Topology imbalance and Relation inauthenticity aware Hierarchical Graph Attention Networks for Fake News Detection

Li Gao[1], **Lingyun Song***[1,2], **Jie Liu**[1], **Bolin Chen**[1,2], **Xuequn Shang***[1,2]
[1]School of Computer Science, Northwestern Polytechnical University, Xi'an, China
[2]Key Laboratory of Big Data Storage and Management,
Ministry of Industry and Information Technology, Xi'an, China
{lysong,shang}@nwpu.edu.com

Abstract

Fake news detection is a challenging problem due to its tremendous real-world political and social impacts. Recent fake news detection works focus on learning news features from News Propagation Graph (NPG). However, little attention is paid to the issues of both authenticity of the relationships and topology imbalance in the structure of NPG, which trick existing methods and thus lead to incorrect prediction results. To tackle these issues, in this paper, we propose a novel **T**opology imbalance and **R**elation inauthenticity aware **H**ierarchical **G**raph **A**ttention **N**etworks (**TR-HGAN**) to identify fake news on social media. Specifically, we design a new topology imbalance smoothing strategy to measure the topology weight of each node. Besides, we adopt a hierarchical-level attention mechanism for graph convolutional learning, which can adaptively identify the authenticity of relationships by assigning appropriate weights to each of them. Experiments on real-world datasets demonstrate that TR-HGAN significantly outperforms state-of-the-art methods.

1 Introduction

Fake news is a news article that is intentionally and verifiably false and could mislead readers (Rashkin et al., 2017; Zhou and Zafarani, 2020). The widespread of fake news can immensely affect cyberspace security and even social stability (Lazer et al., 2018). For example, fake news "all of Walmart's e-commerce stores would include a pay with litecoin option beginning October 1"[1] causes Litecoin's prices surged by over 37% to $236 and Bitcoin's prices fell below $44,000 (Kogan et al., 2021), resulting in huge losses for many investors.

Various fake news detection methods are developed to classify whether the news is fake or not, which can be roughly divided into news content-based methods and social context-based methods. Existing news content-based methods (Qazvinian et al., 2011; Maddock et al., 2015; Jin et al., 2013; Wu et al., 2015; Ma et al., 2017) typically focus on mining lexical and syntactic features (Feng et al., 2012; Potthast et al., 2017; Conroy et al., 2015) from news contents, ignoring the rich structural information of news propagation. To address this limitation, social context-based methods (Yuan et al., 2019; Yang et al., 2021; Shu et al., 2019; Yuan et al., 2020) tend to learn the feature representations by the structure information of News Propagation Graph (NPG), where NPG consists of multiple types of nodes (e.g., news, comments and users) and relationships (e.g., follower, retweet and friendship).

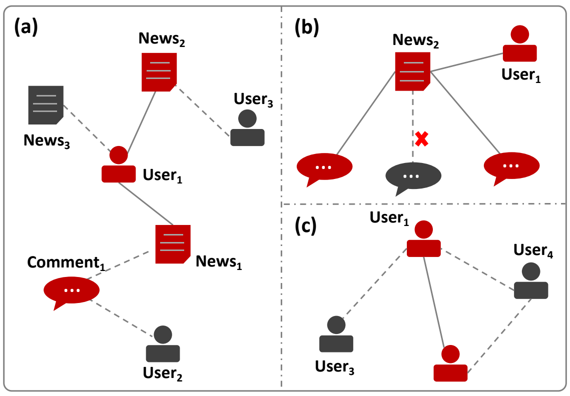

Figure 1: Three types of unauthentic relation scenarios on news propagation graphs (NPG). Nodes representing fake objects (e.g., news or comments) and abnormal users are highlighted in red, while the unauthentic relations are represented by dotted lines. **(a)** An abnormal user $User_1$ forwards a true news $News_3$, and he manipulates or tricks a normal user $User_2$ to create a fake supporting comment on a fake news $News_1$, or he hacks a normal user $User_3$ to forward a fake news $News_2$. **(b)** A fake news producer $User_1$ deletes real opposing comments produced by normal users. **(c)** An abnormal user $User_1$ follows many normal users to disguise himself.

*Corresponding authors: Lingyun Song, Xuequn Shang
[1]https://genesisblockhk.com/fake-news-litecoin-and-walmart-are-not-partnering-in-payments/

Proceedings of the 29th International Conference on Computational Linguistics, pages 4687–4696
October 12–17, 2022.

Although these social context-based methods achieve promising performance, they have an ill-advised assumption that impair the performance of fake news detection. Specifically, they assume a piece of news connecting with a trustworthy user should also have high credit. However, in many cases, the relationships in NPG can be manipulated by users and thus are unauthentic. For example, we present several types of unauthentic relations in Figure 1, where a fake news creator could manipulate other users to create fake comments to support the fake news, or delete real comments opposing the fake news. This leads to inaccurate NPG with unauthentic propagation structure, which confuses existing social context-based methods to make an incorrect prediction.

Besides, NPG usually has the problem of topology imbalance: the imbalance caused by the asymmetric and uneven topology of labeled nodes, where the decision boundaries are driven by the labeled nodes close to the topological class boundaries thus interfering with the model learning (Chen et al., 2021). Ideally, the influence from labeled nodes should decay with the topology distance and also the node influence boundaries should be consistent with the true class boundaries. But the topology imbalance issue would cause the node influence boundaries to deviate from the true class boundaries, resulting in inaccurate results. Nevertheless, most approaches neglect this issue in NPG, interfering with the detection results.

