ECAI 2024
U. Endriss et al. (Eds.)

doi:10.3233/FAIA240932

Global Structural-Temporal Graph Network with Public Opinion for Online Rumor Detection

Tiening Sun[a], Zhong Qian[a], Peifeng Li[a] and Qiaoming Zhu[a,*]

[a]Soochow University

Abstract. Rumors on social media can spread rapidly and widely with the help of the Internet characteristics, causing serious negative impacts on social stability and public life. In order to distinguish rumors from non-rumors, most of the existing methods are based on neural units to encode and observe the content of claims, user comments and rumor propagation patterns. However, these methods only consider the event context information in a single conversation thread, ignoring the public opinion (global contextual information) corresponding to the event in the external news environment. Be aware that users are easily distracted by opinion leaders to false facts and induced to make supportive replies on false claims. In order to address the above-mentioned limitation, we propose a Global Structural-Temporal Graph Network (GSTGN) framework. Specifically, we first construct a multi-modal global opinion graph based on the conversation threads belonging to the same event to capture the external public opinion of the target event. Then to enhance representation learning, we design a Structural-Temporal (ST) unit to encode structural and temporal features of the local conversation graph, and utilize the structural feature of the local graph to guide the learning and encoding of the global opinion graph. Experimental results on two public benchmark datasets prove that our GSTGN method achieves better results than other state-of-the-art models.

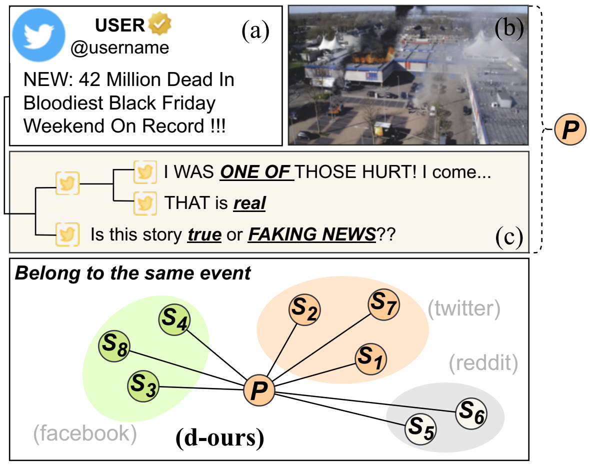

Figure 1. Existing models mainly consider the following features: (a) text, (b) multi-modality, and (c) users' local opinion. But our method exploits (d) global public opinion. P is the conversation thread of the target event containing multi-modal information, and S is the ones with the same event as P in the external information environment.

1 Introduction

Online rumors, leveraging the Internet as their primary medium, exhibit distinct characteristics of digitization, large-scale dissemination, and globalization. These attributes enable them to significantly surpass traditional interpersonal communication methods, thereby exerting profound impacts on various fields including politics [1], finance [7], and public safety [19]. In the context of the new media era, the development of automatic rumor detection technology is imminent.

Rumor detection has evolved from early manual tracking and feature engineering to deep learning-based methods. Most deep learning-based rumor detection frameworks mainly rely on advanced neural network tools to mine and analyze rumor text features [16, 34, 18] (Figure 1(a)). With the development of multimedia, rumor posts have changed from a single plain text to a multi-modal form consisting of text, images and even videos, which also promotes the development of multi-modal rumor detection models such as EANN [31] and MCAN [32] (Figure 1(b)). Recently, propagation-based methods have emerged that consider not only multi-modal content information (including social contexts) but also structural features of rumor propagation [28, 36, 27] (Figure 1(c)). However, these methods only observe the event context information covered in a single conversation thread, ignoring the public opinion (global contextual information) corresponding to the event in the external information environment. More specifically, ordinary users may be emotionally manipulated by persuasive misinformation, or misled by opinion leaders, which may lead them to post some supportive comments. And these comments supporting false claims can confuse detection models and cause misclassification. The introduction of global public opinion can reduce the misclassification of graph-based models (such as BiGCN [2]) caused by partial comments that deviate from the facts, thereby improving the generalization and inference accuracy of rumor detection models.

For a better understanding, we give a case, as shown in Figure 2. It can be observed that users in the left conversation thread blindly trust and support the opinion leader's post. But the views of these users are partial. For this target event, we get the overall perspective and find that many users have given clear objections and explanations about the rumor that "COVID-19 is AIDS", which can be used as key classification clues, as shown in the Figure 2(right)[1].

* Corresponding Author. Email: qmzhu@suda.edu.cn

[1] Note that due to limited space, we only show a set of external opinions on the same event in Figure 2, and in the experiment we will select the top-k ones.

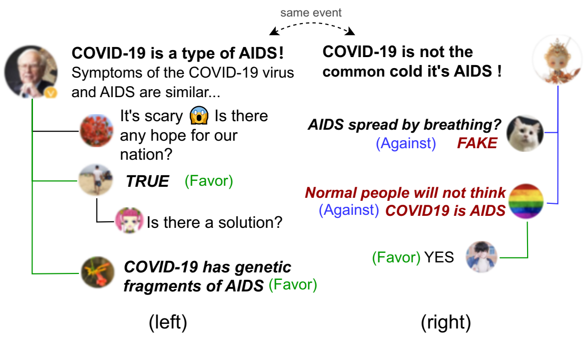

Figure 2. A real case of rumor event. The conversation thread on the (left) is the target event to be detected, and the conversation thread on the (right) is the set of external opinions belonging to the same event. $Favor$ represents the user's support position, and $Against$ represents the user's disagreement position.

