

Towards Propagation Uncertainty: Edge-enhanced Bayesian Graph Convolutional Networks for Rumor Detection

Lingwei Wei^{1,4}, Dou Hu², Wei Zhou^{1*}, Zhaojuan Yue³, Songlin Hu^{1,4*}

¹ Institute of Information Engineering, Chinese Academy of Sciences

² National Computer System Engineering Research Institute of China

³ Computer Network Information Center, Chinese Academy of Sciences

⁴ School of Cyber Security, University of Chinese Academy of Sciences

{weilingwei18, hudou18}@mailsucas.edu.cn

{zhouwei, husonglin}@iie.ac.cn

yuezhaajuan@cnic.cn

Abstract

Detecting rumors on social media is a very critical task with significant implications to the economy, public health, etc. Previous works generally capture effective features from texts and the propagation structure. However, the *uncertainty caused by unreliable relations in the propagation structure is common and inevitable* due to wily rumor producers and the limited collection of spread data. Most approaches neglect it and may seriously limit the learning of features. Towards this issue, this paper makes the first attempt to explore propagation uncertainty for rumor detection. Specifically, we propose a novel Edge-enhanced Bayesian Graph Convolutional Network (EBGCN) to capture robust structural features. The model adaptively rethinks the reliability of latent relations by adopting a Bayesian approach. Besides, we design a new edge-wise consistency training framework to optimize the model by enforcing consistency on relations. Experiments on three public benchmark datasets demonstrate that the proposed model achieves better performance than baseline methods on both rumor detection and early rumor detection tasks.

1 Introduction

With the ever-increasing popularity of social media sites, user-generated messages can quickly reach a wide audience. However, social media can also enable the spread of false rumor information (Vosoughi et al., 2018). Rumors are now viewed as one of the greatest threats to democracy, journalism, and freedom of expression. Therefore, detecting rumors on social media is highly desirable and socially beneficial (Ahsan et al., 2019).

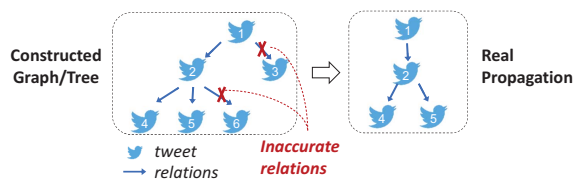


Figure 1: An example of uncertain propagation structure. It includes inaccurate relations, making constructed graph inconsistent with the real propagation.

Almost all the previous studies on rumor detection leverage text content including the source tweet and all user retweets or replies. As time goes on, rumors form their specific propagation structures after being retweeted or replied to. Vosoughi (2015); Vosoughi et al. (2018) have confirmed rumors spread significantly farther, faster, deeper, and more broadly than the truth. They provide the possibility of detecting rumors through the propagation structure. Some works (Ma et al., 2016; Kochkina et al., 2018) typically learn temporal features alone from propagation sequences, ignoring the internal topology. Recent approaches (Ma et al., 2018; Khoo et al., 2020) model the propagation structure as trees to capture structural features. Bian et al. (2020); Wei et al. (2019) construct graphs and aggregate neighbors’ features through edges based on reply or retweet relations.

However, most of them only work well in a narrow scope since they treat these relations as reliable edges for message-passing. As shown in Figure 1, the existence of inaccurate relations brings *uncertainty* in the propagation structure. The neglect of unreliable relations would lead to severe error accumulation through multi-layer message-passing and limit the learning of effective features.

We argue such inherent uncertainty in the propagation structure is inevitable for two aspects: i)

* Corresponding author.

In the real world, rumor producers are always wily. They tend to viciously manipulate others to create fake supporting tweets or remove opposing voices to evade detection (Yang et al., 2020). In these common scenarios, relations can be manipulated, which provides uncertainty in the propagation structure. ii) Some annotations of spread relations are subjective and fragmentary (Ma et al., 2017; Zubiaga et al., 2016). The available graph would be a portion of the real propagation structure as well as contain noisy relations, resulting in uncertainty. Therefore, it is very challenging to handle inherent uncertainty in the propagation structure to obtain robust detection results.

To alleviate this issue, we make the first attempt to explore the uncertainty in the propagation structure. Specifically, we propose a novel Edge-enhanced Bayesian Graph Convolutional Network (EBGCN) for rumor detection to model the uncertainty issue in the propagation structure from a probability perspective. The core idea of EBGCN is to adaptively control the message-passing based on the prior belief of the observed graph to surrogate the fixed edge weights in the propagation graph. In each iteration, edge weights are inferred by the posterior distribution of latent relations according to the prior belief of node features in the observed graph. Then, we utilize graph convolutional layers to aggregate node features by aggregating various adjacent information on the refining edges. Through the above network, EBGCN can handle the uncertainty in the propagation structure and promote the robustness of rumor detection.

Moreover, due to the unavailable of missing or inaccurate relations for training the proposed model, we design a new edge-wise consistency training framework. The framework combines unsupervised consistency training on these unlabeled relations into the original supervised training on labeled samples, to promote better learning. We further ensure the consistency between the latent distribution of edges and the distribution of node features in the observed graph by computing KL-divergence between two distributions. Ultimately, both the cross-entropy loss of each claim and the Bayes by Backprop loss of latent relations will be optimized to train the proposed model.

