An Information-theoretic Propagation Denoising and Fusion Framework for Fake News Detection

Mengyang Chen[1,2] , **Lingwei Wei**[1,*] , **Wei Zhou**[1] and **Songlin Hu**[1,2]
[1]Institute of Information Engineering, Chinese Academy of Sciences
[2]School of Cyber Security, University of Chinese Academy of Sciences
{chenmengyang, weilingwei, zhouwei, husonglin}@iie.ac.cn

Abstract

Incomplete propagation data significantly hinders robust fake news detection. Recent approaches leverage large language models to simulate missing user interactions via role-playing, thereby enriching propagation with synthetic signals. However, such propagation data is intrinsically unreliable, and directly fusing it can lead to biased representations and limited detection performance. In this paper, we alleviate the unreliability of synthetic propagation from the mutual information perspective and propose a novel information-theoretic propagation denoising and fusion (InfoPDF) framework to learn effective representations from both real and synthetic propagation. Specifically, we first generate attribute-specific synthetic propagation using large language models. Then we model each synthetic propagation graph as a probabilistic latent distribution to guide reliability-aware adaptive fusion with real propagation. During training, we design a mutual information-based objective to learn compressed and task-sufficient propagation representations. It jointly suppresses noisy signals across attribute-specific synthetic propagation, maintains consistency between real and synthetic propagation representations, and ensures task sufficiency for fake news detection and attribute prediction. Experiments on three real-world datasets show that InfoPDF consistently achieves superior performance across various fake news detection tasks. Further analysis demonstrates that InfoPDF can estimate attribute-level reliabilities and learn more discriminative propagation representations.

1 Introduction

The proliferation of fake news poses a significant threat to societal stability. Existing detection approaches primarily rely on content features (e.g., articles and comments) [Castillo *et al.*, 2011; Ma *et al.*, 2015; Luvembe *et al.*, 2023] or propagation patterns [Bian *et al.*, 2020; Wei *et al.*, 2021; Jiang *et al.*, 2025]. Compared to content-based methods, propagation-based approaches generally yield superior performance by capturing rich structural features. However, real-world propagation data is often incomplete due to platform limitations or malicious deletions [Wei *et al.*, 2021; Ma *et al.*, 2022], leading to information sparsity that compromises detection accuracy. While some works attempt to enhance robustness in modeling partial propagation [Wei *et al.*, 2022; Jiang *et al.*, 2025], they remain constrained by the availability of initial real data. Recently, large language models (LLMs) have been leveraged to simulate missing interactions through attribute-mixed role-playing [Nan *et al.*, 2024; Wan *et al.*, 2024], offering a promising solution for incomplete propagation data.

However, these LLM-based simulation approaches overlook the inherent unreliability of synthetic data. Crucially, naive attribute-mixed prompting can further exacerbate this issue, i.e., LLMs may fail to fully capture the situational context of discussions, leading to generated interactions that contain hallucinations [Ji *et al.*, 2023], as well as stylistic deviations and unnatural tones [Trott *et al.*, 2023], which deviate from real-world discussions. Moreover, diverse user attributes (e.g., age and politics) impact fake news spread differently [Shu *et al.*, 2019b]. Mixing multiple attributes within a single prompt often yields inconsistent or contradictory signals. As a result, low-quality synthetic signals can contaminate informative ones, reducing the effective information available for learning and ultimately weakening detection. As shown in a preliminary analysis in Figure 1, performance varies significantly across different user attributes and simulation settings. This observation suggests that the reliability of LLM-generated synthetic propagation is attribute-dependent, motivating the explicit modeling of attribute-level reliabilities to distinguish informative signals from noise.

In this paper, we alleviate the unreliability of synthetic propagation from the mutual information perspective, and propose a novel information-theoretic propagation denoising and fusion (InfoPDF) framework to learn effective representations from both real and synthetic propagation. Specifically, InfoPDF integrates real and synthetic propagation for each attribute view to form a fused representation, which is modeled as a stochastic latent variable for uncertainty estimation. The estimated uncertainty is then converted into an attribute-level reliability score. It enables reliability-aware adaptive fusion to emphasize informative views while down-weighting

*Corresponding author.

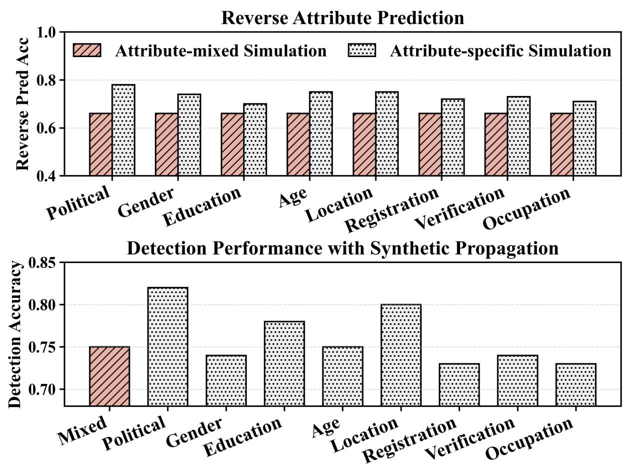

Figure 1: Preliminary analysis about different attributes and simulation settings on Twitter [Lin *et al.*, 2022]. Top: reverse attribute prediction accuracy based on the generated comments using GPT-4 Turbo. Bottom: fake news detection accuracy using synthetic propagation (GCN [Kipf and Welling, 2017] as detector).

noisy ones, thereby mitigating the negative impact of unreliable synthetic propagation on detection. During training, we design a mutual information based objective to learn compressed and task-sufficient propagation representations. It consists of three complementary objectives, i.e., minimizing the mutual information between synthetic propagation and latent representations to suppress redundant noise across different attribute views; maximizing the mutual information between real and synthetic propagation representations to reduce distributional discrepancy for propagation alignment; and maximizing the mutual information between latent representations and task labels to ensure task sufficiency for fake news detection and user attribute prediction. Through this design, InfoPDF effectively denoises and fuses multi-view synthetic propagation, improving the robustness of detection under incomplete propagation scenarios.

We conduct experiments on three public real-world fake news datasets. Experimental results demonstrate that InfoPDF consistently obtains superior performance in general detection as well as robust detection against noisy and incomplete scenarios. Further analysis reveals that our InfoPDF effectively estimates attribute-level reliability weights and extracts more discriminative propagation representations for fake news detection.

