Meta-Path-based Fake News Detection Leveraging Multi-level Social Context Information

Jian Cui
S2W Inc.
Seongnam, South Korea
cj19960819@gmail.com

Kwanwoo Kim
Korea Advanced Institute of Science and Technology
Daejeon, South Korea
kw2128@kaist.ac.kr

Seung Ho Na
Korea Advanced Institute of Science and Technology
Daejeon, South Korea
harry.na@kaist.ac.kr

Seungwon Shin*
Korea Advanced Institute of Science and Technology
Daejeon, South Korea
claude@kaist.ac.kr

CIKM 2022

Reported by Minqin Li

Introduction

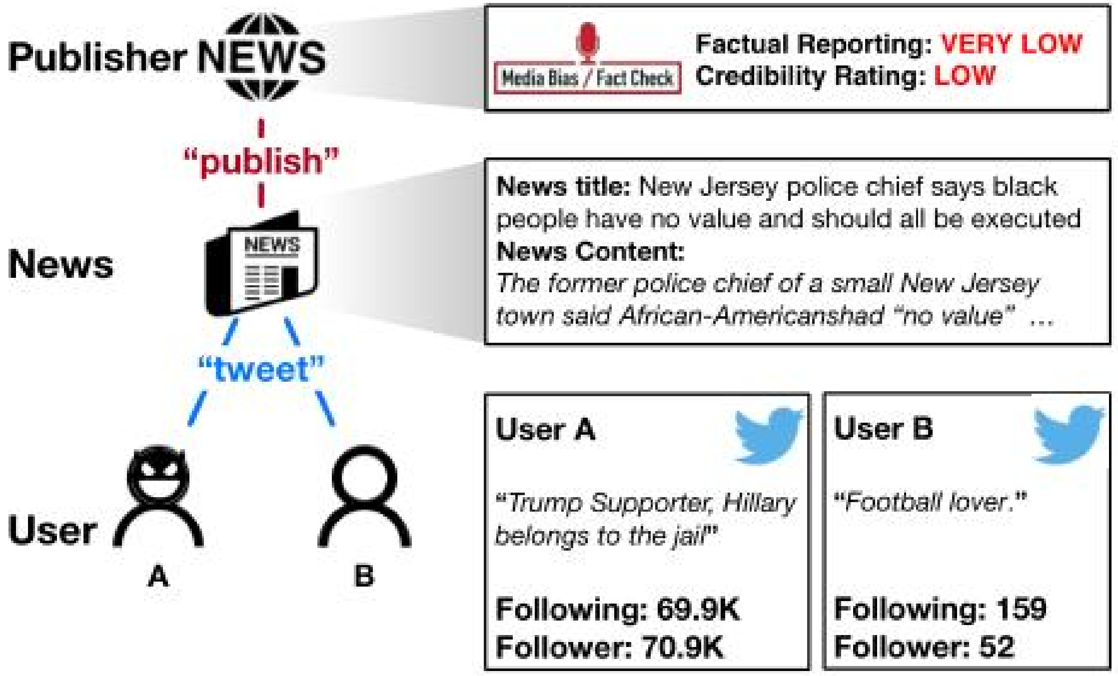


Figure 1: Example of fake news distribution and dissemination. Publishers publish the news, and users tweet the news. Some publishers are regarded as low credibility sources according to the famous fact-checking website, MBFC. User A is an example of an instigator in Twitter, and User B is an example of a regular user.

