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

Jian Cui  
cj19960819@kaist.ac.kr  
Korea Advanced Institute of Science and Technology  
Daejeon, South Korea

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

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

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

ABSTRACT
Fake news, false or misleading information presented as news, has a significant impact on many aspects of society, such as in politics or healthcare domains. Due to the deceiving nature of fake news, applying Natural Language Processing (NLP) techniques to the news content alone is insufficient. Therefore, more information is required to improve fake news detection, such as the multi-level social context (news publishers and engaged users in social media) information and the temporal information of user engagement. The proper usage of this information, however, introduces three chronic difficulties: 1) multi-level social context information is hard to be used without information loss, 2) temporal information is hard to be used along with multi-level social context information, 3) news representation with multi-level social context and temporal information is hard to be learned in an end-to-end manner. To overcome all three difficulties, we propose a novel fake news detection framework, Hetero-SCAN. We use Meta-Path to extract meaningful multi-level social context information without loss. Meta-Path, a composite relation connecting two node types, is proposed to capture the semantics in the heterogeneous graph. We then propose Meta-Path instance encoding and aggregation methods to capture the temporal information of user engagement and learn news representation end-to-end. According to our experiment, Hetero-SCAN yields significant performance improvement over state-of-the-art fake news detection methods.

CCS CONCEPTS
* Computing methodologies → Artificial intelligence; * Information systems → Social networks.

KEYWORDS
Fake News Detection; Graph Representation Learning

1 INTRODUCTION
The wide dissemination of fake news has become a major social problem in the world. The most recent and infamous distribution of fake news was in the 2020 United States presidential election fraud [9] and COVID-19 rumors [1]. Both industry and government are making efforts to prevent the spread of fake news [10]. Nevertheless, fake news verification still relies on human experts and their manual efforts in analyzing the news contents with additional evidence. Therefore, there should be an automatic and efficient way to identify the veracity of the news.

The most typical way to detect fake news is applying Natural Language Processing (NLP) techniques on the news content [15, 18]. Considering that even people struggle in identifying the news authenticity by the news content alone, these NLP solutions are ineffective. Thus, more information is required to improve fake news detection.

The first important information is the users in social media. Social media is one of the most influential mediums to propagate information, and it has become a common practice for people to share their thoughts in social media. Even though regular users use social media as a communication tool, some users, known as instigators, intentionally spread fake news. Instigators usually have a highly partisan-biased personal description and a lot of followers and followings, which is significantly different from the profiles of regular users (See in Figure 1). Therefore, analyzing the users engaged in the news can provide additional evidence for identifying news authenticity. The publisher information can also play an important role because certain partisan-biased publishers are more likely to publish fake news [3, 5, 6]. Information on users and publishers can be viewed as multi-level social context information, and they provide additional clues for fake news detection.

In addition to multi-level social context information, temporal information of user engagement (temporal information for short) is another instrumental information in fake news detection. Fake and real news show different propagation properties in social media: Fake news is periodically mentioned by people and usually lasts longer, but real news receives attention only at the beginning of the news publication [27]. In this context, the temporal information should be included in the news representation along with multi-level social context information.

Using multi-level social context and temporal information, however, leads to three chronic difficulties. Firstly, due to the heterogeneity of multi-level social context information, it is hard to use this information without loss. Secondly, temporal information is hard to be used along with multi-level social context information. The graph is a typical way to present social context and its connectivity to the news, but the graph itself has complications in...
Our major contributions are:

1. We conduct diverse experiments on the two real-world fake news datasets, covering the broad definition of fake news (Section 3), and demonstrate that Hetero-SCAN achieves significant improvement over previous approaches in terms of F1 score, accuracy, and AUC score. Our code with data is released on the GitHub for reproducibility. Our major contributions are:

2. We pose three chronic difficulties in social context aware fake news detection and address them by proposing a novel fake news detection framework, Hetero-SCAN.

3. We conduct diverse experiments on the two real-world fake news datasets, covering the broad definition of fake news (Section 3), and demonstrate that Hetero-SCAN shows better performance than existing solutions.

4. We provide new insights into the differences in the behavior of engaged users between intentional and unintentional fake news.

