TaHiD: Tackling Data Hiding in Fake News Detection with News Propagation Networks

Anonymous ACL submission

Abstract

Fake news with detrimental societal effects has attracted extensive attention and research. Despite early success, state-of-the-art methods fall short of addressing the data hiding challenge, which can be divided into two aspects: disguise and disappearance. Disguise means that fake news publishers may disguise fake news as real ones by imitating the content. Thus, it is not enough to identify fake news only using news content, the different mediums that news propagates through should be taken into account. Disappearance means the related medium information may lose during the propagation of news due to relevant regulations or fake news publishers. It requires the model to capture the propagation features of the news and identify what matters in the propagating process. In this paper, we propose a novel graph-based and heterogeneous-aware fake news detection framework named TaHiD. TaHiD addresses the disguise challenge by encoding multiple mediums during propagation to obtain heterogeneous information. TaHiD aggregates mediums information and measures the influence of different propagation through a propagation transformer module, to handle the disappearance challenge. Experiments demonstrate that TaHiD outperforms state-of-the-art methods on benchmark datasets. Additional studies also show that TaHiD is capable of identifying fake news in the case of data hiding.

1 Introduction

Nowadays, with the development of social media, the spread of misinformation and disinformation is at an all-time high. As Pew Research Center reported, about six in ten of those who at least rarely get news on social media say they expect that news to be largely inaccurate (Elish Shearer, 2021). Fake news contains intentional false information and can disrupt social order. For example, Cui and Lee (2020) mentioned that 77 cell phone towers have been set on fire due to the conspiracy claiming that 5G mobile networks can spread COVID-19.

(a) Disguise (b) Disappearance

Figure 1: The two aspects of the data hiding problem. (a) **Disguise.** The news content is similar to real ones, but people can identify fake news because of the highlighted parts. (b) **Disappearance.** The related mediums' information may be lost due to relevant regulations or fake news publishers.

Since widespread fake news has harmful societal effects, there is an urgent need to develop fake news detection methods. Traditional methods focus on news content including texts or pictures (Wang et al., 2018), to identify fake news with feature engineering (Pennebaker et al., 2015) or neural networks (Shu et al., 2019a). Recent advanced research has suggested leveraging the social network to help identify rumors (Shu et al., 2020b; Nguyen et al., 2020; Dou et al., 2021).

Despite early success, existing models do not consider the challenge of data hiding, which can be divided into two aspects, disguise and disappearance, as figure 1 illustrates. For one thing, the content of some fake news, such as texts and pictures, is very similar to real ones to confuse people (Shu et al., 2019b). For example, malicious actors tend to disguise fake news as authentic reporting by adding made-up sources such as "CNN"

or "Rachel Maddow" (Narwal, 2018). Methods focused on the news content may not find out this fake news. Fortunately, because news propagates in social media, which consists of various related mediums (*e.g.*, comments, users, and source), people can judge the real and unreal through related mediums of news and spread it to more people. As a result, a model that can leverage related mediums to identify fake news is required. For another, the news may lose information in the process of propagation. For example, users may delete their comments that mentioned fake news (Guess et al., 2019). To make matters worse, fake news publishers may refute debunking comments (Jung et al., 2020) or use social bots (Heidari et al., 2021) to confuse people. Fake news propagates differently from real news (Zhao et al., 2020), people can identify fake news through the characteristics of news propagation, which previous social network based methods do not consider. Meanwhile, Xu et al. (2022) claim that the social influence of debunkers may not translate to wide mitigation of information propagation as expected. There is a need for a method that can analyze the characteristics of news propagation and identify fake news.

In light of the two challenges, we propose TaHiD (**Ta**ckling Data **Hi**ding in Fake News **D**etection with News Propagation Networks), a fake news detection framework. TaHiD models the news propagating process as a heterogeneous dynamic graph dubbed the news propagation network. TaHiD adopts heterogeneous information to address the disguise problem. Specifically, TaHiD encodes critical entities of the propagation process, including news, sources, comments, users, and temporal information. TaHiD adopts a propagation attention module to aggregate medium information and identifies important propagation in the process to capture the propagation features, which ensures that TaHiD can address the disappearance challenge. Our main contributions are summarized as follows:

- We propose an end-to-end fake news detection framework named TaHiD. TaHiD is graph based and heterogeneous aware. It can extract heterogeneous information to address the disguise challenge.
- TaHiD aggregates multiple mediums information and measures the contributions of each propagation through the propagation attention module to address the disappearance challenge.
- We conduct extensive experiments on real-world datasets to evaluate TaHiD and 11 competitive baselines. The results show the excellent ability of TaHiD to identify fake news. Further experiments verify the ability to address the data hiding problem.

2 Related Work

The widespread dissemination of fake news on social media has brought serious harm to politics, the economy, and other fields. Researchers have adopted various methods for fake new detection (Zellers et al., 2019; Fung et al., 2021; Dementieva and Panchenko, 2021; Jin et al., 2022; Silva et al., 2021). Previous works on fake news detection mainly focus on news content. Castillo et al. (2011) use features from user's posting and retweeting behavior, tweet content and URLs. Popat et al. (2016) propose a classifier that uses factors to assess the credibility of the claim from different sources. Deep Neural Networks are also adopted recently and significantly outperform traditional methods (Hartl and Kruschwitz, 2022). Karimi and Tang (2019) utilize automatic document structure learning and learn structurally rich representations for news documents. Ma et al. (2018) integrate both structure and content semantics based on tree-structured recursive neural networks for detecting rumors. Tan et al. (2020) introduce the task of detecting machine-generated news that includes images. Dun et al. (2021) propose a knowledge-aware attention network for fake news detection. Pelrine et al. (2021) find that traditional NLP baselines are competitive with and can outperform novel transformer-based methods. However, content-based fake news detection methods are not able to handle fake news which disguises itself as genuine news. Graph-based models that leverage additional information from mediums during the news propagation process are needed.