Introduction

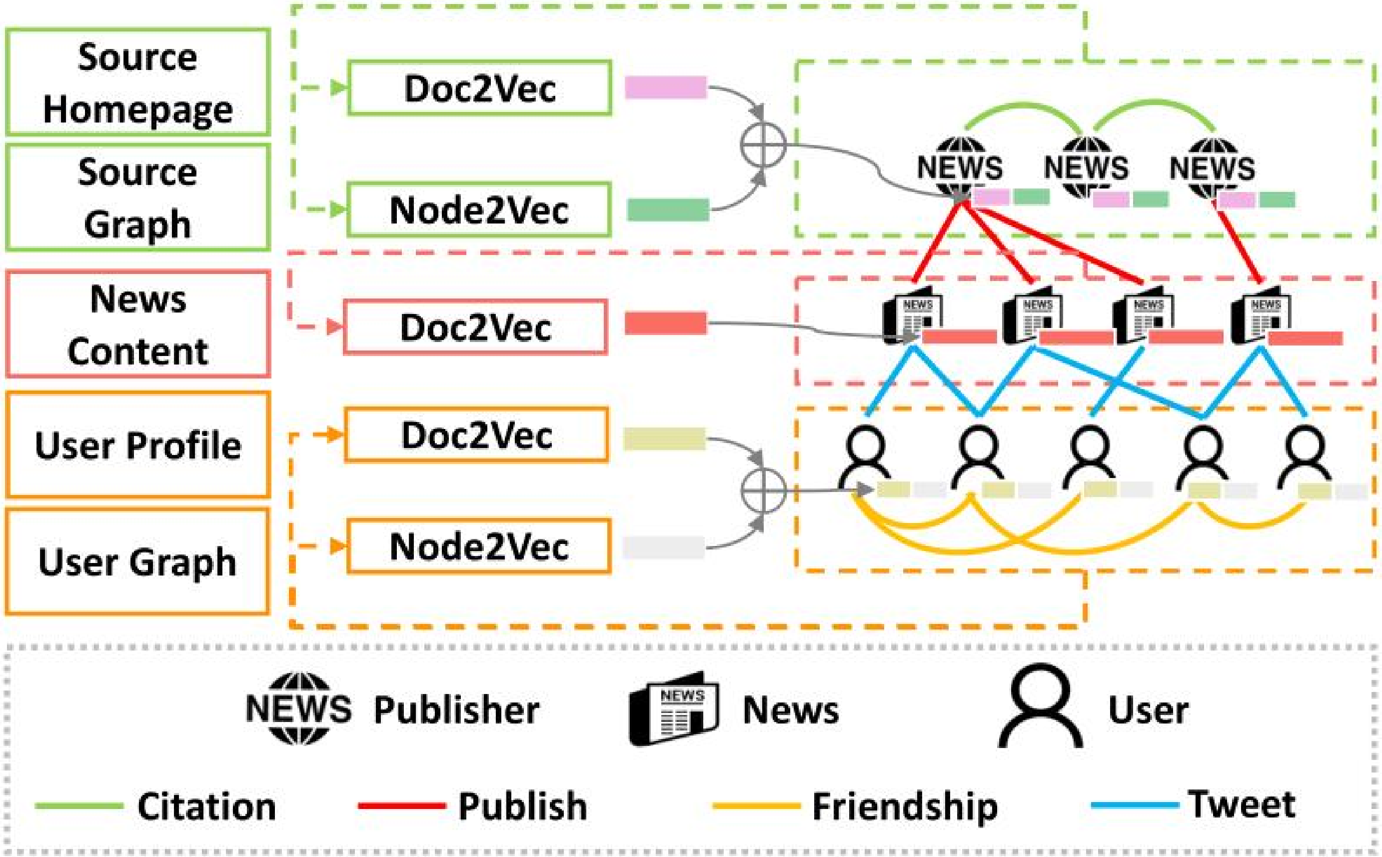


Figure 2: Heterogeneous Graph of News and Node Feature Engineering.

