Towards Real-World Rumor Detection: Anomaly Detection Framework with Graph Supervised Contrastive Learning

Chaoqun Cui, Caiyan Jia[*]

School of Computer Science and Technology & Beijing Key Lab of Traffic Data
Analysis and Mining Beijing Jiaotong University, Beijing 100044, China
ccqun19990728@gmail.com
cyjia@bjtu.edu.cn

Abstract

Current rumor detection methods based on propagation structure learning predominately treat rumor detection as a class-balanced classification task on limited labeled data. However, real-world social media data exhibits an imbalanced distribution with a minority of rumors among massive regular posts. To address the data scarcity and imbalance issues, we construct two large-scale conversation datasets from Weibo and Twitter and analyze the domain distributions. We find obvious differences between rumor and non-rumor distributions, with non-rumors mostly in entertainment domains while rumors concentrate in news, indicating the conformity of rumor detection to an anomaly detection paradigm. Correspondingly, we propose the Anomaly Detection framework with Graph Supervised Contrastive Learning (AD-GSCL). It heuristically treats unlabeled data as non-rumors and adapts graph contrastive learning for rumor detection. Extensive experiments demonstrate AD-GSCL's superiority under class-balanced, imbalanced, and few-shot conditions. Our findings provide valuable insights for real-world rumor detection featuring imbalanced data distributions.

1 Introduction

Numerous studies (Bian et al., 2020; Cui and Jia, 2024) highlight the value of propagation structures in revealing inter-post relations for rumor detection. Datasets like Weibo (Ma et al., 2016), DRWeibo (Cui and Jia, 2024), Twitter15, Twitter16 (Ma et al., 2017), and PHEME (Zubiaga et al., 2017) are utilized, containing source posts and replies. Propagation structure based methods convert a claim into a tree structure, with source posts as roots, replies as nodes, and reply relations as edges (see Appendix A.1 for examples). Graph Neural Networks (GNNs) then learn this tree to detect authenticity.

Contemporary rumor detection methods based on propagation structure learning typically treat rumor detection as a classification task, predominantly run on small-scale, class-balanced datasets. However, these techniques encounter several prevailing challenges: (1) Owing to effective post-screening measures conducted by social media platforms, numerous rumor posts are promptly identified and eradicated upon their publication. This results in difficulties in securing a sufficiently large corpus of labeled data; (2) Practically, rumor detection on social media platforms inherently entails the identification of a relatively minute quantity of rumor posts within a massive pool of posts, the vast majority of which are non-rumors. Therefore, the extant paradigm, founded on class-balanced classification tasks, lacks alignment with the genuine application scenarios; (3) Posts circulating on social media platforms traverse multiple diverse domains, each showcasing unique characteristics. Consequently, the data cross-domain nature introduces added complexities to the task. Our ambition is to address these three predominant issues.

In this study, we initially construct two large-scale unlabeled conversation datasets from Weibo and Twitter using web crawler, aligning with real-world data distributions. We then investigate the domain distribution of data on these platforms. Our findings lead us to heuristically frame rumor detection as an anomaly detection task rather than a conventional class-balanced classification task, more accurately reflecting real-world scenarios. Specifically, we propose the **A**nomaly **D**etection framework with **G**raph **S**upervised **C**ontrastive **L**earning (AD-GSCL). Guided by our findings, AD-GSCL treats large-scale unlabeled data as non-rumor and smartly upgrades existing graph contrastive learning methods (Khosla et al., 2020; Sun et al., 2019) to adapt to rumor detection's domain distribution.

In summary, this study contributes as follows:

---

[*]Corresponding author.

- We built two large-scale unlabeled conversation datasets to mitigate labeled data scarcity.

- We analyze social media's data domain distribution, highlighting significant discrepancies between rumor and non-rumor class data.

- We propose AD-GSCL, which employs an anomaly detection framework to align with real-world scenarios.

- Experiments demonstrate AD-GSCL's superior performance in multiple scenarios.

2 Related Work

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

2.1 Rumor Detection

Early studies in rumor detection used traditional classification with hand-crafted features (Castillo et al., 2011; Zhao et al., 2015; Vicario et al., 2019). Deep learning's success has spawned numerous methods that significantly enhance rumor detection performance. These methods can be roughly divided into four classes, including time-series based methods (Yu et al., 2017; Lu and Li, 2020; Bing et al., 2022) which model text content or user profiles as time series, propagation structure learning methods (Ma et al., 2018; Ma and Gao, 2020; Cui and Jia, 2024; Qiao et al., 2024) which consider the propagation structures of source claims and their replies, multi-source integation methods (Karimi et al., 2018; Yuan et al., 2019; Huang et al., 2020; 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 (Jin et al., 2017; Zhou et al., 2020; Peng and Xintong, 2022) which use posts' content and their related images to debunk rumors.

Propagation information's role is increasingly acknowledged. Numerous state-of-the-art (SOTA) models (Lin et al., 2021; Sun et al., 2022b) bank on learning the representations of rumor propagation trees utilizing GNNs. Traditional approaches treat rumor detection as a class-balanced task, diverging from real-world scenarios. Recent studies recognize the inherent data imbalance in practical settings and use large volumes of unlabeled data through anomaly detection or Positive-Unlabeled (P-U) learning frameworks. For example, Tam et al. (2019) unveiled a multi-modal approach that was capable of incrementally processing platform data streams. Nguyen et al. (2021) formulated a detection schema predicated on anomaly signal scoring. de Souza et al. (2021, 2022) leveraged P-U learning on heterogeneous networks. Moreover, Nguyen et al. (2023) introduced an efficacious streaming framework to comply with the latency bounds inherent in the real-time processing of large-scale social media data streams. Nonetheless, these methods neglected to effectively employ the propagation structure from user feedback (i.e., comments). In light of the above considerations, AD-GSCL tackles imbalanced data distribution with anomaly detection paradigm, learning from claim propagation structure information via GNNs, and adapting SOTA graph contrastive learning for rumor detection.

2.2 Graph Contrastive Learning

Deep learning advancements have led to progress in neural message passing algorithms (Gilmer et al., 2017), which set new SOTA benchmarks in various tasks (Kipf et al., 2018; Xie and Grossman, 2018; Chen et al., 2019) through supervised learning of graph representations. However, the challenge of acquiring labeled data has shifted focus to graph self-supervised learning methods (Hassani and Khasahmadi, 2020; Zhu et al., 2021; Hou et al., 2022), which utilize unlabeled data to learn robust node or graph representations. Most of them are graph contrastive learning methods including methods based on mutual information (MI) maximization (Velickovic et al., 2019; Sun et al., 2019) and graph augmentation (You et al., 2020, 2021). The former are usually trained by maximizing the MI between local representations and global representation. The latter firstly use different graph augmentation strategies to obtain different views of a given graph, and then construct positive and negative samples. Finally, graph representations are learned by minimizing the contrastive loss. These methods provide powerful ways to extract the discriminative features from rumor propagation trees. In order to adapt to different types of graph datasets, there are also some researches on adaptive graph augmentation (Zhu et al., 2021; You et al., 2021; Yin et al., 2022), which are able to automatically select optimal augmentation strategies for different datasets. AD-GSCL does not necessitate the design of intricate network architectures. Rather, it combines the existing supervised contrastive learning (Khosla et al., 2020) and unsupervised contrastive

learning (Sun et al., 2019) paradigms, and innovates upon these frameworks in alignment with the domain distribution of the datasets.