Our contributions can be summarized as follows: 1) We identify the unreliability of LLM-generated synthetic propagation for fake news detection, demonstrating that synthetic propagation can be noisy and its quality varies across attributes, which may reduce the effective information for detection. 2) We propose an LLM-generated propagation denoising and fusion framework to learn effective representations from both original and synthetic propagation. It denoises unreliable synthetic propagation via a compression-based denoising objective, and maintains the consistency between real and synthetic propagation representations as well as task sufficiency for fake news detection and user attribute prediction. 3) Experiments on three public datasets demonstrate the superiority of our proposed framework, validating its effectiveness and robustness in handling incomplete propagation scenarios.

2 Task Formulation

Fake news detection aims to verify the authenticity of a news item, which can be formulated as a binary classification task.

Formally, each sample is represented by a graph $\mathcal{G} = (V, E)$, where $V = \{n, c_1, \dots, c_N\}$ denotes the news n and its comments, and $\mathcal{E}$ denotes explicit interactions (e.g., reply).

Due to missing interactions in practice, the observed real propagation graph is often incomplete, denoted as $G_r = (V_r, E_r)$. To enrich propagation signals, we construct K attribute-specific synthetic graphs $\{G_s^{(k)}\}_{k=1}^{K}$, where each $G_s^{(k)} = (V_s^{(k)}, E_s^{(k)})$ is conditioned on an attribute $a_k \in A = \{a_1, \dots, a_K\}$. This allows precise assessment of each attribute's contribution. The task goal can be formulated as,

$$f : (G_r, \{G_s^{(k)}\}_{k=1}^{K}) \longrightarrow y, \quad y \in \{0, 1\}.$$

3 Methodology

In this section, we propose a principled InfoPDF framework to learn effective representations from both real and synthetic propagation for fake news detection.

3.1 Overview of InfoPDF

The architecture of InfoPDF is shown in Figure 2. We first leverage large language models (LLMs) to generate synthetic propagation based on each user attribute. We encode the incomplete real propagation graph and these attribute-specific synthetic propagation graphs by graph neural networks. We model the fused representation as a stochastic latent variable to estimate the reliability of synthetic propagation for each attribute, which is converted to guide the adaptive fusion where informative views are emphasized and noisy ones are down-weighted. During the training, we design a mutual-information based objective to jointly optimize the representations via minimizing the mutual information between synthetic propagation and latent representations to suppress redundant noise across different attribute views, maximizing the mutual information between real and synthetic propagation representations to reduce distributional discrepancy for propagation alignment, and maximizing the mutual information between latent representations and task labels to ensure task sufficiency for fake news detection and user attribute prediction. InfoPDF learns compressed and task-sufficient representations by exploiting both synthetic and real propagation for robust fake news detection.

3.2 Reliability-Aware Graph Fusion Network

Since the quality of synthetic propagation may vary across attributes, we model the reliability of multiple attribute-specific synthetic propagation graphs via variational estimation and use it to guide adaptive fusion.

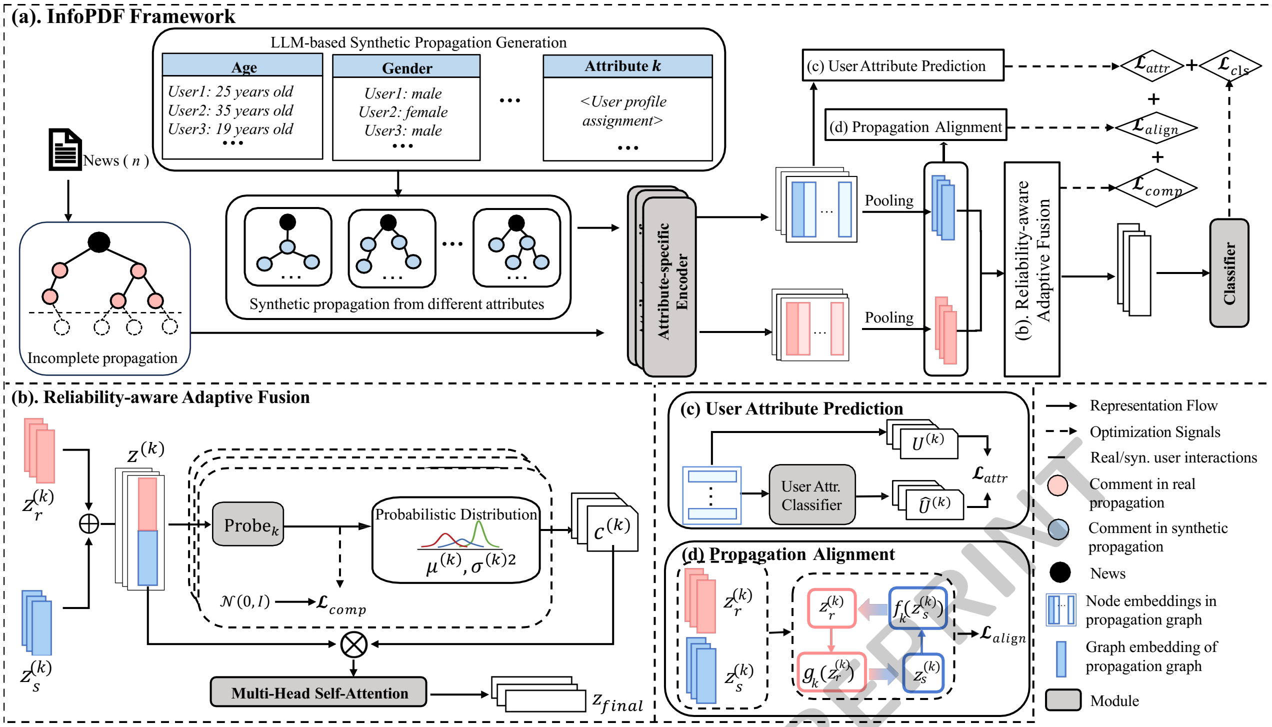

Figure 2: Overview of the InfoPDF framework. InfoPDF generates attribute-specific synthetic propagation graphs, encodes them with incomplete real propagation, estimates reliability via probabilistic modeling, and performs reliability-aware fusion for fake news detection.