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

| Method     | Social Context | Information Preserving | Temporal Information | End-to-end |
|------------|----------------|------------------------|----------------------|------------|
| CSI [39]   | ✓              | ✓                      | ✓                    | ✓          |
| SAFER [13] | ✓              | ✓                      | ✓                    | ✓          |
| FANG [32]  | ✓              | ✓                      | ✓                    | ✓          |
| AA-HGNN [37] | ✓              | ✓                      | ✓                    | ✓          |
| Hetero-SCAN | ✓              | ✓                      | ✓                    | ✓          |

2 RELATED WORK

2.1 Fake News Detection

Fake news detection methods can be categorized into two types: content-based and graph-based approaches.

The content-based approach models the content of the news, such as headline or body text, to detect news authenticity. Some research on content-based approaches utilizes linguistic features such as stylometry, psycholinguistic properties, and rhetorical relations [12, 34, 35, 38]. Researchers also use multi-modal approaches, the combination of visual and linguistic features to verify the news authenticity [20, 24, 36, 46, 48].

The graph-based approach, also known as the social context aware approach, adds auxiliary information of the user or publisher to model the news. CSI [39] is a framework that aims to capture the information of users and their temporal engagements. CSL however, does not consider publishers, and the connection between users and news was also ignored. Bi-GCN [11] and SAFER [13] use Graph Convolution Network (GCN) [25] to obtain the news representation with user information. However, they suffer from a severe information loss since they present news and user information in a homogeneous graph. In other words, they fail to taking the node and relation types into account. Most recently, FANG [32] is proposed to preserve information by dividing the fake news detection task into several sub-tasks, such as textual encoding and stance detection. Nonetheless, dividing into sub-tasks causes the error propagation problem: If the sub-tasks have errors, the errors can propagate up to the final news representation and thereby deteriorate the detection performance. AA-HGNN [37] uses adversarial active learning and extends Graph Attention Network (GAT) [45] into the heterogeneous graph to learn the news representation with limited training data. Information of users and their temporal engagement information, however, are not considered in AA-HGNN.

Table 1 compares Hetero-SCAN and existing fake news detection methods.

2.2 Graph Neural Network

Graph Neural Network, the extension of the deep learning method into graphs, shows its effectiveness in graph-represented data. The first method proposed is Graph Convolutional Network (GCN) [25] which aggregates the features from the adjacent nodes in the graph. To further improve it, some methods adopt the attention mechanism and random walk with restart sampling strategy, namely Graph Attention Network (GAT) [45] and GraphSAGE [22].
As these methods are designed for homogeneous graphs, they are not general enough to apply to the heterogeneous graph, so new approaches tailor to heterogeneous graphs are then proposed. To model the multi-relations in the graph, the Relation aware GCN (R-GCN) [40] is proposed first. HetGNN [51] uses a sampling strategy based on random walk with restart and Bi-LSTM to aggregate the node features in the heterogeneous graph. Later, the methods based on Meta-Path and attention mechanism, such as HAN [23] and MAGNN [19], are proposed.

3 PRELIMINARIES

Definition 3.1 (Broad Definition of Fake News). Fake news is false news.

Definition 3.2 (Narrow Definition of Fake News). Fake news is intentionally false news published by a news outlet.

Contrary to the amount of research done, the term fake news has only just been defined by the recent work of Zhou, Xinyi and Reza Zafarani [52]. They define fake news in two scopes, broad and narrow. The broad definition emphasizes the authenticity of the information, and the narrow one emphasizes the intentions of the author. Most research on fake news detection has employed a broad definition of fake news. We experiment on the two dataset (with and without intention) following broad definition and analyze how intention affect the performance of the detection (in Section 5.3).

Definition 3.3 (Heterogeneous Graph). A heterogeneous graph is defined as a graph \( G = (V, E) \) associated with a node type mapping function \( \phi: V \rightarrow \mathcal{A} \) and an edge type mapping function \( \psi: E \rightarrow \mathcal{R} \). \( \mathcal{A} \) and \( \mathcal{R} \) denotes the predefined sets of node types and edge types, respectively, with \(|\mathcal{A}| + |\mathcal{R}| > 2\).