Unfortunately, existing methods ignore the global public opinion of the target event. To overcome this limitation, we propose a Global Structural-Temporal Graph Network (GSTGN) method. Specifically, given the target event to be detected, we first capture its corresponding local opinion and global public opinion, and construct the local and global graphs based on their natural thread structure respectively, which can provide more richer classification clues. Next, we design a Structural-Temporal (ST) module to extract and encode the structural and temporal features of the local opinion graph, and utilize the structural features of the local graph as a guide for participating in the learning and encoding of the global opinion graph. Finally, we combine the local and global opinion features for online rumor detection.

Our contributions are summarized as follows:

- **Idea**: To the best of our knowledge, this is the first study to introduce global opinion summarization in the rumor detection task, which aims to prevent the model from being misled by local misinformation. Furthermore, this concept can also be successfully applied to other detection models.
- **Method**: We propose the GSTGN framework, which comprehensively considers local opinion and global public opinion of the given conversation thread, and exploits the structural-temporal features of propagation paths for rumor classification.
- **Experiments**: We experimentally demonstrate that our model outperforms state-of-the-art baselines on two real-world datasets.

2 Related Work

In the context of the new media era, the characteristics of the Internet make rumors spread wantonly, which also prompts more and more researchers to devote themselves to the analysis and research of rumor identification. Scholars have used different methods to study rumors from different angles. In the early days, researchers mainly relied on manual tracking, screening, and fact-checking, and built rumor-dispelling websites such as snopes.com.

However, such methods are quickly eliminated due to low efficiency in the face of large-scale data. In order to speed up the process of automatic rumor recognition research, methods based on feature engineering and machine learning have become popular, among which models such as support vector machine [33] and random forest [13] were widely used. Although these machine learning-based methods have improved efficiency compared with traditional manual screening, there are also some common problems, such as the complexity of feature engineering, data sparsity, and noise interference. With the development of deep learning technology, some new ideas and progress have emerged in subsequent rumor detection methods, such as attention mechanisms [30], graph neural networks [11], and pre-trained models [4].

The works related to rumor detection based on deep learning can be roughly divided from four perspectives: text, multi-modality, propagation structure, and user information. The number of works around rumor texts is the largest, and the research and analysis are relatively mature. For example, Ma et al. [16] and Dun et al. [6] utilize RNN and BERT for post encoding, respectively. Sheng et al. [24] introduce news item to assist the model. With the maturity of the language pre-training model, the work of researchers has also changed from generating better text semantic features to integrating multi-modal information, including learning image-text interaction features [31, 10, 22, 3, 29], and mining the inconsistency of image-text [26, 21]. At the same time, the structural information of rumor propagation path has also been valued recently. Bian et al. [2] utilize GCN to encode conversational threads of rumors. Sun et al. [28] employ graph contrastive learning to analyze the difference in rumor and non-rumor propagation structures. In addition, user characteristics are often incorporated into the model framework from different perspectives [35, 14, 8].

It should be noted that although the GLAN method proposed by Yuan et al. [35] also considers global information, it starts from the user's perspective and mines potential tags of target source posts based on the concept of shared neighbors. However, our method takes an event-centric method, providing comprehensive clues to the model by integrating conversation structures associated with the same event.

3 Problem Definition

Rumor detection is usually defined as a classification task whose purpose is to learn a classifier from a set of labeled training events (where each event contains multi-modal features , context and propagation structure), and then use it to predict the label of the test events in this paper. Specifically, we represent the event set as $P = \{p_1, p_2, ..., p_n\}$, where p_i is i-th event and n is the number of events. Each event $p = (y, G)$ consists of the ground-truth label $y \in \{R, N\}$ of the event (i.e. Rumor or Non-rumor) and the graph $G = (V, E)$ referring to local or global propagation structure, where V is the set of graph nodes and E is the set of edges. In some cases, rumor detection is defined as a four-class classification task, correspondingly, $y \in \{N, F, T, U\}$ (i.e., Non-rumor, False Rumor, True Rumor, and Unverified Rumor). Our goal is to learn a model $f(\cdot)$ to classify the event p_i into the ground-truth label y.

4 Method

In this section, we propose a supervised GSTGN framework for rumor classification tasks, which aims to empower the model to mine global contextual information, thereby improving detection accuracy. As shown in Figure 3, the architecture of GSTGN mainly consists of three modules: a) *Local opinion graph encoding*. We first construct the conversation thread graph for the target event p. Then we design the ST encoder to capture the spatial-temporal features of the local