LLM-based Synthetic Propagation Generation. Given a news item, we generate an attribute-specific synthetic propagation graph $G_s^{(k)}$ for each user attribute $a_k \in A$ using a frozen LLM, following the previous method [Wan *et al.*, 2024]. Beyond them, we consider eight user attributes including age, education [Gaillard *et al.*, 2021], gender, occupation [Lim and Perrault, 2021], politics, registration time, verification status, and location [Shu *et al.*, 2019b].

We then encode the incomplete real propagation graph G_r and each synthetic graph $G_s^{(k)}$ into latent representations. To ensure that real and synthetic views are comparable, we adopt a Siamese-style encoding strategy [Chopra *et al.*, 2005] to compute the fused representations. Specifically, for the k-th attribute, we use a graph neural network encoder denoted as $\text{Enc}_k(\cdot)$ to compute the graph-level representations of real and synthetic propagation, i.e.,

$$\mathbf{z}_r^{(k)} = \text{Pooling}(\text{Enc}_k(G_r)), \tag{1}$$

$$\mathbf{z}_s^{(k)} = \text{Pooling}(\text{Enc}_k(G_s^{(k)})). \tag{2}$$

Then we build a unified representation for each attribute view by concatenating the real and synthetic branches, i.e.,

$$\mathbf{z}^{(k)} = [\mathbf{z}_r^{(k)}; \mathbf{z}_s^{(k)}]. \tag{3}$$

Reliability-aware Adaptive Fusion. To model the reliability of synthetic propagation, we introduce a distributional uncertainty estimator for each attribute-specific view $\mathbf{z}^{(k)}$. We implement it with a Gaussian distribution and sample from it via the reparameterization trick [Kingma and Welling, 2013]:

$$[\boldsymbol{\mu}^{(k)}, \log \boldsymbol{\sigma}^{(k)2}] = \text{Probe}_k(\mathbf{z}^{(k)}), \tag{4}$$

$$\mathbf{m}^{(k)} = \boldsymbol{\mu}^{(k)} + \boldsymbol{\sigma}^{(k)} \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}), \tag{5}$$

where $\text{Probe}_k(\cdot)$ is a lightweight MLP.

Since the standard deviation of the Gaussian posterior is given by $\boldsymbol{\sigma}^{(k)}$, we estimate the uncertainty of each attribute view analytically and convert it into a reliability weight:

$$c^{(k)} = 1 - \text{Softmax}\Big(\frac{1}{d}\sum_{i=1}^{d} \boldsymbol{\sigma}_i^{(k)}\Big)_k, \tag{6}$$

where d is the feature dimension. Other probabilistic distributions, such as the Beta distribution, can also be used for reliability estimation [Han *et al.*, 2022; Ma *et al.*, 2024]; we further discuss this in Section 4.4.

Finally, we perform reliability-weighted aggregation over attribute-specific views, followed by multi-head self-attention to obtain the final representation:

$$\mathbf{h}_{final} = \text{Attention}([c^{(1)}, \ldots, c^{(K)}] \cdot [\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(K)}]^{\mathrm{T}}). \tag{7}$$

Fake News Detection. The final prediction of fake news is computed by:

$$\hat{\mathbf{y}} = \text{Softmax}(\mathbf{W}_{cls}\mathbf{h}_{final} + \mathbf{b}_{cls}), \tag{8}$$

where $\mathbf{W}_{cls}$ and $\mathbf{b}_{cls}$ are learnable parameters.

3.3 Training Objectives based on Mutual Information

We design a mutual information based objective to learn compressed and task-sufficient propagation representations.

Our goal is to i) minimize the mutual information between the input of synthetic propagation and latent representations to suppress redundant noise across different attribute views, ii) maximize the mutual information between real and synthetic propagation representations to reduce distributional discrepancy for propagation alignment, and iii) maximize the mutual information between latent representations and task labels to ensure task sufficiency for both fake news detection and user attribute prediction.

Formally, let X denote the input graph (i.e., the unified view of real and synthetic propagation), Y is the ground-truth task label (i.e., graph-level label Y_G for fake news detection or node-level label Y_V for user attribution label), node-level and graph-level representations are Z_V and Z_G. Given the Markov chain constraint $(Y_G, Y_V) \rightarrow X \rightarrow Z_V \rightarrow Z_G$, the overall objective of InfoPDF is formulated as:

$$\max \underbrace{I(Z_G; Y_G) + \alpha I(Z_V; Y_V)}_{\text{Task Sufficiency}} - \gamma \underbrace{I(X; Z_G)}_{\text{Compression}} + \lambda \underbrace{I(Z_G^r; Z_G^s)}_{\text{Alignment}}, \tag{9}$$

where Z_G^r and Z_G^s are graph-level representations learned from real and synthetic propagation. α, λ and γ are trade-off hyperparameters controlling the weights of task supervision, propagation alignment, and information compression.

Task Sufficiency. We jointly optimize **fake news detection** and **user attribute prediction** via supervised learning objectives, which serve as variational lower bounds of the corresponding mutual information terms $I(Z_G; Y_G)$ and $I(Z_V; Y_V)$. First, maximizing $I(Z_G; Y_G)$ encourages the learned graph-level representation Z_G to be maximally informative for fake news classification task, which is implemented by minimizing the binary cross-entropy loss,

$$\mathcal{L}_{cls} = -\log p_{\theta_{cls}}(Y_G | Z_G), \tag{10}$$

where θ_{cls} refers to trainable parameters for fake news classifier. Minimizing $\mathcal{L}_{cls}$ corresponds to maximizing the conditional log-likelihood $\log p(Y_G|Z_G)$ and thus a lower bound of $I(Z_G; Y_G)$.

Additionally, to encourage node-level representations to retain attribute-discriminative information, we introduce an auxiliary node-level attribute prediction objective to maximize $I(Z_V; Y_V)$. We minimize the following negative log-likelihood loss, i.e.,

$$\mathcal{L}_{attr} = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{\log |\mathcal{C}_k|} \left(-\log p_{\theta_{attr}^{(k)}} \left(Y_V^{(k)} \mid Z_V^{(k)} \right) \right), \tag{11}$$

where $|\mathcal{C}_k|$ is the number of classes for attribute a_k. $\theta_{attr}^{(k)}$ refers to trainable parameters for the k-th attribution classifier. We use the log-scale factor to balance attribute spaces with different labels. Minimizing $\mathcal{L}_{attr}$ maximizes the conditional likelihood of attribute labels given node representations, serving as a variational lower bound for $I(Z_V; Y_V)$.