3   Analysis on Domain Distribution

Social media platforms, striving for a healthy network environment, regulate posts via automated and manual methods. The prompt removal of detected rumors results in a scarcity of rumor data, reflecting in the limited size of labeled datasets (Table 1). In our study, we reviewed the most recent 500 user-reported rumors displayed on Weibo rumor-refutation platform[1] from July 29, 2023, to April 5, 2024. Out of these, only 84 corresponding source posts remained undeleted. This indicates that social platforms like Weibo are very stringent in the control of rumors and misinformation.

For the issue of scarce labeled data, unsupervised or semi-supervised methods serve as apt solutions, as they harness massive unlabeled data to learn unsupervised representations. Several studies have embarked on rumor detection leveraging large-scale data streams from social media platforms for semi-supervised rumor detection (Tam et al., 2019; Nguyen et al., 2021, 2023), however, such streams typically do not include the propagation structure information. We utilized web crawlers to create two large-scale unlabeled datasets from Weibo and Twitter, termed Unlabeled Weibo (UWeibo) and Unlabeled Twitter (UTwitter). These datasets, derived from trending posts on both platforms, include complete claim propagation structures. Dataset construction details are in Appendix A.2, and Table 1 presents the statistics of the datasets harnessed in our study.

The intricate nature of the real world ensures that social posts cross multiple domains (like sport, science, culture, etc.), with substantial disparity in attributes and content structures among these domains (Lin et al., 2022; Ran and Jia, 2023). The inability of rumor detection models to efficaciously adapt to these domain-specific differences will markedly influence their performance. We utilize two unlabeled datasets and several labeled datasets to investigate the domain distributions of social media data and varying data classes. Specifically, we further divide major social media domains into three sections: (1) **Entertainment Section**: This includes domains unlikely to propagate misinformation or rumors, like personal life sharing, sports.

Most social media domains reside within this section. These domains typically contain less sensitive content with minimal public opinion impact. Their authenticity is usually non-controversial. (2) **Knowledge Section**: This covers domains disseminating knowledge across various fields like popular science, healthcare. Posts here often share specific knowledge, sometimes containing misinformation or falsified details. (3) **News Section**: This includes domains focusing on news content (e.g., social, financial news). These posts are often news-oriented, occasionally containing rumors.

For Weibo platform data, we ran web scraping to extract domain labels, enabling us to statistically analyze the domain and section distribution. Our statistical analysis of the domain and section distribution of UWeibo dataset is presented in Figure 1, providing an insight into the typical content structure of trending posts on Weibo platform. We also scrutinize the section distribution of non-rumor and rumor classes within Weibo dataset, as depicted in Figure 2. From this analysis, we discern that: (1) A significant majority of posts in UWeibo reside in Entertainment section (82.41%), with a substantial proportion of these being personal life sharing content (57.79%), which is also consistent with the main purpose of social media platforms; (2) A marked discrepancy is observed in the section distribution between the non-rumor and rumor classes in Weibo dataset, with the non-rumor class predominantly in Entertainment section (85.26%), whereas the Rumor class primarily features in News section (66.67%); (3) The section distribution of unlabeled data in UWeibo is very similar to the non-rumor in Weibo, i.e. most posts belong to Entertainment section that is less likely to generate rumors.

Informed by the aforementioned survey and some intuitive considerations, we determined a pivotal configuration in AD-GSCL, namely, the heuristic treating of unlabeled data from UWeibo and UTwitter as the non-rumor class. We justify this decision based on: (1) Intuitively, social media platforms monitor and regulate posts from users, so rumor posts can be deleted after being detected, thus rarely appearing among trending posts; (2) A considerable portion of trending platform posts often come from high-follower-count users, thus their credibility is ensured to some extent; (3) The majority of UWeibo belongs to Entertainment section that is less likely to generate rumors, and its section distribution is highly consistent with the non-rumor class. Moreover, data may be classified

---

[1]https://service.account.weibo.com/?type=5

Statistic	Weibo	DRWeibo	Twitter15	Twitter16	PHEME	UWeibo	UTwitter
language	zh	zh	en	en	en	zh	en
labeled	True	True	True	True	True	False	False
# claims	4664	6037	1490	818	6425	209549	204922
# non-rumors	2351	3185	374	205	4023	-	-
# false rumors	2313	2852	370	205	638	-	-
# true rumors	-	-	372	207	1067	-	-
# unverified rumors	-	-	374	201	697	-	-
# avg reply	803.5	61.8	50.2	49.1	15.4	45.0	83.9
# avg shallow (≤ 2) reply	692.2(86%)	59.1(95%)	41.4(83%)	37.6(76%)	10.8(70%)	46.6(92%)	70.0(85%)
# avg deep (> 2) reply	111.2(14%)	2.7(5%)	8.7(17%)	11.5(24%)	4.6(30%)	4.0(8%)	12.5(15%)

Table 1: Statistics on the distribution of dataset classes and node depth.

in a more fine-grained manner, such as Non-Rumor, True Rumor, False Rumor, and Unverified Rumor. To substantiate whether unlabeled data can be assigned to classes other than non-rumor, we also examined the data distribution of UTwitter and the section distributions of different classes in Twitter15 and Twitter16 (see Appendix A.3). This analysis reveals a similar section distribution in UTwitter to UWeibo (with Entertainment section posts being predominant), and that the section distribution of the non-rumor class significantly diverges from the other three rumor classes in Twitter15 and Twitter16, while being highly consistent with the UTwitter section distribution. These auxiliary findings underscore the broader applicability of our survey results across diverse platforms and support the rationale of designating unlabeled data as non-rumor. By adopting this configuration, AD-GSCL frames rumor detection within an anomaly detection paradigm—identifying the minority of rumor data (anomalous samples) within the majority of non-rumor data (normal samples).

4 Method

In this section, we present the design of AD-GSCL.

4.1 Notation

Rumor detection is a graph classification task whose goal is to predict the class of labeled claims with conversation context. Specifically, we denote the labeled claim set as $\mathbb{C}^L = \{c_1, c_2, \cdots, c_{N_L}\}$. Each 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 conversation context modeling as a propagation graph $G = (V, E)$, where V and E represent the set of nodes and edges, respectively. The set of propagation graphs corresponding to la-

beled claims is $\mathbb{G}^L = \{G_1, G_2, \cdots, G_{N_L}\}$. The initial feature matrix corresponding to V is $X \in \mathbb{R}^{|V| \times d_x}$, and d_x is the feature vector dimension. The initial feature of node can be selected widely such as Word2vec (Mikolov et al., 2013), tf-idf features, or vectors extracted from pre-trained language models (Liu et al., 2019; Clark et al., 2020). In addition, we denote the unlabeled claims set as $\mathbb{C}^U = \{c_{N_L+1}, c_{N_L+2}, \cdots, c_{N_L+N_U}\}$. There is no corresponding y for unlabeled claims, but only the propagation structure G for each claim. The propagation graph set corresponding to unlabeled claims is $\mathbb{G}^U = \{G_{N_L+1}, G_{N_L+2}, \cdots, G_{N_L+N_U}\}$. Our goal is to learn high-quality representations of graphs in $\mathbb{G}^L$ and $\mathbb{G}^U$, simultaneously train a classifier to predict the class of claims in $\mathbb{C}^L$.

