

Propagation Tree Is Not Deep: Adaptive Graph Contrastive Learning Approach for Rumor Detection

Chaoqun Cui, Caiyan Jia*

School of Computer and Information Technology & Beijing Key Lab of Traffic Data Analysis and Mining
Beijing Jiaotong University, Beijing 100044, China
{21120341,cyjia}@bjtu.edu.cn

Abstract

Rumor detection on social media has become increasingly important. Most existing graph-based models presume rumor propagation trees (RPTs) have deep structures and learn sequential stance features along branches. However, through statistical analysis on real-world datasets, we find RPTs exhibit wide structures, with most nodes being shallow 1-level replies. To focus learning on intensive substructures, we propose Rumor Adaptive Graph Contrastive Learning (RAGCL) method with adaptive view augmentation guided by node centralities. We summarize three principles for RPT augmentation: 1) exempt root nodes, 2) retain deep reply nodes, 3) preserve lower-level nodes in deep sections. We employ node dropping, attribute masking and edge dropping with probabilities from centrality-based importance scores to generate views. A graph contrastive objective then learns robust rumor representations. Extensive experiments on four benchmark datasets demonstrate RAGCL outperforms state-of-the-art methods. Our work reveals the wide-structure nature of RPTs and contributes an effective graph contrastive learning approach tailored for rumor detection through principled adaptive augmentation. The proposed principles and augmentation techniques can potentially benefit other applications involving tree-structured graphs.

Introduction

The unprecedented growth of the Internet in recent years has promoted the widespread applications of social media. Digital platforms like Weibo and Twitter have evolved into critical conduits for users to garner information and interact with each other. These platforms, while facilitating information dissemination and diverse opinion expression on a multitude of trending issues, are also breeding grounds for various rumors. Given the massive user base and the ease of use, rumors are disseminated extensively and swiftly via social media, wreaking substantial societal havoc. Therefore, there is an urgent need to establish efficacious and efficient strategies for automated rumor verification on social media.

Currently, a plethora of studies concerning rumor detection exist. Certain studies (Bian et al. 2020; Wei et al. 2021) have demonstrated that the propagation structure of a claim, which fully encapsulates the interrelationship between posts

and harnesses the collective intelligence of the crowd, is invaluable for debunking rumors. In general, rumor detection models built upon rumor propagation structures glean discriminative features of rumors from the interrelation of comments, apprehending specific patterns of reply stances as the basis for claim classification, given that clear disparities exist between the comment stances of rumor claims and those of non-rumor claims. These models are adept at discerning these differences, which constitute one of the fundamental postulates of rumor detection methods based on rumor propagation trees (Ma, Gao, and Wong 2018). This supposition relies, to a certain degree, on the deep structures of rumor propagation trees (RPTs). But are RPTs really deep?

Bearing this question in mind, we undertook a statistical analysis to explore the structural characteristics of RPTs. The findings indicate that, in commonly employed rumor detection datasets and real-world social media platforms, the tree structures of claims are typically shallow. The vast majority of a claim’s comments constitute 1-level replies, with the remainder primarily consisting of 2-level replies, and only a negligible number delve into deeper levels. This essentially implies that RPTs are not characterized by deep tree structures, but instead exhibit wide structures.

As per our statistical analysis, the majority of nodes in RPTs are 1-level replies, all pointing directly to the root node (i.e., the source post), which highlights the significance of the root node. Further, it is plausible that the majority of the 1-level replies, due to their lack of deeper engagement, may contain less informational value in rumor identification compared to nodes with more extensive paths. Based on these findings, we propose the Rumor Adaptive Graph Contrastive Learning (RAGCL) method. RAGCL utilizes node centrality measures to generate augmented views of RPTs and leverages graph contrastive learning methods to facilitate graph neural networks (GNNs) in learning crucial rumor discriminative features from the deep sections of RPTs. Empirical studies demonstrate the effectiveness of RAGCL.

In summary, the contributions of this study are as follows.

- Our statistical survey has unveiled that RPTs primarily exhibit a wide tree structure, breaking the stereotype of a deep tree structure in previous studies. This shifts the understanding of information propagation processes on social media platforms.
- Informed by the structural characteristics of RPTs and in-

*Corresponding authors

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

spired by current research on graph self-supervised learning (Zhu et al. 2021), we propose the RAGCL method to learn discriminative features for rumor detection.

- In light of the unique tree structure of RPTs, we propose three guiding principles to be followed when designing adaptive data augmentation methods for RPTs.
- Our experimental results underscore the superior performance of RAGCL in comparison to the current state-of-the-art (SOTA) methods and substantiate the validity of our three principles.

Related Work

In this section, we will review the related works on rumor detection and graph contrastive learning.

Social Media Rumor Detection

To debunk rumors, various efforts have been made. Among the existing studies, early methods mainly take advantage of traditional classification methods by using hand-crafted features (Castillo, Mendoza, and Poblete 2011; Kwon et al. 2013). In recent years, with the advent of deep learning, more effective approaches have emerged, resulting in significant improvements in rumor detection performance. These approaches can be broadly categorized into four classes, including time-series based methods (Ma et al. 2016; Yu et al. 2017; Liu and Wu 2018) which model text content or user profiles as time series, propagation structure learning methods (Ma, Gao, and Wong 2018; Bian et al. 2020; Wei et al. 2021; Sun et al. 2022b) which consider the propagation structures of source rumors and their replies, multi-source integration methods (Karimi et al. 2018; Birunda and Devi 2021) which combine multiple resources of rumors including post content, user profiles, heterogeneous relations between posts and users, multi-modal fusion methods (Wang et al. 2018; Jin et al. 2017) which incorporate both post content and related images to effectively debunk rumors.