To tackle these issues, we propose a novel **T**opology imbalance and **R**elation inauthenticity aware **H**ierarchical **G**raph **A**ttention **N**etworks (**TR-HGAN**) for fake news detection. We firstly design a topology smoothing strategy to measure the weights of labeled nodes to alleviate the topology imbalance issue. Then we propose a hierarchical-level attention mechanism to identify the authenticity of relations by measuring the appropriate weights to each of them, which can effectively reduce the influence of the inauthentic relationships in NPG. The main contributions of this paper are as follows:

- We study a novel topology smoothing strategy to address the problem of topology imbalance of NPG. To the best of our knowledge, this is the first attempt to solve the topology imbalance issue of the NPG during fake news detection.
- We develop a hierarchical graph attention mechanism which can effectively identify the authenticity of the relationships by assigning appropriate weights to each of them.
- Experiments on three real world datasets demonstrate that TR-HGAN achieves state-of-the-art performances on fake news classification.

2 Related Work

Fake news detection challenges the usage of related information (such as text content, comments, propagation patterns, etc.) to distinguish whether a news article is fake or not. Related works can be divided into two perspectives: i) News content-based methods; ii) Social context-based methods.

2.1 News Content-based Methods

News content-based methods concentrate on designing some textual features such as content writing styles (Shu et al., 2017), lexical and syntactic features (Feng et al., 2012; Potthast et al., 2017; Conroy et al., 2015) to detect the truthfulness of news articles. For instance, Potthast et al. (Potthast et al., 2017) extracts various style features from news contents and predict fake news and media bias. Ma et al. (Ma et al., 2016; Yu et al., 2017) captures news features from low-level to high-level with deep neural networks. Although these approaches achieve good performance on fake news detection, they focus on learning text features alone, rarely considering whether the features can be captured by utilizing the news propagation structure.

2.2 Social Context-based Methods

Social context-based methods principally learn social interactions or information propagation structures through neural networks for further detection. Specifically, Wu et al. (Wu et al., 2015) propose a graph kernel-based SVM classifier that aims to learn high-order news propagation patterns of news articles. RvNN (Ma et al., 2018) and BiGCN (Bian et al., 2020) are developed based on bottom-up and top-down propagation trees for learning the embedding of fake news propagation structure. However, these tree-structured based studies don't utilize the social network structure information, neglecting the fact that the information dissemination on social media is essentially spread in the form of heterogeneous graph.

To tackle these issues, (Lu and Li, 2020) constructs an user interaction graph to model the potential interactions between users, and then develop a dual co-attention mechanism to learn the co-influence features. Recent studies (Yang et al., 2021; Yuan et al., 2019; Nguyen et al., 2020) formulate the news propagation structure as a heterogeneous news propagation graph with various types of nodes (e.g., news, comments, and users), and then apply GNNs model to capture the structure features for fake news detection. However, they don't consider the unauthentic relations and inherent topology imbalance problem in the graph, and thus may fail to detect the intentional fake news.

3 Proposed Model

The overall architecture of TR-HGAN is shown in Figure 2, which involves four main components: i) NPG Construction; ii) Text Embedding; iii) Hierarchical Graph Attention Network that consists of topology imbalance smoothing strategy and hierarchical-level attention strategy; and iv) Fake news classification. Next, we describe each part of TR-HGAN in detail.

3.1 NPG Construction

Let $\mathcal{M} = \{m_1, m_2, ..., m_{|\mathcal{M}|}\}$ be a set of news articles on social media, where m_i is the i-th news articles and $|\mathcal{M}|$ is the number of news. Let $\mathcal{C}(m_i) = \{c_1, c_2, ..., c_n\}$ be a set of comments of m_i, $\mathcal{U} = \{u_1, u_2, ..., u_{|\mathcal{U}|}\}$ be a set of users who create news or comments. To illustrate our motivation, we construct a heterogeneous news propagation graph (NPG) denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \mathcal{M} \cup \mathcal{U} \cup \mathcal{C}$ covers the sets of news articles, users and comments, $\mathcal{E}$ involves different relationships between nodes. An example of the NPG is shown in the left of Figure 2, where two nodes are connected if they have posting/commenting relationships. Specially, there would be an edge between two news nodes if they share similar neighbors but irrelevant content. Users are connected if they have follower/comment the same news/followee relationships.

3.2 Text Embedding

To obtain the text representation for each news article, we apply CNN (Chen, 2015; Grefenstette et al., 2014) and multi-head self-attention (Vaswani et al., 2017) to obtain the text representation for each text, which is the same as the existing state-of-the-art fake news detection approaches, i.e. GLAN (Yuan et al., 2019) and CGAT (Yang et al., 2021).

Given a source news m_i and its comments $C = \{c_1, c_2, ..., c_n\}$. Firstly, we capture initial sequence feature $\boldsymbol{f_{m_i}} \in \mathbb{R}^d$ of news m_i with CNN. By the same way, the feature $\boldsymbol{f_{c_j}} \in \mathbb{R}^d$ of each comment c_j can be extracted. Then we refine the coherence semantic representation between comments and source news by using multi-head self-attention to capture dependencies across news content and comments. From the above operation, the final text feature $\hat{\boldsymbol{f}}_{\boldsymbol{m_i}}$ for each news and $\hat{\boldsymbol{f}}_{\boldsymbol{c_j}}$ for each comment can be extracted. The initial feature $\hat{\boldsymbol{f}}_{\boldsymbol{u_k}}$ of user node u_k can be calculated by the user profile data (such as friends count, followers count, status count, etc.).

