connected nodes. This phenomenon suggests that a small number of nodes in the information dissemination network, i.e., those highly connected nodes, which play a key role in the network, have a significant impact on the overall performance of the network. Further, based on the analysis of the degree distribution graph in Figure 6, we can observe that the node degree distributions of both disinformation and non-disinformation in the information dissemination network exhibit significant long-tail characteristics. That is, there are a few highly connected nodes in the network structure, which are usually called opinion leaders or key communicators, and they play a crucial role in the information dissemination process. Therefore, accurate identification of these key nodes will help to improve the accuracy and efficiency of detection when performing disinformation detection.

4.6. Discussion

In this study, we propose the ICP-BGCN model, an innovative approach to disinformation detection, which integrates original text content, propagation text, and the structural information of message dissemination. Our experiments on the Twitter15, Twitter16, and Pheme datasets demonstrate that the fusion of semantic features extracted via BERT and propagation features learned through a bidirectional graph convolutional network leads to superior detection accuracy. Notably, the model achieves accuracies of 89.7% on Twitter15 and 91.7% on Twitter16, outperforming eight mainstream baselines by 1.1% and 3.7%, respectively, and maintains robust generalization with an 84.4% accuracy on the Pheme dataset. These results strongly support our original hypothesis that leveraging both text semantics and propagation structure can enhance disinformation detection.

Our work makes several key contributions that push the field forward. First, by embedding interactive data, such as user comments and retweets, into a graph structure, ICP-BGCN captures global coupling features that traditional models often overlook. This methodological advancement addresses limitations identified in earlier studies [11,45] and opens new avenues for exploiting network topology in misinformation analysis. Second, the detailed analysis of propagation metrics—such as degree distribution, diameter, and average path length—provides fresh insights into the distinct dissemination patterns of disinformation versus non-disinformation. For example, our observation that disinformation tends to form low-density heterogeneous networks with several highly connected nodes not only explains the superior performance on the Twitter16 dataset but also suggests potential indicators for early detection in real-world applications.

Despite these advances, our study has limitations that must be acknowledged. One notable limitation is that the current model does not account for the temporal decay of post influence, an aspect that requires further investigation.

5. Conclusions

Based on the fusion of original text content, propagation text content and the propagation structure, we propose an effective false information detection model called ICP-BGCN that fuses message content with the propagation path. The BERT model is employed to discern the deep semantic features of the original text and the propagation text. The propagation structure is integrated with the semantic features of the text, and the bidirectional graph convolutional neural network is used to learn the propagation features from the semantic information. The obtained propagation features are combined with the enhanced semantic features of the original text to generate fusion features, upon which

false information detection is performed. To evaluate the rationality and effectiveness of our ICP-BGCN model, comparative experiments are conducted on the public datasets Twitter15 and Twitter16. The experimental results show that, in general, our ICP-BGCN model performs better than eight baseline models, such as DTC. To further assess its generalization capability, we validate the model on the PHEME dataset, achieving robust performance (84.4% accuracy) that demonstrates cross-domain adaptability. We compare the propagation network structures of disinformation and non-disinformation in the Twitter15 and Twitter16 datasets, and find that the disinformation propagation structure exhibits a low-density hetero-collocation network characterized by multiple nodes with high connectivity. In addition, the Twitter16 dataset shows more obvious characteristics in its disinformation propagation networks compared to the Twitter15 dataset, which partially explains the higher performance of our model over this dataset. This study did not take into account the gradual loss of influence of posts over time, which will be considered in future work to more accurately model the dynamic process of information dissemination. In addition, we will focus on the early stages of disinformation propagation, using incomplete propagation structures and limited information for early disinformation detection.