We conduct experiments on three real-world benchmark datasets (*i.e.*, *Twitter15*, *Twitter16*, and *PHEME*). Extensive experimental results demonstrate the effectiveness of our model. EBGCN of-

fers a superior uncertainty representation strategy and boosts the performance for rumor detection. The main contributions of this work are summarized as follows:

- We propose novel Edge-enhanced Bayesian Graph Convolutional Networks (EBGCN) to handle the uncertainty in a probability manner. To the best of our knowledge, this is the first attempt to consider the inherent uncertainty in the propagation structure for rumor detection.
- We design a new edge-wise consistency training framework to optimize the model with unlabeled latent relations.
- Experiments on three real-world benchmark datasets demonstrate the effectiveness of our model on both rumor detection and early rumor detection tasks¹.

2 Related Work

2.1 Rumor Detection

Traditional methods on rumor detection adopted machine learning classifiers based on handcrafted features, such as sentiments (Castillo et al., 2011), bag of words (Enayet and El-Beltagy, 2017) and time patterns (Ma et al., 2015). Based on salient features of rumors spreading, Wu et al. (2015); Ma et al. (2017) modeled propagation trees and then used SVM with different kernels to detect rumors.

Recent works have been devoted to deep learning methods. Ma et al. (2016) employed Recurrent Neural Networks (RNN) to sequentially process each timestep in the rumor propagation sequence. To improve it, many researchers captured more long-range dependency via attention mechanisms (Chen et al., 2018), convolutional neural networks (Yu et al., 2017; Chen et al., 2019), and Transformer (Khoo et al., 2020). However, most of them focused on learning temporal features alone, ignoring the internal topology structure.

To capture topological-structural features, Ma et al. (2018) presented two recursive neural network (RvNN) based on bottom-up and top-down propagation trees. Yuan et al. (2019); Lu and Li (2020); Nguyen et al. (2020) formulated the propagation structure as graphs. Inspired by Graph Convolutional Network (GCN) (Kipf and Welling, 2017), Bian et al. (2020) first applied two GCNs

¹The source code is available at <https://github.com/weilingwei96/EBGCN>.

based on the propagation and dispersion graphs. Wei et al. (2019) jointly modeled the structural property by GCN and the temporal evolution by RNN.

However, most of them treat the edge as the reliable topology connection for message-passing. Ignoring the uncertainty caused by unreliable relations could lead to lacking robustness and make it risky for rumor detection. Inspired by valuable research (Zhang et al., 2019a) that modeled uncertainty caused by finite available textual contents, this paper makes the first attempt to consider the uncertainty caused by unreliable relations in the propagation structure for rumor detection.

2.2 Graph Neural Networks

Graph Neural Networks (GNNs) (Kipf and Welling, 2017; Schlichtkrull et al., 2018; Velickovic et al., 2018) have demonstrated remarkable performance in modeling structured data in a wide variety of fields, *e.g.*, text classification (Yao et al., 2019), recommendation system (Wu et al., 2019) and emotion recognition (Ghosal et al., 2019). Although promising, they have limited capability to handle uncertainty in the graph structure. While the graphs employed in real-world applications are themselves derived from noisy data or modeling assumptions. To alleviate this issue, some valuable works (Luo et al., 2020; Zhang et al., 2019b) provide an approach for incorporating uncertain graph information by exploiting a Bayesian framework (Maddox et al., 2019). Inspired by them, this paper explores the uncertainty in the propagation structure from a probability perspective, to obtain more robust rumor detection results.

3 Problem Statement

This paper develops EBGCN which processes text contents and propagation structure of each claim for rumor detection. In general, **rumor detection** commonly can be regarded as a multi-classification task, which aims to learn a classifier from training claims for predicting the label of a test claim.

Formally, let $\mathcal{C} = \{c^1, c^2, \dots, c^m\}$ be the rumor detection dataset, where c^i is the i -th claim and m is the number of claims. For each claim $c^i = \{r^i, x_1^i, x_2^i, \dots, x_{n_i-1}^i, G^i\}$, G^i indicates the propagation structure, r^i is the source tweet, x_j^i refers to the j -th relevant retweet, and n_i represents the number of tweets in the claim c^i . Specifically, G^i is defined as a propagation graph $G_i = \langle V_i, E_i \rangle$

with the root node r^i (Ma et al., 2018; Bian et al., 2020), where $V_i = \{r^i, x_1^i, x_2^i, \dots, x_{n_i-1}^i\}$ refers to the node set and $E_i = \{e_{st}^i | s, t = 0, \dots, n_i - 1\}$ represent a set of directed edges from a tweet to its corresponding retweets. Denote $\mathbf{A}_i \in \mathbb{R}^{n_i \times n_i}$ as an adjacency matrix where the initial value is

$$\alpha_{st} = \begin{cases} 1, & \text{if } e_{st}^i \in E_i \\ 0, & \text{otherwise} \end{cases}.$$

Besides, each claim c^i is annotated with a ground-truth label $y^i \in \mathcal{Y}$, where \mathcal{Y} represents fine-grained classes. Our goal is to learn a classifier from the labeled claimed set, that is $f: \mathcal{C} \rightarrow \mathcal{Y}$.

4 The Proposed Model

In this section, we propose a novel edge-enhanced bayesian graph convolutional network (EBGCN) for rumor detection in Section 4.2. For better training, we design an edge-wise consistency training framework to optimize EBGCN in Section 4.3.

4.1 Overview

The overall architecture of EBGCN is shown in Figure 2. Given the input sample including text contents and its propagation structure, we first formulate the propagation structure as directed graphs with two opposite directions, *i.e.*, a top-down propagation graph and a bottom-up dispersion graph. Text contents are embedded by the text embedding layer. After that, we iteratively capture rich structural characteristics via two main components, node update module, and edge inference module. Then, we aggregate node embeddings to generate graph embedding and output the label of the claim.

For training, we incorporate unsupervised consistency training on the Bayes by Backprop loss of unlabeled latent relations. Accordingly, we optimize the model by minimizing the weighted sum of the unsupervised loss and supervised loss.