3.3 Hierarchical Graph Attention Network

3.3.1 Topology Imbalance Smoothing Strategy

To tackle the topology imbalance problem in NPG, we propose a strategy of smoothing the topology structure to alleviate the resulting problems. We assume that if a labeled news node $m_i \in \mathcal{V}$ encounters strong influence from the other labeled neighbor nodes, the node m_i owns great influence and is close to topological class boundaries. In other words, the influence of labeled nodes should decay with the topology distance. To measure how topologically close node m_i is to the center of the class it belongs to, we calculate the topology location value $\boldsymbol{T}_m$ by measuring the expectation of message-passing probability between the node m_i and its neighbors when node m_i randomly walks across the entire graph:

$$\boldsymbol{T_m} = \mathbb{E}_{x \sim P_{m,:}}[\sum_{j \in [1,s], j \neq y_m} \frac{1}{|\mathcal{S}_j|} \sum_{i \in \mathcal{S}_j} P_{i,x}], \quad (1)$$

where y_m is the ground-truth label of node m_i. P_m indicates the personalized PageRank probability (Page et al., 1999) vector for the node m_i, which can be viewed as the distribution of influence exerted outward from each m_i. $P_{i,x}$ indicates the probability from m_i to m_x. $\mathcal{S}_j$ represents the training sets for different classes, where s is the number of classes. The normalization item $\frac{1}{|\mathcal{S}_j|}$ is added to make the influence from the different classes comparable when computing conflict.

The larger the topology location value $\boldsymbol{T}_m$ of m_i, the more topologically closer to class boundaries the node m_i. Figure 2.(A) apparently shows that it can decrease the training weights of labeled nodes c_2 and increase the weights of labeled nodes u_4 close to the fake news m_i, thus relieving the topology-imbalance issue. Inspired by the study

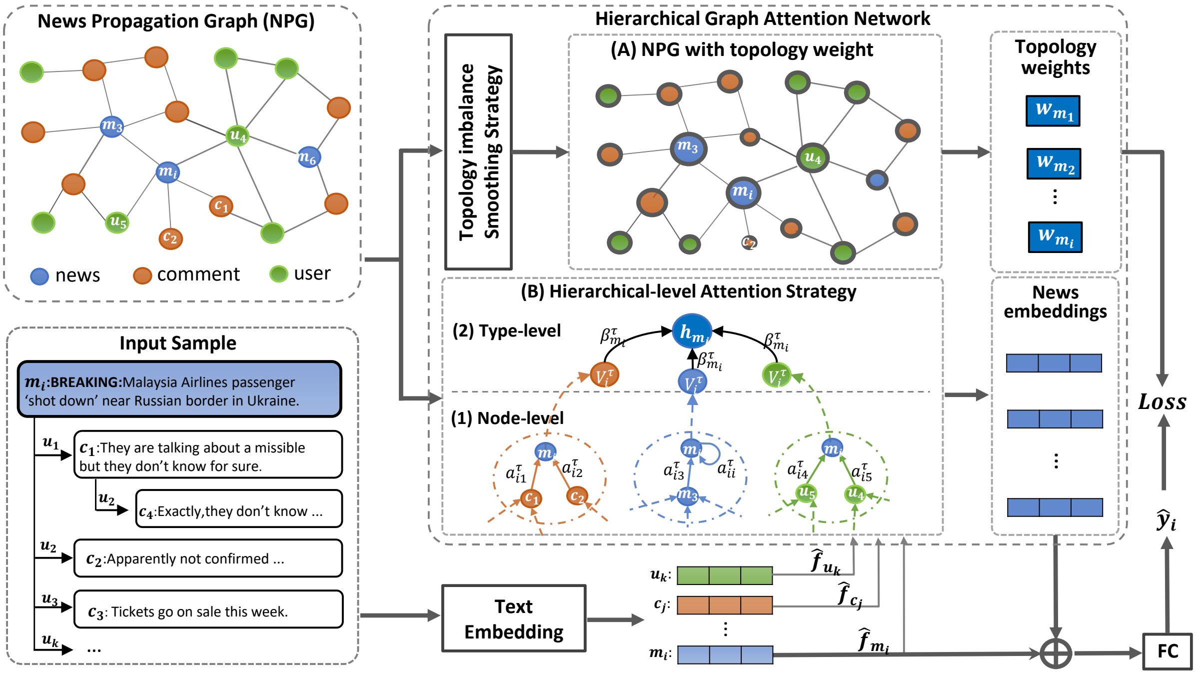

Figure 2: The architecture of TR-HGAN. According to the given *Input Sample*, we construct a *News Propagation Graph (NPG)* which consists of multiple types of nodes and relations. *Topology Imbalance Smoothing Strategy* can increase the weight of great influence node and decrease the weight of week influence one, and finally get the topology weight for each node. *Hierarchical-level Attention Strategy* can adaptively identify the authenticity of the relationship by assigning appropriate weight a_{ij}^{τ} for each edge, where i, j indicate the nodes in NPG, and τ indicates the type of node (i.e. news, comment, user). For simplicity of description, we give the graph learning of one-hop neighbors of m_i as an example in *Hierarchical-level Attention Strategy*. For fake news detection, we only extract the topology weights and features of the news node m_i for final classification. The initial feature of each node is extracted through *Text Embedding*.