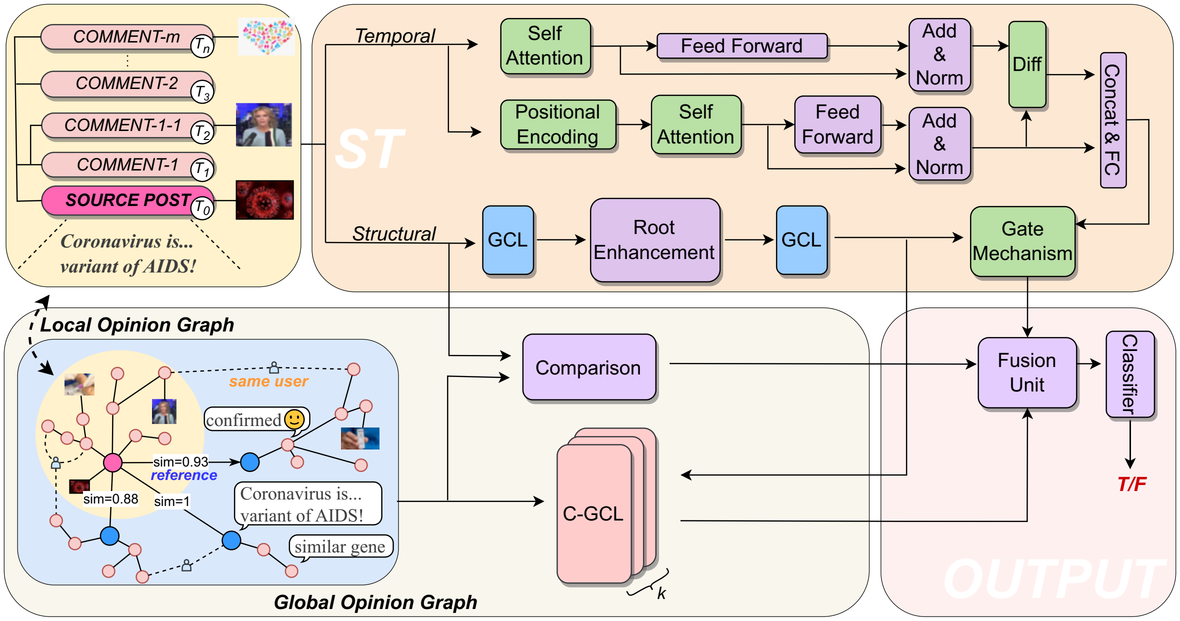

Figure 3. Overview of our GSTGN rumor detection model.

graph and obtain the local vector V^l. b) *Global opinion graph encoding*. From the perspective of clustering the same events to enrich classification clues, we construct a global opinion graph. At the same time, we utilize the spatial features of the local graph to guide the encoding process of the global graph and obtain the global vector V^g. c) *Rumor classification*. The $Comparison$ layer is designed to mine the difference between the local opinion and the global opinion and obtain the vector V^c. Finally V^l, V^g and V^c are fused for rumor classification.

4.1 Local Opinion Graph Encoding

Conversation threads can capture multi-modal contextual information related to rumors and network topology information between participants, which is crucial for judging the authenticity of claims. At the same time, conversation threads can maintain the temporal relationship between claims and responses, helping the model understand how information spreads and evolves. Below we will first introduce the construction of local conversation threads, and then describe the ST module used to encode structural features and temporal features.

4.1.1 Local Opinion Graph Construction

Utilizing the natural tree-like structure of conversation threads, we initially establish connections between the source post and its associated comments. Subsequently, we connect the event header image with the text of the source post. For other comments containing images, we connect the comment text with their respective images. Note that in cases where comments consist solely of images, we directly connect the images with its parent node. This process ultimately results in the formation of a multi-modal tree-like graph, referred to as the local opinion graph G^l, with the adjacency matrix A^l and the node feature matrix X^l, as depicted in Figure 3. For the nodes of the text attribute in the graph, we employ the Bidirectional Encoder Representations from Transformers (BERT) [20] which has been successfully applied in fields such as classification [25], translation, etc., to separately encode the source and comments. For the nodes of image attribute, we use the pre-trained Visual Transformer (ViT) [5] with a similar architecture to BERT to generate their representations.

4.1.2 Local Structural Feature Encoding

Rumor conversation threads inherently contain both structural and temporal information. It is evident that a single feature extractor, such as RNN or CNN, cannot effectively capture both of these features simultaneously. Therefore, we design the Structural-Temporal (ST) encoder, as illustrated in Figure 3, to learn the structural-temporal feature of the local opinion graph G^l. As can be seen in Figure 3, to capture structural features of the local graph, we employ the Graph Convolutional Layers (GCL) [11]. Meanwhile in order to strengthen the importance of source posts, we introduce the Root Enhancement (RE) [2]. Specifically, we first feed the A^l and X^l of G^l to GCL, which can be formulated as follows.

$$H^l_{(1)} = \sigma(\hat{A}^l X^l W^l_{(0)}), \tag{1}$$

where $H^l_{(1)}$ is the output of the GCL, and σ is an activation function such as the ReLU function. $\hat{A}^l = D^{-\frac{1}{2}} A^l D^{-\frac{1}{2}}$ is the normalized adjacency matrix, where D is the degree matrix of A^l. $W^l_{(0)}$ is the trainable parameter matrix. We use RE to strengthen the influence of the target source post in the graph, that is, $\tilde{H}^l_{(1)} = concat(H^l_{(1)}, X^{l,root})$, where $X^{l,root}$ is the feature vector of the source post (root node). Then, we add a new GCL to aggregate node features as follow.

$$H^l_{(2)} = \sigma(\hat{A}^l \tilde{H}^l_{(1)} W^l_{(1)}), \tag{2}$$

where $H^l_{(2)}$ is the output of the second GCL, and $W^l_{(1)}$ is the parameter matrix.