Definition 3.4 (Meta-Path). A Meta-Path \( P \) is defined as a path in the form of \( A_1 \overset{R_1}{\rightarrow} A_2 \overset{R_2}{\rightarrow} \cdots \overset{R_l}{\rightarrow} A_l \) (abbreviated as \( A_1A_2\ldots A_l \)), which describes a composite relation \( R = R_1 \circ R_2 \circ \ldots \circ R_l \) between node types \( A_l \) and \( A_{l+1} \), where \( \circ \) denotes the composition operator on relations.

Definition 3.5 (Meta-Path Instance). Given a Meta-Path \( P \) of a heterogeneous graph, a Meta-Path instance \( p \) of \( P \) is defined as a node sequence in the graph following the schema defined by \( P \).

4 METHODOLOGY

4.1 Graph Construction & Feature Engineering

To integrate multi-level social context information, we build a heterogeneous graph of news (Figure 2). The graph consists of three types of nodes (publisher, news, and users) and four types of edges (citation, publication, tweet, and following). Formally, the heterogeneous graph of news is noted as \( G(V, E) \), and the set of three node types are symbolized as \( \mathcal{A} = \{A_p, A_n, A_u\} \).

Before utilizing this heterogeneous graph, it is necessary to construct initial node features for three types of nodes in the graph. For news nodes, Doc2Vec [28] is applied to the news article to construct their initial features. The user and publisher nodes, however, need additional information to construct their respective initial features. Users’ profiles are used for user nodes since the importance of the user profiles for detecting news authenticity has been proved by

\[
\mathcal{P} \in \{P_U, P_S\} \tag{1}
\]

where \( P_U : \text{News} \rightarrow \text{User} \rightarrow \text{News} \) and \( P_S : \text{News} \rightarrow \text{Publisher} \rightarrow \text{News} \).

After defining a set of Meta-Path, we extract Meta-Path instances \( p \) following each Meta-Path, \( p_S \) or \( p_U \), for each target news node. To efficiently extract Meta-Path instances, we first divide the whole graph into two sub-graphs, which only contain the nodes types specified in the Meta-Path, \( p_S \) or \( p_U \). Then, in each sub-graph, the Meta-Path instances following each Meta-Path are extracted. The corresponding collection of features are fed into Hetero-SCAN to get the final representation of the target news node. The sets of instances following two Meta-Path \( p_S \) and \( p_U \) are denoted as \( P_S \) and \( P_U \) respectively. For instance, if we want to extract the Meta-Path instances of the target news node \( X^N \) in Figure 3, we first
which is have different dimensions since different sources and techniques
addressed chronic difficulties. The yet addressed chronic difficulties.
processes them through four steps as shown in Figure 4 to tackle
instances from the Meta-Path model. In the following sections, we assume that the Meta-Path
Path instances. Thus, the Meta-Path instances from the Meta-Path
randomly sampled for each news node according to a pre-defined
the real world. To cope with this situation, we extract Meta-Path
instances of target node \( x \). We use \( P \) and \( U \) to denote the set of Meta-Path instances follow each Meta-Path.
node \( u \), and \( w \) refer to the nodes along the Meta-Path. Considering the Meta-Path we defined in the Section 4.2,
\( v \in \mathcal{A}_n \) and \( w \in \{ \mathcal{A}_p, \mathcal{A}_u \} \). The \( r \) and \( r^{-1} \) is the relation
between \( u \) and \( w \) or \( v \) respectively, \( h \) is the transformed embedding of
the node as we stated in Section 4.3.1, and \( e \) is the embedding of
the knowledge graph triple.

Several research on knowledge graph domain tackle the triple
embedding problem \[17, 44, 49\]. We use TransE \[49\] as our main
encoding method for the proposed model. TransE \[49\] represents
relations as translations, so the object vector \( e_v \) in the triple is
considered as a translation of subject vector \( e_u \) on predicate vector \( e_p \). Other than TransE, RotatE \[44\] and ConvE \[17\] knowledge
graph embedding methods are also examined in our work. Ablation
study on different knowledge graph triple embedding methods and
their descriptions are provided in the Appendix.