(Chen et al., 2021), in order to promote the training weights of nodes for effective model learning, we also train node weights based on their topology location values by:

$$\boldsymbol{w}_m = w_{min} + \frac{w_{max} - w_{min}}{2}(1 + cos\frac{Rank(\boldsymbol{T_m})}{|Y|}\pi), \quad (2)$$

where $\boldsymbol{w}_m$ is the modified training weight for m_i, w_{min}, w_{max} indicates the lower bound and upper bound of the weight correction factor, Y is the set of labeled news article nodes, $Rank(\boldsymbol{T}_m)$ is the ranking order of $\boldsymbol{T}_m$ from the smallest to the largest.

3.3.2 Hierarchical-level Attention Strategy

Considering nodes u, m, c are three different types of nodes belong to the different semantic spaces, the GCN (Kipf and Welling, 2016) cannot be directly applied to the NPG due to the node heterogeneity issue. Surprised by the recent work on heterogeneous graph convolution (Linmei et al., 2019), we employ a hierarchical attention mechanism to learn the representation of each node. Specifically, to alleviate the negative effects of those unauthentic relations, we adjust the weights between two nodes by type-level attention and node-level attention learning.

1) *Node-level attention.* For each type of nodes (e.g., node with type τ, where $\tau \in \{u, m, c\}$), we firstly design the type-specific transformation matrix $\boldsymbol{M}_{\tau}$ to project the features of different types of nodes into the same feature space. Take the news article node m_i as an example, given the input feature vectors $\hat{\boldsymbol{f}}_{m_i}$, the projection process can be shown as:

$$\boldsymbol{h}_{m_i}^{'} = \boldsymbol{M}_{\tau} \cdot \hat{\boldsymbol{f}}_{m_i}. \quad (3)$$

Similarly, the initial features of each comment and each user can be projected as $\boldsymbol{h}_{c_j}^{'}, \boldsymbol{h}_{u_k}^{'}$, respectively. Thus, the initial input features of the node-level attention layer are the $\boldsymbol{h}_{c_j}^{'}, \boldsymbol{h}_{u_k}^{'}, \boldsymbol{h}_{m_i}^{'}$.

In the face of fake news detection, the target node is the news article node $m_i \in \mathcal{M}$ with the type τ. The neighbors of it belong to $\mathcal{V} \in \{\mathcal{M}, \mathcal{U}, \mathcal{C}\}$ with the type $\tau^{'}$. Given a node pair (m_i, v_j), where $v_j \in \mathcal{V}$, we design the node-level attention to get the weight coefficient different neighboring nodes via:

$$\boldsymbol{\alpha}_{ij}^{\tau} = softmax\left(\sigma(\boldsymbol{\mu_\tau}[\boldsymbol{h}^{'}_{\boldsymbol{m_i}}||\boldsymbol{h}^{'}_{\boldsymbol{v_j}}])\right), \quad (4)$$

where $\boldsymbol{\mu_\tau}$ is the attention parameter for the type τ, $||$ means concatenation operation, $\sigma(\cdot)$ denotes LeakyReLU function. Then, the type-level node $\boldsymbol{\mathcal{V}}^{\boldsymbol{\tau}}_{\boldsymbol{m_i}}$ on graph can be aggregated by the neighbor's projected features with the corresponding weights as follows:

$$\boldsymbol{\mathcal{V}}^{\boldsymbol{\tau}}_{\boldsymbol{m_i}} = \sigma(\sum_{v_j \in \mathcal{V}} \boldsymbol{\alpha}_{ij}^{\tau} \cdot \boldsymbol{h}^{'}_{v_j}). \quad (5)$$

2) *Type-level attention.* Through the node-level attention, we fuse information from neighbor nodes with the same type into the representation of a type-level node. Different types of nodes contain type-specific information, which requires us to learn the importance of different node types. Thus, we adopt the type-level attention to learn the weights of news article nodes m_i from all types of nodes by:

$$\boldsymbol{\beta}_{m_i}^{\tau} = softmax\left(\sigma(\boldsymbol{W}[\boldsymbol{\mathcal{V}}^{\boldsymbol{\tau}}_{\boldsymbol{m_i}}||\boldsymbol{\mathcal{M}}_{\boldsymbol{m_i}}])\right), \quad (6)$$

where $\boldsymbol{W}$ is the weight matrix. Finally, we incorporate all type-level nodes to get the final representation $\boldsymbol{h}_{m_i}$ of the target news node m_i:

$$\boldsymbol{h}_{m_i} = \sum_{\tau} \boldsymbol{\beta}_{m_i}^{\tau} \cdot \boldsymbol{\mathcal{V}}_{m_i}^{\tau}. \quad (7)$$

3.4 Fake News Classification

For the target news m_i, we aim at learning an inference function $f : \mathcal{G} \rightarrow \mathcal{Y}$ to predict whether it is fake or not. After the above procedures, we get the structure features $\boldsymbol{h}_{m_i}$, as well as its coherence-based sentence representation $\hat{\boldsymbol{f}}_{m_i}$, and then concatenate them as final features. Then the final representations of news are fed to softmax classifier based on fully-connected layers to obtain category probability $\hat{\boldsymbol{y}}_i$:

$$\hat{\boldsymbol{y}}_{m_i} = softmax(\boldsymbol{W}[\boldsymbol{h}_{\boldsymbol{m_i}}; \hat{\boldsymbol{f}}_{\boldsymbol{m_i}}] + b). \quad (8)$$

Finally, the cross-entropy loss is used as the optimization objective function for fake news detection:

$$\boldsymbol{\mathcal{L}}(\boldsymbol{\Theta}) = -\frac{1}{|Y|}\sum_{m \in Y} \boldsymbol{w_m} \sum_{j=1}^{s} \boldsymbol{y}_{m_i}^{j} \log \hat{\boldsymbol{y}}_{m_i}^{j}, \quad (9)$$

where y_m is the gold probability of fake news class, s is the number of classes and Θ represents all parameters of the model.