4.1.3 Local Temporal Feature Encoding

Ma et al. [16] are the first to use RNN to learn the temporal relationship between posts, but this method ignores the structural-temporal relationship interaction. At the same time, due to the lack of parallelism, the computational efficiency is reduced. In order to better capture temporal features, we introduce the attention mechanism, and also design a $Diff$ layer and a gate unit in the ST module. Specifically, we first rearrange the node features in X^l in chronological order and form a new time-series-related feature matrix $X^{l,time}$. Then we use the $SinusoidalFunctions$ [30] in the transformer to fuse the order information into $X^{l,time}$ and obtain $T^l = Positional - Encoding(X^{l,time})$. Next, we use self-attention to process the input T^l as follows.

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V, \tag{3}$$

where $Q = K = V = T^l$, and d_k is the their dimension. In addition, the formalization of multi-head self-attention is $MultiHead(Q,K,V) = Concat(Attention_1, ..., Attention_h)W^q$, where h is the number of heads and W^q is the parameter matrix. Subsequently, T^l is converted into $\hat{T}^{l,FFN}$ through the Feed Forward network (FFN) and Normalization (Norm) layers, which is formulated as

$$T^{l,FFN} = max(0, T^l W_1^{FFN} + b_1)W_2^{FFN}, \tag{4}$$

$$\hat{T}^{l,FFN} = Norm(T^{l,FFN} + T^l), \tag{5}$$

where W^{FFN} and b are the weight matrix and bias respectively. We take $X^{l,time}$ as input and obtain $\hat{R}^{l,FFN}$ without temporal information only through Eq. (3-5). To capture the change of features under known temporal priors, a simple $Diff$ unit is designed shown in Eq. 6. Then the obtained new vector $T^{l,diff}$ and $\hat{T}^{l,FFN}$ are concatenated, and fed into a projection (linear) layer and a gate unit. The calculation process is as follows.

$$T^{l,diff} = \hat{T}^{l,FFN} - \hat{R}^{l,FFN}, \tag{6}$$

$$T^{l,P} = Linear(Concat(\hat{T}^{l,FFN}, T^{l,diff})), \tag{7}$$

$$V^l = gate\_value \cdot H^l_{(2)} + (1 - gate\_value) \cdot T^{l,P}, \tag{8}$$

where the $gate\_value$ can be calculated as $sigmoid(W_{gate} \cdot [H^l_{(2)}, T^{l,P}])$.

4.2 Global Opinion Graph Encoding

Modeling local conversational threads of target events is helpful for rumor detection models [2, 28], but its inherent flaw is that it relies too heavily on the authenticity of comments posted by users in these threads. Ordinary participants can easily be misled by the source posters. And supportive comments made by these participants can seriously interfere with the model. Therefore, we should not only pay attention to local voices, but also zoom out to observe a broader and objective external information environment, and introduce global public views on target events. In the following, we will introduce in detail how to construct and encode the global opinion graph.

4.2.1 Global Opinion Graph Construction

Hot events will be widely quoted and disseminated on the Internet. In order to capture the public opinion and propagation path structure of the target event p in the global information environment, we design the following rules: a) Given the target event, the local opinion graph construction is completed first. b) Connect the target source post in the local graph with the source ones with the same event. In order to achieve this goal, we assume that each source post represents a distinct event. We utilize BERT to encode event sentences and then employ cosine similarity to obtain the similarity between the target event sentence and potential event sentences. Sources with cosine similarity values greater than 0.88 (fine-tuning on development set) are retained. At the same time, the similarity values between source posts are used as the edge weights. It is worth noting that we had previously applied the whitening operation to enhance the isotropy of the data, but testing results indicated no significant improvement. Therefore, for a more simple design, this paper directly employs the original BERT embeddings. c) On the basis of b), connect the comments made by the same user. This is mainly because rumor producers may use robot accounts to comment on related rumor posts on a large scale in order to support the source post and mislead the public, so constructing the relationship between comments and users may help to identify this pattern. d) On the basis of c), add directed edges between source posts with reference relationship. The constructed global opinion graph (shown in Figure 3) is defined as G^g with the adjacency matrix A^g and the node feature matrix X^g, where the node features are also represented by BERT or ViT.

4.2.2 Global Structural Feature Encoding

Directly using GCL to encode the global opinion graph is a straightforward strategy. However, the GCL treats every node in the global graph as equally important, implying that during training, both the target conversation thread and other threads related to the same event are given the same weight. Therefore, we design the Cross-GCL (C-GCL) layer, which utilizes the structural features of the local graph to guide the learning and encoding of the global opinion graph. It can be formulated as follows.

$$V^g = sigmoid(W_v H^l_{(2)}) \cdot relu(\hat{A}^g X^g W^g), \tag{9}$$

where $\hat{A}^g = D^{-\frac{1}{2}} A^g D^{-\frac{1}{2}}$ is the normalized adjacency matrix, where D is the degree matrix of A^g. W is the trainable parameter matrix. Now we can obtain the corresponding the graph representation V^g of the public opinion.

4.3 Rumor Classification

If there is a large difference between the local opinion and the global public opinion of the target event, then the local opinion can be regarded as an outlier, which may be one of the important clues to identify the rumor pattern. We therefore design a $Comparison$ layer to learn the interaction of local and global opinions, which is formalized as follows.

$$V^c = MLP((X^g \odot X^l) \oplus (X^g - X^l)). \tag{10}$$

Now, we have obtained the local structural-temporal feature V^l, the global feature V^g and the comparison feature V^c. Then, we concatenate them to merge the information and obtain $h^o = concat(V^l, V^g, V^c)$. Next, h^o is fed into the full-connection layer and a softmax layer, and the output is calculated as follows.