Graph-based methods have many advantages compared to content-based methods. GNNs are neural networks that can be directly applied to graphs, which provide an easy way to perform node-level, edge-level, and graph-level prediction tasks. Many graph-based efforts were made on the task of fake news detection (Gangireddy et al., 2020; Zhong et al., 2020; Benamira et al., 2019; Wang et al., 2020). Monti et al. (2019) propose a novel fake news detection based on deep learning on the graph, with different kinds of node information such as content, profile, social graph, and

news propagation. Hu et al. (2021) and Feng et al. (2022) utilize external knowledge through entities for fake news detection. Han et al. (2020) use the propagation pattern of news on social media and focus on a propagation-based approach for fake news detection. Mehta and Goldwasser (2021) contribute a novel benchmark for fake news detection at the knowledge element level, as well as a solution for this task that incorporates cross-media consistency checking to detect the fine-grained knowledge elements and make news articles misinformative. Mehta et al. (2022) formulate fake news detection as a reasoning problem and propose an inference-based graph representation learning method. Sun et al. (2022) combine Knowledge graphs and the dynamics of the message. Works on fake news detection often simplified social graphs while real-world scenarios are dynamic, heterogeneous, and more complicated. Abouzeid et al. (2022) point out the social network has random dynamics which change over time. Though novel graph-based fake news detection methods have made great achievements in detecting fake news through graph structure, they fall short to address the disappearance problem. In the process of news propagation, much information, such as news content, news source, and images, will gradually disappear due to the passage of time or be deleted by the publisher. There is a need for a model that can dynamically evaluate the contribution of different propagation.

3 Methodology

Figure 2 presents the overview of TaHiD. TaHiD firstly constructs a heterogeneous dynamic graph dubbed the news propagation network and encodes mediums to get heterogeneous information to handle the disguise challenge. Then, TaHiD extracts temporal information with time encoding and topological information by graph neural networks. After that, TaHiD extracts the propagation features and identifies what matters in every propagation by the propagation transformer module to address the disappearance challenge.

3.1 News Propagation Network

TaHiD models the news propagation networks as a heterogeneous dynamic graph. In this sense, TaHiD can identify fake news with the help of multiple mediums in the process of news propagation. Let us first define the propagation network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{C}^v, \mathcal{R}^e, \varphi, \psi, t)$ that consists of news, related mediums and the relations, where $\mathcal{V}$ is the node (medium) set; $\mathcal{E}$ is the edge (relation) set; $\mathcal{C}^v$ is the medium type set; $\mathcal{R}^e$ is the relation types set; φ and ψ are the node type mapping function and the edge type mapping function, respectively; t is the time stamp mapping function of each node. We can now formally define fake news detection using the news propagation network.

Definition. *Given a propagation network $\mathcal{G}$, our task is to construct a classification function f : $f(\mathcal{G}) \rightarrow \hat{y}$, where $\hat{y}$ is the predicted label of each news, such that $\hat{y}$ approximates ground truth y to maximize the prediction accuracy.*

Specifically, TaHiD leverages various mediums to improve the ability to solve the disguise problem. The detailed information is as follows:

- **News.** A title leads the outline and attracts the audience. It is the most important element in news (Horne and Adali, 2017). Besides, as fake news is created to spread inaccurate and harmful information, the content often has a language style different from real ones. TaHiD collects title and textual information from the news.

- **Source.** Source is the news organization, *e.g.*, BBC. A reliable source may not publish fake news, and news published by an unreliable source is often untrustworthy, conversely. The credibility of sources cannot be easily changed. TaHiD collects information about the source from the description on the homepage thanks to the rich textual information (Kulkarni et al., 2018).

- **User.** Users play an important role in spreading fake news (Shu et al., 2019c), and therefore it is significantly necessary to extract information from users. TaHiD leverages textual information from description and property information such as, whether the user is verified, which are widely exploited in different tasks on social media (Feng et al., 2021).

- **Comment.** Users express their opinions towards news through posting comments on social media. People can identify fake news with the assistance of comments (Shu et al., 2019a). For each comment, TaHiD collects the textual information and the property information such as the support count.

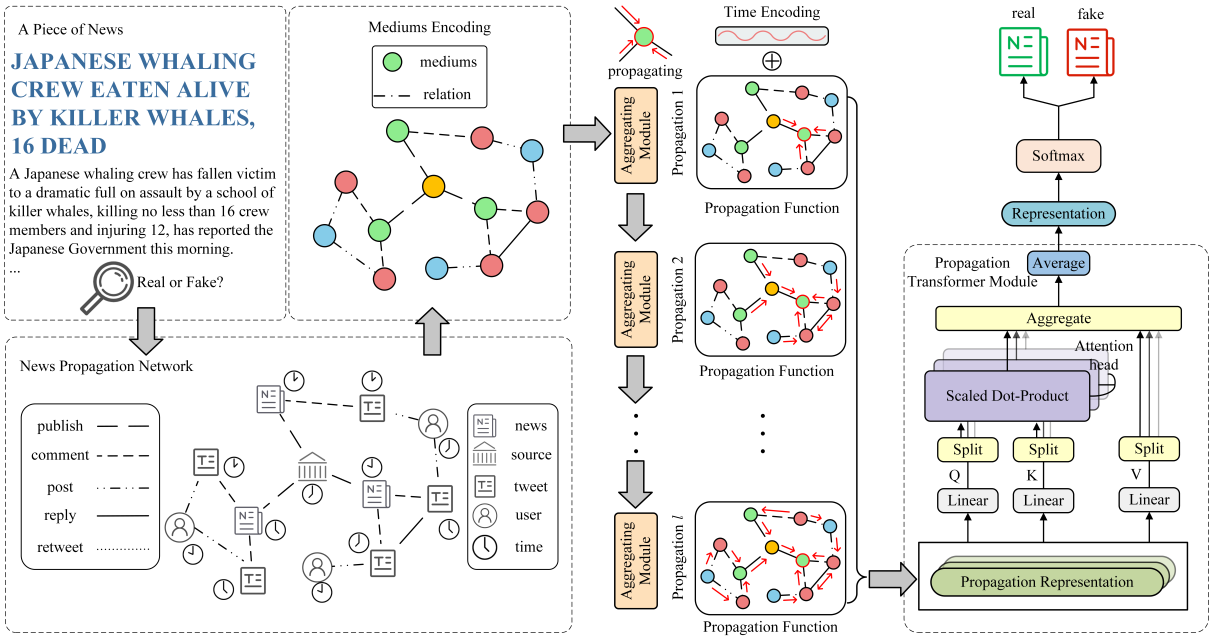


Figure 2: Overview of our proposed fake news framework TaHiD. (a) TaHiD builds a news propagation network and extracts heterogeneous information to address the disguise challenge. (b) TaHiD adopts the propagation transformer module to dynamically evaluate the contribution of different propagations and solve the disappearance challenge.