4.2 Framework

AD-GSCL solely requires modification and integrations of existing contrastive losses, bypassing the need for intricate architectural designs. All claims within the unlabeled dataset are treated as non-rumor instances. Given the cross-domain nature of non-rumor claims, we refrain from constructing positive samples between non-rumor claims within the supervised contrastive loss. Hence, to secure a good representation of non-rumor instances, we employ unsupervised contrastive loss to learn the non-rumor samples within our unlabeled dataset. The framework of AD-GSCL is shown in Figure 3.

4.3 Supervised Contrastive Learning

In AD-GSCL, rumor data is sourced from labeled benchmark datasets, whereas non-rumor data derives from both labeled and unlabeled datasets. AD-GSCL employs supervised contrastive learning to enhance the inter-class representation dissimilarity and intra-class representation similarity of different class claims in the feature space. Specifically, we
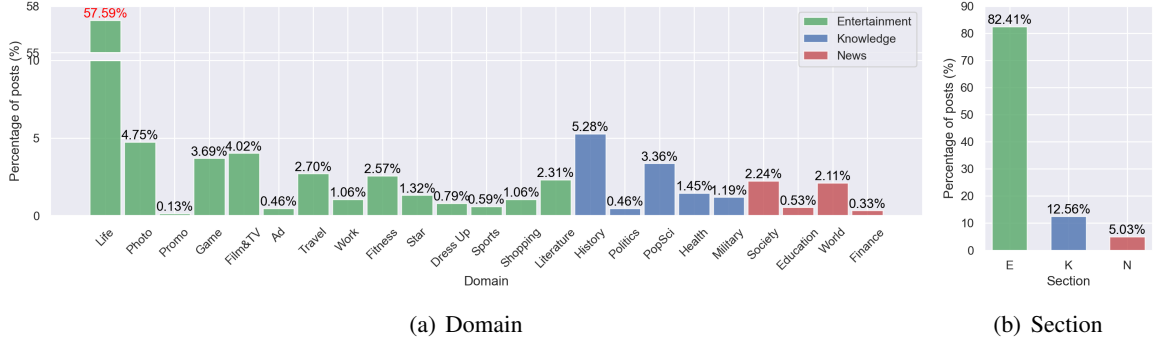
(a) Domain

(b) Section

Figure 1: Domain and section distribution of posts on UWeibo dataset. E, K, and N represent the three sections.

(a) Non-Rumor

(b) Rumor

Figure 2: Post section distribution on Weibo dataset.

designate the claims within each class, excluding those from non-rumor classes, as positive samples for each other, whereas claims spanning different classes are treated as negative samples. Notably, non-rumor class claims are not deemed as positive samples. The rationale behind this methodology of positive sample construction is backed by our survey findings depicted in Figure 1. The UWeibo data mainly belongs to Entertainment section. The vast majority of domains in social media belong to this section, indicating that the UWeibo data has very broad cross-domain characteristics. Consequently, it would be unreasonable to pull these data with severe cross-domain issues closer together as positive samples in the feature space. In contrast, rumor claims tend to mainly fall under News section and to a lesser extent, Knowledge section. These sections typically feature a smaller number of domains, thus implying a lesser degree of cross-domain characteristics. Moreover, user replies to rumor claims frequently exhibit stronger stances and emotional expressions (Ma et al., 2018; Wei et al., 2021), thus yielding more prominent domain invariant features. Consequently, the construction of positive samples for the supervised contrastive loss in AD-GSCL

relies solely on these rumor claims.

Specifically, for labeled dataset $\mathbb{C}^L$ and unlabeled dataset $\mathbb{C}^U$, the graph representation of a propagation graph $G \in \mathbb{G}^L \cup \mathbb{G}^U$ obtained by GCN encoder (Kipf and Welling, 2016) is denoted as h_G. The supervised contrastive loss is

$$\mathcal{L}_{scl}(\mathbb{C}^L \cup \mathbb{C}^U) =$$
$$-\frac{1}{|\mathbb{G}_R^L|} \sum_{G \in \mathbb{G}_R^L} log \left\{ \frac{1}{|P(G)|} \frac{\sum_{G_p} exp(sim(h_G, h_{G_p})\tau)}{\sum_{G_n} exp(sim(h_G, h_{G_n})\tau)} \right\},$$
(1)

where $\mathbb{G}_R^L$ is the set of rumor class claim propagation graphs in dataset $\mathbb{C}^L$; $p \in P(G)$ and $n \in N(G)$ are indices of positive and negative samples respectively; $P(G) \equiv \{p : G_p \in \mathbb{G}^L, y_{G_p} = y_G\}$ and $N(G) \equiv \{n : G_n \in \mathbb{G}^L \cup \mathbb{G}^U, y_{G_n} \neq y_G\}$ are the sets of positive and negative sample indices respectively; $sim(\cdot)$ is the cosine similarity; $\tau \in \mathbb{R}^+$ is a scalar temperature parameter.

4.4 Anchor Representation Learning

Supervised contrastive learning primarily aims to use massive non-rumor claims (normal samples) as anchor samples, reducing the similarity between the representations of rumor claims (anomalous samples) and anchor samples in feature space, thereby aiding rumor classification. Consequently, learning robust representations for anchor samples is crucial. We previously asserted that increasing the representation similarity among non-rumor claims by considering them as mutual positive samples in supervised contrastive learning may not be logical, given the diverse domains these claims may originate from. Consequently, we resort to unsupervised contrastive learning in AD-GSCL to learn the representations of anchor samples. In particular, a graph contrastive learning approach based on MI maximization (Sun et al., 2019) has been employed. The positive samples of unsupervised contrastive
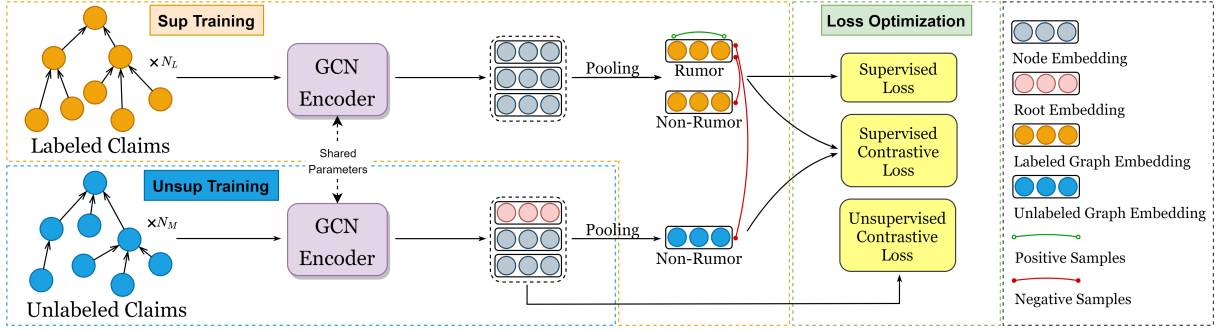
7145

Figure 3: The framework of AD-GSCL.

learning are derived from both global and local representations of an individual instance, ensuring the absence of amplified representation similarity between claims of divergent domains. As reported in Table 1, most nodes are located at shallow positions in the propagation tree of claims in social media, i.e., they are very close to the root node. During the forward propagation in GNN, nodes aggregate their k-hop neighborhood nodes features (Gilmer et al., 2017; Kipf and Welling, 2016). In AD-GSCL, we employ the bottom-up propagation tree, so the root node can aggregate the features of most nodes in the graph. Therefore, we directly use the root representation as the global representation to compute the MI maximization contrastive loss with node representations, which is another specific design in AD-GSCL. Since it is crucial for rumor detection to model the interaction between source posts and replies (Bian et al., 2020; Sun et al., 2022b), this design that emphasizes the importance of the root node is more suitable for this task.