Information Compression. To suppress redundant information introduced by LLM-generated synthetic propagation, we adopt the variational information bottleneck [Alemi *et al.*, 2017] to minimize the mutual information between the latent representation and the input view. Specifically, we approximate each attribute-specific view with a variational Gaussian posterior $q(\mathbf{m}^{(k)}|\mathbf{z}^{(k)}) = \mathcal{N}(\boldsymbol{\mu}^{(k)}, \boldsymbol{\sigma}^{2(k)})$, and impose a standard normal prior $p(\mathbf{m}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. The mutual information $I(X; Z)$ can be upper bounded by the Kullback–Leibler (KL) divergence between the posterior and the prior. That is, we minimize the following compression loss,

$$\mathcal{L}_{comp} = \sum_{k=1}^{K} D_{KL}\big(q(\mathbf{m}^{(k)}|\mathbf{z}^{(k)}) \,\|\, p(\mathbf{m})\big). \tag{12}$$

Minimizing $\mathcal{L}_{comp}$ reduces the upper bound of $I(X; Z)$, encouraging accurate reliability estimation of each attribute-specific view by suppressing task-irrelevant noise.

Propagation Alignment. Since directly estimating mutual information is intractable, we adopt a deterministic alignment objective as a surrogate. Specifically, we introduce a bi-directional alignment loss function to reduce the gap between real and synthetic propagation representations. The intuition is that when two views share high mutual information, one view should be predictive of the other [Yu *et al.*, 2025]. For each attribute view k, we employ two lightweight predictors $f_k(\cdot)$ and $g_k(\cdot)$ (implemented as MLPs) to reconstruct one representation from the other. The loss function is defined as,

$$\mathcal{L}_{align} = \frac{1}{K} \sum_{k=1}^{K} \left(\|f_k(\mathbf{z}_s^{(k)}) - \mathbf{z}_r^{(k)}\|_2^2 + \|g_k(\mathbf{z}_r^{(k)}) - \mathbf{z}_s^{(k)}\|_2^2 \right). \tag{13}$$

Overall, the training loss of InfoPDF is defined as:

$$\mathcal{L} = \mathcal{L}_{cls} + \alpha \mathcal{L}_{attr} + \lambda \mathcal{L}_{align} + \gamma \mathcal{L}_{comp}, \tag{14}$$

where α, λ, and γ are trade-off hyperparameters.

3.4 LLM-free Inference

Inspired by efficient LLM-free deployment [Hu *et al.*, 2024], we provide an inference mode without online propagation simulation. The trained predictor g_k estimates the synthetic attribute representation as

$$\hat{z}_s^{(k)} = g_k(z_r^{(k)}). \tag{15}$$

Then, $\hat{z}^{(k)} = [z_r^{(k)}; \hat{z}_s^{(k)}]$ is fed into reliability-aware fusion to obtain h_{final} for detection.

4 Experiments

4.1 Experimental Setups

Datasets. We conduct experiments on three public datasets, i.e., Twitter [Lin *et al.*, 2022], PHEME [Zubiaga *et al.*, 2016], and CED [Song *et al.*, 2019].

Comparison Methods. We compare InfoPDF with 11 content-based methods including BERT [Devlin *et al.*, 2018] and DeBerta [He *et al.*, 2020], EANN [Wang *et al.*, 2018], dEFEND [Shu *et al.*, 2019a], DualEmo [Zhang *et al.*, 2021],

	Method	Twitter Acc	Twitter F1	CED Acc	CED F1	PHEME Acc	PHEME F1
Content-based	BERT	71.12±2.43	70.86±2.71	88.44±0.96	88.43±0.89	82.96±0.61	80.57±0.82
	DeBerta	75.43±1.92	74.47±1.87	89.97±0.73	89.32±0.79	82.54±1.09	80.24±1.21
	EANN	72.84±1.68	72.41±1.79	89.73±0.76	89.69±0.72	83.27±0.41	81.07±0.27
	dEFEND	75.12±1.41	73.56±1.47	90.12±0.63	89.98±0.71	83.24±0.84	82.45±0.91
	DualEmo	73.27±1.57	73.01±1.66	91.21±0.71	91.20±0.71	83.60±0.88	83.34±0.86
	LLM_{text}	56.00±4.23	54.60±4.41	48.08±2.53	37.22±2.74	33.74±3.35	28.64±3.52
	ARG	78.66±2.35	78.59±2.42	92.92±0.31	92.88±0.35	79.74±0.51	77.52±0.42
	$\text{LLM}_{\text{comment}}$	60.75±3.64	60.47±3.78	80.10±1.41	80.05±1.38	53.51±2.41	52.77±2.58
	TELLER	64.68±3.50	62.27±6.13	74.19±0.71	73.88±0.53	60.88±0.07	38.65±0.38
	GenFEND						
	w/ BERT	78.26±1.38	75.64±1.52	90.23±0.59	90.04±0.63	83.84±0.81	81.23±0.87
	w/ dEFEND	82.25±0.74	82.04±0.81	93.23±0.24	92.96±0.27	85.07±0.39	84.66±0.45
	CAMERED	82.85±0.65	82.52±0.71	93.58±0.22	93.25±0.25	85.44±0.36	85.02±0.42
Propagation-based	GCN	78.02±2.91	77.67±3.08	91.59±1.12	91.17±1.08	80.59±0.30	75.55±1.05
	GAT	79.31±2.73	78.86±2.87	88.23±1.29	87.68±1.31	80.41±0.46	77.13±1.25
	GraphSAGE	81.03±2.58	79.75±2.69	89.26±1.19	88.92±1.26	79.03±1.29	75.70±0.52
	BiGCN	82.76±1.56	82.71±1.49	91.74±0.69	91.63±0.74	82.70±0.35	80.28±0.88
	UPFD	79.74±1.89	79.02±2.01	88.79±1.02	88.38±0.98	81.05±0.41	77.53±1.12
	EBGCN	83.19±0.91	82.60±0.98	93.96±0.36	93.76±0.39	83.89±0.28	80.72±0.52
	UPSR	83.62±0.87	83.22±0.94	93.36±0.33	92.96±0.36	82.95±0.31	82.95±0.37
	RAGCL	84.42±0.72	84.15±0.78	94.10±0.29	93.95±0.32	84.24±0.25	83.58±0.41
	$\text{LLM}_{\text{propagation}}$	48.84±4.42	46.76±4.59	71.15±1.91	67.32±2.04	52.14±2.64	52.04±2.78
	DELL						
	- *single*	80.17±1.62	79.75±1.54	92.12±0.71	91.69±0.68	82.60±0.38	80.75±0.73
	- *vanilla*	80.43±0.58	80.32±0.59	92.23±0.62	91.75±0.66	82.52±0.42	80.60±0.68
	- *confidence*	78.97±1.69	78.10±1.73	91.03±0.68	90.10±0.72	83.03±0.45	81.08±0.76
	- *selective*	81.22±1.43	81.02±1.51	92.28±0.59	91.32±0.65	83.29±0.39	82.75±0.62
	EIN	84.33±0.64	83.97±0.71	94.25±0.26	94.08±0.29	85.27±0.23	84.46±0.38
	InfoPDF (Ours)						
	- GCN	85.63*±0.39	85.32*±0.40	96.36*±0.07	96.24*±0.06	87.65*±0.24	85.47±0.26
	- GAT	85.43*±0.32	85.15*±0.34	**97.40***±0.18	**97.28***±0.18	**88.17***±0.21	85.96*±0.28
	- GraphSAGE	**86.90***±0.34	**86.51***±0.35	97.25*±0.18	97.13*±0.19	88.11*±0.22	**86.11***±0.25
	InfoPDF-Inference (Ours)						
	- GCN	84.78±0.43	84.45±0.46	94.48±0.12	94.31±0.13	85.42±0.28	84.68±0.31
	- GAT	84.62±0.36	84.28±0.39	94.55±0.21	94.38±0.22	85.58±0.24	84.85±0.27
	- GraphSAGE	84.95±0.38	84.72±0.41	94.71±0.20	94.56±0.21	85.71±0.25	84.92±0.28