Method

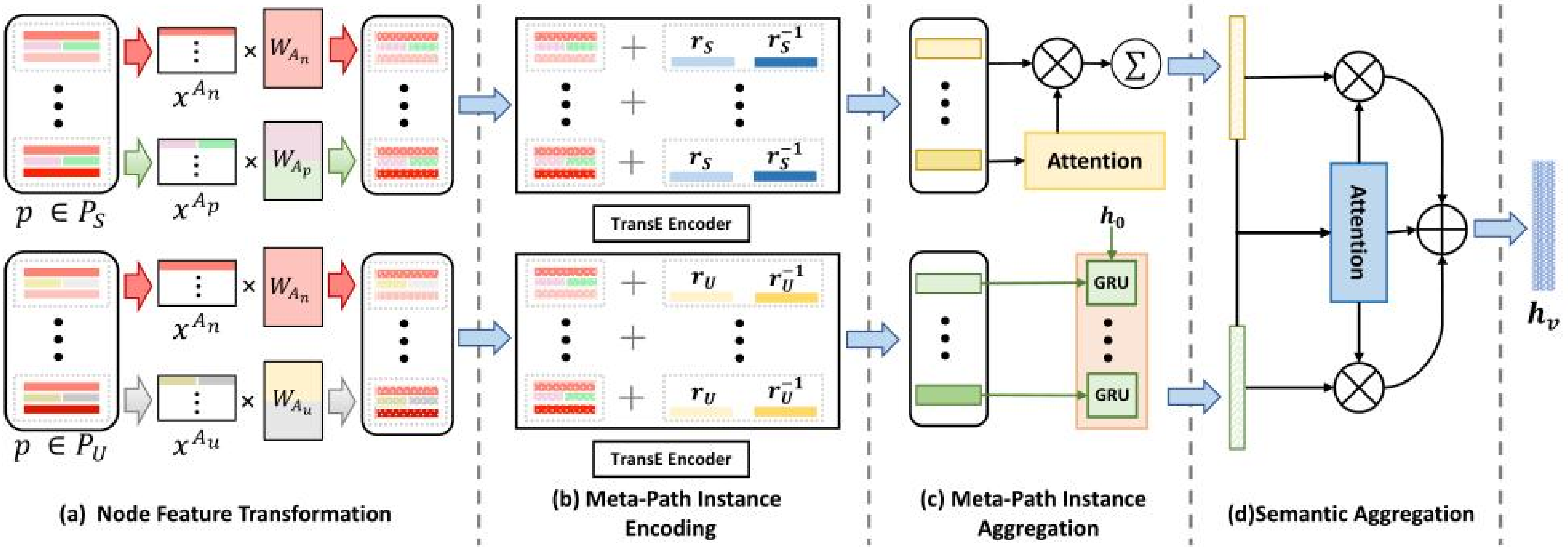


Figure 4: Architecture of *Hetero-SCAN*.