4.2 Edge-enhanced Bayesian Graph Convolutional Networks

4.2.1 Graph Construction and Text Embedding

The initial graph construction is similar to the previous work (Bian et al., 2020), *i.e.*, build two distinct directed graphs for the propagation structure of each claim c^i . The top-down propagation graph and bottom-up dispersion graph are denoted as G_i^{TD} and G_i^{BU} , respectively. Their corresponding initial adjacency matrices are $\mathbf{A}_i^{TD} = \mathbf{A}_i$ and $\mathbf{A}_i^{BU} = \mathbf{A}_i^\top$.

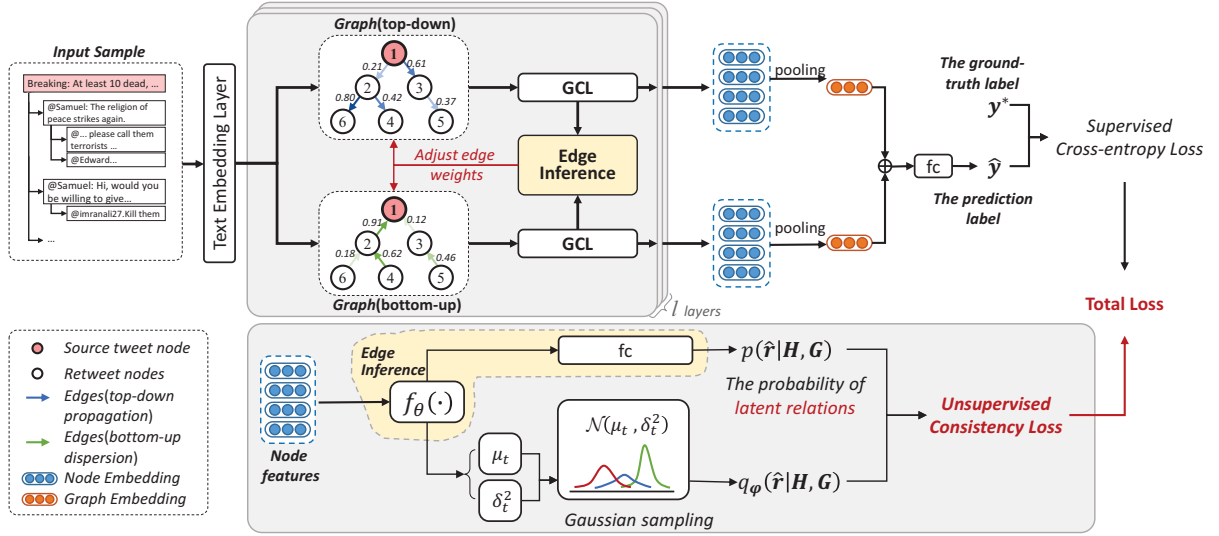


Figure 2: The architecture of the proposed rumor detection model EBGCN.

Here, we leave out the superscript i in the following description for better presenting our method.

The initial feature matrix of postings in the claim c can be extracted Top-5000 words in terms of *TF-IDF* values, denoted as $\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{n-1}] \in \mathbb{R}^{n \times d_0}$, where $\mathbf{x}_0 \in \mathbb{R}^{d_0}$ is the vector of the source tweet and d_0 is the dimensionality of textual features. The initial feature matrices of nodes in propagation graph and dispersion graph are the same, i.e., $\mathbf{X}^{TD} = \mathbf{X}^{BU} = \mathbf{X}$.

4.2.2 Node Update

Graph convolutional networks (GCNs) (Kipf and Welling, 2017) are able to extract graph structure information and better characterize a node's neighborhood. They define multiple Graph Convolutional Layers (GCLs) to iteratively aggregate features of neighbors for each node and can be formulated as a simple differentiable message-passing framework. Motivated by GCNs, we employ the GCL to update node features in each graph. Formally, node features at the l -th layer $\mathbf{H}^{(l)} = [\mathbf{h}_0^{(l)}, \mathbf{h}_1^{(l)}, \dots, \mathbf{h}_{n-1}^{(l)}]$ can be defined as,

$$\mathbf{H}^{(l)} = \sigma(\hat{\mathbf{A}}^{(l-1)} \mathbf{H}^{(l-1)} \mathbf{W}^{(l)} + \mathbf{b}^{(l)}), \quad (1)$$

where $\hat{\mathbf{A}}^{(l-1)}$ represents the normalization of adjacency matrix $\mathbf{A}^{(l-1)}$ (Kipf and Welling, 2017). We initialize node representations by textual features, i.e., $\mathbf{H}^{(0)} = \mathbf{X}$.

4.2.3 Edge Inference

To alleviate the negative effects of unreliable relations, we rethink edge weights based on the cur-

rently observed graph by adopting a soft connection.

Specifically, we adjust the weight between two nodes by computing a transformation $f_e(\cdot; \theta_t)$ based on node representations at the previous layer. Then, the adjacency matrix will be updated, i.e.,

$$\begin{aligned} \mathbf{g}_t^{(l)} &= f_e(\|\mathbf{h}_i^{(l-1)} - \mathbf{h}_j^{(l-1)}\|; \theta_t), \\ \mathbf{A}^{(l)} &= \sum_{t=1}^T \sigma(\mathbf{W}_t^{(l)} \mathbf{g}_t^{(l)} + \mathbf{b}_t^{(l)}) \cdot \mathbf{A}^{(l-1)}. \end{aligned} \quad (2)$$

In practice, $f_e(\cdot; \theta_t)$ consists an convolutional layer and an activation function. T refers to the number of latent relation types. $\sigma(\cdot)$ refers to a *sigmoid* function. $\mathbf{W}_t^{(l)}$ and $\mathbf{b}_t^{(l)}$ are learnable parameters.