Table 1: Results (%) for fake news detection on three datasets. The best results are in **boldface**. InfoPDF-Inference represents the LLM-free inference mode of InfoPDF. * represents statistical significance over state-of-the-art baselines under the t-test ($p < 0.05$).

ARG [Hu *et al.*, 2024], GenFEND [Nan *et al.*, 2024], TELLER [Liu *et al.*, 2024a], CAMERED [Wang *et al.*, 2025], LLM_{text}, and $\text{LLM}_{\text{comment}}$ [Chen *et al.*, 2025], and 11 propagation-based methods including GCN [Kipf and Welling, 2017], GAT [Veličković *et al.*, 2018], GraphSAGE [Hamilton *et al.*, 2017], Bi-GCN [Bian *et al.*, 2020], EBGCN [Wei *et al.*, 2021], UPSR [Wei *et al.*, 2022], RAGCL [Cui and Jia, 2024], UPFD [Dou *et al.*, 2021], DELL [Wan *et al.*, 2024], EIN [Jiang *et al.*, 2025], and $\text{LLM}_{\text{propagation}}$ [Chen *et al.*, 2025].

Evaluation Metrics. We evaluate with the accuracy score (Acc) and macro-F1 score (F1).

Implementation Details. All experiments are implemented on a single NVIDIA Tesla V100 GPU. The size of synthetic propagation graphs is fixed to 30 nodes (1 root news node and 29 comment nodes). For the generation of synthetic propagation, we employ *Mistral-7B-Instruct-v0.2* for Twitter and PHEME, and *Qwen1.5-7B-Chat* for CED.

4.2 Main Results

We evaluate the effectiveness of InfoPDF across different detection scenarios.

General Detection. Table 1 reports the performance comparison on three public datasets. Overall, InfoPDF achieves strong and generally superior performance across datasets and GNN backbones, demonstrating its effectiveness and strong generalization under incomplete propagation. For instance, InfoPDF improves the GraphSAGE backbone by +5.87% accuracy scores on Twitter. Moreover, LLM-free variant InfoPDF-Inference remains competitive with strong baselines while reducing the deployment cost by avoiding online LLM-based propagation simulation. This validates that InfoPDF can distill the benefit of LLM-generated synthetic propagation into an efficient LLM-free detector, making it

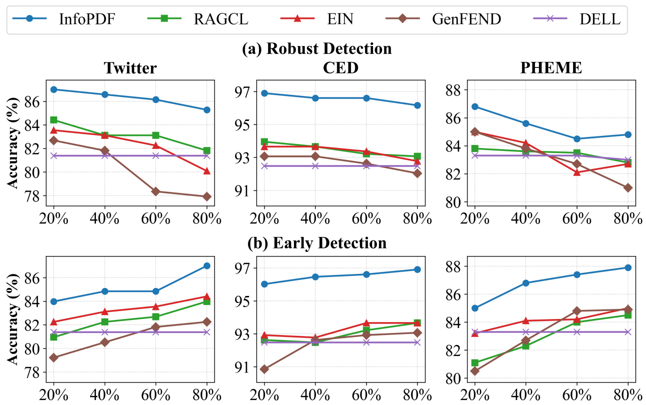

Figure 3: Results (%) of robust and early fake news detection on two incomplete propagation scenarios.

Method	Twitter		CED		PHEME	
	Acc	F1	Acc	F1	Acc	F1
InfoPDF	**86.90**	**86.51**	**97.25**	**97.13**	**88.11**	**86.11**
w/o RP	84.34	84.10	94.05	93.89	86.21	85.97
w/o Attr	84.70	84.23	94.79	94.54	86.04	85.64
w/o RW	83.84	83.33	94.39	94.25	85.78	85.14
w/o Attention	84.48	84.14	94.54	94.36	85.86	85.24
w/o $\mathcal{L}_{align}$	83.20	82.93	93.95	93.85	85.34	84.91
w/o $\mathcal{L}_{comp}$	84.42	84.17	94.40	94.16	85.87	85.48
w/o $\mathcal{L}_{attr}$	84.99	84.83	94.69	94.32	86.12	85.78

Table 2: Ablation results on three datasets. The best performance is highlighted in bold. We apply GraphSAGE as the GNN backbone.

more practical for real-world deployment.

Robust Detection. Figure 3(a) shows the results of InfoPDF and baselines for robust detection. Specifically, we simulate noisy propagation settings by randomly removing edges and masking node features with ratios of 20%, 40%, 60%, and 80%. From the results, InfoPDF shows smaller performance degradation, exhibiting stronger robustness towards different incomplete propagation settings.