In knowledge graph, there are usually explicit features for predi-
cated (\( e_p \) in Equation 3), but in our case, there is no explicit
features for the relations (\( r \) in Equation 3), so we use learnable em-
bedding vectors to present relations. Inverse relationships, such as
Publisher – News and News – Publisher, are represented by
taking the sign inverses. For instance, if we define \( r \) as the em-
bedding of Publisher – News relationship, the inverse relationship,
News – Publisher is \( r^{-1} = -r \). Our encoding function \( f_{enc} \) is defined
as:

\[
h_p = f_{enc}(p) = f_{enc}(h_u, r, h_w, r^{-1})
\]

The existing knowledge graph triple embedding methods ex-
plained above are designed for two nodes and the relation between
them. In our Meta-Path, we have a total of three nodes and two
relations in a Meta-Path instance. We deal with this by slightly
tuning the formulation to fulfill our needs. The original formu-
lation of knowledge graph triple embedding methods and ours are
summarized in Table 2. In this table, the \( h \) means the reshape of
dimension for each type of node in the graph. The transformed feature for a node \( v \in \mathcal{V}_A \) of type \( A \in \mathcal{A} \) is:

\[
h_v^A = W_A \cdot x_v^A
\]

where \( x_v \in \mathbb{R}^{d_A} \) is the initial feature of node \( v \), and \( W_A \in \mathbb{R}^{d \times d_A} \) is the learnable type-specific weight matrix for node type \( A \).

### 4.3 Model Architecture

**Hetero-SCAN** takes in vectors from the previous step as input and
processes them through four steps as shown in Figure 4 to tackle
the yet addressed chronic difficulties.

#### 4.3.1 Node Feature Transformation

The initial node features have different dimensions since different sources and techniques
are used in the feature engineering process (Section 4.1). To make
them lie in the same latent space, we apply the type-specific linear
transform on the features of each type of node. Type-specific trans-
formation refers to the linear projection of a vector into another

![Figure 3: Extracting Meta-Path instances of the target news node](image)
vector $h$ in a 2D form, and the $\odot$ and $\parallel$ represent the element-wise product and concatenation of vector, respectively.

### 4.3.3 Meta-Path Instance Aggregation

The encoded vectors from two different Meta-Paths are aggregated by using different methods.

The encoded vectors from Meta-Path $\mathcal{P}_S: \text{News} \rightarrow \text{Publisher} \rightarrow \text{News}$ contain information of other news from the same publisher. Among the news published by the publisher, not all news will contain valuable information for detection. Thus, the model should ‘focus’ on some of the news published by this publisher and include this information in the aggregated representation. For each Meta-Path instance $p \in P_S$:

$$
e_p = \text{LeakyReLU}(a^T \cdot h_p)$$

$$
\alpha_p = \text{softmax}(e_p) = \frac{\exp(e_p)}{\sum_{p' \in P_S} \exp(e_{p'})}
$$

where $e_p$ is the attention value calculated by multiplying encoded Meta-Path instance $h_p$ with attention vector $a \in \mathbb{R}^{2d}$, and it is normalized by a softmax function over all Meta-Path instances of the target node $v$, the result is denoted as $\alpha_p$ above.

To alleviate the effect of the high variance of the data in a heterogeneous graph, we adopt multi-head attention mechanism. $K$ independent attention mechanisms execute the transformation as shown in Equation 6, and their features are concatenated after they pass the activation function $\sigma$. The output feature representation can be formulated as:

$$h_v^{P_S} = \| \sigma( \sum_{p \in P_S} [\alpha_p]_k \cdot h_p )$$

where $[\alpha_p]_k$ is the normalized attention value of Meta-Path instance $p$ of target node $v$ at the $k$-th attention head.

### 4.3.4 Semantic Aggregation

Two vectors, $h_v^{P_S}$ and $h_v^{P_U}$, from previous step represent two different aspects of the news. The final news representation is produced by fusing these two vectors, which enables us to learn the news representation end-to-end (the third difficulty). As two Meta-Paths show two different aspects of a given news, the model should be able to weigh the importance of the two aspects with different news. To this end, we adopt another attention mechanism. Before applying the attention mechanism, non-linear transformations are applied to summarize $h_v^{P_S}$ and $h_v^{P_U}$.

Thus for $P \in \{P_S, P_U\}$:

$$s_P = \frac{1}{|V|} \sum_{v \in V} tanh(M_A \cdot h_v + b_A)$$

Here, $M_A \in \mathbb{R}^{d_m \times d'}$ and $b \in \mathbb{R}^{d_m}$ is a learnable weight matrix and bias vector. $V$ is the set of news nodes.