Table 1. Statistics of the datasets

Statistic	Twitter	PHEME
# posts	47,832	99,013
# users	41,947	47,531
# source posts	1818	6425
# images	5585	7239
# non-rumors	555	4023
# false rumors	446	2402
# unverified rumors	322	-
# true rumors	495	-

$$\hat{y} = softmax(W^F h^o + b^F), \tag{11}$$

where $\hat{y}$ is the predicted probability distribution. W^F and b^F are the trainable weight matrix and bias respectively. Finally, we use stochastic gradient descent to minimize the cross-entropy loss for model training.

5 Experiments

5.1 Datasets

We evaluate the proposed GSTGN model on two real-world datasets: PHEME [2] [12] and Twitter [3] [17]. The PHEME dataset consists of nine topics, each of which contains a number of subdivided events. In addition, the PHEME contains only two types of tags: Rumor (R) and Non-Rumor (N), which is used for the binary classification of rumors and non-rumors. The Twitter contain four tags: Non-rumor (N), False Rumor (F), True Rumor (T), and Unverified Rumor (U), which are used for quaternary classification. We collected their corresponding propagation threads and images according to the source posts in Twitter. Detailed statistics with invalid data removed are shown in Table 1.

5.2 Experimental Settings

We make comparisons with the following state-of-the-art baselines:

- **SVM-TS** [15] is a linear SVM classifier for rumor detection.
- **GRU-2** [16] is a RNN-based model that can learn the temporal relationship between rumor posts.
- **EANN**[4] [31] uses VGG-19 and Text-CNN to encode visual and text information respectively.
- **UDGCN** [2] directly employs GCN for rumor detection, in which the root feature enhancement strategy is used.
- **BiGCN**[5] [2] is a GCN-based method that considers bottom-up and top-down structure information of rumor trees.
- **GLAN** [35] designs an attention network to jointly encode local semantic and global structural information.

[2] https://figshare.com/search?q=pheme
[3] The Twitter dataset is an integrated compilation built upon the Twitter15/16 datasets with enhancements in visual and structural information. Download from https://www.dropbox.com/s/7ewzdrbelpmrnxu/.
[4] https://github.com/yaqingwang/EANN-KDD18
[5] https://github.com/TianBian95/BiGCN

Table 2. Rumor detection results on PHEME dataset

Method	Class	Acc.	Prec.	Rec.	F_1
SVM-TS	R	0.685	0.553	0.539	0.539
	N		0.758	0.762	0.757
GRU-2	R	0.791	0.752	0.659	0.692
	N		0.813	0.867	0.835
EANN	R	0.806	0.739	0.685	0.702
	N		0.829	0.863	0.842
UDGCN	R	0.816	0.748	0.771	0.750
	N		0.867	0.835	0.847
BiGCN	R	0.817	0.754	0.758	0.751
	N		0.854	0.854	0.851
GLAN	R	0.828	0.770	0.776	0.765
	N		0.843	0.844	0.841
HMCAN	R	0.837	0.775	0.745	0.748
	N		0.866	0.878	0.868
GACL	R	0.841	0.768	0.764	0.762
	N		0.873	**0.884**	0.878
GSTGN	R	**0.856**	**0.796**	**0.813**	**0.798**
	N		**0.890**	0.883	**0.884**

- **HMCAN**[6] [22] is a attention-based method that can learn multi-modal high-order complementary information and hierarchical semantics of text.
- **GACL**[7][28] is a GCN-based method that designs a graph adversarial contrastive learning algorithm to capture the differences in rumor patterns.

The proposed GSTGN method is implemented by PyTorch [9]. We split the datasets for training, validation, and testing with a ratio of 6:2:2. Meanwhile, the learning rate is initialized to 0.0002, and the batch size is set to 32. To prevent overfitting, the early stopping strategy is introduced. Then, the Adam is adopted to optimize our objective function. Finally, the Accuracy (Acc.), Precision (Prec.), Recall (Rec.) and F_1-measure (F_1 [8]) are used as evaluation metrics in this paper.

5.3 Results

Tables 2 and 3 show the performance of all comparison methods on the two datasets. Our GSTGN is superior to the other baselines and the improvement is significant (p<0.03). In addition, we also have the following noteworthy observations:

1) The outdated SVM-TS model gets the worst results, and the performance of GRU-2 and EANN is close. HMCAN is an improved version of EANN with better performance. Similar to GSTGN, both UDGCN and BiGCN are GCN-based methods. The comprehensive performance of these two models is comparable to HMCAN. Their algorithmic core primarily involves aggregating features of neighboring nodes within local conversation threads to explore the interaction between source posts and comments. Although GLAN's approach is promising, its performance lags behind GSTGN due to the use of relatively outdated encoders for semantic extraction and structural encoding. It is also important to note that while GLAN employs a "global strategy", it is fundamentally different from our GSTGN. GLAN's approach focuses on clustering other conversation threads from the perspective of

[6] https://github.com/wangjinguang502/HMCAN
[7] https://github.com/agangbe/GACL
[8] $avgF_1$ represents the average of F_1 values of all classes.