Specifically, for a claim in $\mathbb{C}^U$ with propagation tree $G = (V, E)$, the node representations and the corresponding root representation obtained from the encoder are denoted as h_v ($v \in V$) and h_r. We use Jensen-Shannon MI estimator (Nowozin et al., 2016) to calculate the MI as below.

$$
\begin{aligned}
I(h_v(G); h_G) := \ &\mathbb{E}_{\mathbb{P}}[-sp(-T(h_v(G), h_r(G)))] \\
&-\mathbb{E}_{\mathbb{P} \times \tilde{\mathbb{P}}}[sp(T(h_v(G'), h_r(G)))],
\end{aligned} \tag{2}
$$

where $\mathbb{P}$ is the distribution followed by $\mathbb{G}^U$; G is an input graph sampled from $\mathbb{P}$; G' is a negative sample sampled from $\tilde{\mathbb{P}} = \mathbb{P}$; T is a discriminator; $sp(z) = log(1 + e^z)$ is the softplus function. In practice, we use all combinations of root representations and node representations of all graph instances within a batch to generate negative samples. The unsupervised contrastive loss on $\mathbb{C}^U$ is

$$
\mathcal{L}_{ucl}(\mathbb{C}^U) = -\frac{1}{N_U} \sum_{G \in \mathbb{G}^U} \sum_{v \in V} I(h_v(G); h_r(G)). \tag{3}
$$

Algorithm 1 Pre-training Strategy Optimization

Input: initial model parameter $\theta^{(0)}$, pre-training step N, fine-tuning step M.
Output: optimized model parameter $\theta^{(N+M)}$.
1: // Pre-training on $\mathbb{C}^U$.
2: **for** $n = 1$ to N **do**
3:     Minimize $\mathcal{L}_{pre}$ to update $\theta^{(n)}$.
4: **end for**
5: // Fine-tuning on $\mathbb{C}^L$ and $\mathbb{C}^U$.
6: **for** $m = 1$ to **M do**
7:     Minimize $\mathcal{L}_{ft}$ to update $\theta^{(N+m)}$.
8: **end for**
9: **return** $\theta^{(N+M)}$.

4.5 Training Strategy

AD-GSCL can be trained with pre-training strategy and semi-supervised strategy. Different training strategies will minimize $\mathcal{L}_{sup}(\mathbb{C}^L)$, $\mathcal{L}_{scl}(\mathbb{C}^L \cup \mathbb{C}^U)$ and $\mathcal{L}_{ucl}(\mathbb{C}^U)$ in different ways.

4.5.1 Pre-Training Strategy

AD-GSCL can be pretrained on unlabeled dataset $\mathbb{C}^U$ to minimize $\mathcal{L}_{ucl}(\mathbb{C}^U)$. The process is aimed at learning more robust representations of non-rumor claims so that they can better serve as anchors:

$$
\mathcal{L}_{pre} = \mathcal{L}_{ucl}(\mathbb{C}^U). \tag{4}
$$

AD-GSCL will continue to minimize $\mathcal{L}_{scl}(\mathbb{C}^L \cup \mathbb{C}^U)$ after pre-training to learn discriminative features of different class claims, and also minimize cross-entropy classification loss $\mathcal{L}_{sup}(\mathbb{C}^L)$, in which $\mathcal{L}_{scl}(\mathbb{C}^L \cup \mathbb{C}^U)$ is a regularization term. This process is similar to fine-tuning, whose loss is

$$
\mathcal{L}_{ft} = \mathcal{L}_{sup}(\mathbb{C}^L) + \lambda \cdot \mathcal{L}_{scl}(\mathbb{C}^L \cup \mathbb{C}^U), \tag{5}
$$

where λ is an adjustable hyperparameter. The whole process is shown in Algorithm 1.

4.5.2 Semi-Supervised Strategy

AD-GSCL uses the combination of all the loss functions for semi-supervised learning. In this case, $\mathcal{L}_{sup}(\mathbb{C}^L)$ is the main loss, $\mathcal{L}_{scl}(\mathbb{C}^L \cup \mathbb{C}^U)$ and $\mathcal{L}_{ucl}(\mathbb{C}^U)$ are regularization terms, then the loss function of semi-supervised training strategy is

$$\mathcal{L}_{total} = \mathcal{L}_{sup}(\mathbb{C}^L) + \alpha \cdot \mathcal{L}_{scl}(\mathbb{C}^L \cup \mathbb{C}^U) \\ + \beta \cdot \mathcal{L}_{ucl}(\mathbb{C}^U), \quad (6)$$

where α and β are adjustable hyperparameters.

5 Experiments

We present main experiments in this section. See Appendix C for further Evaluations.

5.1 Experimental Settings

We conducted experiments on datasets Weibo, DR-Weibo, Twitter15, Twitter16, and PHEME. Table 1 shows the statistics of all the datasets. UWeibo and UTwitter are available at `https://github.com/CcQunResearch/UWeibo` and `https://github.com/CcQunResearch/UTwitter`.

We compare with the following baselines:

PLAN (Khoo et al., 2020) uses Transformer architecture to detect rumors.

BiGCN (Bian et al., 2020) leverages two bidirectional GCN encoders and root node feature enhancement strategy.

GACL (Sun et al., 2022b) utilizes contrastive learning and adversarial training to classify rumor.

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

RAGCL (Cui and Jia, 2024) uses contrastive learning with adaptive data augmentation.

Experiment setting details are in Appendix B. Weibo and DRWeibo datasets are used alongside UWeibo, and Twitter15, Twitter16, and PHEME are paired with UTwitter. The code is at `https://github.com/CcQunResearch/AD-GSCL`.

5.2 Class-Balanced Classification

We conducted classification experiments on class-balanced datasets of Weibo, DRWeibo, Twitter15, and Twitter16, maintaining class-balanced test sets. Average results from 10 random splits of labeled datasets are reported in Table 2 and 3. We evaluated AD-GSCL's performance under two training strategies: pre-training ('p') and semi-supervised training ('s'). AD-GSCL outperformed the compared baseline methods across all datasets, achieving an accuracy improvement of 1.2-2.8%. PLAN,

using the Transformer architecture, underperform and consume significant GPU resources, highlighting the need for GNN architecture. While GACL, utilizing BERT (Devlin et al., 2018) for initial feature extraction, exhibits negligible improvement over other baselines. This observation suggests that the mechanism of initial feature extraction might be less critical to rumor detection models compared to the high-level model's capacity to learn node interactions. Traditional methods using GNNs and contrastive learning face issues like overfitting, over-smoothing, and cross-domain discrepancies on limited-scale datasets. In stark contrast, AD-GSCL utilizes large-scale unlabeled data to mitigate the limitations of small data scale, adjusts the data distribution to align with real-world scenarios, and addresses cross-domain discrepancies, thus contributing to improved performance.