Besides, TaHiD adopts five types of edges between different mediums to represent different ways of news propagation, which are 'publish' relation between source and news, 'comment' relation between comment and news, 'post' relation between user and comment, 'reply' relation between comment and comment, and 'retweet' relation between comment and comment.

3.2 Mediums Encoding

After determining the mediums and edges in the news propagation network, TaHiD encodes each medium and aggregates heterogeneous information against the disguise challenge. The information that TaHiD collects can be divided into textual and property information.

- **Textual Information.** TaHiD adopts hierarchical encoding, which can extract different levels of semantic information (Yang et al., 2016), to encode a paragraph of text A. TaHiD firstly leverages NLTK (Bird et al., 2009) to tokenize A into sentences $\{a_i\}_{i=1}^{Q}$, where a_i denotes the $i-$th sentence and Q is sentence count. Then TaHiD adopts pre-trained language model RoBERTa (Liu et al., 2019) to encode each word in each sentence a_i, *i.e.*,

$$\{\overline{a}_{i,j}\} = \text{RoBERTa}(\{a_{i,j}\}_{j=1}^{Q_i}),$$

where $a_{i,j}$ is $j-$th word in $i-$th sentence and Q_i is the token count in $i-$th sentence. After that, TaHiD average the token encodes to obtain sentence-level encode, *i.e.*,

$$\overline{a}_i = \frac{1}{Q_i}\sum_{j=1}^{Q_i}\overline{a}_{i,j}.$$

Finally, TaHiD feeds the paragraph-level encode, which is obtained by averaging sentence-level encode, into a linear layer to get the semantic representation, *i.e.*,

$$\overline{a} = W_a \cdot (\frac{1}{Q}\sum_{i=1}^{Q}\overline{a}_i) + b_a,$$

where W_a and b_a are learnable parameters.

- **Property Information.** TaHiD firstly adopts z-score normalization to normalize numerical properties and then feeds all processed properties into a linear layer to obtain the property representation.

TaHiD then concatenates each encoding to obtain the initial encoding of a medium.

3.3 Aggregating Module

TaHiD aggregates temporal and topological information of each medium by time encoding and

graph neural networks against the disappearance challenge. TaHiD adopts different time encoding for different ways of news propagation, *i.e.*, for each medium s, which representation is r_s, TaHiD adds the time encoding by

$$\overline{r}_s = r_s + \sum_{(s,e)\in\mathcal{E}} \text{Time}_{\psi(s,e)}(\|t(s) - t(e)\|),$$

where $\text{Time}_{\psi(s,e)}$ is the time encoding of type $\psi(s,e)$. TaHiD modifies the RTE model (Hu et al., 2020) to adapt to the different types of edges.

TaHiD adopts one of (i) **GCN** (Kipf and Welling, 2017); (ii) **GAT** (Velickovic et al., 2017); (iii) **HGT** (Hu et al., 2020); (iv) **RGCN** (Schlichtkrull et al., 2018) as the propagation function to extract the topological information over the news propagation network.

3.4 Propagation Transformer Module

To address the disappearance challenge, TaHiD needs to extract the propagation features and understand what matters in the propagation process. TaHiD considers the initial encoding, namely $\overline{r}_s$, of news as the first propagation representation. Every time a propagation function is applied, TaHiD obtains a new propagation representation. Assume there are l times propagation, TaHiD gets l propagation representations $\{x_i\}_{i=1}^{l}$. Inspired by Transformer (Vaswani et al., 2017), TaHiD adopts the propagation transformer module to find out the important propagation during the propagating process. TaHiD firstly concatenates a learnable parameter x_0 to the propagation representations (Devlin et al., 2019). TaHiD then obtains query, key, and value for the c-th attention head with regard to x_i, formulated as

$$\begin{aligned} q_{c,i} &= W_{c,q} \cdot x_i + b_{c,q}, \\ k_{c,i} &= W_{c,k} \cdot x_i + b_{c,k}, \\ v_{c,i} &= W_{c,v} \cdot x_i + b_{c,v}, \end{aligned}$$

where q, k and v are query, key, and value of the attention mechanism, and all W and b are learnable parameters with regard to attention heads. After that, TaHiD models the importance of each propagation by calculating attention weights by

$$\alpha_{c,ij} = \frac{\langle q_{c,i}, k_{c,j}\rangle}{\sum_{k=0}^{l}\langle q_{c,i}, k_{c,k}\rangle},$$

where $\alpha_{c,ij}$ denotes the attention weight between propagation i and j; $\langle q, k\rangle = \exp(\frac{q^T k}{\sqrt{d}})$ is the exponential scale dot-product function, where d refers to the hidden dimension of each attention head. TaHiD then aggregates over attention heads to obtain propagation representation, *i.e.*,

$$z_i = \frac{1}{C}\sum_{c=1}^{C}\sum_{j=0}^{l}\alpha_{c,ij} \cdot v_{c,j},$$

where C is the number of attention heads. TaHiD adopts z_0 as the final representation z of news.

3.5 Learning and Optimization

TaHid identifies fake news according to the news representation z. The goal is to minimize the following loss function

$$L(\theta) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y}) + \lambda\sum_{\omega\in\theta}\omega^2,$$

where θ collects the learnable parameters of TaHiD; λ is the regularization hyperparameters; y refers to the ground truth of this news; $\hat{y}$ is the predicted probability of fake news.

4 Experiment

4.1 Experiment Settings

Datasets TaHiD is graph based and heterogeneity aware, which requires datasets that provide news, source, comment, user, and relations among them. We make use of a widely used dataset, FakeNewsNet (Shu et al., 2020a) dataset and its publicly available version at the official GitHub repository[1]. Detailed information can be found in the appendix. We experiment with the following three settings: (i) **Merged**, the whole dataset; (ii) **Politifact**, the news from *Politifact* platform; (iii) **Gossipcop**, the news from *Gossipcop* platform. We randomly conduct a 6:2:2 partition for each setting as training, validation, and test set. We reproduce baselines under the same setting to make the results comparable.