We perform share parameters to the edge inference layer in two graphs G^{TD} and G^{BU} . After the stack of transformations in two layers, the model can effectively accumulate a normalized sum of features of the neighbors driven by latent relations, denoted as \mathbf{H}^{TD} and \mathbf{H}^{BU} .

4.2.4 Classification

We regard the rumor detection task as a graph classification problem. To aggregate node representations in the graph, we employ aggregator to form the graph representations. Given the node representations in the propagation graph \mathbf{H}^{TD} and the node representations in the dispersion graph \mathbf{H}^{BU} , the graph representations can be computed as:

$$\begin{aligned} \mathbf{C}^{TD} &= \text{meanpooling}(\mathbf{H}^{TD}), \\ \mathbf{C}^{BU} &= \text{meanpooling}(\mathbf{H}^{BU}), \end{aligned} \quad (3)$$

where $\text{meanpooling}(\cdot)$ refers to the mean-pooling aggregating function. Based on the concatenation of two distinct graph representations, label probabilities of all classes can be defined by a full connection layer and a softmax function, *i.e.*,

$$\hat{\mathbf{y}} = \text{softmax}(\mathbf{W}_c[\mathbf{C}^{TD}; \mathbf{C}^{BU}] + \mathbf{b}_c), \quad (4)$$

where \mathbf{W}_c and \mathbf{b}_c are learnable parameter matrices.

4.3 Edge-wise Consistency Training Framework

For the supervised learning loss \mathcal{L}_c , we compute the cross-entropy of the predictions and ground truth distributions $C = \{c_1, c_2, \dots, c_m\}$, *i.e.*,

$$\mathcal{L}_c = - \sum_i^{|Y|} \mathbf{y}^i \log \hat{\mathbf{y}}^i, \quad (5)$$

where \mathbf{y}^i is a vector representing distribution of ground truth label for the i -th claim sample.

For the unsupervised learning loss \mathcal{L}_e , we amortize the posterior distribution of the classification weight $p(\varphi)$ as $q(\varphi)$ to enable quick prediction at the test stage and learn parameters by minimizing the average expected loss over latent relations, *i.e.*, $\varphi^* = \arg \min_{\varphi} \mathcal{L}_e$, where

$$\mathcal{L}_e = \mathbb{E} \left[D_{KL} \left(p(\hat{\mathbf{r}}^{(l)} | \mathbf{H}^{(l-1)}, G) \| q_{\varphi}(\hat{\mathbf{r}}^{(l)} | \mathbf{H}^{(l-1)}, G) \right) \right], \quad (6)$$

$$\varphi^* = \arg \max_{\varphi} \mathbb{E} [\log \int p(\hat{\mathbf{r}}^{(l)} | \mathbf{H}^{(l-1)}, \varphi) q_{\varphi}(\varphi | \mathbf{H}^{(l-1)}, G) d\varphi],$$

where $\hat{\mathbf{r}}$ is the prediction distribution of latent relations. To ensure likelihood tractably, we model the prior distribution of each latent relation $r_t, t \in [1, T]$ independently. For each relation, we define a factorized Gaussian distribution for each latent relation $q_{\varphi}(\varphi | \mathbf{H}^{(l-1)}, G; \Theta)$ with means μ_t and variances δ_t^2 set by the transformation layer,

$$q_{\varphi}(\varphi | \mathbf{H}^{(l-1)}, G; \Theta) = \prod_{t=1}^T q_{\varphi}(\varphi_t | \{\mathbf{g}_t^{(l)}\}_{t=1}^T) \\ = \prod_{t=1}^T \mathcal{N}(\mu_t, \delta_t^2), \quad (7)$$

$$\mu_t = f_{\mu}(\{\mathbf{g}_t^{(l)}\}_{t=1}^T; \theta_{\mu}), \delta_t^2 = f_{\delta}(\{\mathbf{g}_t^{(l)}\}_{t=1}^T; \theta_{\delta}),$$

where $f_{\mu}(\cdot; \theta_{\mu})$ and $f_{\delta}(\cdot; \theta_{\delta})$ refer to compute the mean and variance of input vectors, parameterized by θ_{μ} and θ_{δ} , respectively. Such that amounts to set the weight of each latent relation.

Besides, we also consider the likelihood of latent relations when parameterizing the posterior

distribution of prototype vectors. The likelihood of latent relations from the l -th layer based on node embeddings can be adaptively computed by,

$$p(\hat{\mathbf{r}}^{(l)} | \mathbf{H}^{(l-1)}, \varphi) = \prod_{t=1}^T p(\hat{\mathbf{r}}_t^{(l)} | \mathbf{H}^{(l-1)}, \varphi_t), \quad (8)$$

$$p(\hat{\mathbf{r}}_t^{(l)} | \mathbf{H}^{(l-1)}, \varphi_t) = \frac{\exp(\mathbf{W}_t \mathbf{g}_t^{(l)} + \mathbf{b}_t)}{\sum_{t=1}^T \exp(\mathbf{W}_t \mathbf{g}_t^{(l)} + \mathbf{b}_t)}.$$

In this way, the weight of edges can be adaptively adjusted based on the observed graph, which can thus be used to effectively pass messages and learn more discriminative features for rumor detection.

To sum up, in training, we optimize our model EBGCN by minimizing the cross-entropy loss of labeled claims \mathcal{L}_c and Bayes by Backprop loss of unlabeled latent relations \mathcal{L}_e , *i.e.*,

$$\Theta^* = \arg \min_{\Theta} \gamma \mathcal{L}_c + (1 - \gamma) \mathcal{L}_e, \quad (9)$$

where γ is the trade-off coefficient.