5.3 Class-Imbalanced Classification

To validate the efficacy of AD-GSCL in practical applications, we ran class-imbalanced experiments. The class-imbalanced PHEME was selected for the tests. Moreover, to construct class-imbalanced test sets for Weibo and DRWeibo, we expanded the non-rumor class scale by incorporating 3% of UWeibo data (which is excluded from training) into these two dataset's test sets. The results are shown in Table 4, including the Area Under the Curve (AUC) and macro F1 score. AD-GSCL outperformed baseline methods, showing enhanced adaptability to real-world scenarios due to our design catering to real-world data domain distribution. AD-GSCL shows a larger improvement over baseline methods in class-imbalanced scenarios compared to class-balanced scenarios. We posit that rumor detection research should utilize datasets reflecting actual data distributions, as it's essential for developing effective models that perform efficiently in real-world environments. From our experiments, we noted that the semi-supervised strategy performs better under class-balanced conditions, while the pre-training strategy proves more effective under class-imbalanced conditions. Therefore, the class distribution within a dataset can significantly guide the selection of the training strategy.

5.4 Ablation Study

Some ablation studies have been performed to examine the role of various loss components and unlabeled data, as shown in Table 5. Experiments were conducted under class-balanced condition and

Method	Class	Weibo				DRWeibo			
		Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
PLAN	R	91.5±0.7	90.8	92.3	91.5	78.8±0.5	78.6	76.0	77.1
	N		92.3	90.7	91.4		79.3	81.3	80.2
BiGCN	R	94.2±0.8	91.9	96.8	94.2	86.6±1.0	86.9	84.9	85.8
	N		96.7	91.8	94.2		86.3	88.2	87.2
GACL	R	93.8±0.6	93.6	94.0	93.8	87.0±0.9	86.5	85.6	86.0
	N		94.0	93.6	93.8		87.4	88.2	87.8
DDGCN	R	94.8±0.4	92.4	**97.9**	95.1	87.8±0.5	87.2	86.4	86.8
	N		**97.6**	91.7	94.6		88.3	89.1	88.7
RAGCL	R	95.9±0.4	95.3	96.5	95.9	89.4±0.4	89.4	87.7	88.5
	N		96.6	95.4	96.0		89.5	90.9	90.2
AD-GSCL(p)	R	**97.1±0.2**	**96.9**	97.1	**97.0**	91.3±0.4	91.6	89.6	90.6
	N		97.2	**97.0**	**97.1**		91.2	92.9	92.0
AD-GSCL(s)	R	97.0±0.3	96.7	97.1	96.9	**92.2±0.3**	**92.3**	90.7	91.5
	N		97.2	96.8	97.0		92.1	93.5	92.8

Table 2: Weibo and DRWeibo experimental results on class-balanced test sets.

Method	Twitter15					Twitter16				
	Acc.	N F1	F F1	T F1	U F1	Acc.	N F1	F F1	T F1	U F1
PLAN	81.9±0.4	83.9	85.4	81.7	75.9	84.3±0.5	**85.5**	85.1	85.8	80.5
BiGCN	84.4±0.5	85.6	84.4	86.3	80.9	88.0±0.9	79.3	91.2	94.7	84.9
GACL	84.6±0.7	85.9	84.5	86.6	81.2	89.1±0.4	80.2	92.9	94.5	87.2
DDGCN	83.5±0.6	84.0	85.0	85.6	79.1	89.3±0.4	80.7	93.1	94.6	87.1
RAGCL	85.9±0.5	88.3	85.9	85.1	**83.7**	90.0±0.3	83.1	91.8	95.8	87.7
AD-GSCL(p)	86.7±0.5	**90.2**	88.3	84.6	83.1	90.8±0.3	82.8	**93.5**	95.6	**89.7**
AD-GSCL(s)	**87.4±0.4**	88.6	**89.1**	**88.1**	83.6	**91.3±0.3**	84.5	92.6	**97.9**	89.3

Table 3: Twitter15 and Twitter16 experimental results on class-balanced test sets.

Method	Weibo		DRWeibo		PHEME	
	AUC	F1	AUC	F1	AUC	F1
PLAN	84.5	86.6	79.0	83.2	64.5	67.1
BiGCN	93.5	93.0	80.5	87.1	76.5	78.5
GACL	89.6	91.7	82.0	86.8	74.0	78.2
DDGCN	90.4	93.6	79.6	86.5	77.5	79.1
RAGCL	91.8	92.2	79.4	83.6	74.4	78.5
AD-GSCL(p)	**96.7**	95.6	**90.6**	**92.1**	**78.1**	**82.3**
AD-GSCL(s)	96.2	**96.0**	89.6	91.7	78.0	81.8

Table 4: Results on class-imbalanced test sets.

$\mathcal{L}_{scl}$	$\mathcal{L}_{ucl}$	Unlabeled Data	Dataset	
			Weibo	DRWeibo
			93.2±0.6	86.5±0.7
✓			95.2±0.7	89.8±0.6
✓		✓	96.1±0.5	91.1±0.4
✓	✓	✓	**97.0±0.3**	**92.2±0.3**

Table 5: The influence of loss and unlabeled data.

semi-supervised strategy. The results show that the supervised contrastive loss and unlabeled data can improve the performance of GNNs. The unsupervised contrastive loss can further improve performance by making massive non-rumor class claims serve as better anchors. These results demonstrate that the modifications made to existing graph contrastive losses in AD-GSCL make it better suited for the domain distribution in rumor detection.

5.5 Few-shot Rumor Detection

In a class-balanced setting, we tested BiGCN, GACL, and AD-GSCL on Weibo dataset for few-shot learning (Figure 4). As rumors are typically removed once detected, the accumulation of a large-scale labeled dataset becomes challenging, thereby rendering the exploration of rumor detection in few-shot settings as highly significant. Our experiments were conducted with varying numbers of labeled samples, denoted as k, including 10, 20, 40, 80, 100, 200, 300, and 500. An initial pre-training
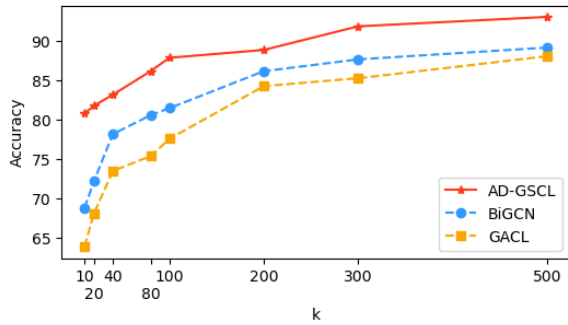
Figure 4: Results of few-shot experiments.

phase was conducted on UWeibo, followed by fine-tuning on the labeled samples. The results suggest that, even with a few labeled samples, AD-GSCL can effectively utilize unlabeled data to learn superior claim representations. This leads to appreciable performance, reinforcing the potential utility of AD-GSCL in practical, data-scarce scenarios.

6 Conclusion

In this study, we analyzed the distribution characteristics of rumor detection datasets and real scenarios. We introduced AD-GSCL, adapting graph contrastive learning for rumor detection. Our extensive experiments demonstrated AD-GSCL's superiority over baselines. The proposed paradigm provides an impactful approach and modeling strategies for handling imbalanced distributions.

Ethical Statement

In this research, we place a high emphasis on ethical considerations. We utilized web crawling tools to collect data from publicly posted content by users on the Weibo and Twitter platforms, which is visible to any user of these platforms. We will process the final public data by removing personally identifiable information to ensure that no individual can be identified from our dataset. We explicitly state that the sole purpose of collecting and analyzing data is for academic research, aimed at improving the quality of information on social media and reducing the spread of rumors. By employing semi-supervised learning methods, we are able to enhance the performance of models under conditions of scarce labeled data, which holds significant academic and social value for rumor detection research. We commit to adhering to ethical principles throughout the research process, ensuring that the conduct of the research complies with legal requirements and moral standards, and treating our data,

participants, and society with responsibility.