Author Contributions: Conceptualization, M.Y., J.H. (Jie Hu) and J.H. (Jianjun Hu); methodology, M.Y.; software, M.Y.; validation, M.Y., B.T. and J.H. (Jianjun Hu); data curation, M.Y. and B.T.; writing—original draft, M.Y.; writing—review and editing, J.H. (Jie Hu), B.T. and J.H. (Jianjun Hu); visualization, M.Y., B.T. and J.H. (Jianjun Hu); supervision, J.H. (Jie Hu). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guizhou Provincial Science and Technology Fund (Qian Kehe Basic-ZK [2021] General 337), supported by the Fund of the State Key Laboratory of Public Big Data, Guizhou University (No. PBD2023-35), and the Graduate Program of Guizhou University of Finance and Economics (2022ZXS036).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in the experiments were based on the two publicly available Twitter15 and Twitter16 datasets released by Ma et al. [38]. The raw datasets can be respectively downloaded from <https://www.dropbox.com/s/7ewzdrbelpmrnxu/rumdetect2017.zip?dl=0> (accessed on 1 May 2023).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. DataReportal. Digital 2023: Global Overview Report. Available online: <https://datareportal.com/reports/digital-2023-global-overview-report> (accessed on 10 December 2023).
2. Li, Q.; Cheng, L.; Wang, W.; Li, X.; Li, S.; Zhu, P. Influence maximization through exploring structural information. *Appl. Math. Comput.* **2023**, *442*, 127721.
3. Vosoughi, S.; Roy, D.; Aral, S. The spread of true and false news online. *Science* **2018**, *359*, 1146–1151. [PubMed]
4. Tan, L.; Wang, G.; Jia, F.; Lian, X. Research status of deep learning methods for rumor detection. *Multimed. Tools Appl.* **2023**, *82*, 2941–2982. [PubMed]
5. Castillo, C.; Mendoza, M.; Poblete, B. Predicting information credibility in time-sensitive social media. *Internet Res.* **2013**, *23*, 560–588.
6. Yang, F.; Liu, Y.; Yu, X.; Yang, M. Automatic detection of rumor on sina weibo. In Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, Beijing China, 12–16 August 2012; pp. 1–7.
7. Zhao, Z.; Resnick, P.; Mei, Q. Enquiring minds: Early detection of rumors in social media from enquiry posts. In Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 18–22 May 2015; pp. 1395–1405.
8. Ma, J.; Gao, W.; Mitra, P.; Kwon, S.; Jansen, B.J.; Wong, K.-F.; Cha, M. Detecting rumors from microblogs with recurrent neural networks. In Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016), New York, NY, USA, 9–15 July 2016.