The significance of propagation structure information has been increasingly recognized in the field of rumor detection research. Numerous SOTA models bank on learning the representations of RPTs utilizing GNNs. Ma, Gao, and Wong (2018) designed a bottom-up and top-down tree-structured recursive neural network to extract information from RPTs. In a similar vein, Bian et al. (2020) applied a bidirectional GCN alongside a root node feature enhancement technique to address rumor detection tasks. Furthermore, Sun et al. (2022b) incorporated contrastive loss with adversarial training to learn representations robust to rumor noise. These studies bear testament to the efficacy of propagation structure learning in accurately identifying rumors.

Graph Contrastive Learning

The advancement of deep learning has instigated progress across numerous studies predicated on neural message passing algorithms (Gilmer et al. 2017). These algorithms learn graph representations in a supervised manner and have attained SOTA results across a wide array of tasks (Kipf et al. 2018; Xie and Grossman 2018; Chen et al. 2019). In recent

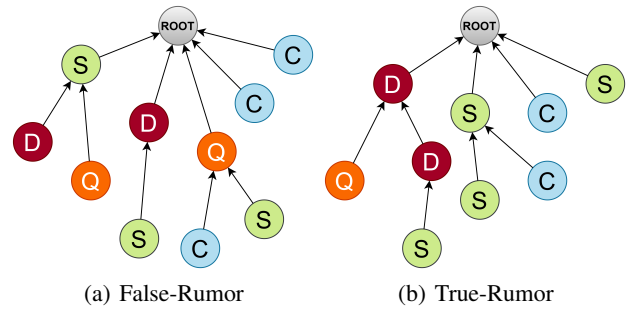


Figure 1: The stances in rumor propagation trees.

years, graph self-supervised learning methods have gradually emerged to leverage unlabeled data for addressing the problem of scarce labeled data, with most of them being graph contrastive learning methods. Contrastive learning methods have been widely applied in the domain of image representation learning (He et al. 2020; Chen et al. 2020), and subsequently extended to the realms of text (Giorgi et al. 2020; Shi et al. 2019; Fang et al. 2020) and graph data (Velickovic et al. 2019; Sun et al. 2019). Graph contrastive learning methodologies have evolved from initial methods premised on mutual information maximization (Velickovic et al. 2019; Sun et al. 2019) to contemporary methods based on graph augmentation (Hassani and Khasahmadi 2020; You et al. 2020, 2021). Graph contrastive learning methods based on graph augmentation first employ diverse graph augmentation strategies (such as node drop, edge perturbation, etc.) to acquire varying views of a given graph, thereafter constructing positive and negative samples in the contrastive loss. Ultimately, graph representations are learned by minimizing the contrastive loss. To accommodate different types of graph datasets, several research studies have focused on adaptive graph augmentation (Zhu et al. 2021; You et al. 2021; Yin et al. 2022). RAGCL represents a novel adaptive graph augmentation approach for rumor detection to learn robust and discriminative representations of RPTs.

Analysis on Propagation Tree

As shown in Figure 1, current propagation structure learning based rumor detection methods are devoted to collect support (S), deny (D), question (Q), comment (C) and other stances between a reply and its source post, and pairs of replies (Ma, Gao, and Wong 2018). For different classes of claims, there exist noticeable differences in their stance patterns, which can serve as discriminative features for rumor identification. For instance, the true stance of a D-D relation is S, whereas the true stance of a D-S relation is D. Current propagation structure learning methods exploit stance features between sequential nodes on the same branch of the tree structure for rumor detection (Bian et al. 2020; Sun et al. 2022b). Nonetheless, these features are dependent on the depth structure of RPTs. But, are RPTs truly deep?

In the present study, we conduct a thorough exploration of the structural properties inherent to RPTs, deploying a statistical approach. The datasets under scrutiny consist of

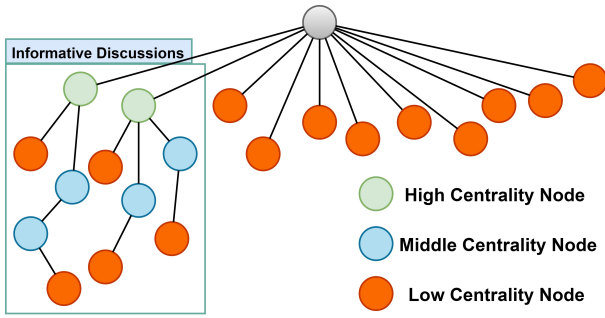


Figure 2: A rumor propagation tree.

Weibo (Ma et al. 2016), DRWeibo¹, Twitter15, and Twitter16 (Ma, Gao, and Wong 2017). Further, we survey two large-scale unlabeled public datasets, namely UWeibo² and UTwitter³. Data from these datasets originate from popular posts on Weibo and Twitter platforms, mirroring the universal traits of claims within social media environments. The statistical results are shown in Table 1. The entries beneath the dotted line in Table 1 denote the mean count of replies, 1-level replies, 2-level replies, deeper (>2) replies, and 1-level replies with subsequent replies per claim in the dataset, respectively. The statistics lead us to the ensuing conclusions.