Then we apply the attention mechanism to aggregate two vectors to obtain our final news representation $h_v$.

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

$$\beta_p = \frac{\exp(e_p)}{\sum_{p' \in P} \exp(e_{p'})}$$

$$h_v = \sum_{p \in P} \beta_p \cdot h_v^p$$

Temporal information of user engagement is another critical feature to determine the veracity of the given news, and incorporating this information is the second difficulty to resolve. To capture the temporal information, we aggregate the Meta-Path instances follow $\mathcal{P}_T: \text{News} \rightarrow \text{User} \rightarrow \text{News}$ through Recurrent Neural Network (RNN). Since Meta-Path instances are already encoded in the previous step, we can directly feed them into the RNN. There are usually a large number of users engaged per news, so we choose GRU [14] as our RNN unit to avoid the vanishing or exploding gradients problem.

$$h_v^{P_U} = \text{GRU}(h_{p_1}, h_{p_2}, ..., h_{p_n}), p_i \in P_U$$

The last hidden state of the GRU is used for the downstream task as it is the high-level representation that summarizes the temporal information of the user engagement.

### Table 2: Formulation of Encoding Method.

| Method    | Original                          | In Our Paper                          |
|-----------|-----------------------------------|---------------------------------------|
| TransE   | $e_p + \epsilon_p$               | $\text{MEAN}[(h_v + r + r^{-1}), (h_v + r^{-1})]$ |
| ConvE    | $[e_p \parallel \epsilon_p] \ast W$ | $[h_v \parallel \epsilon_v \ast W]$ |
| RotatE   | $e_p \odot \epsilon_p$          | $\text{MEAN}[(h_v \odot r \ast r^{-1}, (h_v \odot r^{-1})]$ |
where \( q \in \mathbb{R}^{d_m} \) is the attention vector and \( \beta P \) is the normalized importance of Meta-Path \( P \).

4.4 Training

The final representation of the target news vector is passed to the classification layer to get the classification result. During training, our predictions and labels are used to calculate the loss, and we update the learnable parameters of the model by using the back-propagation algorithm. The loss function used in Hetero-SCAN is cross-entropy loss, which is:

\[
\mathcal{L} = -\sum y \log P_{fake} + (1 - y) \log P_{real}
\]

The overall all learning algorithm is summarized in Algorithm 1 (Appendix).

5 EXPERIMENTAL RESULT AND ANALYSIS

5.1 Dataset and Settings

To test the effectiveness of our method, we conducted our experiments with two real-world datasets: FANG [32] and FakeHealth [16]. The dataset FANG was composed in a study by Nguyen et al. [32] based on the datasets collected by related work on rumor and news classification [26, 29, 41]. The original news content was obtained through the provided news url, and for the 100 news urls that did not have the news content available, resorted to manually searching the news title for the content. From provided tweet ids, users and their profiles on Twitter could be found through the Twitter API [8]. The labels of the news in FANG are obtained from two well-known fact-checking websites: Snopes [7] and PolitiFact [4]. FakeHealth is another publicly available benchmark dataset for fake news detection, mainly focused on the healthcare domain. The dataset consists of two subsets, HealthStory and HealthRelease; HealthStory was used in our study due to the number of news articles in HealthRelease being too small. HealthStory is collected from the healthcare information review website HealthNewsReviews [2]. On this website, the professional reviewers gave scores of 1 to 5 for each news. Similar to the original study that published the FakeHealth dataset, an article is considered as fake if the score is less than three and real otherwise. The detailed statistics of the dataset used in our experiment are listed in Table 4.

In each dataset, we used 70% of news articles as our training set, and the remaining 30% of news articles are further divided into equal sizes of validation and test set. For the hyper-parameters, the transformed hidden dimension and the learning rate are set to 512 and 0.0001, respectively. The early-stopping training strategy with patience 20 is adopted to avoid overfitting. Since fake news detection is a binary classification problem, the real class was treated as positive and the fake class as negative.