Method

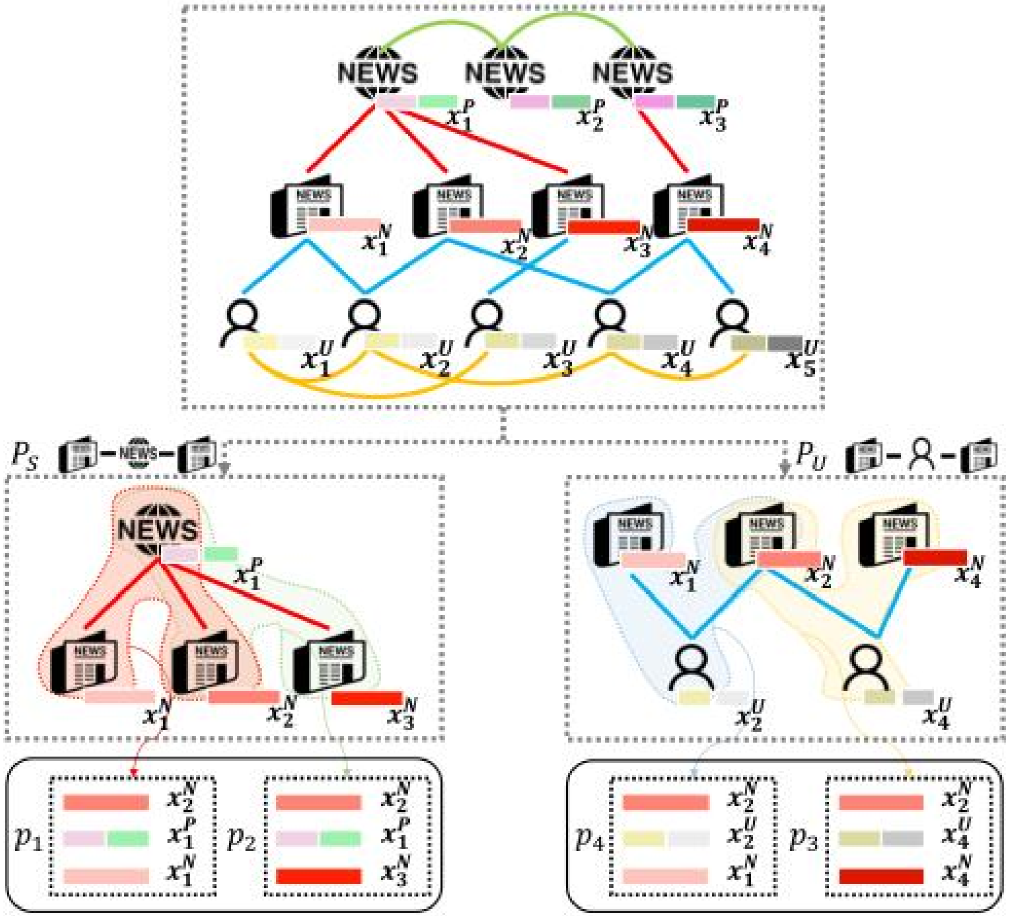


Figure 3: Extracting Meta-Path instances of the target news node x_2^N.

$$\mathcal{P} \in \{\mathcal{P}_U, \mathcal{P}_S\} \tag{1}$$

,where $\mathcal{P}_U$: $News \rightarrow User \rightarrow News$ and $\mathcal{P}_S$: $News \rightarrow Publisher \rightarrow News$.

$$\mathbf{h}_v^A = \mathbf{W}_A \cdot \mathbf{x}_v^A \tag{2}$$

Knowledge graph triple:$\mathbf{e}_s \xrightarrow{\mathbf{e}_p} \mathbf{e}_o$

$$\text{Meta-Path:}\mathbf{h}_u \xrightarrow{r} \mathbf{h}_w \xrightarrow{r^{-1}} \mathbf{h}_v \tag{3}$$

$$\mathbf{h}_p = f_{enc}(p) = f_{enc}(\mathbf{h}_u, r, \mathbf{h}_w, r^{-1}) \tag{4}$$