- **RPTs resemble wide trees rather than deep ones.** 1-level replies constitute the majority of all replies within RPTs, with proportions of 65.1%, 77.8%, 70.7%, and 64.2% for the four labeled datasets respectively.
- **Only a minimal portion of 1-level replies within RPTs spawn subsequent replies.** Amongst all 1-level replies in RPTs, only a fraction give rise to further replies, with percentages of 9.7%, 6.4%, 10.4%, and 10.8%.
- **Deep replies within a RPT are seldom observed.** Deep replies make up a tiny fraction of all replies in RPTs, with percentages of 13.8%, 4.4%, 17.3%, and 23.4%. This infers that the model is constrained to learning the aforesaid stance features from a limited set of replies.

Both UWeibo and UTwitter datasets also exhibit these three characteristics, signifying that these traits are pervasive attributes of claims on social media platforms. The above observations paint the generalized structure of a RPT, as illustrated in Figure 2. It’s noticeable that only nodes enclosed within the box in Figure 2 carry the aforementioned stance features, while most of the nodes in the tree are 1-level nodes without further reply (no deep structure).

Based on the above observations, we can conceptualize a RPT as a highly imbalanced graph, with imbalances reflected in the following two aspects:

- The root node of a RPT features highly dense connections, whereas connections at the remaining nodes are exceedingly sparse.
- The intensive discussions and informative portions of a RPT are predominantly found within a limited number

¹<https://github.com/CcQunResearch/DRWeibo>

²<https://github.com/CcQunResearch/UWeibo>

³<https://github.com/CcQunResearch/UTwitter>

of 1-level replies (the two green nodes in Figure 2). In contrast, the majority of the 1-level replies that lack further deeper responses also lack discriminative features that can aid in rumor identification.

Such characteristics are determined by users’ habits of using social media and the order in which platforms display comments. In general, users are inclined to reply directly to source posts rather than to other users’ comments. Additionally, platforms such as Weibo and Twitter tend to sort replies based on popularity rather than the chronological order of posting. This contributes to the imbalance in the information distribution within a propagation tree. With the aim of enhancing our model’s focus on the intense and informative discussions of RPTs and reducing the influence of a large number of unresponded 1-level replies, we put forward our RAGCL method. The objective of RAGCL is to stress the importance of comments within RPTs that have intensive replies, while also focusing on root nodes by directing the aggregation of information from other nodes towards these roots, considering the wide structures of RPTs.

Method

We will present the design of RAGCL in this section.

Notation

The rumor detection task can be defined as a graph-level classification task. Specifically, we denote a labeled claim dataset as $\mathbb{C} = \{c_1, c_2, \dots, c_m\}$, where c_i represents the i -th claim and m represents the number of labeled claims. Each labeled claim $c = (y, G)$ consists of its ground-truth label $y \in \{N, R\}$ (i.e., Non-rumor or Rumor) or fine-grained label $y \in \{N, F, T, U\}$ (i.e., Non-rumor, False Rumor, True Rumor, Unverified Rumor) and its propagation structure $G = (V, E)$, where V and E represent the set of nodes (a source post and comments of the claim) and edges (the relations between pairs of replies or source post and a reply), respectively. The set of propagation structure graphs corresponding to all claims is $\mathbb{G} = \{G_1, G_2, \dots, G_m\}$. The goal of rumor detection task is to learn a classifier $f : \mathbb{G} \rightarrow Y$ ($Y = \{y_1, y_2 \dots y_m\}$) from dataset \mathbb{C} .

Framework

From the aforementioned analysis, it is important to learn discriminative features from nodes with deep structures (e.g., the nodes in the box in Figure 2). These nodes and their corresponding edges possess a conspicuously higher importance compared to the nodes located outside the box. Based on this idea, we introduce RAGCL, an adaptive graph contrastive learning framework purposefully engineered for rumor detection. RAGCL assigns varying levels of importance to nodes and edges within a RPT based on a selected node centrality measure. Subsequently, varying probabilities of drop or mask, informed by these scores, are employed to adaptively generate two graph augmented views of the RPT, utilizing node drop, attribute mask, or edge drop operators. The contrastive loss is subsequently minimized to learn the tree’s representation. A comprehensive illustration of the RAGCL process is presented in Figure 3.

Statistic	Weibo	DRWeibo	Twitter15	Twitter16	UWeibo	UTwitter
language	zh	zh	en	en	zh	en
# claims	4664	6037	1490	818	209549	204922
# non-rumors	2351	3185	374	205	-	-
# false rumors	2313	2852	370	205	-	-
# true rumors	-	-	372	207	-	-
# unverified rumors	-	-	374	201	-	-
# avg reply	803.5	61.8	50.2	49.1	50.5	82.5
# avg 1-level reply	522.9(65%)	48.1(78%)	35.5(71%)	31.6(64%)	36.4(72%)	48.5(59%)
# avg 2-level reply	169.3(21%)	11.0(17%)	5.9(12%)	6.0(12%)	10.2(20%)	21.5(26%)
# avg deeper reply	111.2(14%)	2.7(5%)	8.7(17%)	11.5(24%)	4.0(8%)	12.5(15%)
# avg responded 1-level reply	50.7	3.1	3.7	3.4	4.1	8.0

Table 1: Statistics of the datasets.