4 Results and Discussion

4.1 Datasets

In order to evaluate the performance of TR-HGAN, we conduct experiments on three benchmark datasets: Weibo (Ma et al., 2016), Twitter15 (Liu et al., 2015) and Twitter16 (Ma et al., 2017). Table 1 gives statistics of the three datasets. The Weibo dataset contains binary labels, i.e., fake news (FR) and non-fake news (NR), whereas Twitter15 and Twitter16 datasets have four types of labels, i.e., fake news (FR), non-fake news (NR), unverified news (UR), and true news (TR), where the label true news denotes a news article that debunks the fake news. For a fair comparison, for each dataset, we randomly select 10% of the dataset as the validation subset, and divide the rest data into training and testing subsets with a ratio of 3:1.

Statistic	Weibo	Twitter15	Twitter16
source news	4664	1490	818
fake news	2313	370	205
non-fake news	2351	374	205
unverified news	0	374	203
true news	0	372	205
users	2,746,818	276,663	173,487
comments	3,805,656	331,612	204,820

Table 1: Dataset statistics.

4.2 Baselines

To highlight performance superiority of the proposed TR-HGAN, we select a series of state-of-the-art methods as baselines. The first two methods try to capture fake news features by using content-based structure. The next six methods extract fake news features through social contexts-based structure, among which, the first four methods are tree-structured models and the last two models are pure heterogeneous graph-based approaches. They are described as follows:

- **SVM-TS** (Ma et al., 2015): An SVM model that utilizes time-series to model the variation of hand-crafted features of news.
- **GRU** (Ma et al., 2016): A RNN-based model that captures the temporal contextual information of relevant retweets or comments.
- **RvNN** (Ma et al., 2018): A tree-structured recursive neural method that learns fake news features via the news propagation structure.

- **PPC** (Liu and Wu, 2018): A propagation-based approach that detect fake news with a combination of recurrent and convolutional networks.

- **Bi-GCN** (Bian et al., 2020): A novel Bi-directional model that explores fake news characteristics by operating on both top-down and bottom-up propagation of fake news.

- **EBGCN** (Wei et al., 2021): A propagation-based method that adaptively rethinks the reliability of latent edge-wise relations.

- **CGAT** (Yang et al., 2021): An end-to-end graph-based framework that jointly exploits text and structure information by using graph adversarial learning framework.

- **GLAN** (Yuan et al., 2019): A graph-based method that encodes contextual information and global structural information by adopting a global-local attention network.

4.3 Experimental Setup

For fair comparison, we adopt the same evaluation metrics used in the prior studies (Yuan et al., 2019; Bian et al., 2020; Wei et al., 2021), we also add accuracy (Acc.), precision (Prec.), recall (Rec.) and $\boldsymbol{F_1}$ score as the evaluation metrics.

For text embedding step, for each source news m_i, we truncate the text if its length is larger than 150 words and pad zero if the length is smaller than 150. All word embeddings of the model are initialized with the 300-dimensional word vectors, which is released by (Yuan et al., 2019). The convolutional kernel size k is set to (4, 5, 6) with 100 kernels for each kind of size. Therefore, the final text representation $\boldsymbol{f}_{m_i} \in \mathbb{R}^d$ of news m_i are concatenated by all feature vectors $\boldsymbol{f}_{m_i}^k$ obtained by different filters. Besides, the parameters w_{min}, w_{max} at topology imbalance smoothing strategy is set to [0.25, 0.5, 0.75], [1.25, 1.5, 1.75], respectively.

4.4 Performance Efficiency

Table 2 and Table 3 show the results of fake news detection on *Weibo*, *Twitter15* and *Twitter16* datasets. We bold the best performance of each column in tables, from where we can observe social context-based methods outperform those news content-based methods using only textual features, which reveals the superiority of learning high-level representations for detecting fake news.

Method	Class	Acc.	Prec.	Rec.	$\boldsymbol{F_1}$
SVM-TS	NR	0.857	0.878	0.830	0.857
	FR		0.839	0.885	0.861
GRU	NR	0.910	0.952	0.864	0.906
	FR		0.876	0.956	0.914
RvNN	NR	0.908	0.912	0.897	0.905
	FR		0.904	0.918	0.911
PPC	NR	0.921	0.949	0.889	0.918
	FR		0.896	**0.962**	0.923
BiGCN	NR	0.935	0.925	0.943	0.933
	FR		0.951	0.887	0.917
CGAT	NR	0.939	0.938	0.942	0.941
	FR		0.939	0.938	0.935
GLAN	NR	0.948	0.937	0.957	0.947
	FR		**0.967**	0.934	0.950
TR-HGAN	NR	**0.963**	**0.957**	**0.964**	**0.960**
	FR		0.962	0.961	**0.960**

Table 2: Fake news detection results on *Weibo* dataset.

In addition, TR-HGAN performs better than the tree-structured based methods (e.g., RvNN, PPC, BiGCN, EBGCN). It can attribute that tree-structured methods neglect the messages are spread by a graph structure (i.e., constructed with source news, comments and users) rather than a tree structure (i.e., only constructed with source news and retweets), which limit the learning of high-level features.