Table 3. Rumor detection results on Twitter dataset

Method	Acc.	N F_1	F F_1	T F_1	U F_1
SVM-TS	0.659	0.746	0.400	0.669	0.567
GRU-2	0.776	0.865	0.679	0.756	0.731
EANN	0.781	0.875	0.692	0.763	0.745
UDGCN	0.828	0.858	0.773	0.840	0.757
BiGCN	0.830	0.902	0.759	0.832	0.755
GLAN	0.823	0.892	0.751	0.834	0.739
HMCAN	0.826	0.881	0.766	0.830	**0.791**
GACL	0.839	**0.963**	0.779	0.813	0.775
GSTGN	**0.857**	0.925	**0.782**	**0.862**	0.775

Table 4. Results of ablation study on the PHEME and Twitter

Model	PHEME		Twitter	
	Acc.	$avgF_1$	Acc.	$avgF_1$
GSTGN	**0.856**	**0.841**	**0.857**	**0.836**
w/o G	0.838	0.820	0.844	0.811
w/o ST	0.816	0.793	0.821	0.806
w/o C	0.842	0.825	0.846	0.835
w/o I	0.851	0.830	0.849	0.815

users. The motivation is that even if two tweets are unrelated in content, they might have the same label if they share similar commenting users. Clearly, this approach is different from our objective, despite both methods referencing the term "global".

2) The GSTGN shows a modest 1.5% improvement in accuracy over the state-of-the-art GACL on the PHEME dataset. But our GSTGN has a notable improvement of identifying rumors (R), approximately a 3.6% improvement in F_1 score, which may have more practical implications. The reason for this phenomenon is mainly due to the greater number of non-rumor samples in the PHEME dataset compared to the number of rumor samples. GACL is better suited for non-rumor features and neglects to mine rumor features during the training process. However, our GSTGN complements classification clues by incorporating information from other conversation threads, mitigating the misclassification of rumors as non-rumors. It is also worth mentioning that the target rumor event to be detected may already exist in our training database. Aggregating conversation thread information that belongs to the same event as the target can assist the model in making classification judgments quickly. This is one of the reasons why GSTGN can improve the rumor F_1. The phenomenon of old rumors being spread newly occurs from time to time, and Sheng et al. [23] are also actively studying this subdivision.

3) Our GSTGN lags behind in N_F_1 and U_F_1 metrics compared to GACL and HMCAN on the Twitter dataset . This may be because when GSTGN tests N/U class instances, noise is mixed in by aggregating conversation thread information from other classes. This observation motivates our commitment to exploring more effective event clustering methods in the future. However, it's worth noting that our GSTGN model exhibits a significant overall accuracy improvement compared to GACL and HMCAN, especially surpassing the GACL method by 5% and the HMCAN method by 3% in the T_F_1 metric.

5.4 Ablation Study

To verify the effectiveness of the different modules of GSTGN, we compare it with the following variants:

- **w/o G**: We do not consider global public opinions, and only adopts conversational threads composed of local comments for training and prediction.
- **w/o ST**: We remove the ST encoder module and only use the original GCN to encode local and global opinion graphs.
- **w/o C**: We remove the $Comparison$ unit, making the model unable to capture the bias between the local opinion and the global opinion of the target event.
- **w/o I**: Image features are not considered.

The experimental results are shown in Table 4. It can be observed: a) Every module of GSTGN is effective, among which the ST encoder and public opinion features have the greatest impact on the model performance. In particular, public opinion can provide more classification cues for the model (more detailed discussion in Section 5.7). b) Visual features can sometimes provide classification cues for models. For instance, false news stories often tend to use visually striking (sensational) images to enhance the spread of rumors. c) The global public opinion feature contributes more to the accuracy improvement on the PHEME dataset than on Twitter, which may be because the event topics in PHEME are more focused, similar events are more abundant, and the annotation quality of PHEME is higher.

5.5 Idea Transplantation

Global public opinion can not only be used in GSTGN, but also can be transplanted into other methods to help the model understand more objective public attitudes. Taking the widely cited benchmark BiGCN as an example, we introduce the public opinion into the model. Specifically, we keep the top-down and bottom-up structure of BiGCN unchanged, and then cluster other posts with the same event as the target source post. Subsequently, the top-down and bottom-up global propagation graphs with public opinion are constructed for BiGCN in a manner similar to GSTGN, as shown in Figure 4(a). For more structural details of BiGCN, please refer to [2]. The experimental results are shown in Figure 4(b). Obviously, compared with the plain BiGCN, the average accuracy of BiGCN with public opinion on the two datasets is improved by 1.7%. In addition, the way of introducing public opinion is not limited to constructing the global public opinion graph (like GSTGN), and other more efficient transplantation strategies have yet to be developed.

5.6 Early Rumor Detection

The sooner rumors are detected, the more timely the spread of harm can be curbed. Our GSTGN is good at early rumor detection tasks, because GSTGN can aggregate conversation threads belonging to the same event before a specified time point to assist in detection. For example, assuming that the authenticity tag of $claimA$ already exists in the database, when we detect $claimB$ which belongs to the same event as $claimA$, the GSTGN method will automatically connect the previously fact-checked $claimA$. In this way, the model will easily determine the label of $claimB$. In addition, even if $claimA$ does not exist in the fact-checked database, its corresponding conversation information can also be used as the important clue to classify

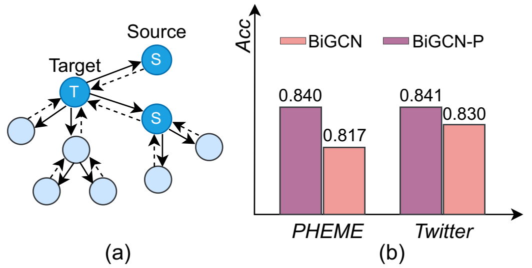

Figure 4. (a) Top-down/bottom-up graph structure under the perspective of public opinion. (b) Performance of BiGCN and BiGCN-P (which incorporates global public opinion features) on PHEME and Twitter.