Baselines We compare TaHiD with **dEFEND** (Shu et al., 2019a), **RoBERTa** (Liu et al., 2019), **UPF** (Shu et al., 2019c), **HPN** (Shu et al., 2020b), **Bi-GCN** (Bian et al., 2020), **FANG** (Nguyen et al., 2020), **UPFN** (Dou et al., 2021), **Cross-D** (Silva et al., 2021), **CMTR** (Hartl and Kruschwitz, 2022), **NEP** (Sheng et al., 2022), and **InfOp** (Mehta et al., 2022). We re-implement and evaluate them and detailed information can be found in the appendix.

[1] https://github.com/KaiDMML/FakeNewsNet

Table 1: Average fake news detection performance and standard deviation of five runs of TaHiD and baselines. 'ACC' denotes the accuracy score while '$F1_{fake}$' denotes the binary f1-score. **Bold** and underline indicate the highest and second highest performance. The 'N', 'S', 'C', and 'U' in the 'Type' column indicate whether a model adopts news, source, comment, or user medium. '-G' denotes a model leverages graph information of news propagation. We conduct the significant tests with an unpaired t-test and report the p-value (* for $p < 0.0005$, ** for $p < 0.005$, *** for $p < 0.05$).

Models	Type	Merged		Politifact		Gossipcop	
		ACC	$F1_{fake}$	ACC	$F1_{fake}$	ACC	$F1_{fake}$
RoBERTa	N	85.45 (±0.19)*	67.59 (±1.11)*	91.79 (±0.26)**	89.38 (±0.28)**	86.27 (±0.29)*	69.67 (±1.04)*
NER	N	81.80 (±0.14)*	58.80 (±1.54)*	87.51 (±0.42)*	83.99 (±0.48)*	82.09 (±0.07)*	59.94 (±1.42)*
UPF	U	94.37 (±0.15)*	86.04 (±0.37)*	86.76 (±0.60)*	84.35 (±0.70)*	95.17 (±0.03)*	87.98 (±0.12)*
dEFEND	NC	84.13 (±4.13)*	67.76 (±6.85)*	84.43 (±0.95)*	78.54 (±2.45)*	87.43 (±2.76)*	72.96 (±6.45)*
CMTR	NCU	91.83 (±0.70)*	82.13 (±2.99)*	83.40 (±1.87)*	77.32 (±3.28)*	93.32 (±0.62)*	86.17 (±0.54)*
HPN	NC-G	87.28 (±0.21)*	82.39 (±0.28)*	78.08 (±0.72)*	81.21 (±0.60)*	87.09 (±0.09)*	81.42 (±0.16)*
Bi-GCN	NCU-G	90.35 (±0.50)*	90.00 (±0.71)*	77.14 (±1.42)*	74.73 (±2.21)*	90.84 (±0.28)*	90.44 (±0.34)*
UPFN	NCU-G	93.62 (±0.99)*	93.70 (±0.84)***	89.52 (±2.13)*	87.15 (±2.62)*	95.80 (±0.83)*	**95.82** (±0.81)***
Cross-D	NCU-G	89.76 (±0.32)*	86.08 (±0.44)*	88.30 (±0.91)*	87.82 (±0.86)*	90.32 (±0.32)*	87.08 (±0.45)*
InfOp	NSU-G	95.77 (±0.08)*	91.54 (±0.22)*	85.76 (±1.30)*	81.72 (±1.58)*	96.09 (±0.09)*	92.22 (±0.07)*
FANG	NSCU-G	81.32 (±0.18)*	53.71 (±1.49)*	80.28 (±1.45)*	75.37 (±3.07)*	82.73 (±0.13)*	51.20 (±1.17)*
TaHiD	NSCU-G	**97.03** (±0.19)	**93.94** (±0.29)	**92.92** (±0.48)	**90.80** (±0.55)	**97.40** (±0.10)	94.69 (±0.19)

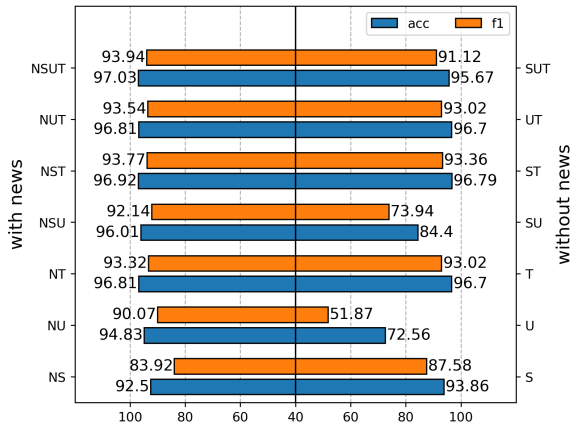


Figure 3: The performance of TaHiD mediums limitation study under the Merged setting. The 'N', 'S', 'U', and 'C' denote the news, source, user, and comment. Every medium information is helpful and TaHiD can detect fake news with the additional information.

Evaluation metrics We adopt accuracy and the binary f1-score, which is calculated for the positive class of fake news, as evaluation metrics of different fake news detection methods.

Implementation The hyperparameters can be found in the appendix. We submit codes of TaHiD and baselines as supplementary material.

4.2 Main Results

We first check out which mediums the baselines adopt and whether leverage graph information. We then evaluate these baselines under the three settings and present the results in Table 1, which demonstrates that:

- TaHiD achieves the best accuracy with very statistically significant over all of the settings ($p < 0.005$). For the f1-score, TaHiD outperforms all baselines except under the gossipcop setting and has only minor inadequacies (1.13%), which is considered to be statistically significant ($p < 0.05$), compared to UPFN.
- Under the Merged setting, which is the most comprehensive, baselines with graph information are generally more effective. It proves graph information can address the disguise challenge. Besides, TaHiD achieves a 2.82% improvement in accuracy ($p < 0.0005$) compared to UPF which is the best model without graph information. TaHiD can make the best use of heterogeneous information from the news propagation network.
- Due to the disappearance challenge, we can now obtain only 20,896 news (910 for Politifact and 19,986 for Gossipcop) with the content (about 26.51% loss). Compared to the models that adopt graph information, TaHiD achieves a 1.32% improvement from InfOp in accuracy ($p < 0.0005$). It proves that TaHiD can capture the propagation features of news to improve performance, which verifies the ability of TaHiD to address the disappearance challenge (more analysis in Section 4.4).