Limitations

AD-GSCL is a semi-supervised method that applies large-scale unlabeled datasets. Compared with the existing methods based on propagation structure learning, it needs to process more large-scale data, so it will occupy more computing resources. Furthermore, our research only investigated the distribution of data on social media platforms during typical periods. However, in certain countries or regions during special circumstances (such as pandemics, elections, or wars), the distribution of data may change.

Acknowledgments

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.

Changsong Bing, Yirong Wu, Fangmin Dong, Shouzhi Xu, Xiaodi Liu, and Shuifa Sun. 2022. Dual co-attention-based multi-feature fusion method for rumor detection. *Information*, 13(1):25.

S Selva Birunda and R Kanniga Devi. 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)*, pages 406–414. IEEE.

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684.

Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. 2019. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*, 31(9):3564–3572.

Kevin Clark, Minh-Thang Luong, Quoc V Le, and Christopher D Manning. 2020. Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555*.

Chaoqun Cui and Caiyan Jia. 2024. Propagation tree is not deep: Adaptive graph contrastive learning approach for rumor detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 73–81.

Mariana C de Souza, Bruno M Nogueira, Rafael G Rossi, Ricardo M Marcacini, and Solange O Rezende. 2021. A heterogeneous network-based positive and unlabeled learning approach to detect fake news. In *Brazilian Conference on Intelligent Systems*, pages 3–18. Springer.

Mariana Caravanti de Souza, Bruno Magalhães Nogueira, Rafael Geraldeli Rossi, Ricardo Marcondes Marcacini, Brucce Neves Dos Santos, and Solange Oliveira Rezende. 2022. A network-based positive and unlabeled learning approach for fake news detection. *Machine learning*, 111(10):3549–3592.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. 2017. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR.

Kaveh Hassani and Amir Hosein Khasahmadi. 2020. Contrastive multi-view representation learning on graphs. In *International Conference on Machine Learning*, pages 4116–4126. PMLR.

Zhenyu Hou, Xiao Liu, Yuxiao Dong, Chunjie Wang, Jie Tang, et al. 2022. Graphmae: Self-supervised masked graph autoencoders. *arXiv preprint arXiv:2205.10803*.

Qi Huang, Junshuai Yu, Jia Wu, and Bin Wang. 2020. Heterogeneous graph attention networks for early detection of rumors on twitter. In *2020 international joint conference on neural networks (IJCNN)*, pages 1–8. IEEE.

Zhiwei Jin, Juan Cao, Han Guo, Yongdong Zhang, and Jiebo Luo. 2017. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 795–816.

Hamid Karimi, Proteek Roy, Sari Saba-Sadiya, and Jiliang Tang. 2018. Multi-source multi-class fake news detection. In *Proceedings of the 27th international conference on computational linguistics*, pages 1546–1557.

Ling Min Serena Khoo, Hai Leong Chieu, Zhong Qian, and Jing Jiang. 2020. Interpretable rumor detection in microblogs by attending to user interactions. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8783–8790.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Thomas Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, and Richard Zemel. 2018. Neural relational inference for interacting systems. In *International Conference on Machine Learning*, pages 2688–2697. PMLR.

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

Hongzhan Lin, Jing Ma, Liangliang Chen, Zhiwei Yang, Mingfei Cheng, and Guang Chen. 2022. Detect rumors in microblog posts for low-resource domains via adversarial contrastive learning. *arXiv preprint arXiv:2204.08143*.

Hongzhan Lin, Jing Ma, Mingfei Cheng, Zhiwei Yang, Liangliang Chen, and Guang Chen. 2021. Rumor detection on twitter with claim-guided hierarchical graph attention networks. *arXiv preprint arXiv:2110.04522*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

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 and Wei Gao. 2020. Debunking rumors on twitter with tree transformer. ACL.

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.

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.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Thanh Tam Nguyen, Thanh Trung Huynh, Hongzhi Yin, Matthias Weidlich, Thanh Thi Nguyen, Thai Son Mai, and Quoc Viet Hung Nguyen. 2023. Detecting rumours with latency guarantees using massive streaming data. *The VLDB Journal*, 32(2):369–387.

Thanh Toan Nguyen, Thanh Tam Nguyen, Thanh Thi Nguyen, Bay Vo, Jun Jo, and Quoc Viet Hung Nguyen. 2021. Judo: Just-in-time rumour detection in streaming social platforms. *Information Sciences*, 570:70–93.

Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. 2016. f-gan: Training generative neural samplers using variational divergence minimization. *Advances in neural information processing systems*, 29.

Xu Peng and Bao Xintong. 2022. An effective strategy for multi-modal fake news detection. *Multimedia Tools and Applications*, 81(10):13799–13822.

Yuhan Qiao, Chaoqun Cui, Yiying Wang, and Caiyan Jia. 2024. A debiased self-training framework with graph self-supervised pre-training aided for semi-supervised rumor detection. *Neurocomputing*, 604:128314.

Hongyan Ran and Caiyan Jia. 2023. Unsupervised cross-domain rumor detection with contrastive learning and cross-attention. *arXiv preprint arXiv:2303.11945*.

Fan-Yun Sun, Jordan Hoffmann, Vikas Verma, and Jian Tang. 2019. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. *arXiv preprint arXiv:1908.01000*.

Mengzhu Sun, Xi Zhang, Jiaqi Zheng, and Guixiang Ma. 2022a. Ddgcn: Dual dynamic graph convolutional networks for rumor detection on social media.

Tiening Sun, Zhong Qian, Sujun Dong, Peifeng Li, and Qiaoming Zhu. 2022b. Rumor detection on social media with graph adversarial contrastive learning. In *Proceedings of the ACM Web Conference 2022*, pages 2789–2797.

Nguyen Thanh Tam, Matthias Weidlich, Bolong Zheng, Hongzhi Yin, Nguyen Quoc Viet Hung, and Bela Stantic. 2019. From anomaly detection to rumour detection using data streams of social platforms. *Proceedings of the VLDB Endowment*, 12(9):1016–1029.

Petar Velickovic, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. 2019. Deep graph infomax. *ICLR (Poster)*, 2(3):4.

Michela Del Vicario, Walter Quattrociocchi, Antonio Scala, and Fabiana Zollo. 2019. Polarization and fake news: Early warning of potential misinformation targets. *ACM Transactions on the Web (TWEB)*, 13(2):1–22.

Lingwei Wei, Dou Hu, Wei Zhou, Zhaojuan Yue, and Songlin Hu. 2021. Towards propagation uncertainty: Edge-enhanced bayesian graph convolutional networks for rumor detection. *arXiv preprint arXiv:2107.11934*.

Jiaying Wu and Bryan Hooi. 2022. Probing spurious correlations in popular event-based rumor detection benchmarks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 274–290. Springer.

Tian Xie and Jeffrey C Grossman. 2018. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120(14):145301.

Yihang Yin, Qingzhong Wang, Siyu Huang, Haoyi Xiong, and Xiang Zhang. 2022. Autogcl: Automated graph contrastive learning via learnable view generators. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8892–8900.

Yuning You, Tianlong Chen, Yang Shen, and Zhangyang Wang. 2021. Graph contrastive learning automated. In *International Conference on Machine Learning*, pages 12121–12132. PMLR.

Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. 2020. Graph contrastive learning with augmentations. *Advances in Neural Information Processing Systems*, 33:5812–5823.