Early Detection. Figure 3(b) shows the results for early detection on three datasets. We preserve the temporal order of propagation and use only partial propagation observed at early time steps for prediction. InfoPDF achieves competitive performance even when only limited early propagation is available, showing its effectiveness in early detection.

4.3 Ablation Study

We conduct ablation studies to examine the contribution of each component in InfoPDF. The ablation results are shown in Table 2. For propagation modeling for fake news detection, removing the incomplete real propagation (*w/o* RP) consistently degrades performance, showing the necessity of real propagation data. Replacing attribute-specific simulation with attribute-mixed generation (*w/o* Attr) leads to clear drops, indicating that modeling attributes separately yields more reliable synthetic signals. Moreover, removing either reliability-aware weighting (*w/o* RW) or multi-head self-attention fusion causes performance degradation, verifying their effectiveness for detection. For ablation variants of training objectives, removing the propagation alignment objective (*w/o* $\mathcal{L}_{align}$) results in a large performance drop, highlighting the importance of reducing the representation gap between real and synthetic propagation. Removing the compression regularization (*w/o* $\mathcal{L}_{comp}$) also consistently harms performance, showing that the compression helps filter redundant features in representations. Finally, removing attribute supervision (*w/o* $\mathcal{L}_{attr}$) leads to further degradation, demonstrating that preserving the sufficiency of representations for user attributes is beneficial for detection.

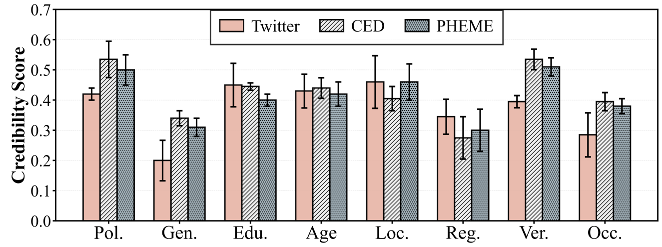

Figure 4: Attribute-level reliability distributions with user variability across three datasets.

Dataset	Method	Supervised Metrics			Unsupervised Metrics		
		ARI↑	NMI↑	FMI↑	SC↑	CHI↑	DBI↓
Twitter	RP	30.14	23.51	65.08	0.24	85.30	1.52
	DF	39.34	30.87	69.56	0.23	77.60	1.59
	InfoPDF	**43.82**	**34.67**	**71.79**	**0.56**	**443.54**	**0.59**
CED	RP	65.28	57.60	82.70	0.56	1887.89	0.50
	DF	60.13	50.60	80.10	0.38	543.31	1.06
	InfoPDF	**89.09**	**81.56**	**94.54**	**0.67**	**2564.49**	**0.43**
PHEME	RP	41.47	28.77	74.58	0.11	139.66	2.53
	DF	40.36	28.71	73.07	0.10	133.82	2.77
	InfoPDF	**51.12**	**39.52**	**77.64**	**0.49**	**1699.84**	**0.74**

Table 3: Representation evaluation results (%) against the propagation data from different strategies. Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and Fowlkes-Mallows Index (FMI) evaluate the accuracy of clustering. Silhouette Coefficient (SC), Calinski-Harabasz Index (CHI), and Davies-Bouldin Index (DBI) evaluate the separation and compactness of clustering. SC, CHI, and DBI are evaluated based on K-Means, and we define the number of clusters K as the true number of categories.

4.4 Further Analysis

We provide more analysis on the reliability of synthetic propagation, representation quality under different fusion strategies, reliability estimation, and hyperparameter sensitivity.

Reliability Analysis of Synthetic Propagation. Figure 4 visualizes the attribute-view reliability scores estimated by the Beta variant of InfoPDF across three datasets. Overall, we observe clear differences in the estimated reliability across attribute views, indicating that LLM-generated propagation signals exhibit varying levels of consistency and usefulness. Moreover, the reliability scores also vary across instances within the same attribute view, suggesting that InfoPDF performs instance-wise reliability estimation rather than relying on static attribute-level priors. These results support the ef-

Method	Twitter		CED		PHEME	
	Acc	F1	Acc	F1	Acc	F1
InfoPDF (Gaussian)						
w/ GCN	85.63	85.32	96.36	96.24	87.65	85.47
w/ GAT	85.43	85.15	**97.40**	**97.28**	**88.17**	85.96
w/ SAGE	**86.90**	**86.51**	97.25	97.13	88.11	**86.11**
InfoPDF (Beta)						
w/ GCN	85.34	85.01	96.08	95.94	87.68	85.36
w/ GAT	85.35	85.04	95.89	95.74	87.93	85.61
w/ SAGE	85.34	85.03	96.17	96.01	87.41	85.15

Table 4: Comparison results (%) of InfoPDF using different distributions for reliability estimation. Best results are in **bold**.

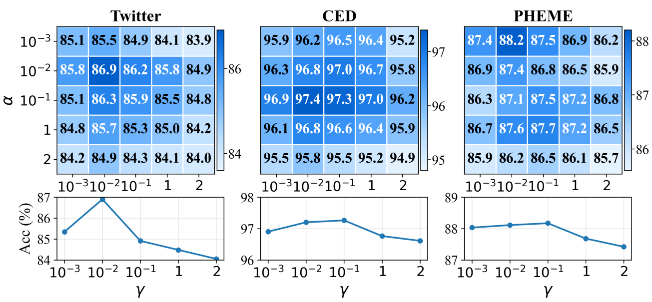

Figure 5: Performance analysis with varying α, λ and γ parameters. We use accuracy score (%) for evaluation.

fectiveness of reliability-aware fusion in adaptively leveraging informative synthetic views.

Representation Quality Analysis. We analyze the quality of representations learned by InfoPDF and several variants including: 1) Real Propagation (RP), which encodes only the incomplete real propagation; and 2) Direct Fusion (DF), which fuses the real propagation representation with synthetic representations directly. Following existing studies [Hu *et al.*, 2023], we compute supervised clustering metrics (ARI, NMI, FMI) and unsupervised clustering metrics (SC, CHI, DBI) to evaluate the representation quality. Table 3 reports results of our method and comparison variants. The more accurate supervised clustering results demonstrate the class discrimination capability of InfoPDF. Better results in terms of unsupervised metrics indicate the high structural quality of representations learned by our method, reflected in compact and well-separated clustering patterns.