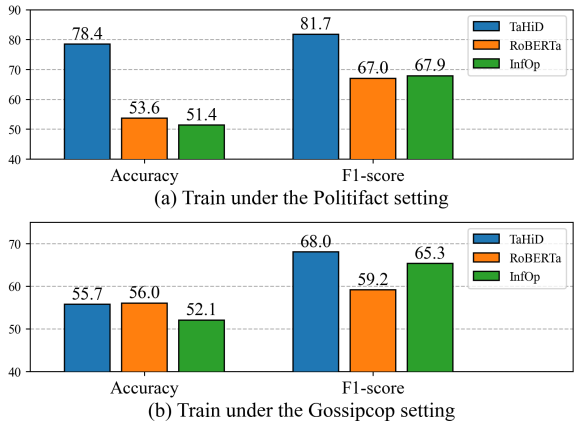


Figure 4: Train TaHiD and two baselines under one setting and test on another. TaHiD can identify different types of fake news even if not trained on them directly.

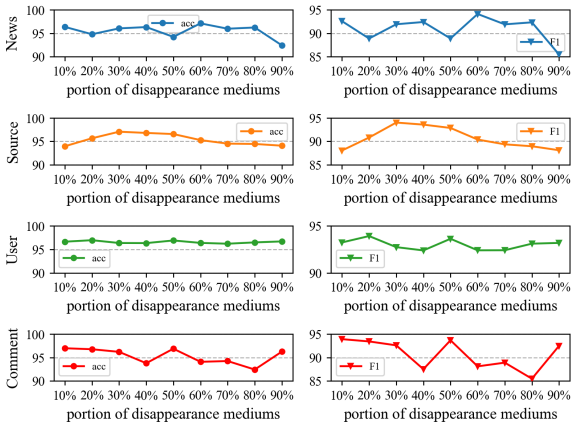


Figure 5: Performance of TaHiD when a portion of the specific medium is gradually removed, which demonstrates the robustness of TaHiD.

4.3 Disguise Challenge Study

The disguise problem demands detection models to identify fake news according to not only the news content but also the additional information.

Mediums Limitation Study TaHiD builds the dynamic heterogeneous graph to simulate the process of news propagation and extract heterogeneous information to solve the disguise problem. In the early propagation of news, only a few mediums can be accessed. We conduct experiments that remove kinds of mediums by deleting the node and related edges of the networks to figure out whether TaHiD can address the disguise challenge with limited mediums. From the results in Figure 3, we observe that: (i) Removing any mediums in the news propagation network leads to a drop in performance. It is sensible to leverage heterogeneous information to address the disguise challenge. (ii) TaHiD can identify fake news even without news mediums (compare the left part and right part in Figure 3). It proves that TaHiD can capture the propagation features to identify fake news.

Transfer Study There are different types of fake news and the content varies widely. Models that focus on content cannot identify different types of fake news and fail to address the disguise challenge. To prove TaHiD can make use of additional information to identify different fake news, we train TaHiD and two competitive baselines, RoBERTa and InfOp, on news from *Politifact* platform and test on news from *Gossipcop* Platform. After that, we swap the train set and test set. We sample 314 news from *Politifact* and 5,464 news from *Gossipcop* to make sure the balance of each label. Results from Figure 4 illustrate that TaHiD could better identify other types of news even when not trained on them. It establishes the claim that TaHiD can capture the news propagation features and additional information to address the disguise challenge.

4.4 Disappearance Challenge Study

The detection model should keep its performance when the information during propagation gradually disappears to address the disappearance challenge.

Ability to Address the Disappearance Challenge We remove a part of news propagation information by setting the initial representation of medium to zero to simulate the disappearance problem and check out if TaHiD can keep performance. The results from Figure 5 illustrate that TaHiD can keep the performance when the portion of a specific medium (even the news) gradually disappears. The accuracy drops to a minimum of 93.79%, but TaHiD still outperforms several baselines.

To measure the missing information sensitivity of TaHiD and other baselines, we simulate disappearance in the reality by randomly gradually removing a portion of the graph under the Politifact setting (that has more complete original information). We compare TaHiD with RoBERTa (content-based) and InfOp (graph-based and shares the network). Results from Figure 6 show that TaHiD outperforms other baselines with any portion of retaining information. Besides, when retaining 20% or more of the graph information, the performance of TaHiD does not decrease significantly.

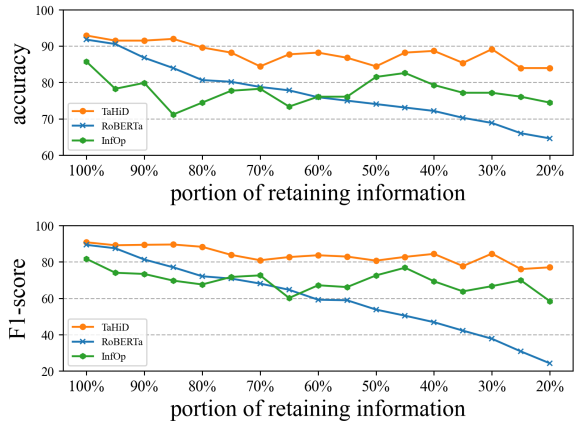


Figure 6: Performance of TaHiD and baselines when a portion of the whole graph information is gradually removed. TaHiD outperforms and keeps the performance with only 20% portion of the information.

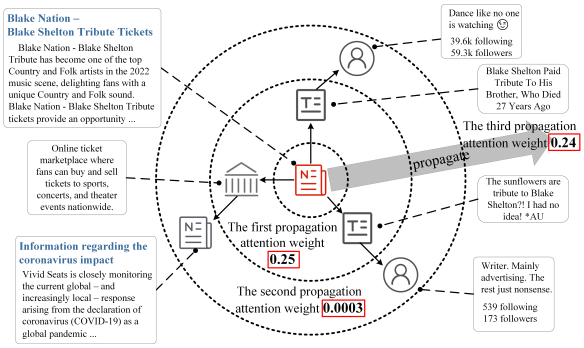


Figure 7: Example propagation network of a piece of fake news, related mediums, and the propagation attention weight. The first and the third propagation are important during the propagation process.