5.2 Evaluation of ML Algorithms on News Embedding

We trained Hetero-SCAN by connecting the output representation to a fully connected layer to classify the news. After training, we evaluated our news representation with five classical machine learning baselines, such as Naive Bayes, Logistic Regression, etc. The metrics used for comparison are precision, recall, accuracy, F1 score, and AUC score, and the evaluation results are summarized in Table 3.

As shown in Table 3, the trained classification layer gives relatively better results than other machine learning algorithms in terms of F1 score and accuracy because the classification layer is optimized by classification objective (cross-entropy loss). In terms of AUC score, SVM gives a better result, but in terms of standard deviation, random forest generally gives more stable results. Based on this, random forest is chosen as the classification algorithm for upcoming evaluations. Regardless of downstream classification methods, Hetero-SCAN surpass any existing fake news detection methods (details in Section 5.4). In the dataset - HealthStory, Hetero-SCAN does not give an ideal result. The explanation for the result on the HealthStory dataset is discussed in the next section.

5.3 Misinformation vs Disinformation

Wardle et al. [50] published a report about information disorder on the Council of Europe in 2017. The report intends to examine information disorder and its related challenges. The authors argue that a large portion of the word 'fake news' consists of three concepts: misinformation, disinformation, and malinformation. They point out the importance of distinguishing the fake news in accordance with creators’ intention and provide the definition of three terms:

Definition 5.1 (Disinformation). Information that is false and deliberately created to harm a person, social group, organization or country.

Definition 5.2 (Misinformation). Information that is false, but not created with the intention of causing harm.

Definition 5.3 (Malinformation). Information that is based on reality, used to inflict harm on a person, organization or country.

According to the definition of malinformation, it is the information based on reality, while the fake news we talk about in this paper is false information. Therefore, we mainly consider disinformation and misinformation here, which are classified according to the news creator’s intention. Considering the definition of fake news given in Section 3, the narrow definition of fake news only covers disinformation, but the broad definition of fake news covers both disinformation and misinformation.

The dataset FANG is mainly composed of checked news from PolitiFact and Snopes, which are political-related fact-checking websites. Thus, the fake news in this dataset is either partisan-biased news or some false information to demean certain politicians, which are considered as information intended to harm the specific person or the organizations. Hence, the fake news in this dataset can be considered as disinformation. The news in HealthStory is collected and fact-checked from Health News Review where evaluates and rates the completeness, accuracy, and balance of news stories that include claims about medical treatments, health care journalism, etc. Most of this information is not spread deliberately to harm anyone, so the fake news in the HealthStory dataset can be regarded as misinformation.

Figure 5 compares the number of engaged users along with the time to see how people react to disinformation, misinformation, and real news. As shown in Figure 5, the disinformation (fake in the left) has many periodic spikes, which means the users periodically talk about disinformation. On the contrary, the misinformation (fake in the right) does not have any periodic spikes and converges...
Table 3: Detection result of 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).

| 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 | 0.834±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±0.041 |
|             | 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±0.036 | 0.831±0.037 | 0.831±0.037 | 0.831±0.037 | 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±0.093 | 0.717±0.003 | 0.599±0.008 | 0.717±0.003 | 0.500±0.003 |
|             | 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 |

Table 4: Dataset Statistics.

| Dataset     | FANG                  | HealthStory           |
|-------------|-----------------------|-----------------------|
| # Users     | 52,357 (sampled)      | 63,723 (sampled)      |
| # News      | 1,054                 | 1,638                 |
| # of Users per News | 71.9                 | 227.26                |
| # Fake News | 448                   | 469                   |
| # Real News | 606                   | 1,178                 |
| # Publishers| 442                   | 31                    |

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

to zero not long after the news is published, which is similar to the real news. As such, disinformation behaves significantly differently from real information, but misinformation behaves in a similar manner to real news.

To see the impact of temporal information in Hetero-SCAN, we replace the RNN in Hetero-SCAN with attention mechanism. In other words, we checked the detection performance difference between the Hetero-SCAN with and without temporal information. We set the hyperparameters the same for both approaches for a fair comparison, with Random Forest chosen as the classification algorithm. The evaluation result on the datasets can be found in Table 5.