Method

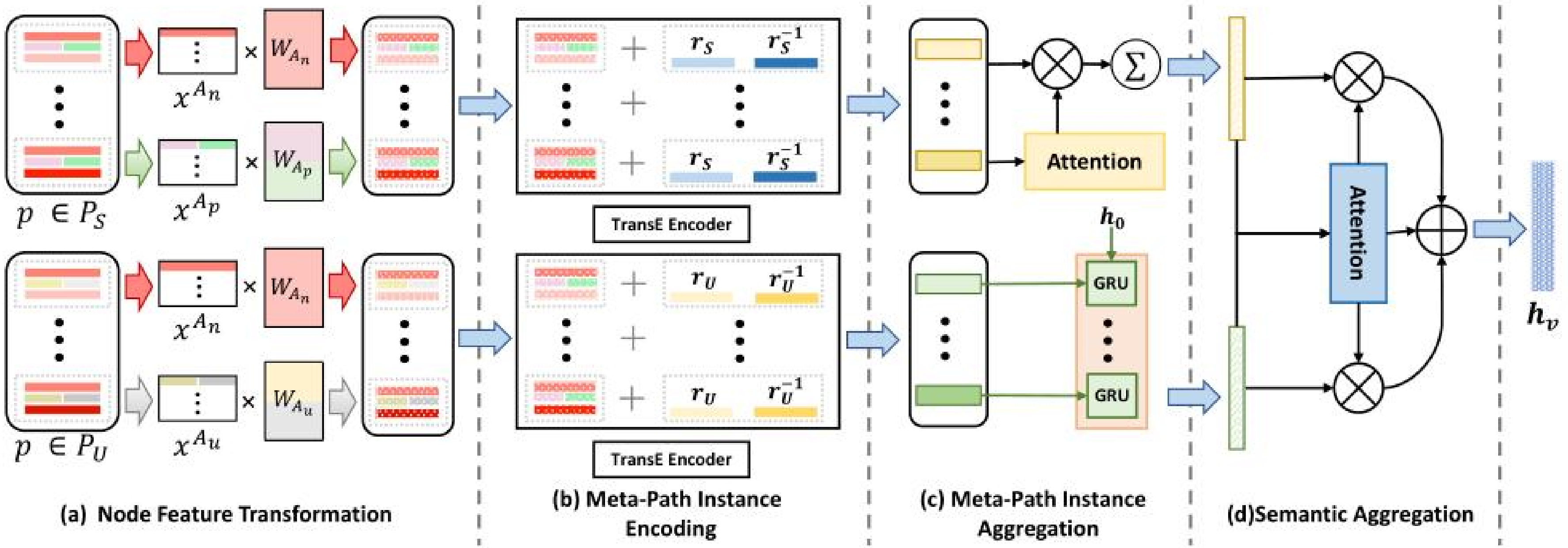


Figure 4: Architecture of *Hetero-SCAN*.

$$\mathbf{h}_v^{\mathcal{P}_S} = \Big\Vert_{k=1}^{K} \sigma\Big(\sum_{p \in \mathrm{P}_S} [\alpha_p]_k \cdot \mathbf{h}_p\Big) \tag{6}$$

$$e_p = LeakyReLU(\mathbf{a}^T \cdot \mathbf{h}_p)$$
$$\alpha_p = softmax(e_p) = \frac{exp(e_p)}{\sum_{p' \in \mathrm{P}_S} exp(e_{p'})} \tag{5}$$

$$\mathbf{h}_v^{\mathcal{P}_U} = \mathbf{GRU}(\mathbf{h}_{p_1}, \mathbf{h}_{p_2}, ..., \mathbf{h}_{p_n}), p_i \in \mathrm{P}_\mathrm{U} \tag{7}$$