Reliability Estimation with Different Distributions. We study reliability estimation with different probabilistic distributions. Besides the Gaussian distribution, we implement InfoPDF with a Beta variant by converting the logits produced by each attribute-specific probe into non-negative evidence and using the resulting uncertainty to compute reliability weights [Han *et al.*, 2022; Ma *et al.*, 2024]. As shown in Table 4, InfoPDF achieves competitive performance with both Gaussian and Beta distributions, demonstrating that our probabilistic reliability estimation remains effective under different distributional choices. These results further suggest that distributional uncertainty provides a useful signal for modeling the reliability of synthetic propagation views.

Hyperparameter Analysis. We analyze the sensitivity of InfoPDF to α, λ, and γ in Figure 5. A small λ (e.g., 0.01) generally performs well, suggesting that mild alignment helps reduce the real–synthetic gap. The optimal α varies across datasets, likely due to different attribute distributions and propagation patterns. Moderate γ values better balance denoising and information preservation, while excessive compression may remove task-relevant signals. Overall, InfoPDF shows reasonable stability under different hyperparameter settings.

5 Related Work

Fake News Detection. Existing methods primarily focus on content and propagation. Content-based methods extract semantic patterns via feature engineering [Castillo *et al.*, 2011; Ma *et al.*, 2015], deep neural networks [Ruchansky *et al.*, 2017; Yu *et al.*, 2017], and PLMs [Kaliyar *et al.*, 2021]. Auxiliary tasks like sentiment analysis [Luvembe *et al.*, 2023] and comment utilization [Shu *et al.*, 2019a] are also employed to complement textual features. Propagation-based methods model dissemination structures using time series [Liu and fang Brook Wu, 2018] or graphs [Bian *et al.*, 2020; Wei *et al.*, 2021; Wei *et al.*, 2022], sometimes incorporating multi-relational interactions [Dou *et al.*, 2021]. However, incomplete propagation remains a challenge [Wei *et al.*, 2021; Ma *et al.*, 2022]. While contrastive learning [Ma *et al.*, 2022; Cui and Jia, 2024] and uncertainty-aware modeling [Wei *et al.*, 2021; Wei *et al.*, 2022; Lian *et al.*, 2026] improve robustness, existing propagation-based methods are still vulnerable to the incomplete propagation scenarios.

LLM-based Propagation Generation. LLMs demonstrate significant potential in simulating human behavior and social knowledge [Liu *et al.*, 2024b; Wang *et al.*, 2025], having recently been established as a verifiable and scalable research method for capturing complex social dynamics [Anthis *et al.*, 2025]. Studies on fake news detection have moved beyond simple discussion simulation [Nan *et al.*, 2024; Wan *et al.*, 2024] to modeling intricate dissemination behaviors such as liking, sharing, and content regulation within multi-agent social environments [Liu *et al.*, 2025]. However, these methods often overlook the inherently unreliable risks in synthetic propagation. We mitigate the unreliability in LLM-generated propagation from the mutual information perspective.

6 Conclusion

This paper alleviates the unreliability in LLM-generated propagation from the mutual information perspective. We propose a novel InfoPDF to learn effective representations from both real and multiple LLM-generated synthetic propagation for fake news detection. InfoPDF denoises unreliable synthetic propagation via a variational information bottleneck and maintains the consistency between real and synthetic propagation representations as well as the task sufficiency of fake news detection and user attribute prediction. Experiments on three datasets demonstrate that InfoPDF achieves superior performance and robustness in data-sparse scenarios, while its LLM-free inference mode offers a practical solution for efficient deployment.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No.U24A20335), the China Postdoctoral Science Foundation (No.2024M753481), and Youth Innovation Promotion Association CAS.

References

[Alemi *et al.*, 2017] Alexander A Alemi, Ian Fischer, Joshua V Dillon, et al. Deep variational information bottleneck. In *International Conference on Learning Representations*, 2017.

[Anthis *et al.*, 2025] Jacy Reese Anthis, Ryan Liu, Sean M Richardson, et al. Position: Llm social simulations are a promising research method. In *ICML Position Paper Track*, 2025.

[Bian *et al.*, 2020] Tian Bian, Xi Xiao, Tingyang Xu, et al. 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, 2020.

[Castillo *et al.*, 2011] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In *WWW*, pages 675–684, 2011.

[Chen *et al.*, 2025] Mengyang Chen, Lingwei Wei, Han Cao, et al. Explore the potential of llms in misinformation detection: An empirical study. In *AAAI 2025 Workshop on Preventing and Detecting LLM Misinformation (PDLM)*, 2025.

[Chopra *et al.*, 2005] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, volume 1, pages 539–546. IEEE, 2005.

[Cui and Jia, 2024] Chaoqun Cui and Caiyan Jia. 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, 2024.

[Devlin *et al.*, 2018] Jacob Devlin, Ming-Wei Chang, Kenton Lee, et al. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[Dou *et al.*, 2021] Yingtong Dou, Kai Shu, Congying Xia, et al. User preference-aware fake news detection. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2051–2055, 2021.

[Gaillard *et al.*, 2021] Stefan Gaillard, Zoril A Oláh, Stephan Venmans, et al. Countering the cognitive, linguistic, and psychological underpinnings behind susceptibility to fake news: A review of current literature with special focus on the role of age and digital literacy. *Frontiers in Communication*, 6:661801, 2021.

[Hamilton *et al.*, 2017] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *NIPS*, 30, 2017.

[Han *et al.*, 2022] Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification with dynamic evidential fusion. *IEEE transactions on pattern analysis and machine intelligence*, 45(2):2551–2566, 2022.

[He *et al.*, 2020] Pengcheng He, Xiaodong Liu, Jianfeng Gao, et al. Deberta: Decoding-enhanced bert with disentangled attention. *ArXiv*, abs/2006.03654, 2020.

[Hu *et al.*, 2023] Dou Hu, Yinan Bao, Lingwei Wei, et al. Supervised adversarial contrastive learning for emotion recognition in conversations. In *Proceedings of ACL*, pages 10835–10852, 2023.

[Hu *et al.*, 2024] Beizhe Hu, Qiang Sheng, Juan Cao, et al. Bad actor, good advisor: Exploring the role of large language models in fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 22105–22113, 2024.

[Ji *et al.*, 2023] Ziwei Ji, Tiezheng Yu, Yan Xu, et al. Towards mitigating llm hallucination via self reflection. In *Findings of the Association for Computational Linguistics: EMNLP*, pages 1827–1843, 2023.