Effectiveness of Propagation Transformer Module TaHiD captures the propagation features through the propagation transformer module. We compare this module with the following six modules which can also aggregate features: (i) **First**: We consider the first propagation representation as the final news representation. (ii) Last: Similar to (i), while we adopt the last. (iii) Mean: We average each representation. (iv) **Max**: The max operation is replaced with mean. (v) **MLP**: We stack each representation and apply a MLP layer. (vi) **GRU**: We apply a GRU layer to get the representation. The results are presented in Table 2. The propagation transformer module achieves the best performance, showing a significant advantage in accuracy and f1-score against first, mean, and max ($p < 0.0005$). Compare to MLP and GRU which aggregate features in a deep way, our proposed module outperforms as well ($p < 0.005$).

Table 2: We compare the propagation transformer module with other modules under the Merged setting. Our proposed module achieves the best performance, which shows the ability to capture the propagation features. We conduct the significant tests with an unpaired t-test and report the p-value (* for $p < 0.0005$, ** for $p < 0.005$, *** for $p < 0.05$).

Module	Acc	Diff.	$F1_{fake}$	Diff.
First	84.89 (±0.33)*	-12.14	64.86 (±1.34)*	-29.08
Last	96.15 (±0.49)**	-0.88	92.28 (±0.73)**	-1.66
Mean	85.67 (±0.99)*	-11.36	75.97 (±1.21)*	-17.97
Max	95.85 (±0.42)*	-1.18	91.80 (±0.74)*	-2.14
MLP	96.03 (±0.62)**	-1.00	92.00 (±1.08)**	-1.94
GRU	94.34 (±0.87)*	-2.69	89.18 (±1.43)*	-4.76
TaHiD	**97.03**(±0.19)	0	**93.94**(±0.29)	0

4.5 Case Study

We study a piece of specific news and related mediums with the propagation attention weight in Figure 7. It is illustrated that the first and the third propagation have a larger attention weight (0.25 and 0.24), but the second one (0.0003) has smaller in this specific fake news. Moreover, we can hardly identify this fake news depending on its content, but we can identify it according to the related comments. The rest of the mediums, the two users and one news, do not help judge the veracity of this news. TaHiD could extract the propagation features and identify what matters in the propagation network to solve the data hiding problem.

5 Conclusion and Future Work

We propose TaHiD, a fake news detection framework that can address the two aspects of the data hiding problem, disguise, and disappearance. The disguise challenge means that fake news publishers tend to disguise misinformation as true by imitating the patterns of genuine news. The disappearance challenge means that information might disappear in related mediums during the news propagation process. TaHiD models the news propagation process as a heterogeneous dynamic graph named news propagation network and encodes mediums including news, sources, users, and comments. The heterogeneous information helps TaHiD to address the disguise problem. TaHiD captures the propagation features and identifies what matters during the propagation process, to address the challenge of disappearance. Extensive experiments demonstrate that TaHiD achieves excellent performance and can address the data hiding problem.

6 Limitations

TaHiD achieves excellent performance on fake news detection and tackles the data hiding challenge. TaHiD has two minor limitations:

- TaHiD leverages a news propagation network to identify fake news and achieve excellent performance. In the early days of news propagation, it could not form a relatively large-scale propagation network. Namely, TaHiD's performance in identifying early fake news may drop.
- Extensive experiments show that TaHiD considers the contribution of different propagations. TaHiD cannot give quantitative indicators of the importance of each propagation, which leads to low explainability.

7 Ethics Statement

Our framework does not present direct societal consequences in general. This frame aims to help identify fake news. It can serve as a detection module for fake news detection systems, especially when given additional medium information including the source, comment, and user mediums. To the best of our knowledge, no code of ethics was violated throughout the experiments done in this paper. We reported all hyper-parameters and submit the code of TaHiD and chosen baselines to reproduce the results. Our data is mostly based on existing datasets, FakeNewsNet. Due to legal reasons, we use the official GitHub repository to obtain a lightly different, copy of the dataset.

References

Ahmed Abouzeid, Ole-Christoffer Granmo, Christian Webersik, and Morten Goodwin. 2022. Socially fair mitigation of misinformation on social networks via constraint stochastic optimization. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 11801–11809. AAAI Press.

Adrien Benamira, Benjamin Devillers, Etienne Lesot, Ayush K Ray, Manal Saadi, and Fragkiskos D Malliaros. 2019. Semi-supervised learning and graph neural networks for fake news detection. In *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 568–569. IEEE.

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.

Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.".

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.

Limeng Cui and Dongwon Lee. 2020. Coaid: COVID-19 healthcare misinformation dataset. *CoRR*, abs/2006.00885.

Daryna Dementieva and Alexander Panchenko. 2021. Cross-lingual evidence improves monolingual fake news detection. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Student Research Workshop*, pages 310–320.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Yingtong Dou, Kai Shu, Congying Xia, Philip S. Yu, and Lichao Sun. 2021. User preference-aware fake news detection. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Yaqian Dun, Kefei Tu, Chen Chen, Chunyan Hou, and Xiaojie Yuan. 2021. Kan: Knowledge-aware attention network for fake news detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 81–89.

Amy Mitchell Elish Shearer. 2021. News use across social media platforms in 2020. 2021, Jan 12.

Shangbin Feng, Zhaoxuan Tan, Wenqian Zhang, Zhenyu Lei, and Yulia Tsvetkov. 2022. Kalm: Knowledge-aware integration of local, document, and global contexts for long document understanding. *arXiv preprint arXiv:2210.04105*.

Shangbin Feng, Herun Wan, Ningnan Wang, and Minnan Luo. 2021. Botrgcn: Twitter bot detection with relational graph convolutional networks. In *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 236–239.

Matthias Fey and Jan Eric Lenssen. 2019. Fast graph representation learning with pytorch geometric. *CoRR*, abs/1903.02428.

Yi Fung, Christopher Thomas, Revanth Gangi Reddy, Sandeep Polisetty, Heng Ji, Shih-Fu Chang, Kathleen McKeown, Mohit Bansal, and Avirup Sil. 2021. Infosurgeon: Cross-media fine-grained information consistency checking for fake news detection. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1683–1698.