Table 5: Performance of the Hetero-SCAN with and without temporal information.

| Dataset     | Hetero-SCAN w/ temporal | Hetero-SCAN w/o temporal | F1 | Accuracy | AUC |
|-------------|--------------------------|--------------------------|----|----------|----|
| FANG        | w/ temporal              | w/o temporal             | 0.831 | 0.831 | 0.900 |
| HealthStory | w/ temporal              | w/o temporal             | 0.526 | 0.520 | 0.513 |

The results show that the RNN based approach performs better than the other one in FANG dataset, but for the HealthStory dataset, the performance is better when the attention is applied. This means the existence of temporal information is not helpful in detecting misinformation. Furthermore, in FANG dataset, the validation loss of Hetero-SCAN with RNN converges much faster than the one with attention mechanism; by contrast, the convergence speed of the two approaches is similar in the HealthStory dataset. (See Figure 6 in Appendix)

To sum up, in a dataset has temporal behavior difference between real and fake class (i.e., disinformation dataset), Hetero-SCAN with RNN not only improves the performance of the fake news detection but also accelerates the learning speed.

5.4 Comparison with Existing Methods

To show that Hetero-SCAN is superior to other fake news detection, we compared Hetero-SCAN with other existing fake news detection methods. The bench-marked detection methods can be categorized into text-based approaches and graph-based approaches. For text-based approach, we use three different document embedding methods, TF-IDF, LIWC [33], and Doc2Vec [28], combined with SVM as baselines; and several representative graph-based fake news detection frameworks [13, 32, 37, 39] are also compared in this experiment.

Hetero-SCAN is also compared with some Graph Neural Network (GNN) methods to show that Hetero-SCAN is better than just simply
Table 6: Comparison with other methods. The AUC score of the CSI is from FANG, the F1 score and AUC score are not reported in this paper.

| Category       | Method   | F1   | Accuracy | AUC  |
|----------------|----------|------|----------|------|
| Text-based     | TF-IDF + SVM | 0.746 | 0.750    | 0.735|
|                | LIWC + SVM  | 0.512 | 0.550    | 0.511|
|                | Doc2Vec + SVM | 0.561 | 0.560    | 0.554|
| Graph-based    | CSI      | -    | -        | 0.741|
|                | SAFER    | 0.678 | 0.680    | 0.669|
|                | FANG     | 0.676 | 0.687    | 0.750|
|                | AA-HGNN  | 0.726 | 0.662    | 0.654|
|                | GCN      | 0.645 | 0.650    | 0.633|
|                | GAT      | 0.642 | 0.650    | 0.630|
| GNN-baselines  | GraphSAGE | 0.779 | 0.780    | 0.773|
|                | R-GCN    | 0.765 | 0.770    | 0.753|
|                | HAN      | 0.662 | 0.660    | 0.658|
| Hetero-SCAN    |          | 0.831 | 0.831    | 0.900|

5.5 Limited training data

Normally, the fake news dataset has limited training data due to the large-scale requirement of human labor, so the model should work well in the circumstance of limited training samples. To show that Hetero-SCAN outperforms existing methods given the circumstance of scarce training data, we gradually enlarge the training data, from 10% to 90%, and compare the fake news detection result with existing methods. Table 7 shows the comparison result.

Table 7: Comparison of AUC score against other fake news detection methods by varying the size of the training data.

| Method          | 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_{w/o time} | 0.594| 0.707| 0.776| 0.749| 0.751|
| Hetero-SCAN_{w/ time} | 0.764| 0.835| 0.878| 0.889| 0.900|

The AUC score of Hetero-SCAN achieves over 0.8 with only 30% of training data and even outperforms the rest of the methods with 90% of the training data. AA-HGNN is designed to overcome the scarcity of training data issues in the fake news detection task, but Hetero-SCAN is still better than AA-HGNN even when the size of training data is small.

6 DISCUSSION

6.1 Inductiveness of Hetero-SCAN

A deep learning based approach dealing with graph-structured data should have generality to produce practical predictions for unseen data. A method is an inductive approach if it can generate embeddings for the nodes that were not seen during training. In contrast, it is called a transductive approach if the method cannot generate embeddings for the nodes appearing in the testing phase for the first time. For example, GCN is inductive, whereas Node2Vec is transductive.