9. Kotteti, C.M.M.; Dong, X.; Qian, L. Ensemble deep learning on time-series representation of tweets for rumor detection in social media. *Appl. Sci.* **2020**, *10*, 7541. [\[CrossRef\]](#)
10. Khoo, L.M.S.; Chieu, H.L.; Qian, Z.; Jiang, J. Interpretable rumor detection in microblogs by attending to user interactions. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 8783–8790.
11. Bian, T.; Xiao, X.; Xu, T.; Zhao, P.; Huang, W.; Rong, Y.; Huang, J. Rumor detection on social media with bi-directional graph convolutional networks. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 549–556.
12. Rahimi, M.; Roayaei, M. A multi-view rumor detection framework using dynamic propagation structure, interaction network, and content. *IEEE Trans. Signal Inf. Process. Netw.* **2024**, *10*, 48–58.
13. Li, J.; Li, W.; Gao, F.; Cai, M.; Zhang, Z.; Liu, X.; Wang, W. Social contagions on higher-order community networks. *Appl. Math. Comput.* **2024**, *478*, 128832.
14. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
15. Tschitschek, S.; Singla, A.; Gomez Rodriguez, M.; Merchant, A.; Krause, A. Fake news detection in social networks via crowd signals. In Proceedings of the Companion Proceedings of the the Web Conference 2018, Lyon, France, 23–27 April 2018; pp. 517–524.
16. Hussain, M.G.; Hasan, M.R.; Rahman, M.; Protim, J.; Al Hasan, S. Detection of bangla fake news using mnb and svm classifier. In Proceedings of the 2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE), Bangalore, India, 2–4 July 2020; pp. 81–85.
17. Antony Vijay, J.; Anwar Basha, H.; Arun Nehru, J. A dynamic approach for detecting the fake news using random forest classifier and NLP. In *Computational Methods and Data Engineering: Proceedings of ICMDE 2020*; Springer: Berlin/Heidelberg, Germany, 2020; Volume 2, pp. 331–341.
18. Choudhury, D.; Acharjee, T. A novel approach to fake news detection in social networks using genetic algorithm applying machine learning classifiers. *Multimed. Tools Appl.* **2023**, *82*, 9029–9045.
19. Nasir, J.A.; Khan, O.S.; Varlamis, I. Fake news detection: A hybrid CNN-RNN based deep learning approach. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100007.
20. Shelke, S.; Attar, V. Rumor detection in social network based on user, content and lexical features. *Multimed. Tools Appl.* **2022**, *81*, 17347–17368. [\[PubMed\]](#)
21. Yang, J.; Pan, Y. COVID-19 Rumor Detection on Social Networks Based on Content Information and User Response. *Front. Phys.* **2021**, *9*, 763081.
22. Li, Y.C.; Qian, L.F.; Ma, J. Early Detection of Micro Blog Rumors Based on BERT-RCNN Model. *Inf. Stud. Theory Appl.* **2021**, *44*, 173–177.
23. Feng, L.Z.; Liu, F.R.; Wang, Y.W. Rumor detection method based on graph convolution network and AttentionMechanism. *Data Anal. Knowl. Discov.* **2024**, *8*, 125–136.
24. Chen, Y.; Sui, J.; Hu, L.; Gong, W. Attention-residual network with CNN for rumor detection. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 1121–1130.
25. Weng, C.-H.; Lin, K.-C.; Ying, J.-C. Detection of chinese deceptive reviews based on pre-trained language model. *Appl. Sci.* **2022**, *12*, 3338. [\[CrossRef\]](#)
26. Lai, J.; Yang, X.; Luo, W.; Zhou, L.; Li, L.; Wang, Y.; Shi, X. RumorLLM: A Rumor Large Language Model-Based Fake-News-Detection Data-Augmentation Approach. *Appl. Sci.* **2024**, *14*, 3532. [\[CrossRef\]](#)
27. Salini, Y.; Harikiran, J. Multiplicative vector fusion model for detecting deepfake news in social media. *Appl. Sci.* **2023**, *13*, 4207. [\[CrossRef\]](#)
28. Chen, T.; Chen, H.; Li, X. Rumor detection via recurrent neural networks: A case study on adaptivity with varied data compositions. In Proceedings of the Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2018 Workshops, BDASC, BDM, ML4Cyber, PAISI, DaMEMO, Revised Selected Papers 22, Melbourne, VIC, Australia, 3 June 2018; pp. 121–127.
29. Feng, W.; Li, Y.; Li, B.; Jia, Z.; Chu, Z. BiMGCL: Rumor detection via bi-directional multi-level graph contrastive learning. *PeerJ Comput. Sci.* **2023**, *9*, e1659.
30. Hu, D.; Wei, L.; Zhou, W.; Huai, X.; Han, J.; Hu, S. A Rumor Detection Approach Based on Multi-Relational Propagation Tree. *J. Comput. Res. Dev.* **2021**, *58*, 1395–1411.
31. Qiang, Z.; Gu, Y. Rumor Detection Model Based on Dynamic Propagation and Community Structure. *Comput. Eng. Appl.* **2024**, *60*, 198.
32. Lin, H.; Yi, P.; Ma, J.; Jiang, H.; Luo, Z.; Shi, S.; Liu, R. Zero-shot rumor detection with propagation structure via prompt learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; pp. 5213–5221.
33. Lu, Y.-J.; Li, C.-T. GCAN: Graph-aware co-attention networks for explainable fake news detection on social media. *arXiv* **2020**, arXiv:2004.11648.