Siva Charan Reddy Gangireddy, Cheng Long, and Tanmoy Chakraborty. 2020. Unsupervised fake news detection: A graph-based approach. In *Proceedings of the 31st ACM conference on hypertext and social media*, pages 75–83.

Andrew Guess, Jonathan Nagler, and Joshua Tucker. 2019. Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Science advances*, 5(1):eaau4586.

Yi Han, Shanika Karunasekera, and Christopher Leckie. 2020. Graph neural networks with continual learning for fake news detection from social media. *CoRR*, abs/2007.03316.

Philipp Hartl and Udo Kruschwitz. 2022. Applying automatic text summarization for fake news detection. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference, LREC 2022, Marseille, France, 20-25 June 2022*, pages 2702–2713. European Language Resources Association.

Maryam Heidari, Samira Zad, Parisa Hajibabaee, Masoud Malekzadeh, SeyyedPooya HekmatiAthar, Ozlem Uzuner, and James H Jones. 2021. Bert model for fake news detection based on social bot activities in the covid-19 pandemic. In *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pages 0103–0109. IEEE.

Benjamin D Horne and Sibel Adali. 2017. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Eleventh international AAAI conference on web and social media*.

Linmei Hu, Tianchi Yang, Luhao Zhang, Wanjun Zhong, Duyu Tang, Chuan Shi, Nan Duan, and Ming Zhou. 2021. Compare to the knowledge: Graph neural fake news detection with external knowledge. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 754–763.

Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. 2020. Heterogeneous graph transformer. In *Proceedings of The Web Conference 2020*, pages 2704–2710.

Yiqiao Jin, Xiting Wang, Ruichao Yang, Yizhou Sun, Wei Wang, Hao Liao, and Xing Xie. 2022. Towards fine-grained reasoning for fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5746–5754.

Anna-Katharina Jung, Björn Ross, and Stefan Stieglitz. 2020. Caution: Rumors ahead—a case study on the debunking of false information on twitter. *Big Data & Society*, 7(2):2053951720980127.

Hamid Karimi and Jiliang Tang. 2019. Learning hierarchical discourse-level structure for fake news detection. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 3432–3442. Association for Computational Linguistics.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.

Vivek Kulkarni, Junting Ye, Steven Skiena, and William Yang Wang. 2018. Multi-view models for political ideology detection of news articles. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3518–3527.

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. *CoRR*, abs/1907.11692.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Rumor detection on twitter with tree-structured recursive neural networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1980–1989.

Nikhil Mehta and Dan Goldwasser. 2021. Tackling fake news detection by interactively learning representations using graph neural networks. In *Proceedings of the First Workshop on Interactive Learning for Natural Language Processing*, pages 46–53.

Nikhil Mehta, María Leonor Pacheco, and Dan Goldwasser. 2022. Tackling fake news detection by continually improving social context representations using graph neural networks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1363–1380.

Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M. Bronstein. 2019. Fake news detection on social media using geometric deep learning. *CoRR*, abs/1902.06673.

Bhawna Narwal. 2018. Fake news in digital media. In *2018 international conference on advances in computing, communication control and networking (icacccn)*, pages 977–981. IEEE.

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

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Kellin Pelrine, Jacob Danovitch, and Reihaneh Rabbany. 2021. The surprising performance of simple baselines for misinformation detection. In *Proceedings of the Web Conference 2021*, pages 3432–3441.

James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015. Technical report.

Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2016. Credibility assessment of textual claims on the web. In *Proceedings of the 25th ACM international on conference on information and knowledge management*, pages 2173–2178.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer.

Qiang Sheng, Juan Cao, Xueyao Zhang, Rundong Li, Danding Wang, and Yongchun Zhu. 2022. Zoom out and observe: News environment perception for fake news detection. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 4543–4556. Association for Computational Linguistics.

Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu. 2019a. defend: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 395–405.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020a. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big data*, 8(3):171–188.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, and Huan Liu. 2020b. Hierarchical propagation networks for fake news detection: Investigation and exploitation. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 626–637.

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

Kai Shu, Xinyi Zhou, Suhang Wang, Reza Zafarani, and Huan Liu. 2019c. The role of user profiles for fake news detection. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining*, pages 436–439.

Amila Silva, Ling Luo, Shanika Karunasekera, and Christopher Leckie. 2021. Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 557–565.

Mengzhu Sun, Xi Zhang, Jiaqi Zheng, and Guixiang Ma. 2022. DDGCN: dual dynamic graph convolutional networks for rumor detection on social media. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 4611–4619. AAAI Press.

Reuben Tan, Bryan A. Plummer, and Kate Saenko. 2020. Detecting cross-modal inconsistency to defend against neural fake news. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 2081–2106. Association for Computational Linguistics.

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

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *stat*, 1050:20.

Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 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*, pages 849–857.

Youze Wang, Shengsheng Qian, Jun Hu, Quan Fang, and Changsheng Xu. 2020. Fake news detection via knowledge-driven multimodal graph convolutional networks. In *Proceedings of the 2020 International Conference on Multimedia Retrieval*, pages 540–547.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen,

Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Xiaofei Xu, Ke Deng, and Xiuzhen Zhang. 2022. Identifying cost-effective debunkers for multi-stage fake news mitigation campaigns. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 1206–1214.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489.

Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending against neural fake news. *Advances in neural information processing systems*, 32.

Zilong Zhao, Jichang Zhao, Yukie Sano, Orr Levy, Hideki Takayasu, Misako Takayasu, Daqing Li, Junjie Wu, and Shlomo Havlin. 2020. Fake news propagates differently from real news even at early stages of spreading. *EPJ data science*, 9(1):7.

Wanjun Zhong, Duyu Tang, Zenan Xu, Ruize Wang, Nan Duan, Ming Zhou, Jiahai Wang, and Jian Yin. 2020. Neural deepfake detection with factual structure of text. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 2461–2470. Association for Computational Linguistics.