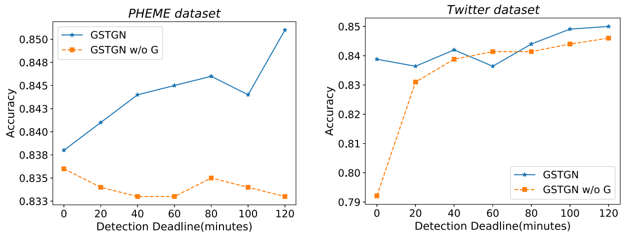

Figure 5. Results of rumor early detection. Different moments (i.e. 0, 20, ..., 120 minutes) are set to verify whether the model can correctly identify rumors based on the limited information carried at the current early moment.

$claimB$. Figure 5 shows the performance of GSTGN at different moments. Obviously, when the rumors have not started to spread (at 0 minutes), GSTGN can obtain a higher detection accuracy, and its performance greatly exceeds that of GSTGN without global public opinion feature. Moreover, as time increases, the information in the conversation thread will be more abundant, and the prediction accuracy of GSTGN will gradually increase.

5.7 Case Analysis

To further illustrate the benefits of introducing global public opinion, we provide a rumor case correctly classified by the GSTGN model, as depicted in Figure 6. In the detection of the source post "*BREAKING UPDATE: The latest footage from Sydney siege shows a man holding a female hostage as a human shield*," the corresponding comments did not provide refutational or indicative information. Instead, they mostly comprised exclamatory statements such as "COWARD!" and "total coward." However, upon introducing viewpoints from other conversation threads, such as "*stop reporting photos Or are you guys immune to police directions*" and "*They could be misinterpreted and cause things to escalate*," the model could clearly indicate the falsity of the target source post.

5.8 Error Analysis

In the case of detecting the source post A, "*When the 5 hostages escaped, the gunman could be seen from here getting extremely agitated, shouting at remaining hostages* (label: non-rumor)", our model erroneously classified it as "rumor". This misclassification may have arisen due to our method initially aggregating the dissemination of the target event in the external environment, thereby capturing sources B (i.e., "*#SYDNEYSIEGE: 5 hostages escape from #Lindt Cafe, more trapped, police in contact with gunman*") and C (i.e.,

Figure 6. A rumor case from the PHEME dataset. Two posters named $PzFeed$ and $9NewsSyd$, retweeted the event and led users to discuss it. The conversation thread of $PzFeed$ is the target object to be detected.

"*Update - Five hostages have escaped from the Lindt cafe in Sydney and an unconfirmed number of hostages remain inside #sydneysiege*"), which are "rumors". Source posts B and C both contain elements of rumor, leading the model to potentially categorize source post A similarly. However, the actual situation of sources B and C is complex: while "*5 hostages escape*" is true, other aspects such as "*police in contact with gunman*" and "*unconfirmed number of hostages*" are false. Consequently, sources B and C present a mixed veracity scenario. Our model lacks the capability to parse source post information into finer-grained events, which is a direction for future exploration. Training the model to disaggregate source posts into one or multiple events and subsequently discerning their veracity individually would align better with real-world circumstances and represents one avenue for future research.

5.9 Limitations

In addition to the limitation mentioned in Section 5.8, there are still noteworthy issues that deserve discussion. On the one hand, as the global graph expands, the computational costs increase linearly. Without the development of a parallel processing algorithm, our method may face certain impediments in practical implementation. On the other hand, due to budget constraints, we are unable to construct large-scale cross-platform multi-modal corpus, limiting the optimization of the model. In addition, when rumor producers have enough financial resources to hire large-scale online supporters and let them complete overwhelming supportive posts on target rumor events, the public's attention will shift from facts to false information. At this moment our model will also fail.

6 Conclusion

In this paper, we propose a novel multi-modal rumor detection framework named GSTGN, which not only considers the multi-media conversation structure, but also introduces the global public opinion as an objective indicator for the model to classify claims. More specifically, we first construct local and global conversational propagation graphs. We then propose the ST encoder and C-GCL to mine the structural and temporal relationships among posts, and design a comparison unit to learn the difference between local and global opinions. Experimental results on PHEME and Twitter show that our proposed GSTGN method outperforms the SOTA baselines.

Acknowledgements

The authors would like to thank the three anonymous reviewers for their comments on this paper. This research was supported by the National Natural Science Foundation of China (Nos. 62276177 and 62376181), and Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