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.

Zhe Zhao, Paul Resnick, and Qiaozhu Mei. 2015. Enquiring minds: Early detection of rumors in social media from enquiry posts. In *Proceedings of the 24th international conference on world wide web*, pages 1395–1405.

Xinyi Zhou, Jindi Wu, and Reza Zafarani. 2020. Safe: Similarity-aware multi-modal fake news detection. In *Pacific-Asia Conference on knowledge discovery and data mining*, pages 354–367. Springer.

Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. 2021. Graph contrastive learning with adaptive augmentation. In *Proceedings of the Web Conference 2021*, pages 2069–2080.

Arkaitz Zubiaga, Maria Liakata, and Rob Procter. 2017. Exploiting context for rumour detection in social media. In *International conference on social informatics*, pages 109–123. Springer.

A Dataset Construction and Statistics

In this section, we will elaborate on the construction of large-scale unlabeled datasets from the Weibo and Twitter platforms, as well as the methodology employed for the statistical analysis of domain distributions within the datasets. Additionally, we will supplement the domain distribution statistics for UTwitter, Twitter15, and Twitter16 datasets.

A.1 Claim Propagation Trees

Figure 5 illustrates two examples of claim propagation trees sourced from the Twitter platform. In these examples, the responses to rumor claims and non-rumor claims reveal notable differences in both stance and sentiment. These characteristics are crucial for effective rumor identification.
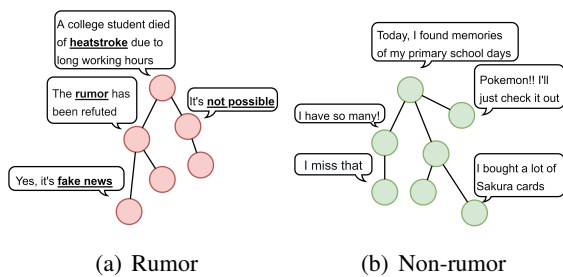

(a) Rumor   (b) Non-rumor

Figure 5: Examples of propagation trees. Comments under rumor thread typically express more heated stances.

A.2 Unlabeled Dataset Construction

For the UWeibo dataset, we employed web crawler techniques to randomly collect trending posts and their complete propagation structures from the homepage of popular Weibo posts[2]. To ensure the dataset's integrity and independence from platform recommendation algorithms, we utilized multiple newly created accounts to extract data. This approach aimed to mitigate potential biases that might arise from the platform's algorithms and to reflect the genuine domain distribution of social media content. Regarding UTwitter dataset, we initially utilized multiple newly created accounts to randomly follow high-follower count influencers. Subsequently, we conducted random crawling of posts and their propagation structures from the Twitter homepage[3]. Due to the fact that UTwitter dataset is exclusively sourced from users with a substantial number of followers,

the authenticity of the posts is more likely to be ensured. The data in the UTwitter dataset spans from October 2022 to April 2023, while the data in the UWeibo dataset covers from March 2022 to April 2023. The data is uniformly distributed over time. The code for the web scraping program can be found at `https://github.com/CcQunResearch/WeiboPostAndCommentCrawl` and `https://github.com/CcQunResearch/TwitterPostAndCommnetCrawl`.

Due to the stringent regulation imposed by platforms on the dissemination of rumors, acquiring a sufficiently large-scale labeled dataset for rumor detection proves to be exceptionally challenging. Conversely, obtaining extensive amounts of unlabeled data is relatively simpler, especially with the availability of platform data APIs offered by certain mainstream social media platforms (e.g., Twitter API). Consequently, we posit that future research should place greater emphasis on semi-supervised rumor detection methods.

A.3 Domain Statistical Methodology

For the Twitter platform data, we manually performed the statistical analysis of its distribution. For UTwitter dataset, we specifically examined the domain distribution of a subset comprising 1000 data instances. Additionally, we analyzed the section distribution of the union of Twitter15 and Twitter16 datasets. Given that these two datasets operate at the event level(Wu and Hooi, 2022), with multiple claims associated with the same event, we implemented a deduplication process based on overlapping claims and events found within the two datasets.

The domain and section distributions for the UTwitter dataset, as well as the section distribution for the union of Twitter15 and Twitter16 datasets, are illustrated in Figure 6 and Figure 7, respectively. Specifically, similar to the domain distribution observed on Weibo platform, the following trends were identified: (1) As for UTwitter Dataset, the majority of posts on the Twitter platform belong to Entertainment section (79.11%), with a substantial portion consisting of personal life-sharing content (46.91%). (2) The section distribution between non-rumor and rumor data in the labeled Twitter15 and Twitter16 datasets shows significant differences. Specifically, the posts in all three Rumor classes are predominantly associated with News section (77.78%, 77.05%, 86.84%). (3) The section distribution of the unlabeled data in UTwitter exhibits

---

striking similarity to the section distribution of non-rumor data in Twitter15 and Twitter16. In other words, the majority of posts in both cases belong to Entertainment section that is less likely to generate rumors or misinformation. These statistical findings further validate the rationale behind utilizing large-scale unlabeled data as non-rumor class instances and provide an explanation for why it is unreasonable to classify them as other Rumor classes.

Indeed, the analysis of unlabeled data from both Weibo and Twitter platforms reveals some subtle differences. For instance, the presence of sports-related posts is relatively lower on the Weibo platform compared to the Twitter platform, while history-related posts are more prevalent on Weibo but less so on Twitter. These disparities could potentially be attributed to cultural differences between the Eastern and Western regions. Cultural variations play a significant role in shaping the content and user preferences on different social media platforms. The observed differences in content themes might reflect the diverse interests and priorities of the respective user bases. Factors such as language, geographic location, historical background, and societal norms can influence the types of content shared and engaged with on social media platforms.

B   Experiment Details

In this section, we mainly introduce some specific settings in our experiments, including the way of data preprocessing and the hyperparameter configuration when training the model.

B.1   Data Preprocessing

For the texts in all datasets, we first standardize the different fonts present in the texts, then identify user mentions and web/url links as special tokens, <@user> and <url>. Next, we use the Tweet-Tokenizer from the NLTK toolkit and jieba word segmentation engine to tokenize the raw texts in English and Chinese datasets, respectively. Additionally, we use the emoji package[4] to translate the emojis in the texts into text string tokens.

B.2   Hyperparameter Configuration

All models are implemented by PyTorch and the baseline methods are re-implemented. GACL uses BERT (Devlin et al., 2018) to extract the initial

feature vector of each post in the propagation tree. In addition to GACL, other models use 200-dimensional word2vec word embeddings (Mikolov et al., 2013) as initial feature vectors.

We set the controlling hyperparameters for the loss functions λ, α, and β uniformly to a small value of 1e-3. The τ of supervised contrastive loss is set to 0.3, while the batch size is set to 32, and the learning rate is set to 1e-3. The GCN encoder consists of 3 layers, and we employ sum-pooling for node representations to obtain graph representation. We optimize the loss function using the Adam optimizer (Kingma and Ba, 2014). The entire training process for all models is conducted on a single Nvidia GeForce RTX 3090 GPU.

Note that BiGCN and GACL employ early stopping to observe the performance attainable by the models. However, due to oscillations during the early stages of model training, the observed model performance becomes unstable. To ensure a fairer comparison of different models' performance, we conduct experiments using the same dataset on AD-GSCL and multiple baseline methods. Furthermore, all models are trained for 100 epochs until convergence. We consider the average results of the last 10 epochs among these 100 epochs as the stable results achievable by the models. This approach helps mitigate the effects of training fluctuations and provides a more reliable basis for comparing the performance of the models.