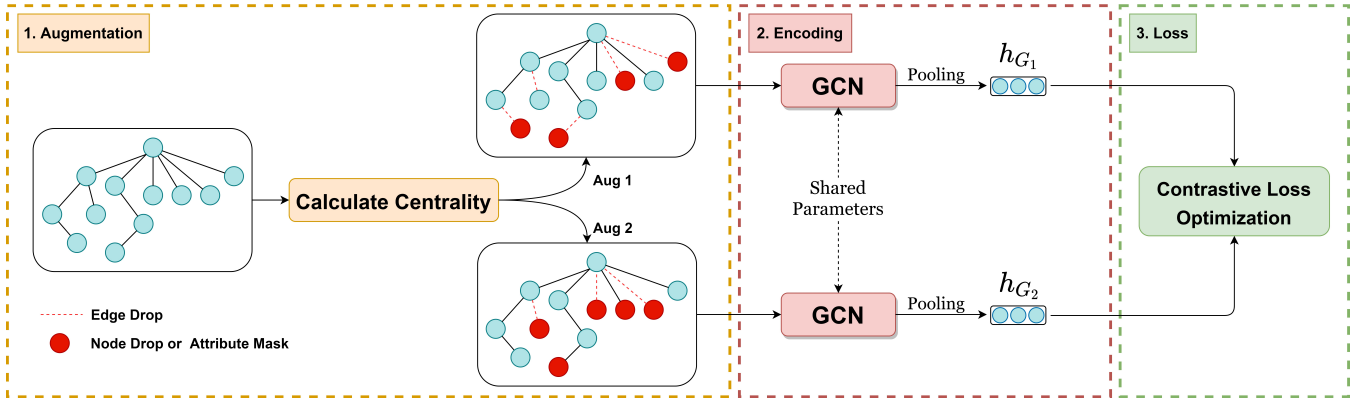


Figure 3: The framework of RAGCL.

Augmentation Principle

Node centrality is an index to measure the importance of nodes in a graph. There are three recommended node centrality measures used by RAGCL, including degree centrality (Shaw 1954), betweenness centrality (Freeman 1977) and PageRank centrality (Brin and Page 1998).

- **Degree centrality** takes the degree of nodes as the measure of node centrality. The idea is that a post with multiple replies is important in a RPT. RAGCL uses the node out-degree of top-down graph of a RPT as the measure of degree centrality.
- **Betweenness centrality** calculates all shortest paths of any two nodes in a graph. A node becomes prominent in terms of betweenness centrality if a multitude of these paths transit through it. RAGCL utilizes either top-down or bottom-up graphs to ascertain betweenness centrality.
- **PageRank centrality** is commonly used in web page ranking. Its basic idea is that the importance of a page on Internet depends on the quantity and quality of inbound links. RAGCL leverages the bottom-up graph of a RPT to compute PageRank centrality.

Given our prior analysis on the structural characteristics of RPTs, we have summarized the following three principles for assigning importance scores to nodes and edges.

- **Principle 1:** Given the pivotal role of source posts (Bian

et al. 2020; Sun et al. 2022b), the root nodes of RPTs are exempt from the data augmentation procedure.

- **Principle 2:** Nodes and edges with deep replies within RPTs (referenced within the boxed portion of Figure 2) should be preserved to the greatest extent feasible.
- **Principle 3:** In the deep parts of RPTs, the nodes in low-level should be retained in data augmentation more than its deeper successor nodes, because the successor nodes are basically discussed around their parent nodes, so they should hold relatively lower importance.

Other node centrality measures, such as eigenvector centrality (Bonacich 1972), Katz centrality (Katz 1953), and closeness centrality (Sabidussi 1966), are deemed unsuitable for RPTs due to inherent characteristics which preclude adherence to the aforementioned principles. The node colors in Figure 2 show the magnitude of node centrality that should be obtained according to the above principles. Furthermore, to ensure compliance with Principle 2, RAGCL assigns the root node of a RPT the minimum value from among all node centralities within the graph, given its dense characteristic. Throughout the data augmentation process, the importance of an edge in RAGCL is gauged by the centrality of the edge’s two constituting nodes. An excessively high root node centrality could artificially inflate the importance of edges connecting the root node with unresponded 1-level replies, thereby contravening Principle 2.

Adaptive Graph Augmentation

RAGCL conducts adaptive data augmentation according to node centrality, yielding two augmented views of the RPT. It primarily utilizes three unique data augmentation operators: node dropping, attribute masking, and edge dropping. During the training phase, two out of these three operators are selected. RAGCL employs node centrality to assign importance scores to nodes and edges, after which it computes the probability of dropping or masking for data augmentation.

Node Dropping Consider a propagation graph G of any given claim within dataset \mathbb{C} . Given a node centrality measure $\varphi_c(\cdot) : \mathcal{V} \rightarrow \mathbb{R}^+$, \mathcal{V} is the space where node v is located, and the final node centrality value of node v is represented by $\varphi_c(v)$. Node dropping in RAGCL involves assigning a drop probability p_v^n to each node v and removing a portion of nodes (along with the edges connected to these nodes) from the node set V in accordance with this probability to yield an augmented view. It is noteworthy that the root node is never dropped in this operation. The node importance score w_v^n is set as node centrality value, that is, $w_v^n = \varphi_c(v)$. Given that the value of node centrality might vary across several orders of magnitude, $s_v^n = \log w_v^n$ is set to alleviate the influence of densely connected nodes. The node drop probability is derived following the subsequent normalization procedure.

$$p_v^n = \frac{s_{max}^n - s_v^n}{s_{max}^n - u_s^n} \cdot p_n, \quad (1)$$

where p_n is a hyperparameter governing the overall probability of node dropping, and s_{max}^n and u_s^n represent the maximum and mean values of s_v^n , respectively.