[1] H. Allcott and M. Gentzkow. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36, 2017.
[2] T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, and J. Huang. Rumor detection on social media with bi-directional graph convolutional networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 549–556, 2020.
[3] Y. Chen, D. Li, P. Zhang, J. Sui, Q. Lv, L. Tun, and L. Shang. Cross-modal ambiguity learning for multimodal fake news detection. In *Proceedings of the ACM Web Conference 2022*, pages 2897–2905, 2022.
[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the Association for Computational Linguistics*, pages 4171–4186, 2019.
[5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *Proceedings of the International Conference on Learning Representations*, pages 1–21, 2021.
[6] Y. Dun, K. Tu, C. Chen, C. Hou, and X. Yuan. KAN: Knowledge-aware attention network for fake news detection. In *Proceedings of the AAAI conference on artificial intelligence*, pages 81–89, 2021.
[7] D. ElBoghdady. Market quavers after fake ap tweet says obama was hurt in white house explosions. *The Washington Post*, 23, 2013.
[8] Z. Huang, Z. Lv, X. Han, B. Li, M. Lu, and D. Li. Social bot-aware graph neural network for early rumor detection. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6680–6690, 2022.
[9] N. Ketkar. Introduction to pytorch. In *Deep learning with python*, pages 195–208. Springer, 2017.
[10] D. Khattar, J. S. Goud, M. Gupta, and V. Varma. MVAE: Multimodal variational autoencoder for fake news detection. In *The world wide web conference*, pages 2915–2921, 2019.
[11] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *Proceedings of the International Conference on Learning Representations*, pages 1–14, 2017.
[12] E. Kochkina, M. Liakata, and A. Zubiaga. All-in-one: Multi-task learning for rumour verification. In *Proceedings of the International Conference on Computational Linguistics*, pages 3402–3413, 2018.
[13] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang. Prominent features of rumor propagation in online social media. In *2013 IEEE 13th international conference on data mining*, pages 1103–1108. IEEE, 2013.
[14] Y.-J. Lu and C.-T. Li. GCAN: Graph-aware co-attention networks for explainable fake news detection on social media. In *Proceedings of the Association for Computational Linguistics*, pages 505–514, 2020.
[15] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.-F. Wong. Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 1751–1754, 2015.
[16] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3818–3824, 2016.
[17] J. Ma, W. Gao, and K.-F. Wong. Detect rumors in microblog posts using propagation structure via kernel learning. In *Proceedings of the Association for Computational Linguistics*, pages 708–717, 2017.
[18] J. Ma, W. Gao, and K.-F. Wong. Detect rumors on twitter by promoting information campaigns with generative adversarial learning. In *The World Wide Web Conference*, pages 3049–3055, 2019.
[19] S. B. Naeem and R. Bhatti. The covid-19 'infodemic': a new front for information professionals. *Health Information & Libraries Journal*, 37 (3):233–239, 2020.
[20] D. Q. Nguyen, T. Vu, and A. T. Nguyen. BERTweet: A pre-trained language model for english tweets. In *Proceedings of the Empirical Methods in Natural Language Processing*, pages 9–14, 2020.
[21] P. Qi, J. Cao, X. Li, H. Liu, Q. Sheng, X. Mi, Q. He, Y. Lv, C. Guo, and Y. Yu. Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 1212–1220, 2021.
[22] S. Qian, J. Wang, J. Hu, Q. Fang, and C. Xu. Hierarchical multi-modal contextual attention network for fake news detection. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 153–162, 2021.
[23] Q. Sheng, J. Cao, X. Zhang, X. Li, and L. Zhong. Article reranking by memory-enhanced key sentence matching for detecting previously fact-checked claims. In *Proceedings of the Association for Computational Linguistics and International Joint Conference on Natural Language Processing*, pages 5468–5481, 2021.
[24] Q. Sheng, J. Cao, X. Zhang, R. Li, D. Wang, and Y. Zhu. Zoom out and observe: News environment perception for fake news detection. In *Proceedings of the Association for Computational Linguistics*, pages 4543–4556, 2022.
[25] C. Sun, X. Qiu, Y. Xu, and X. Huang. How to fine-tune bert for text classification? In *China National Conference on Chinese Computational Linguistics*, pages 194–206. Springer, 2019.
[26] M. Sun, X. Zhang, J. Ma, and Y. Liu. Inconsistency matters: A knowledge-guided dual-inconsistency network for multi-modal rumor detection. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1412–1423, 2021.
[27] M. Sun, X. Zhang, J. Zheng, and G. Ma. DDGCN: Dual dynamic graph convolutional networks for rumor detection on social media. In *Proceedings of the AAAI conference on artificial intelligence*, pages 4611–4619, 2022.
[28] T. Sun, Z. Qian, S. Dong, P. Li, and Q. Zhu. Rumor detection on social media with graph adversarial contrastive learning. In *Proceedings of the ACM Web Conference 2022*, pages 2789–2797, 2022.
[29] T. Sun, Z. Qian, P. Li, and Q. Zhu. Graph interactive network with adaptive gradient for multi-modal rumor detection. In *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval*, pages 316–324, 2023.
[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
[31] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, pages 849–857, 2018.
[32] Y. Wu, P. Zhan, Y. Zhang, L. Wang, and Z. Xu. Multimodal fusion with co-attention networks for fake news detection. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2560–2569, 2021.
[33] F. Yang, Y. Liu, X. Yu, and M. Yang. Automatic detection of rumor on sina weibo. In *Proceedings of the ACM SIGKDD workshop on mining data semantics*, pages 1–7, 2012.
[34] F. Yu, Q. Liu, S. Wu, L. Wang, T. Tan, et al. A convolutional approach for misinformation identification. In *IJCAI*, pages 3901–3907, 2017.
[35] C. Yuan, Q. Ma, W. Zhou, J. Han, and S. Hu. Jointly embedding the local and global relations of heterogeneous graph for rumor detection. In *2019 IEEE International Conference on Data Mining (ICDM)*, pages 796–805. IEEE, 2019.
[36] J. Zheng, X. Zhang, S. Guo, Q. Wang, W. Zang, and Y. Zhang. MFAN: Multi-modal feature-enhanced attention networks for rumor detection. In *IJCAI*, pages 2413–2419, 2022.