Method

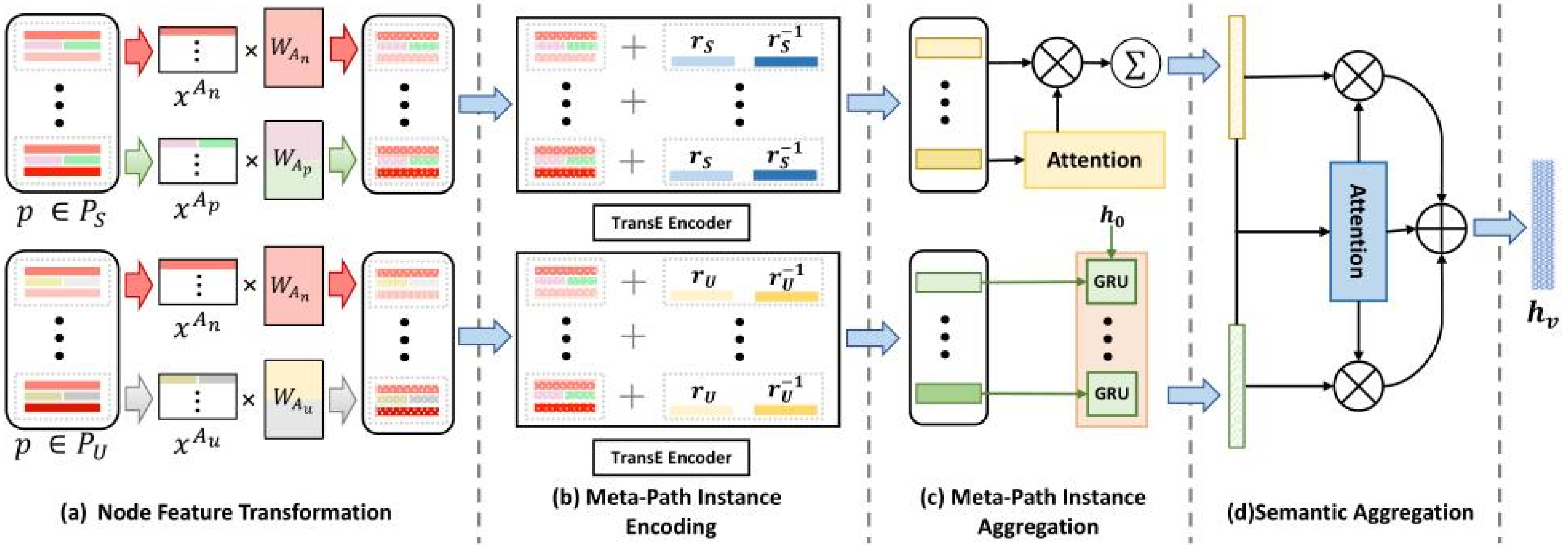


Figure 4: Architecture of *Hetero-SCAN*.

$$s_P = \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} tanh(\mathbf{M}_A \cdot \mathbf{h}_v^P + \mathbf{b}_A) \tag{8}$$

$$e_P = tanh(q^T \cdot s_P)$$

$$\beta_P = \frac{exp(e_P)}{\sum_{P' \in \mathcal{P}} exp(e_{P'})} \tag{9}$$

$$\mathbf{h}_v = \sum_{P \in \mathcal{P}} \beta_P \cdot \mathbf{h}_v^P$$

$$\mathcal{L} = -\sum y \log \mathbf{P}_{fake} + (1-y) \log \mathbf{P}_{real} \tag{10}$$

Experiments

Table 1: Comparison of *Hetero-SCAN* with exiting graph-based fake news detection methods.

	Multi-level Social Context	Information Preserving	Temporal Information	End-to -end
CSI [38]	✗	✓	✓	✓
SAFER [13]	✗	✗	✗	✓
FANG [31]	✓	✓	✓	✗
AA-HGNN [36]	✗	✓	✗	✓
Hetero-SCAN	✓	✓	✓	✓

Experiments

Table 2: Formulation of Encoding Method.

Method	Original	In Our Paper
TransE	$\mathbf{e}_s + \mathbf{e}_p$	$MEAN[(\mathbf{h}_u + r + r^{-1}), (\mathbf{h}_w + r^{-1})]$
ConvE	$[\mathbf{e}_s \parallel \mathbf{e}_p] * \mathbf{W}$	$[\tilde{\mathbf{h}}_u \parallel \tilde{r} \parallel \tilde{\mathbf{h}}_w \parallel \tilde{r}^{-1}] * \mathbf{W}$
RotatE	$\mathbf{e}_s \odot \mathbf{e}_p$	$MEAN[(\mathbf{h}_u \odot r \odot r^{-1}), (\mathbf{h}_w \odot r^{-1})]$

Table 3: Dataset Statistics.

	FANG	HealthStory
# News	1,054	1,638
# Fake News	448	460
# Real News	606	1,178
# Users	52,357	63,723 (sampled)
# of Users per News	71.9	227.26
# Publishers	442	31

Experiments

Table 4: Detection result of *Hetero-SCAN* on two real-word dataset: FANG and FakeHealth. Bold numbers denote the best value in average, and underscored numbers denote the smallest variation (± stands for 95% confidence interval). The classification method with highest AUC score, was pointed out by ⋆ and was selected for the subsequent evaluation.