Attribute Masking Attribute masking in RAGCL is defined as substituting the feature vectors of a fraction of the nodes in V with a zero vector. The root node is exempt from this operation. Attribute masking does not entail node removal; hence, edges connected to masked nodes are retained. The mask probability for a node v is also p_v^n .

Edge Dropping We adopt the top-down graph of RPTs in RAGCL. Edge dropping involves setting a drop probability p_{uv}^e for each edge (u, v) , subsequently utilizing this probability to remove certain edges from the edge set E to produce an augmented view. p_{uv}^e should reflect the edge’s importance, implying that the p_{uv}^e of an essential edge should be lower than that of a less critical edge. Note that the centrality of a root node is assigned the minimum centrality value among all nodes in a graph. The importance score w_{uv}^e is defined as the mean centralities of its two connecting nodes.

$$w_{uv}^e = (\varphi_c(u) + \varphi_c(v))/2. \quad (2)$$

The drop probability is then derived based on the importance score of edge (u, v) . Analogously, we set $s_{uv}^e = \log w_{uv}^e$ to mitigate the impact of densely connected nodes. The probability is then ascertained similarly as follows.

$$p_{uv}^e = \frac{s_{max}^e - s_{uv}^e}{s_{max}^e - u_s^e} \cdot p_e, \quad (3)$$

where p_e is a hyperparameter utilized to regulate the overall probability of edge dropping, and s_{max}^e and u_s^e represent the maximum and mean values of s_{uv}^e , respectively.

Contrastive Loss Optimization

The data augmentation of a propagation graph G yields two augmented views, namely G_1 and G_2 . These views are processed through a GCN (Kipf and Welling 2016) encoder to obtain two representations: h_{G_1} and h_{G_2} . Within RAGCL, the unsupervised contrastive loss on the graph set \mathbb{G} , corresponding to the dataset \mathbb{C} , is formulated as follows:

$$\begin{aligned} \mathcal{L}_{unsup} = & -\mathbb{E}_{\mathbb{P}}[\text{sim}(h_{G_1}, h_{G_2})] \\ & + \mathbb{E}_{\mathbb{P}}[\log(\mathbb{E}_{\tilde{\mathbb{P}}} \exp(\text{sim}(h_{G_1}, h'_{G_2})))] \\ & + \log(\mathbb{E}_{\tilde{\mathbb{P}}} \exp(\text{sim}(h'_{G_1}, h_{G_2}))), \end{aligned} \quad (4)$$

where \mathbb{P} denotes the distribution adhered to by \mathbb{G} ; G represents an input sample drawn from \mathbb{P} ; G' is a negative sample drawn from $\tilde{\mathbb{P}} = \mathbb{P}$; $\text{sim}(x_1, x_2) = x_1^T x_2 / (\|x_1\| \|x_2\|)$ is the cosine similarity.

RAGCL employs \mathcal{L}_{unsup} as the regularization term of the supervised loss \mathcal{L}_{sup} (calculated by h_G), and optimizes the following loss function during the training phase.

$$\mathcal{L} = \mathcal{L}_{sup} + \lambda \cdot \mathcal{L}_{unsup}, \quad (5)$$

where λ is a tunable hyperparameter.

Experiments

In this section, we present main experimental results. Experiments on the effects of hyperparameters are in the supplementary material.

Experimental Configuration

We conducted experiments on four real-world benchmark datasets, Weibo, DRWeibo, Twitter15 and Twitter16, to evaluate RAGCL’s performance. Weibo and DRWeibo are Chinese binary classification datasets, and Twitter15 and Twitter16 are English multiple classification datasets. Table 1 shows the statistics of the datasets.

We make comparisons with the following baselines.

PLAN (Khoo et al. 2020) is based on Transformer. Its StA-PLAN version incorporates RPT structural information.

BiGCN (Bian et al. 2020) leverages two GCN encoders, a top-down and a bottom-up, and root node feature enhancement strategy to classify rumor.

UDGCN is a variant of BiGCN, it takes undirected graph of a RPT as the model input, and only one GCN encoder that applies root node feature enhancement strategy is used.

GACL (Sun et al. 2022b) performs rumor classification based on contrastive learning and adversarial training.

DDGCN (Sun et al. 2022a) can model multiple types of information in one unified framework.

The experiment setting details will be explained in the supplementary material. The experimental results are the average results of 10 random split of the datasets. We report the best performance of RAGCL that can be achieved with different node centrality and data augmentation combination. The source code of RAGCL is available at <https://github.com/CcQunResearch/RAGCL>.