5 Experimental Setup

5.1 Datasets

We evaluate the model on three real-world benchmark datasets: *Twitter15* (Ma et al., 2017), *Twitter16* (Ma et al., 2017), and *PHEME* (Zubiaga et al., 2016). The statistics are shown in Table 1. **Twitter15** and **Twitter16**² contain 1,490 and 818 claims, respectively. Each claim is labeled as Non-rumor (NR), False Rumor (F), True Rumor (T), or Unverified Rumor (U). Following (Ma et al., 2018; Bian et al., 2020), we randomly split the dataset into five parts and conduct 5-fold cross-validation to obtain robust results. **PHEME** dataset³ provides 2,402 claims covering nine events and contains three labels, False Rumor (F), True Rumor (T), and Unverified Rumor (U). Following the previous work (Wei et al., 2019), we conduct leave-one-event-out cross-validation, *i.e.*, in each fold, one event's samples are used for testing, and all the rest are used for training.

5.2 Baselines

For *Twitter15* and *Twitter16*, we compare our proposed model with the following methods. **DTC**

²<https://www.dropbox.com/s/7ewzdrbelpmrnxu/rumdetect2017.zip?dl=0>

³https://figshare.com/articles/dataset/PHEME_dataset_for_Rumour_Detection_and_Veracity_Classification/6392078

Dataset	<i>Twitter15</i>	<i>Twitter16</i>	<i>PHEME</i>
# of claims	1,490	818	2,402
# of false rumors	370	205	638
# of true rumors	374	205	1,067
# of unverified rumors	374	203	697
# of non-rumors	372	205	-
# of postings	331,612	204,820	105,354

Table 1: Statistics of the datasets.

(Castillo et al., 2011) adopted a decision tree classifier based on information credibility. **SVM-TS** (Ma et al., 2015) leveraged time series to model the chronological variation of social context features via a linear SVM classifier. **SVM-TK** (Ma et al., 2017) applied an SVM classifier with a propagation tree kernel to model the propagation structure of rumors. **GRU-RNN** (Ma et al., 2016) employed RNNs to model the sequential structural features. **RvNN** (Ma et al., 2018) adopted two recursive neural models based on a bottom-up and a top-down propagation tree. **StA-PLAN** (Khoo et al., 2020) employed transformer networks to incorporate long-distance interactions among tweets with propagation tree structure. **BiGCN** (Bian et al., 2020) utilized bi-directional GCNs to model bottom-up propagation and top-down dispersion.

For *PHEME*, we compare with several representative state-of-the-art baselines. **NileTMRG** (Enayet and El-Beltagy, 2017) used linear support vector classification based on bag of words. **BranchLSTM** (Kochkina et al., 2018) decomposed the propagation tree into multiple branches and adopted a shared LSTM to capture structural features. **RvNN** (Ma et al., 2018) consisted of two recursive neural networks to model propagation trees. **Hierarchical GCN-RNN** (Wei et al., 2019) modeled structural property based on GCN and RNN. **BiGCN** (Bian et al., 2020) consisted of propagation and dispersion GCNs to learn structural features from propagation graph.

5.3 Evaluation Metrics

For *Twitter15* and *Twitter16*, we follow (Ma et al., 2018; Bian et al., 2020; Khoo et al., 2020) and evaluate the **accuracy** (Acc.) over four categories and **F1** score (F_1) on each class. For *PHEME*, following (Enayet and El-Beltagy, 2017; Kochkina et al., 2018; Wei et al., 2019), we apply the **accuracy** (Acc.), **macro-averaged F1** (mF_1) as evaluation metrics. Also, we report the **weighted-averaged F1** (wF_1) because of the imbalanced class problem.

5.4 Parameter Settings

Following comparison baselines, the dimension of hidden vectors in the GCL is set to 64. The number of latent relations T and the coefficient weight γ are set to $[1, 5]$ and $[0.0, 1.0]$, respectively. we train the model via backpropagation and a widely used stochastic gradient descent named Adam (Kingma and Ba, 2015). The learning rate is set to $\{0.0002, 0.0005, 0.02\}$ for *Twitter15*, *Twitter16*, and *PHEME*, respectively. The training process is iterated upon 200 epochs and early stopping (Yuan et al., 2007) is applied when the validation loss stops decreasing by 10 epochs. The optimal set of hyperparameters are determined by testing the performance on the fold-0 set of *Twitter15* and *Twitter16*, and the class-balanced *charlie hebdo* event set of *PHEME*.

Besides, on *PHEME*, following (Wei et al., 2019), we replace TF-IDF features with word embeddings by skip-gram with negative sampling (Mikolov et al., 2013) and set the dimension of textual features to 200. We implement this variant of BiGCN and EBGCN, denoted as BiGCN(SKP) and EBGCN(SKP), respectively.

For results of baselines, we implement BiGCN according to their public project⁴ under the same environment. Other results of baselines are referenced from original papers (Khoo et al., 2020; Wei et al., 2019; Ma et al., 2018).

6 Results and Analysis

6.1 Performance Comparison with Baselines

Table 2 shows results of rumor detection on *Twitter15*, *Twitter16*, and *PHEME* datasets. Our proposed model **EBGCN** obtains the best performance among baselines. Specifically, for *Twitter15*, **EBGCN** outperforms state-of-the-art models 2.4% accuracy and 3.6% F1 score of false rumor. For *Twitter16*, our model obtains 3.4% and 6.0% improvements on accuracy and F1 score of non-rumor, respectively. For *PHEME*, EBGCN significantly outperforms previous work by 40.2% accuracy, 34.7% mF_1 , and 18.0% wF_1 .