C   Supplementary Experiments

We conducted a series of additional ablation experiments to delve into the impact of different modules in AD-GSCL more comprehensively. Additionally, we extended the few-shot experiments on DRWeibo dataset.

C.1   Extended Ablation Study

We conducted a series of extended ablation experiments to verify the influence of different factors on the model performance. Experiments were conducted under class-balanced condition and semi-supervised training strategy.

C.1.1   Adaptation Measures

In order to enable the supervised contrastive learning method and the unsupervised contrastive learning method of MI maximization adopted by AD-GSCL to adapt to the distribution and structural characteristics of social media data, we have made the following designs respectively: (1) Modified

---

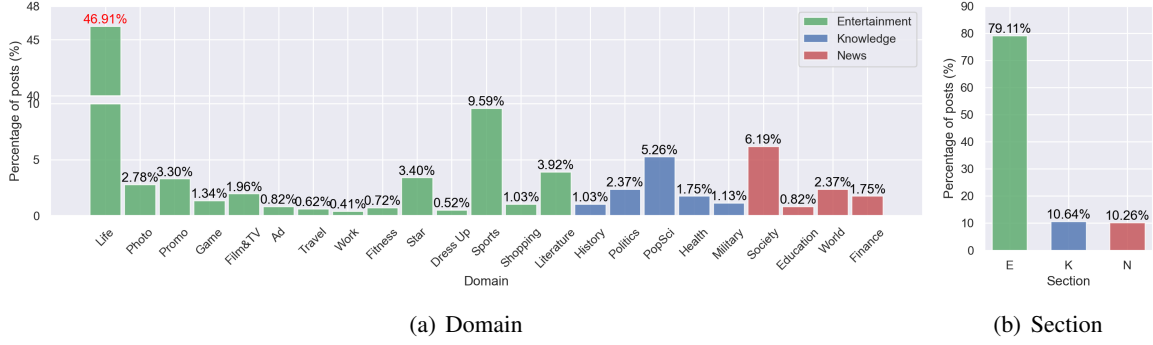[4]https://pypi.org/project/emoji

(a) Domain

(b) Section

Figure 6: Distribution of post counts by domain and section on UTwitter dataset. E, K and N represent the three sections Entertainment, Knowledge and News respectively.
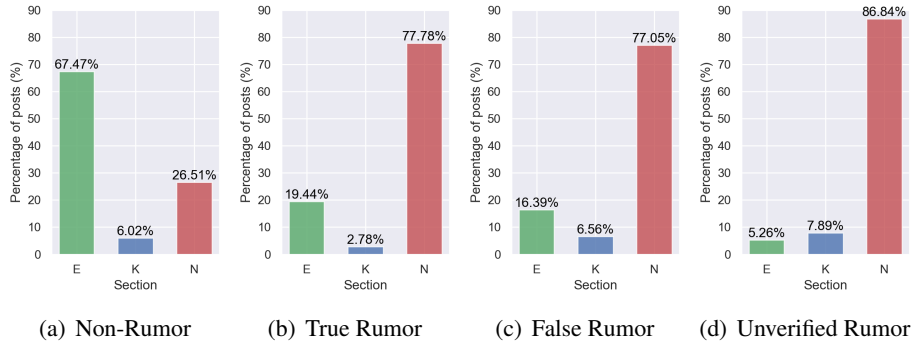

(a) Non-Rumor (b) True Rumor (c) False Rumor (d) Unverified Rumor

Figure 7: Post section distribution on Twitter15 and Twitter16 dataset.

	Weibo	DRWeibo
AD-GSCL	97.0±0.3	92.2±0.3
w/ Standard SCL	95.4±0.6	89.4±0.5
w/ Standard MI Max.	95.9±0.4	90.5±0.4

Table 6: The influence of adaptation measures.

(a) Weibo (b) DRWeibo

Figure 8: The influence of information flow direction.

C.1.2 Graph Directionality

In Figure 8, we explored the effects of different graph directionality on performance on Weibo and DRWeibo datasets. The results show that undirected graphs lead to performance degradation. This may be because in the forward propagation of GNNs, nodes densely connected at the root node can see each other in their neighborhood views, aggregating each other's information and eventually losing the uniqueness of node features, causing over-smoothing. In contrast, top-down and

the process of constructing positive and negative samples in standard supervised contrastive learning (Khosla et al., 2020), restricting non-rumor data from participating in positive sample construction; (2) Changed the standard MI maximization contrastive loss (maximizing MI between node representations and graph representations) to maximizing the MI between node representations and root node representations. We explored the effects of these two designs in Table 6. Experimental results on Weibo and DRWeibo showed that both designs improved model performance to some extent. This indicates that for existing contrastive learning methods, adapting them according to the characteristics of social media data when applying them to rumor detection tasks is very important.
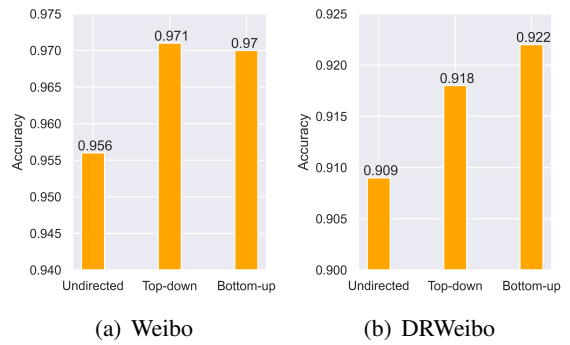
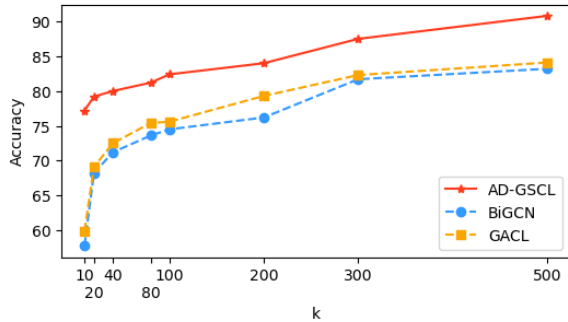| Measures	Weibo	DRWeibo
JS	97.0±0.3	92.2±0.3
KL	96.6±0.6	92.0±0.4
DV	97.1±0.3	91.6±0.5

Table 7: The influence of MI estimators.

Figure 9: Results of few-shot experiments.

bottom-up directed graphs can effectively block excessive information flow between nodes at the root. For AD-GSCL, since we use the root node representation as the global representation in unsupervised training, we adopt a bottom-up directed graph, which allows the root node to aggregate features from most nodes in the graph.

C.1.3 Mutual Information Estimators

We investigated the effects of MI measures in unsupervised training on Weibo and DRWeibo datasets, with experimental results shown in Table 7. The results demonstrate that AD-GSCL is relatively robust to the choice of MI estimator. We select the Jensen-Shannon MI estimator (Nowozin et al., 2016) which gives the most stable performance.

C.2 Extended Few-shot Experiments

We conducted few-shot experiments on DRWeibo dataset with the same settings as the Weibo dataset. The experimental results are shown in Figure 9. Similar results are observed on DRWeibo as on Weibo dataset, that is, the fewer labeled samples, the more significant improvement of AD-GSCL over baselines. This demonstrates the superiority of semi-supervised methods under few-shot conditions.