Dataset	Classification Method	Precision	Recall	F1 Score	Accuracy	AUC Score
FANG	Classification Layer	**0.845**±0.052	**0.843**±0.054	**0.843**±0.053	**0.843**±0.054	0.839±0.048
	Naive Bayes	0.839±0.053	0.837±0.058	0.835±0.057	0.837±0.058	0.840±<u>0.041</u>
	Logistic Regression	0.835±0.054	0.835±0.054	0.835±0.054	0.835±0.054	0.907±0.058
	⋆ **SVM**	0.832±0.036	0.839±0.053	0.840±0.053	0.839±0.053	**0.910**±0.047
	Random Forest	0.832±<u>0.036</u>	0.831±<u>0.037</u>	0.831±<u>0.037</u>	0.831±<u>0.037</u>	0.900±0.057
	AdaBoost	0.811±0.070	0.807±0.076	0.808±0.075	0.807±0.076	0.881±0.056
HealthStory	Classification Layer	0.529±<u>0.093</u>	**0.717**±<u>0.003</u>	0.599±<u>0.008</u>	**0.717**±<u>0.003</u>	0.500±<u>0.003</u>
	Naive Bayes	0.662±0.139	0.600±0.244	0.573±0.289	0.633±0.131	0.508±0.177
	⋆ **Logistic Regression**	0.660±0.065	0.595±0.206	0.594±0.185	0.584±0.180	**0.557**±0.076
	SVM	0.649±0.094	0.620±0.137	**0.612**±0.089	0.623±0.137	0.536±0.108
	Random Forest	**0.674**±0.117	0.550±0.272	0.526±0.327	0.520±0.269	0.513±0.134
	AdaBoost	0.656±0.129	0.539±0.302	0.492±0.303	0.540±0.301	0.554±0.076

Experiments

Table 5: Comparison of AUC scores with existing methods. The AUC scores of CSI and FANG are from Nguyen, Van-Hoang, et al. [31]. FANG experiment on HealthStory dataset cannot be conducted since it needs additional labels.

Category	Method	FANG	HealthStory
Text-based	TF.IDF + SVM	0.735	0.526
	LIWC + SVM	0.511	0.534
	Doc2Vec + SVM	0.554	0.582
Graph-based	CSI	0.741	-
	SAFER	0.669	0.615
	FANG	0.750	-
	AA-HGNN	0.654	0.559
GNN-baselines	GCN	0.633	0.528
	GAT	0.630	0.541
	GraphSAGE	0.773	0.589
	R-GCN	0.753	0.500
	HAN	0.658	0.600
Hetero-SCAN	*w/ temporal*	**0.910**	0.557
	w/o temporal	0.823	**0.636**

Experiments

Table 6: Comparison of AUC score against other fake news detection methods by varying the size of the training data. (-t) and (t) refer to *Hetero-SCAN* without and with temporal information, respectively.

	10%	30%	50%	70%	90%
CSI	0.636	0.671	0.670	0.689	0.691
SAFER	0.546	0.689	0.666	0.692	0.669
FANG	0.669	0.704	0.717	0.723	0.752
AA-HGNN	0.573	0.598	0.656	0.657	0.642
Hetero-SCAN$_{(-t)}$	0.594	0.707	0.776	0.749	0.751
Hetero-SCAN$_{(t)}$	**0.764**	**0.835**	**0.878**	**0.889**	**0.900**

Experiments

Table 7: Performance of detection result when apply different Meta-Path encoding method. Bold texts indicate the highest value.

	F1 Score	Accuracy	AUC
TransE	**0.840**±0.053	**0.839**±0.053	**0.910**±0.047
RotatE	0.799±0.035	0.799±0.036	0.862±0.035
ConvE	0.532±0.174	0.526±0.079	0.665±0.021

Table 8: Performance of the *Hetero-SCAN* with and without temporal information.