Deep learning-based (**RvNN**, **StA-PLAN**, **BiGCN** and **EBGCN**) outperform conventional methods using hand-crafted features (**DTC**, **SVM-TS**), which reveals the superiority of learning high-level representations for detecting rumors.

⁴<https://github.com/TianBian95/BiGCN>

Twitter15					
Method	Acc.	NR	F	T	U
		F_1	F_1	F_1	F_1
DTC	45.5	73.3	35.5	31.7	41.5
SVM-TS	54.4	79.6	47.2	40.4	48.3
GRU-RNN	64.1	68.4	63.4	68.8	57.1
SVM-TK	66.7	61.9	66.9	77.2	64.5
RvNN	72.3	68.2	75.8	82.1	65.4
StA-PLAN	85.2	84.0	84.6	88.4	83.7
BiGCN	87.1	86.0	86.7	91.4	85.4
EBGCN	89.2	86.9	89.7	93.4	86.7

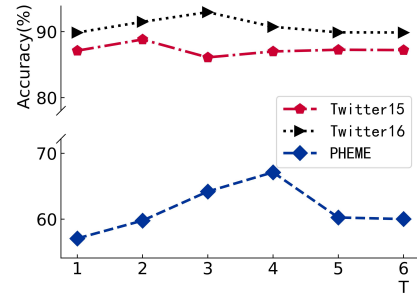
Twitter16					
Method	Acc.	NR	F	T	U
		F_1	F_1	F_1	F_1
DTC	46.5	64.3	39.3	41.9	40.3
SVM-TS	54.4	79.6	47.2	40.4	48.3
GRU-RNN	63.6	61.7	71.5	57.7	52.7
SVM-TK	66.7	61.9	66.9	77.2	64.5
RvNN	72.3	68.2	75.8	82.1	65.4
StA-PLAN	85.2	84.0	84.6	88.4	83.7
BiGCN	88.5	82.9	89.9	93.2	88.2
EBGCN	91.5	87.9	90.6	94.7	91.0

PHEME				
Method	Acc.	mF_1	wF_1	
NileTMRG	36.0	29.7	-	
BranchLSTM	31.4	25.9	-	
RvNN	34.1	26.4	-	
Hierarchical GCN-RNN	35.6	31.7	-	
BiGCN	49.2	46.7	63.2	
BiGCN(SKP)	56.9	48.3	66.8	
EBGCN	69.0	62.9	74.6	
EBGCN(SKP)	71.5	57.5	79.1	

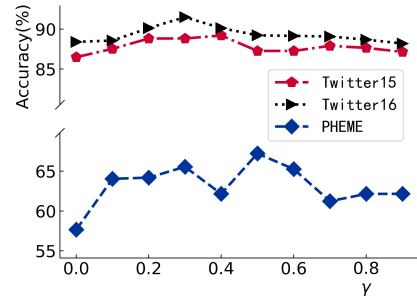
Table 2: Results (%) of rumor detection.

Moreover, compared with sequence-based models **GRU-RNN**, and **StA-PLAN**, **EBGCN** outperform them. It can attribute that they capture temporal features alone but ignore internal topology structures, which limit the learning of structural features. **EBGCN** can aggregate neighbor features in the graph to learn rich structural features.

Furthermore, compared with state-of-the-art graph-based **BiGCN**, **EBGCN** also obtains better performance. We discuss the fact for two main reasons. First, **BiGCN** treats relations among tweet nodes as reliable edges, which may introduce inaccurate or irrelevant features. Thereby their performance lacks robustness. **EBGCN** considers the inherent uncertainty in the propagation structure. In the model, the unreliable relations can be refined



(a) The effect of edge inference



(b) The effect of unsupervised relation learning loss

Figure 3: Results of model analysis on three datasets.

in a probability manner, which boosts the bias of express uncertainty. Accordingly, the robustness of detection is enhanced. Second, the edge-wise consistency training framework ensures the consistency between uncertain edges and the current nodes, which is also beneficial to learn more effective structural features for rumor detection.

Besides, **EBGCN(SKP)** and **BiGCN(SKP)** outperforms **EBGCN** and **BiGCN** that use TF-IDF features in terms of Acc. and wF_1 . It shows the superiority of word embedding to capture textual features. Our model consistently obtains better performance in different text embedding. It reveals the stability of **EBGCN**.

6.2 Model Analysis

In this part, we further evaluate the effects of key components in the proposed model.

The Effect of Edge Inference. The number of latent relation types T is a critical parameter in the edge inference module. Figure 3(a) shows the accuracy score against T . The best performance is obtained when T is 2, 3, and 4 on *Twitter15*, *Twitter16*, and *PHEME*, respectively. Besides, these best settings are different. An idea explanation is that complex relations among tweets are various in different periods and gradually tend to be more sophisticated in the real world with the development

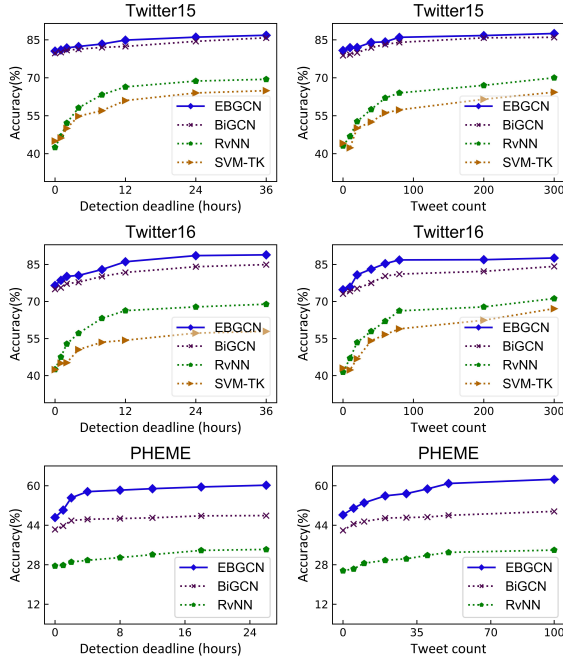


Figure 4: Performance of early rumor detection.

of social media. The edge inference module can adaptively refine the reliability of these complex relations by the posterior distribution of latent relations. It enhances the bias of uncertain relations and promotes the robustness of rumor detection.