Moreover, the proposed TR-HGAN outperforms state-of-the-art graph-based GLAN on three datasets. Specifically, as Table 2 shows, TR-HGAN achieves improvement of 1.5% on Weibo, comparing with the best baseline GLAN in terms of accuracy. We can also find TR-HGAN obtains 2.5% and 1.7% improvements than the best model on accuracy over all metrics across Twitter15 and Twitter16, respectively. We discuss the fact for two main reasons. First, TR-HGAN considers the inherent unauthentic relations and rich structural features in the news propagation graph. Second, unlike CGAT and GLAN, TR-HGAN pays more attention to the node topology imbalance problem on NPG, which helps improve our models much more.

4.5 Ablation Study

In this part, we test the performance of TR-HGAN variants with different configurations, including:

- ***TR-HGAN w/o Text***: it only uses source news texts for fake news classification.

- ***TR-HGAN w/o C***: it removes the news com-

Method	Twitter15					Twitter16				
	Acc.	NR F_1	FR F_1	TR F_1	UR F_1	Acc.	NR F_1	FR F_1	TR F_1	UR F_1
SVM-TS	0.544	0.796	0.472	0.404	0.483	0.574	0.755	0.420	0.571	0.526
GRU	0.646	0.792	0.574	0.608	0.592	0.633	0.772	0.489	0.686	0.593
RvNN	0.723	0.682	0.758	0.821	0.654	0.737	0.662	0.743	0.835	0.708
PPC	0.842	0.811	0.875	0.818	0.790	0.863	0.820	0.898	0.843	0.837
BiGCN	0.886	0.891	0.860	0.930	0.864	0.880	0.847	0.869	**0.937**	0.865
EBGCN	0.891	0.864	0.892	0.916	0.867	0.915	0.868	0.899	0.924	0.901
GLAN	0.904	0.922	0.915	0.857	0.923	0.895	0.910	0.875	0.851	0.941
TR-HGAN	**0.929**	**0.930**	**0.928**	**0.925**	**0.935**	**0.932**	**0.923**	**0.910**	0.932	**0.949**

Table 3: Fake news detection results on *Twitter15* and *Twitter16* datasets.

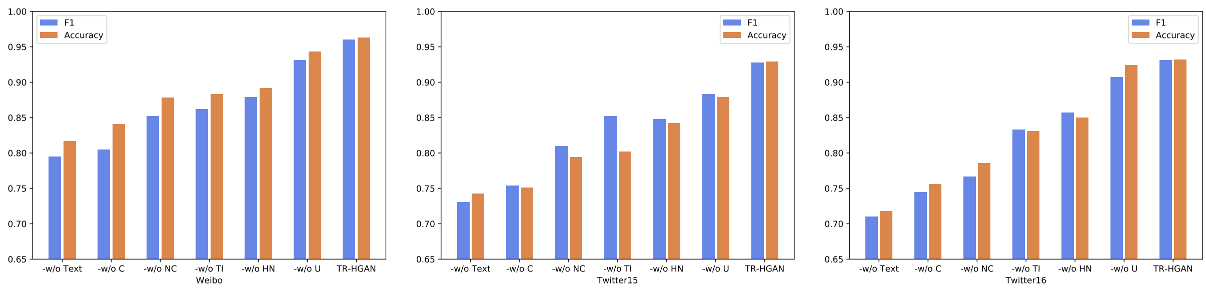

Figure 3: TR-HGAN ablation analysis results on three datasets in terms of Accuracy and F_1.

ments, leaving the source news nodes and users nodes in NPG for fake news detection.

- ***TR-HGAN w/o U***: it removes the users, leaving the source news and comments in graph for fake news detection.
- ***TR-HGAN w/o NC***: it removes the edges between source news and comments.
- ***TR-HGAN w/o HN***: it uses graph convolutional network (GCN) (Kipf and Welling, 2016) to encode the graph structure NPG.
- ***TR-HGAN w/o TI***: it removes the topology imbalance smoothing strategy module for fake news classification.

The comparison results of F_1 and accuracy are shown in Figure 3. We can observe that all the other methods outperform *TR-HGAN w/o Text*, indicating the structure features are important for fake news detection. When we try to remove the information in NPG, the performance apparently decreases (i.e. *TR-HGAN w/o C* and *TR-HGAN w/o NC*) which indicating that the social-aware structure features indeed benefits the performance of fake news detection.

Moreover, *TR-HGAN w/o TI* models the graph structure without considering the topology imbalance problem, resulting the performance drops rapidly, which shows the importance of measuring topology imbalance. Furthermore, the performance of *TR-HGAN w/o HN* gets worse when don't consider the heterogeneity and the unauthentic relations of NPG.

4.6 Early Detection

Early fake news detection aims to detect fake news at the early stage of propagation, which is especially critical for restricting the dissemination scope of fake news. The earlier the detection deadline, the less propagation information such as comments and users are available. To evaluate the performance on early fake news detection, we set up a series of detection deadlines [$0h$, $2h$, $4h$, $6h$, $8h$, $12h$, $24h$].