Results and Discussion

Results in Table 2 and 3 show that RAGCL outperforms the baselines on all datasets. PLAN performs relatively poorly on all datasets and consumes more GPU resources due to Transformer architecture, which points to the necessity of adopting GNN architecture. BiGCN is a typical model built on the deep structure of RPT, which presupposes that the information flow in RPTs presents as a top-down propagation and a bottom-up dispersion process. However, our research findings indicate that the RPT actually manifests as a wide structure. This suggests that, for tree structures like RPT, in addition to the depth-directional information flow, the imbalanced distribution of information in the width direction is also an important characteristic, which is currently overlooked by existing techniques. Although GACL uses BERT (Devlin et al. 2018) to extract initial feature vectors, it does not improve significantly over other baselines. This may suggest that rumor detection models are insensitive to the way initial features are extracted, and what is more crucial is the high-level model’s ability to learn the interactions between nodes. Additionally, GACL utilizes supervised contrastive learning to learn the claim representation, while RAGCL, which adopts unsupervised contrastive loss, also achieves superior performance. The application of unsupervised loss allows the model to learn good representations without relying on labels. This suggests that it is feasible to use RAGCL to further enhance the rumor detection capability of the model by pretraining on large-scale, unlabeled dataset from social media platforms (such as UWeibo and UTwitter). We leave this for future research.

Ablation Study

We conducted a series of ablation experiments to verify the influence of different factors on the model performance.

Unresponded 1-level Replies In order to validate the impact of unresponded 1-level replies within RPTs, we conducted experiments on the four datasets depicted in Figure 4. We eliminated $\alpha\%$ of unresponded 1-level replies in each RPT, subsequently utilizing BiGCN (Bian et al. 2020) for classification. With the increase of α , it is observed that the model performance remains steady, or even improves to some extent. This indicates that these unresponded 1-level replies, as we previously conjectured, have less significance or may even serve as noise within rumor classification process, thus RAGCL is justified in dropping them.

Data Augmentation Combinations Table 4 presents the impact of different data augmentation combinations, where we report the accuracy for each dataset. The experimental results show that using attribute masking in Chinese datasets (Weibo and DRWeibo) will reduce the model performance. For English datasets, various data augmentation combinations have minimal effect on the results. Different data augmentation combinations all achieve significant performance gains over using only GCN for supervised classification without applying contrastive loss. Furthermore, the results also indicate that adaptive data augmentation outperforms random data augmentation, providing further validation of the reliability of our theory.

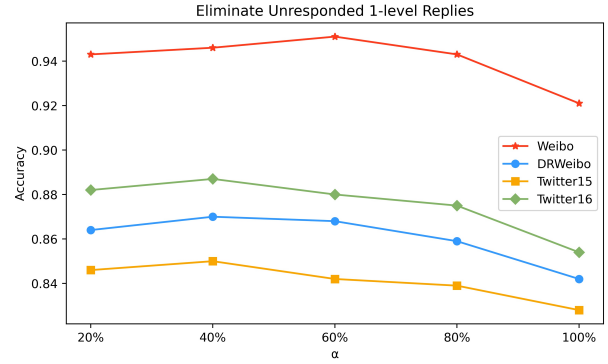


Figure 4: The influence of unresponded 1-level replies.

Node Centrality Measures We conducted the experiments in Table 5 to explore the influence of different node centrality measures. We report the accuracy that RAGCL achieves with different node centrality measures and the average time cost (in seconds) to calculate each RPT centrality. Degree centrality can be calculated rapidly, thus yielding efficient determination of node centrality within sizable datasets. However, its exclusive focus on edge number fails to satisfy Principle 3, thereby highlighting a limitation of degree centrality. For instance, a parent node and one of its children possessing identical reply counts will be assigned the same centrality. In fact, degree centrality also achieves relatively poor performance. Betweenness centrality aligns well with the three principles. For RPTs, the betweenness centrality is a very intuitive index to measure the importance of nodes. A node with numerous successor nodes will have many shortest paths traversing through it, leading to a correspondingly elevated betweenness centrality. However, the computation of betweenness centrality is more complex and time-intensive than the other measures. PageRank centrality, on the other hand, not only aligns well with the basic principles but also benefits from a relatively swift calculation process, making it more conducive to RAGCL’s training phase. We also examined the effect of eigenvector centrality, Katz centrality, and closeness centrality to verify the validity of our three guiding principles. Given their individual characteristics, these measures fail to meet Principle 2 and 3, resulting in subpar performance. Additionally, their computational complexity is relatively high. Therefore, we do not recommend using these centrality measures in RAGCL.

Graph Direction RAGCL is compatible with top-down and bottom-up directed graphs as well as undirected graphs. We investigated the impact of different types of graph in Figure 5. The results show that using undirected graphs leads to a performance decline. This could be due to the fact that during the forward propagation process of GNNs, densely connected nodes at the root node will see each other in their neighboring field of view. These nodes mutually aggregate each other’s information, ultimately resulting in a loss of node feature uniqueness, causing an over-smoothing problem (Li, Han, and Wu 2018; Cai and Wang 2020; Oono and Suzuki 2019). On the other hand, top-down and bottom-up

Method	Class	Weibo				DRWeibo			
		Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
PLAN	R	0.915±0.007	0.908	0.923	0.915	0.788±0.005	0.786	0.760	0.771
	N		0.923	0.907	0.914		0.793	0.813	0.802
BiGCN	R	0.942±0.008	0.919	0.968	0.942	0.866±0.010	0.869	0.849	0.858
	N		0.967	0.918	0.942		0.863	0.882	0.872
UDGCN	R	0.940±0.007	0.914	0.971	0.942	0.861±0.010	0.839	0.871	0.855
	N		0.969	0.910	0.938		0.882	0.852	0.867
GACL	R	0.938±0.006	0.936	0.940	0.938	0.870±0.009	0.865	0.856	0.860
	N		0.940	0.936	0.938		0.874	0.882	0.878
DDGCN	R	0.948±0.004	0.924	0.979	0.951	0.878±0.005	0.872	0.864	0.868
	N		0.976	0.917	0.946		0.883	0.891	0.887
RAGCL	R	0.962±0.005	0.956	0.968	0.962	0.894±0.004	0.893	0.877	0.885
	N		0.969	0.957	0.963		0.895	0.909	0.902

Table 2: Experimental results on Weibo and DRWeibo dataset.