[Jiang *et al.*, 2025] Wei Jiang, Tong Chen, Xinyi Gao, et al. Epidemiology-informed network for robust rumor detection. In *Proceedings of the ACM Web Conference*, pages 3618–3627, 2025.

[Kaliyar *et al.*, 2021] Rohit Kumar Kaliyar, Anurag Goswami, and Pratik Narang. Fakebert: Fake news detection in social media with a bert-based deep learning approach. *Multimedia tools and applications*, 80(8):11765–11788, 2021.

[Kingma and Welling, 2013] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[Kipf and Welling, 2017] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

[Lian *et al.*, 2026] Xinglin Lian, Chengtai Cao, Ting Zhong, Yong Wang, Kai Chen, and Fan Zhou. Decompose to understand, fuse to detect: Frequency-decoupled anomaly detection for encrypted network traffic. *arXiv: 2605.02970*, 2026.

[Lim and Perrault, 2021] Gionnieve Lim and Simon Tangi Perrault. Local perceptions and practices of news sharing and fake news. In *Companion Publication of the Conference on Computer Supported Cooperative Work and Social Computing*, pages 117–120, 2021.

[Lin *et al.*, 2022] Hongzhan Lin, Jing Ma, Liangliang Chen, Zhiwei Yang, Mingfei Cheng, and Chen Guang. Detect rumors in microblog posts for low-resource domains via adversarial contrastive learning. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 2543–2556, 2022.

[Liu and fang Brook Wu, 2018] Yang Liu and Yi fang Brook Wu. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.

[Liu *et al.*, 2024a] Hui Liu, Wenya Wang, Haoru Li, et al. Teller: A trustworthy framework for explainable, generalizable and controllable fake news detection. In *Findings of ACL*, pages 15556–15583, 2024.

[Liu *et al.*, 2024b] Yuhan Liu, Xiuying Chen, Xiaoqing Zhang, et al. From skepticism to acceptance: Simulating the attitude dynamics toward fake news. In *International Joint Conference on Artificial Intelligence*, 2024.

[Liu *et al.*, 2025] Genglin Liu, Vivian T Le, Salman Rahman, et al. Mosaic: Modeling social ai for content dissemination and regulation in multi-agent simulations. In *Conference on Empirical Methods in Natural Language Processing*, pages 6401–6428, 2025.

[Luvembe *et al.*, 2023] Alex Munyole Luvembe, Weimin Li, Shaohua Li, et al. Dual emotion based fake news detection: A deep attention-weight update approach. *IPM*, 60(4):103354, 2023.

[Ma *et al.*, 2015] Jing Ma, Wei Gao, Zhongyu Wei, et al. Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 1751–1754, 2015.

[Ma *et al.*, 2022] Guanghui Ma, Chunming Hu, Ling Ge, et al. Towards robust false information detection on social networks with contrastive learning. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 1441–1450, 2022.

[Ma *et al.*, 2024] Zihan Ma, Minnan Luo, Hao Guo, et al. Event-radar: Event-driven multi-view learning for multimodal fake news detection. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 5809–5821, 2024.

[Nan *et al.*, 2024] Qiong Nan, Qiang Sheng, Juan Cao, et al. Let silence speak: Enhancing fake news detection with generated comments from large language models. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 1732–1742, 2024.

[Ruchansky *et al.*, 2017] Natali Ruchansky, Sungyong Seo, and Yan Liu. Csi: A hybrid deep model for fake news detection. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 797–806, 2017.

[Shu *et al.*, 2019a] Kai Shu, Limeng Cui, Suhang Wang, et al. defend: Explainable fake news detection. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 395–405, 2019.

[Shu *et al.*, 2019b] Kai Shu, Xinyi Zhou, Suhang Wang, et al. The role of user profiles for fake news detection. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 436–439, 2019.

[Song *et al.*, 2019] Changhe Song, Cheng Yang, Huimin Chen, et al. Ced: Credible early detection of social media rumors. *TKDE*, 33(8):3035–3047, 2019.

[Trott *et al.*, 2023] Sean Trott, Cameron Jones, Tyler Chang, et al. Do large language models know what humans know? *Cognitive Science*, 47(7):e13309, 2023.

[Veličković *et al.*, 2018] Petar Veličković, Guillem Cucurull, Arantxa Casanova, et al. Graph attention networks. In *International Conference on Learning Representations*, 2018.

[Wan *et al.*, 2024] Herun Wan, Shangbin Feng, Zhaoxuan Tan, Heng Wang, Yulia Tsvetkov, and Minnan Luo. Dell: Generating reactions and explanations for llm-based misinformation detection. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 2637–2667, 2024.

[Wang *et al.*, 2018] Yaqing Wang, Fenglong Ma, Zhiwei Jin, et al. Eann: Event adversarial neural networks for multimodal fake news detection. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 849–857, 2018.

[Wang *et al.*, 2025] Bing Wang, Bingrui Zhao, Ximing Li, et al. Collaboration and controversy among experts: Rumor early detection by tuning a comment generator. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 468–478, 2025.

[Wei *et al.*, 2021] Lingwei Wei, Dou Hu, Wei Zhou, et al. Towards propagation uncertainty: Edge-enhanced Bayesian graph convolutional networks for rumor detection. In *Annual Meeting of the Association for Computational Linguistics*, pages 3845–3854, August 2021.

[Wei *et al.*, 2022] Lingwei Wei, Dou Hu, Wei Zhou, et al. Uncertainty-aware propagation structure reconstruction for fake news detection. In *International Conference on Computational Linguistics*, pages 2759–2768, 2022.

[Yu *et al.*, 2017] Feng Yu, Qiang Liu, Shu Wu, et al. A convolutional approach for misinformation identification. In *International Joint Conference on Artificial Intelligence*, pages 3901–3907, 2017.

[Yu *et al.*, 2025] Xuejiao Yu, Guoqing Chao, Yi Jiang, et al. Incomplete multi-view clustering via mutual information. *IEEE Transactions on Multimedia*, 2025.

[Zhang *et al.*, 2021] Xueyao Zhang, Juan Cao, Xirong Li, et al. Mining dual emotion for fake news detection. In *Proceedings of the ACM Web Conference*, pages 3465–3476, 2021.

[Zubiaga *et al.*, 2016] Arkaitz Zubiaga, Maria Liakata, and Rob Procter. Learning reporting dynamics during breaking news for rumour detection in social media. *arXiv preprint arXiv:1610.07363*, 2016.