Figure 4 shows that the proposed TR-HGAN reaches relatively high accuracy at a very early period after the source news initial broadcast. Specifically, TR-HGAN achieves 94% accuracy on Weibo, 87.2% accuracy on Twitter15 and 84.9% on Twitter16 within 2 hours, which is much faster than other models. This is because TR-HGAN takes the reliability of relations in NPG into account by mea-

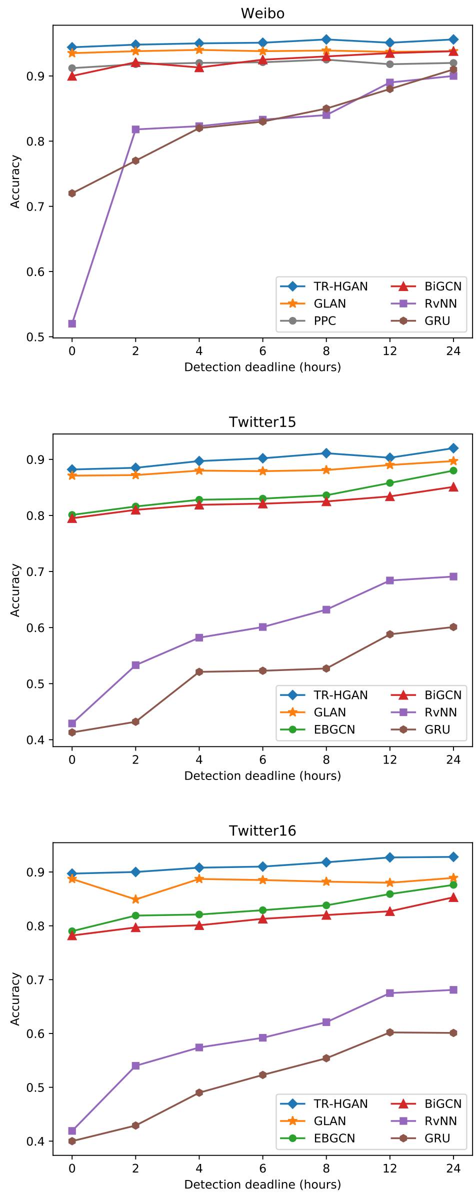

Figure 4: The performance of early fake news detection on three datasets in terms of Accuracy.

suring the importance between different nodes. Besides, TR-HGAN consistently achieves relatively high accuracy score on all datasets than other models at each deadline. This is because TR-HGAN considers the problem of topology imbalance and unauthentic relations in NPG. In the model, the importance of each node can be measured by topology imbalance smoothing strategy, which boosts the performance of detecting results. Second, the unauthentic relations can be refined by the hierarchical attention mechanism, which helps identify unreliable relationships and reduce the noisy of problem nodes in time, so as to detect the authenticity of news as soon as possible.

4.7 Case Study

To further illustrate why our model outperforms state-of-the-art baseline GLAN (Yuan et al., 2019), we randomly sample two fake news from Twitter15 dataset. As depicted in Figure 5, the news, comments and corresponding users are formulated as nodes and relations are modeled as edges in NPG. As shown in the left of Figure 5, we observe that comment c_3 is irrelevant with news m_1 although replying, which reveals the ubiquity of unauthentic relations among news in the NPG and it is necessary to consider the inauthenticity caused by these unauthentic relations.

The right of Figure 5 indicates the constructed weighted NPG. For a target node m_1, existing graph-based models (e.g., GLAN, CGAT) always generate the feature representation of m_1 by aggregating the information of its all neighbors according to seemingly authentic edges. However, edge between node m_1 and c_3 would bring noise features and limit the learning of useful features for fake news detection. The proposed TR-HGAN can successfully weaken the negative effect of this unauthentic edge by hierarchical-level attention network. Besides, the target fake news node m_1 will affect the class boundary shift of its neighbors due to the topology imbalance of NPG. Thus, the topology imbalance smoothing strategy is adopted to decrease the weights of those useless nodes and strengthen the weights of important nodes close to target node. Accordingly, the TR-HGAN is capable of learning more conducive features and can enhance the robustness of results.

5 Conclusion

In this paper, we have studied the unauthentic relations and topology imbalance issue in the news propagation structure from a weight learning perspective on fake news detection. We propose a topology imbalance and relation inauthenticity aware hierarchical graph attention networks (TR-HGAN) to capture robust structural features. Specifically, we design a topology imbalance smoothing strategy to address the problem of topology imbalance in NPG by increasing the weight of nodes with great influence and decreasing the weight of nodes with week influence. Besides, we develop a hierarchical graph attention mechanism for graph convolutional learning, which can adaptively measure the authenticity of the relationships by assigning appropriate weight to each relationship,