34. Huang, Q.; Yu, J.; Wu, J.; Wang, B. Heterogeneous graph attention networks for early detection of rumors on twitter. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8.
35. Tao, X.; Wang, L.; Liu, Q.; Wu, S.; Wang, L. Semantic evolvement enhanced graph autoencoder for rumor detection. In Proceedings of the ACM Web Conference 2024, Singapore, 13–17 May 2024; pp. 4150–4159.
36. Chang, Q.; Li, X.; Duan, Z. A novel approach for rumor detection in social platforms: Memory-augmented transformer with graph convolutional networks. *Knowl.-Based Syst.* **2024**, *292*, 111625.
37. Li, Q.; Han, Z.; Wu, X.-M. Deeper insights into graph convolutional networks for semi-supervised learning. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.
38. Ma, J.; Gao, W.; Wong, K.-F. Detect rumors in microblog posts using propagation structure via kernel learning. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), Vancouver, BC, Canada, 30 July–4 August 2017.
39. Zubiaga, A.; Liakata, M.; Procter, R. Exploiting context for rumour detection in social media. In Proceedings of the Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, Part I 9, 13–15 September 2017; Proceedings; pp. 109–123.
40. Derczynski, L.; Bontcheva, K.; Liakata, M.; Procter, R.; Hoi, G.W.S.; Zubiaga, A. SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours. *arXiv* **2017**, arXiv:1704.05972.
41. Castillo, C.; Mendoza, M.; Poblete, B. Information credibility on twitter. In Proceedings of the 20th International Conference on World Wide Web, Hyderabad, India, 28 March–1 April 2011; pp. 675–684.
42. Ma, J.; Gao, W.; Wong, K.-F. Rumor detection on twitter with tree-structured recursive neural networks. In Proceedings of the 2018 56th Annual Meeting of the Association for Computational Linguistics (ACL), Melbourne, Australia, 26–28 March 2018.
43. Liu, Y.; Wu, Y.-F. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.
44. Cai, X.; Tohti, T.; Hamdulla, A. A Rumor Detection Method Incorporating Correlation Features. In Proceedings of the 2022 3rd International Conference on Pattern Recognition and Machine Learning (PRML), Chengdu, China, 22–24 July 2022; pp. 328–333.
45. Xu, S.; Liu, X.; Ma, K.; Dong, F.; Riskhan, B.; Xiang, S.; Bing, C. Rumor detection on social media using hierarchically aggregated feature via graph neural networks. *Appl. Intell.* **2023**, *53*, 3136–3149.
46. Kochkina, E.; Liakata, M.; Zubiaga, A. All-in-one: Multi-task learning for rumour verification. *arXiv* **2018**, arXiv:1806.03713.
47. Askarizade, M. Enhancing rumor detection with data augmentation and generative pre-trained transformer. *Expert Syst. Appl.* **2025**, *262*, 125649.
48. Vu, D.T.; Jung, J.J. Rumor detection by propagation embedding based on graph convolutional network. *Int. J. Comput. Intell. Syst.* **2021**, *14*, 1053–1065.
49. Ma, J.; Li, J.; Gao, W.; Yang, Y.; Wong, K.-F. Improving rumor detection by promoting information campaigns with transformer-based generative adversarial learning. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 2657–2670.
50. Yu, J.; Jiang, J.; Khoo, L.M.S.; Chieu, H.L.; Xia, R. Coupled hierarchical transformer for stance-aware rumor verification in social media conversations. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, Virtual Conference, 16–20 November 2020.
51. Luo, N.; Xie, D.; Mo, Y.; Li, F.; Teng, C.; Ji, D. Joint rumour and stance identification based on semantic and structural information in social networks. *Appl. Intell.* **2024**, *54*, 264–282.
52. Cui, C.; Jia, C. Propagation Tree Is Not Deep: Adaptive Graph Contrastive Learning Approach for Rumor Detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; pp. 73–81.
53. Juul, J.L.; Ugander, J. Comparing information diffusion mechanisms by matching on cascade size. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2100786118.
54. Hou, D.; Yin, S.; Gao, C.; Li, X.; Wang, Z. Propagation Dynamics of Rumor vs. Non-rumor across Multiple Social Media Platforms Driven by User Characteristics. *arXiv* **2024**, arXiv:2401.17840.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.