Dataset	*Hetero-SCAN*	F1	Accuracy	AUC
FANG	*w/ temporal*	**0.840**	**0.839**	**0.910**
	w/o temporal	0.759	0.760	0.823
HealthStory	*w/ temporal*	0.594	0.584	0.557
	w/o temporal	**0.614**	**0.595**	**0.636**

Experiments

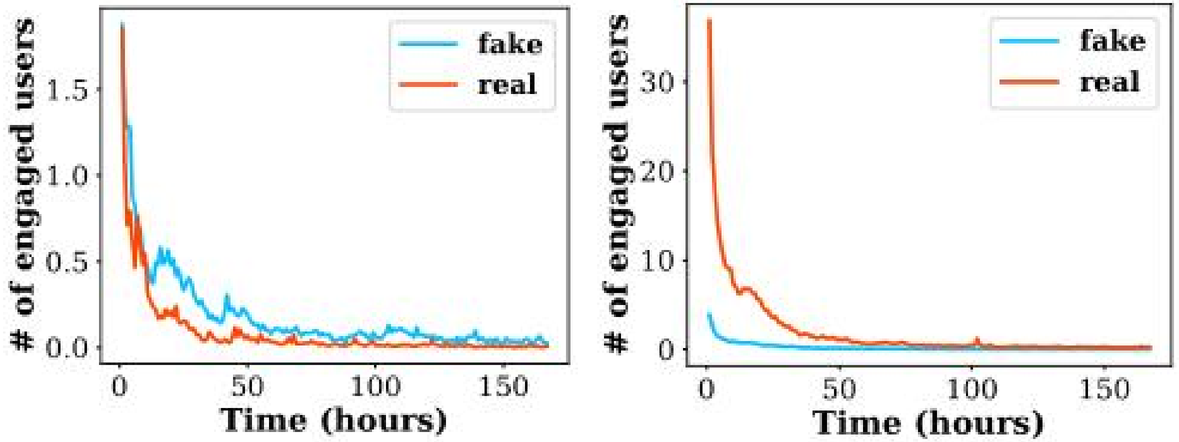
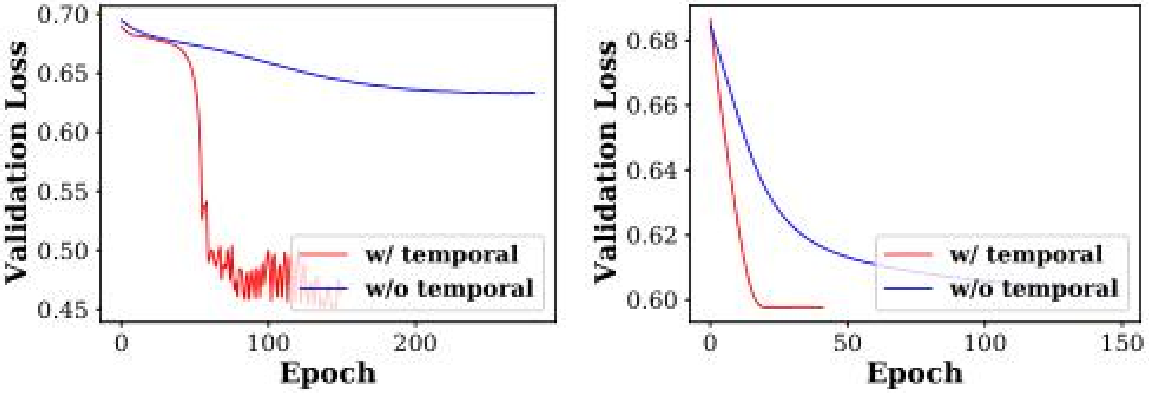


Figure 5: Comparison of temporal behaviors on two datasets. Both figures show the # of engagements (tweets) per news vs. time (hours) for FANG (left) and HealthStory (right).



(a) FANG (b) HealthStory

Figure 6: Validation loss during training. (Red and blue lines indicate the validation loss of *Hetero-SCAN* with and without temporal information, respectively.)

Thanks