A Dataset Details

TaHiD is graph-based and heterogeneity-aware, which requires datasets that provide news, source, comment, user, and relations among them. We make use of a widely used dataset, FakeNewsNet (Shu et al., 2020a) dataset. We use the official GitHub repository to obtain a lightly different, copy of the dataset. This dataset consists of 28,433 news, 2,165 sources, 2,446,885 comments, 805,842 users, and 4,966,149 relations. In order to evaluate the ability to solve the disguise and disappearance problems, we experiment with the following three settings. The statistics of these settings are listed in Tabel 3.

Table 3: Benchmark settings statistics.

Settings	News			Fake News	Real News	News with content
	Train	Valid	Test			
Merged	13,916	4,638	4,640	5,755	17,439	20,896
Politifact	632	210	212	432	622	910
Gossipcop	13,284	4,428	4,428	5,323	16,817	19,986

Table 4: Hyperparameter settings of TaHiD across various settings. We find our hyperparameter settings robust across all datasets and all hyperparameters are chosen by the performance on the validation set.

Hyperarameter	Value
RoBERTa dim D_s	768
batch size	64
hidden dim D	512
dropout rate	0.5
optimizer	Adam
learning rate	1e-4
weight decay	1e-5
attention head q	4
num of propagation l	8
propagation transformer layers	1
propagation function	GAT

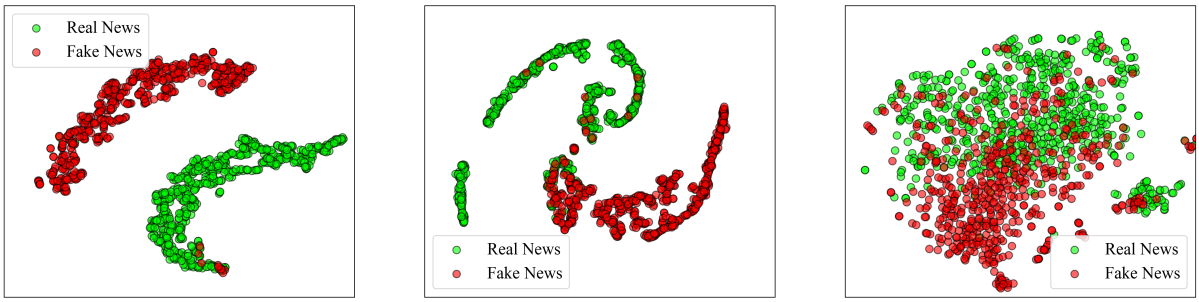
B Implementation of TaHiD

We implement TaHiD framework with pytorch (Paszke et al., 2019), pytorch geometric (Fey and Lenssen, 2019), and the transformers library (Wolf et al., 2020). We present our hyperparameter settings in Table 4. Our implementation is trained on a GeForce RTX 2080 Ti GPU with 12GB of memory. Under these settings, TaHiD runs a batch for about 60 seconds (Merged setting) and the size of the parameters is about 29.5 MB. We submit the code of TaHiD as supplementary material.

C Implementation of Baselines

We re-implement chosen baselines to have a fair comparison of the dataset under three settings. We submit the codes as supplementary material. The detailed information on baselines and implementation details are as followed:

- **dEFEND** (Shu et al., 2019a) develops a sentence-comment co-attention sub-network to exploit both news content and user comments for fake news detection. We set the max sentence length as 120, max sentence



(a) TaHiD (V-Measure: 0.869) (b) UPFN (V-Measure: 0.656) (c) RoBERTa (V-Measure: 0.204)

Figure 8: 2D t-SNE plot of the news representations of TaHiD, UPFN, and RoBERTa and the corresponding V-Measure. TaHiD achieves the best score and shows excellent collocation for the group of fake and real news.

count as 50, max comment length as 120, and max comment count as 150.

- **RoBERTa** (Liu et al., 2019) is a pre-trained language model. We apply MLP directly on news textual embeddings encoded by RoBERTa as a baseline, namely, the encoding is the same as the initial news representation of the news propagation networks.
- **UPF** (Shu et al., 2019c) adopts explicit and implicit profile features between user groups, which reveals their potential to identify fake news. We remove the news that does not have the related user.
- **HPN** (Shu et al., 2020b) builds hierarchical propagation networks and extracts structural, temporal, and linguistic features to detect fake news. We remove the news that can not construct the hierarchical propagation networks.
- **Bi-GCN** (Bian et al., 2020) proposes a bi-directional graph model, which can explore both characteristics by operating on both top-down and bottom-up propagation of rumors. We remove the news that does not exist in the original graph of the paper.
- **FANG** (Nguyen et al., 2020) is a novel graphical social context representation and learning framework for fake news detection. FANG adopts unsupervised proximity loss, self-supervised stance loss, and supervised fake news loss to optimize the parameters of the model. We modify the stance detection target from 5 stances into 2 stances.
- **UPFN** (Dou et al., 2021) proposes a new framework that simultaneously captures various signals from users to identify fake news. We remove the news that does not have related users.
- **Cross-D** (Silva et al., 2021) jointly preserves domain-specific and cross-domain knowledge in news records to detect fake news. We adopt the same hyperparameters as the original paper.
- **CMTR** (Hartl and Kruschwitz, 2022) combines multiple text representations, with the goal of circumventing sequential limits and related loss of information the underlying transformer architecture typically suffers from. We adopt the same hyperparameters as the original paper.
- **NEP** (Sheng et al., 2022) constructs the macro and micro news environment of specific news. NEP then designs a popularity-oriented and a novelty-oriented module to perceive useful signals and further assist final prediction from recent mainstream news. We adopt the same news environment and set the macro env days as 30.
- **InfOp** (Mehta et al., 2022) views fake news detection as reasoning over the relations between sources, articles they publish, and engaging users on social media in a graph framework. We use the original code from this paper to build the graph and evaluate the results.

D Representation Learning Study

TaHiD learns the representation of news to identify fake news. The representation can be used for other downstream tasks like political stance detection. To examine the quality of representation learning with our proposal, we cluster representations using k-means and calculate the V-Measure, which is an

external cluster evaluation. Figure 8 visualizes the representations and we can obviously observe that the representations of TaHiD show excellent collocation for the group of fake and real news, while UPFN and RoBERTa show little. Quantitatively, TaHiD's clusters achieve the best V-Measure score of 0.869, compared to a 0.656 score for the UPFN clusters and 0.204 for the RoBERTa.