Method	Twitter15					Twitter16				
	Acc.	N	F	T	U	Acc.	N	F	T	U
		F1	F1	F1	F1		F1	F1	F1	F1
PLAN	0.819±0.004	0.839	0.854	0.817	0.759	0.843±0.005	0.855	0.851	0.858	0.805
BiGCN	0.844±0.005	0.856	0.844	0.863	0.809	0.880±0.009	0.793	0.912	0.947	0.849
UDGCN	0.840±0.005	0.848	0.847	0.864	0.799	0.875±0.009	0.783	0.902	0.954	0.839
GACL	0.846±0.007	0.859	0.845	0.866	0.812	0.891±0.004	0.802	0.929	0.945	0.872
DDGCN	0.835±0.006	0.840	0.850	0.856	0.791	0.893±0.004	0.807	0.931	0.946	0.871
RAGCL	0.867±0.005	0.891	0.867	0.869	0.835	0.905±0.003	0.836	0.923	0.963	0.882

Table 3: Experimental results on Twitter15 and Twitter16 dataset.

	Aug1	Aug2	Weibo	DRWeibo	Twitter15	Twitter16
-	-	-	0.927	0.844	0.822	0.846
Node Dropping (random)	Attr Masking (random)		0.940	0.861	0.837	0.865
Node Dropping	Attr Masking		0.953	0.892	0.867	0.896
Node Dropping	Edge Dropping		0.962	0.894	0.864	0.902
Attr Masking	Edge Dropping		0.952	0.888	0.864	0.905

Table 4: The influence of combinations of data augmentation.

Centrality	T(n)	Weibo		Twitter15	
		Acc.	Time	Acc.	Time
Degree	$O(1)$	0.953	0.82	0.860	0.13
Betweenness	$O(n^3)$	0.958	8.12	0.867	1.34
PageRank	$O(n)$	0.962	1.37	0.865	0.22
Eigenvector	$O(n^3)$	0.939	9.23	0.850	1.62
Katz	$O(n^3)$	0.943	9.37	0.849	1.67
Closeness	$O(n^3)$	0.935	7.72	0.841	1.44

Table 5: The influence of node centrality measures.

directed graphs are able to effectively block excessive information flow between nodes at the root node.

Conclusion

This study introduces RAGCL, an adaptive graph contrastive learning method specifically for rumor detection. By taking into consideration the structural characteristics of

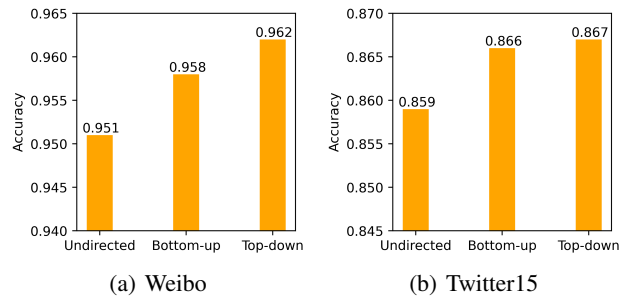


Figure 5: The impact of information flow direction.

RPTs, we propose three adaptive data augmentation methods based on node centrality and provide guiding principles for designing these methods. Our experimental results demonstrate that RAGCL surpasses current SOTA methods on all datasets, showcasing its superior performance.

Acknowledgments

The authors would like to thank all the anonymous reviewers for their help and insightful comments. This work is supported in part by the National Key R&D Program of China (2018AAA0100302) and the National Natural Science Foundation of China (61876016).