The Effect of Unsupervised Relation Learning Loss. The trade-off parameter γ controls the effect of the proposed edge-wise consistency training framework. $\gamma = 0.0$ means this framework is omitted. The right in Figure 3 shows the accuracy score against γ . When this framework is removed, the model gains the worst performance. The optimal γ is 0.4, 0.3, and 0.3 on *Twitter15*, *Twitter16*, and *PHEME*, respectively. The results proves the effectiveness of this framework. Due to wily rumor producers and limited annotations of spread information, it is common and inevitable that datasets contains unreliable relations. This framework can ensure the consistency between edges and the corresponding node pairs to avoid the negative features.

6.3 Early Rumor Detection

Rumor early detection is to detect a rumor at its early stage before it wide-spreads on social media so that one can take appropriate actions earlier. It is especially critical for a real-time rumor detection system. To evaluate the performance on rumor early detection, we follow (Ma et al., 2018) and control the detection deadline or tweet count since the source tweet was posted. The earlier the detec-

tion deadline or the less the tweet count, the less propagation information can be available.

Figure 4 shows the performance of early rumor detection. First, all models climb as the detection deadline elapses or tweet count increases. Particularly, at each deadline or tweet count, our model **EBGCN** reaches a relatively high accuracy score than other comparable models.

Second, compared with **RvNN** that captures temporal features alone and **STM-TK** based on handcrafted features, the superior performance of **EBGCN** and **BiGCN** that explored rich structural features reveals that structural features are more beneficial to the early detection of rumors.

Third, **EBGCN** obtains better early detection results than **BiGCN**. It demonstrates that **EBGCN** can learn more conducive structural features to identify rumors by modeling uncertainty and enhance the robustness for early rumor detection.

Overall, our model not only performs better long-term rumor detection but also boosts the performance of detecting rumors at an early stage.

6.4 The Case Study

In this part, we perform the case study to show the existence of uncertainty in the propagation structure and explain why **EBGCN** performs well. We randomly sample a false rumor from *PHEME*, as depicted in Figure 5. The tweets are formulated as nodes and relations are modeled as edges in the graph, where *node 1* refers to the source tweet and *node 2-8* refer to the following retweets.

As shown in the left of Figure 5, we observe that *tweet 5* is irrelevant with *tweet 1* although replying, which reveals the ubiquity of unreliable relations among tweets in the propagation structure and it is reasonable to consider the uncertainty caused by these unreliable relations.

Right of Figure 5 indicates constructed graphs where the color shade indicates the value of edge weights. The darker the color, the greater the edge weight. The existing graph-based models always generate the representation of *node 1* by aggregating the information of its all neighbors (*node 2, 5, and 6*) according to seemingly reliable edges. However, *edge between node 1 and 5* would bring noise features and limit the learning of useful features for rumor detection. Our model **EBGCN** successfully weakens the negative effect of this edge by both the edge inference layer under the ingenious edge-wise consistency training framework. Accordingly, the

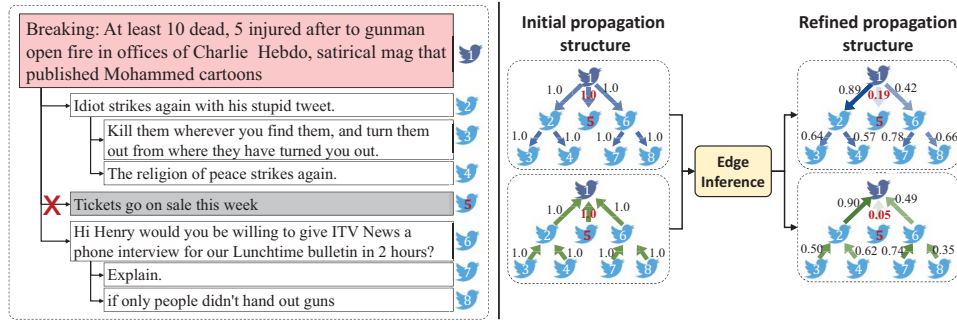


Figure 5: The case study. Left shows a false rumor sampled from *PHEME*. The gray-highlighted tweet is the irrelevant one towards this rumor propagation but included in. Right is the constructed directed graphs in top-down and bottom-up directions based on the propagation structure. Our model iteratively adjusts the weights of edges in each graph to strength the effect of reliable edges and weaken the effect of unreliable edges.

model is capable of learning more conducive characteristics and enhances the robustness of results.

7 Conclusion

In this paper, we have studied the uncertainty in the propagation structure from a probability perspective for rumor detection. Specifically, we propose Edge-enhanced Bayesian Graph Convolutional Networks (EBGCN) to handle uncertainty with a Bayesian method by adaptively adjusting weights of unreliable relations. Besides, we design an edge-wise consistency training framework incorporating unsupervised relation learning to enforce the consistency on latent relations. Extensive experiments on three commonly benchmark datasets have proved the effectiveness of modeling uncertainty in the propagation structure. EBGCN significantly outperforms baselines on both rumor detection and early rumor detection tasks.