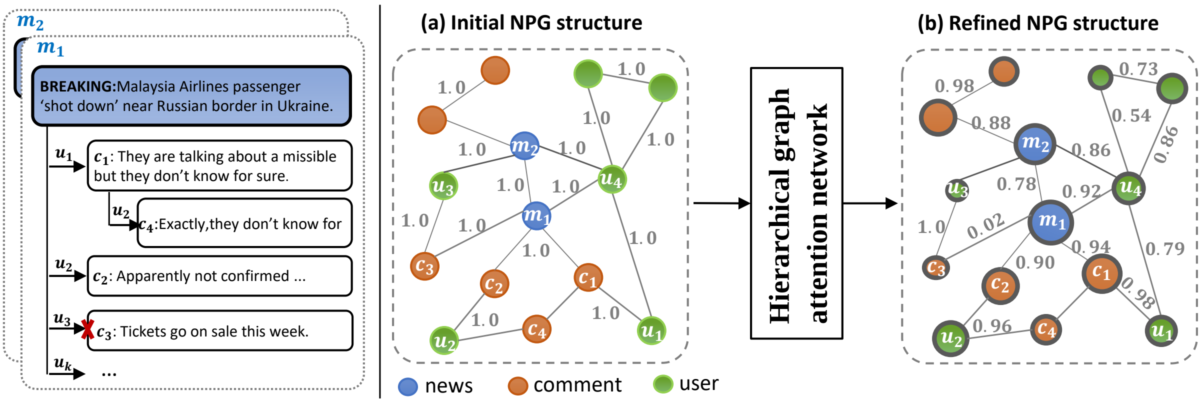

Figure 5: The case study. Left shows two fake news sampled from Twitter15. The comment c_3 is the irrelevant one towards propagation of news m_1 but included in. **(a)** is the constructed NPG based on the news propagation structure. **(b)** shows the result of case study. The proposed TR-HGAN can measure the topology weight of each node to increase the weights of important nodes close to the class centers and decrease the weights of nodes close to class boundary. In addition, TR-HGAN adaptively adjust the weights of edges in NPG to strengthen the effect of authentic edges and weaken the effect of unauthentic edges.

thus effectively reduce the influence of the unauthentic relations. Extensive experiments conducted on three commonly benchmark datasets demonstrate that our model can significantly surpass the state-of-the-art baselines on both fake news classification and early detection tasks.

Acknowledgements

The research is supported by the National Key Research and Development Program of China under Grant 2020AAA0108504, National Nature Science Foundation of China under Grant Nos.62102321, Fundamental Research Funds for the Central Universities under Grant No. D5000200146.

References

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 *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 549–556.

Deli Chen, Yankai Lin, Guangxiang Zhao, Xuancheng Ren, Peng Li, Jie Zhou, and Xu Sun. 2021. Topology-imbalance learning for semi-supervised node classification. *Advances in Neural Information Processing Systems*, 34.

Yahui Chen. 2015. Convolutional neural network for sentence classification. Master's thesis, University of Waterloo.

Nadia K Conroy, Victoria L Rubin, and Yimin Chen. 2015. Automatic deception detection: Methods for finding fake news. *Proceedings of the association for information science and technology*, 52(1):1–4.

Song Feng, Ritwik Banerjee, and Yejin Choi. 2012. Syntactic stylometry for deception detection. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 171–175.

Edward Grefenstette, Phil Blunsom, et al. 2014. A convolutional neural network for modelling sentences. In *ACL*.

Fang Jin, Edward Dougherty, Parang Saraf, Yang Cao, and Naren Ramakrishnan. 2013. Epidemiological modeling of news and rumors on twitter. In *Proceedings of the 7th workshop on social network mining and analysis*, pages 1–9.

Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Shimon Kogan, Tobias J. Moskowitz, and Marina Niessner. 2021. Social media and financial news manipulation. *CommRN: Communication & Computational Linguistics (Topic)*.

David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science*, 359(6380):1094–1096.

Hu Linmei, Tianchi Yang, Chuan Shi, Houye Ji, and Xiaoli Li. 2019. Heterogeneous graph attention networks for semi-supervised short text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the*

9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 4821–4830.

Xiaomo Liu, Armineh Nourbakhsh, Quanzhi Li, Rui Fang, and Sameena Shah. 2015. Real-time rumor debunking on twitter. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 1867–1870.

Yang Liu and Yi-Fang Wu. 2018. 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*, volume 32.

Yi-Ju Lu and Cheng-Te Li. 2020. Gcan: Graph-aware co-attention networks for explainable fake news detection on social media. *arXiv preprint arXiv:2004.11648*.

Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard Jim Jansen, Kam-Fai Wong, and M. Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. In *IJCAI*.

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 *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 1751–1754.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2017. Detect rumors in microblog posts using propagation structure via kernel learning. Association for Computational Linguistics.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Rumor detection on twitter with tree-structured recursive neural networks. Association for Computational Linguistics.

Jim Maddock, Kate Starbird, Haneen J Al-Hassani, Daniel E Sandoval, Mania Orand, and Robert M Mason. 2015. Characterizing online rumoring behavior using multi-dimensional signatures. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 228–241.

Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. 2020. Fang: Leveraging social context for fake news detection using graph representation. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 1165–1174.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2017. A stylometric inquiry into hyperpartisan and fake news. *arXiv preprint arXiv:1702.05638*.

Vahed Qazvinian, Emily Rosengren, Dragomir Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1589–1599.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2931–2937.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36.

Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the twelfth ACM international conference on web search and data mining*, pages 312–320.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Lingwei Wei, Dou Hu, Wei Zhou, Zhaojuan Yue, and Songlin Hu. 2021. Towards propagation uncertainty: Edge-enhanced bayesian graph convolutional networks for rumor detection.

Ke Wu, Song Yang, and Kenny Q Zhu. 2015. False rumors detection on sina weibo by propagation structures. In *2015 IEEE 31st international conference on data engineering*, pages 651–662. IEEE.

Xiaoyu Yang, Yuefei Lyu, Tian Tian, Yifei Liu, Yudong Liu, and Xi Zhang. 2021. Rumor detection on social media with graph structured adversarial learning. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 1417–1423.

Feng Yu, Qiang Liu, Shu Wu, Liang Wang, Tieniu Tan, et al. 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 *2019 IEEE international conference on data mining (ICDM)*, pages 796–805. IEEE.

Chunyuan Yuan, Qianwen Ma, Wei Zhou, Jizhong Han, and Songlin Hu. 2020. Early detection of fake news by utilizing the credibility of news, publishers, and users based on weakly supervised learning. *arXiv preprint arXiv:2012.04233*.

Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40.