References

- Bian, T.; Xiao, X.; Xu, T.; Zhao, P.; Huang, W.; Rong, Y.; and Huang, J. 2020. Rumor detection on social media with bi-directional graph convolutional networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 549–556.
- Birunda, S. S.; and Devi, R. K. 2021. A Novel Score-Based Multi-Source Fake News Detection using Gradient Boosting Algorithm. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 406–414. IEEE.
- Bonacich, P. 1972. Factoring and weighting approaches to status scores and clique identification. *Journal of mathematical sociology*, 2(1): 113–120.
- Brin, S.; and Page, L. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7): 107–117.
- Cai, C.; and Wang, Y. 2020. A note on over-smoothing for graph neural networks. *arXiv preprint arXiv:2006.13318*.
- Castillo, C.; Mendoza, M.; and Poblete, B. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, 675–684.
- Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; and Ong, S. P. 2019. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*, 31(9): 3564–3572.
- Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, 1597–1607. PMLR.
- Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Fang, H.; Wang, S.; Zhou, M.; Ding, J.; and Xie, P. 2020. Cert: Contrastive self-supervised learning for language understanding. *arXiv preprint arXiv:2005.12766*.
- Freeman, L. C. 1977. A set of measures of centrality based on betweenness. *Sociometry*, 35–41.
- Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; and Dahl, G. E. 2017. Neural message passing for quantum chemistry. In *International conference on machine learning*, 1263–1272. PMLR.
- Giorgi, J.; Nitski, O.; Wang, B.; and Bader, G. 2020. Declutr: Deep contrastive learning for unsupervised textual representations. *arXiv preprint arXiv:2006.03659*.
- Hassani, K.; and Khasahmadi, A. H. 2020. Contrastive multi-view representation learning on graphs. In *International Conference on Machine Learning*, 4116–4126. PMLR.
- He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9729–9738.
- Jin, Z.; Cao, J.; Guo, H.; Zhang, Y.; and Luo, J. 2017. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In *Proceedings of the 25th ACM international conference on Multimedia*, 795–816.
- Karimi, H.; Roy, P.; Saba-Sadiya, S.; and Tang, J. 2018. Multi-source multi-class fake news detection. In *Proceedings of the 27th international conference on computational linguistics*, 1546–1557.
- Katz, L. 1953. A new status index derived from sociometric analysis. *Psychometrika*, 18(1): 39–43.
- Khoo, L. M. S.; Chieu, H. L.; Qian, Z.; and Jiang, J. 2020. Interpretable rumor detection in microblogs by attending to user interactions. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 8783–8790.
- Kipf, T.; Fetaya, E.; Wang, K.-C.; Welling, M.; and Zemel, R. 2018. Neural relational inference for interacting systems. In *International Conference on Machine Learning*, 2688–2697. PMLR.
- Kipf, T. N.; and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Kwon, S.; Cha, M.; Jung, K.; Chen, W.; and Wang, Y. 2013. Prominent features of rumor propagation in online social media. In *2013 IEEE 13th international conference on data mining*, 1103–1108. IEEE.
- Li, Q.; Han, Z.; and Wu, X.-M. 2018. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.
- Liu, Y.; and Wu, Y.-F. 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.
- Ma, J.; Gao, W.; Mitra, P.; Kwon, S.; Jansen, B. J.; Wong, K.-F.; and Cha, M. 2016. Detecting rumors from microblogs with recurrent neural networks.
- Ma, J.; Gao, W.; and Wong, K.-F. 2017. Detect rumors in microblog posts using propagation structure via kernel learning. Association for Computational Linguistics.
- Ma, J.; Gao, W.; and Wong, K.-F. 2018. Rumor detection on twitter with tree-structured recursive neural networks. Association for Computational Linguistics.
- Oono, K.; and Suzuki, T. 2019. Graph neural networks exponentially lose expressive power for node classification. *arXiv preprint arXiv:1905.10947*.
- Sabidussi, G. 1966. The centrality index of a graph. *Psychometrika*, 31(4): 581–603.
- Shaw, M. E. 1954. Group structure and the behavior of individuals in small groups. *The Journal of psychology*, 38(1): 139–149.

- Shi, J.; Liang, C.; Hou, L.; Li, J.; Liu, Z.; and Zhang, H. 2019. Deepchannel: Saliency estimation by contrastive learning for extractive document summarization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 6999–7006.
- Sun, F.-Y.; Hoffmann, J.; Verma, V.; and Tang, J. 2019. Info-graph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. *arXiv preprint arXiv:1908.01000*.
- Sun, M.; Zhang, X.; Zheng, J.; and Ma, G. 2022a. DDGCN: Dual Dynamic Graph Convolutional Networks for Rumor Detection on Social Media.
- Sun, T.; Qian, Z.; Dong, S.; Li, P.; and Zhu, Q. 2022b. Rumor Detection on Social Media with Graph Adversarial Contrastive Learning. In *Proceedings of the ACM Web Conference 2022*, 2789–2797.
- Velickovic, P.; Fedus, W.; Hamilton, W. L.; Liò, P.; Bengio, Y.; and Hjelm, R. D. 2019. Deep Graph Infomax. *ICLR (Poster)*, 2(3): 4.
- Wang, Y.; Ma, F.; Jin, Z.; Yuan, Y.; Xun, G.; Jha, K.; Su, L.; and Gao, J. 2018. 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*, 849–857.
- Wei, L.; Hu, D.; Zhou, W.; Yue, Z.; and Hu, S. 2021. Towards propagation uncertainty: Edge-enhanced bayesian graph convolutional networks for rumor detection. *arXiv preprint arXiv:2107.11934*.
- Xie, T.; and Grossman, J. C. 2018. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120(14): 145301.
- Yin, Y.; Wang, Q.; Huang, S.; Xiong, H.; and Zhang, X. 2022. Autogcl: Automated graph contrastive learning via learnable view generators. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 8892–8900.
- You, Y.; Chen, T.; Shen, Y.; and Wang, Z. 2021. Graph contrastive learning automated. In *International Conference on Machine Learning*, 12121–12132. PMLR.
- You, Y.; Chen, T.; Sui, Y.; Chen, T.; Wang, Z.; and Shen, Y. 2020. Graph contrastive learning with augmentations. *Advances in Neural Information Processing Systems*, 33: 5812–5823.
- Yu, F.; Liu, Q.; Wu, S.; Wang, L.; Tan, T.; et al. 2017. A Convolutional Approach for Misinformation Identification. In *IJCAI*, 3901–3907.
- Zhu, Y.; Xu, Y.; Yu, F.; Liu, Q.; Wu, S.; and Wang, L. 2021. Graph contrastive learning with adaptive augmentation. In *Proceedings of the Web Conference 2021*, 2069–2080.