References

- Mohammad Ahsan, Madhu Kumari, and T. P. Sharma. 2019. Rumors detection, verification and controlling mechanisms in online social networks: A survey. *Online Soc. Networks Media*, 14.
- Tian Bian, Xi Xiao, Tingyang Xu, Peilin Zhao, Wenbing Huang, Yu Rong, and Junzhou Huang. 2020. Rumor detection on social media with bi-directional graph convolutional networks. In *AAAI*, pages 549–556. AAAI Press.
- Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *WWW*, pages 675–684. ACM.
- Tong Chen, Xue Li, Hongzhi Yin, and Jun Zhang. 2018. Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. In *PAKDD (Workshops)*, volume 11154 of *Lecture Notes in Computer Science*, pages 40–52. Springer.
- Yixuan Chen, Jie Sui, Liang Hu, and Wei Gong. 2019. Attention-residual network with CNN for rumor detection. In *CIKM*, pages 1121–1130. ACM.
- Omar Enayet and Samhaa R. El-Beltagy. 2017. Niletmr at semeval-2017 task 8: Determining rumour and veracity support for rumours on twitter. pages 470–474. Association for Computational Linguistics.
- Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander F. Gelbukh. 2019. Dialoguecn: A graph convolutional neural network for emotion recognition in conversation. In *EMNLP/IJCNLP (1)*, pages 154–164. Association for Computational Linguistics.
- Ling Min Serena Khoo, Hai Leong Chieu, Zhong Qian, and Jing Jiang. 2020. Interpretable rumor detection in microblogs by attending to user interactions. In *AAAI*, pages 8783–8790. AAAI Press.
- Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *ICLR (Poster)*.
- Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *ICLR (Poster)*. OpenReview.net.
- Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task learning for rumour verification. In *COLING*, pages 3402–3413. Association for Computational Linguistics.
- Yi-Ju Lu and Cheng-Te Li. 2020. GCAN: graph-aware co-attention networks for explainable fake news detection on social media. In *ACL*, pages 505–514. Association for Computational Linguistics.
- Yadan Luo, Zi Huang, Zheng Zhang, Ziwei Wang, Mahsa Baktashmotlagh, and Yang Yang. 2020. Learning from the past: Continual meta-learning with bayesian graph neural networks. In *AAAI*, pages 5021–5028. AAAI Press.

- Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. In *IJCAI*, pages 3818–3824. IJCAI/AAAI Press.
- Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. 2015. Detect rumors using time series of social context information on microblogging websites. In *CIKM*, pages 1751–1754. ACM.
- Jing Ma, Wei Gao, and Kam-Fai Wong. 2017. Detect rumors in microblog posts using propagation structure via kernel learning. In *ACL (I)*, pages 708–717. Association for Computational Linguistics.
- Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Rumor detection on twitter with tree-structured recursive neural networks. In *ACL (I)*, pages 1980–1989. Association for Computational Linguistics.
- Wesley J. Maddox, Pavel Izmailov, Timur Garipov, Dmitry P. Vetrov, and Andrew Gordon Wilson. 2019. A simple baseline for bayesian uncertainty in deep learning. In *NeurIPS*, pages 13132–13143.
- Tomás Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*, pages 3111–3119.
- Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. 2020. FANG: leveraging social context for fake news detection using graph representation. In *CIKM*, pages 1165–1174. ACM.
- Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *ESWC*, volume 10843 of *Lecture Notes in Computer Science*, pages 593–607. Springer.
- Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *ICLR (Poster)*. OpenReview.net.
- Soroush Vosoughi. 2015. Automatic detection and verification of rumors on twitter.
- Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science*, 359(6380):1146–1151.
- Penghui Wei, Nan Xu, and Wenji Mao. 2019. Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity. In *EMNLP/IJCNLP (I)*, pages 4786–4797. Association for Computational Linguistics.
- Ke Wu, Song Yang, and Kenny Q. Zhu. 2015. False rumors detection on sina weibo by propagation structures. In *ICDE*, pages 651–662.
- Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-based recommendation with graph neural networks. In *AAAI*, pages 346–353. AAAI Press.
- Xiaoyu Yang, Yuefei Lyu, Tian Tian, Yifei Liu, Yudong Liu, and Xi Zhang. 2020. Rumor detection on social media with graph structured adversarial learning. In *IJCAI*, pages 1417–1423. ijcai.org.
- Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. Graph convolutional networks for text classification. In *AAAI*, pages 7370–7377. AAAI Press.
- Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2017. A convolutional approach for misinformation identification. In *IJCAI*, pages 3901–3907.
- Chunyuan Yuan, Qianwen Ma, Wei Zhou, Jizhong Han, and Songlin Hu. 2019. Jointly embedding the local and global relations of heterogeneous graph for rumor detection. In *ICDM*, pages 796–805. IEEE.
- Yao Yuan, Lorenzo Rosasco, and Andrea Caponnetto. 2007. On early stopping in gradient descent learning. *Constructive Approximation*, 26(2):289 – 315.
- Qiang Zhang, Aldo Lipani, Shangsong Liang, and Emine Yilmaz. 2019a. Reply-aided detection of misinformation via bayesian deep learning. In *WWW*, pages 2333–2343. ACM.
- Yingxue Zhang, Soumyasundar Pal, Mark Coates, and Deniz Üstebay. 2019b. Bayesian graph convolutional neural networks for semi-supervised classification. In *AAAI*, pages 5829–5836. AAAI Press.
- Arkaitz Zubiaga, Geraldine Wong Sak Hoi, Maria Liakata, Rob Procter, and Peter Tolmie. 2016. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PLoS ONE*, 11(3):e0150989.