In graph-based fake news detection, unseen nodes can appear in the testing phase. It might be newly published news, new publishers, or new users. Some approaches using matrix decomposition [39, 42] are not able to generate embedding for newly published news with social context information. In Hetero-SCAN, however, the learnable parameters in our model are used after Meta-Path extraction with random sampling, and they are shared by all nodes. Therefore, our method is highly inductive, that is, Hetero-SCAN can generate news embeddings that are not seen during the training.

6.2 Limitation and Future Work

As expected, a single news article is engaged with by a large number of users. Using every single user’s information as a feature is therefore impractical, and we eventually used simple random sampling to select a certain number of users. Therefore, an improved method of screening important users is necessary for fake news detection to overcome the limitation. In addition, to apply the proposed method, we must first identify the relevant tweets for particular news. Since this paper focuses primarily on identifying the news in the context...
in which news and related tweets are given, finding relevant tweets for particular news is left as future work.

7 CONCLUSIONS

Fake news is a critical social problem threatening many aspects of the lives of the general public. We pose three difficulties in social context aware fake news detection and address them by proposing a novel fake news detection framework Hetero-SCAN. Our model overcomes the shortcomings of the previous graph-based approaches and exhibits state-of-the-art performance. We also provide insight about misinformation and disinformation by clarifying their different propagation properties. Hetero-SCAN can be of aid in future studies not only residing to fake news detection but also various events concerning disinformation.

ACKNOWLEDGMENTS

REFERENCES

[1] 2020. Coronavirus: The viral rumours that were completely wrong. https://www.bbc.com/news/blogs-trending-53640964.
[2] 2020. Health News Review. https://www.healthnewseview.org/.
[3] 2020. Left bias Publishers checked by MBFC. https://mediabiasfactcheck.com/left/.
[4] 2020. Politifact. https://www.politifact.com/.
[5] 2020. Questionable Publishers checked by MBFC. https://mediabiasfactcheck.com/fake-news/.
[6] 2020. Right bias Publishers checked by MBFC. https://mediabiasfactcheck.com/right/.
[7] 2020. Snopes. https://www.snopes.com/.
[8] 2020. Twitter API. https://developer.twitter.com/en/docs/twitter-api.
[9] 2020. US election 2020: Fact-checking Trump team’s main fraud claims. https://www.bbc.com/news/blogs-trending-53640964.
[10] 2021. Facebook Media: Working to Stop Misinformation and Fake News. https://www.facebook.com/formedia/blog/working-to-stop-misinformation-and-false-news/.
[11] Tian Bian, Xi Xiao, Tingyang Xu, Peilin Zhao, Wenbing Huang, Yu Rong, and Junzhou Hu. 2020. Rumor detection on social media with bi-directional graph convolutional networks. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34. 549–556.
[12] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In Proceedings of the 20th international conference on World wide web 675–684.
[13] Shantanu Chandra, Pushkar Mishra, Helen Yannakoudakis, and Ekaterina Shutova. 2020. Graph-based Modeling of Online Communities for Fake News Detection. arXiv preprint arXiv:2008.06274 (2020).
[14] Junyong Chung, Caglar Gucluer, KyungHyon Cho, and Yoshiha Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint arXiv:1412.5355 (2014).
[15] Niall J Conroy, Victoria L Rubin, and Yimin Chen. 2015. Automatic deception detection: methods for finding fake news. In Proceedings of the 7th ASSET&T Annual Meeting: Information science with Impact: Research in and for the Community. 1–4.
[16] Enyay Dai, Yuwei Sun, and Suhang Wang. 2020. Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. In Proceedings of the International AAAI Conference on Web and Social Media, Vol. 14. 853–862.
[17] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32.
[18] Song Feng, Retwik Banerjee, and Yijin Chiou. 2012. Syntactic stylometry for deception detection. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 171–175.
[19] Xinyu Fu, Jiawei Zhang, Ziaqiao Meng, and Irwin King. 2020. MAGNN: metapath aggregated graph neural network for heterogeneous graph embedding. In Proceedings of The Web Conference 2020. 2331–2341.
[20] Anastasia Gaianou, Guobao Zhang, and Paolo Rosso. 2020. Multimodal Fake News Detection with Textual, Visual and Semantic Information. In International Conference on Text, Speech, and Dialogue. Springer, 30–38.
[21] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. 855–864.
[22] William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In NIPS.
[23] Yi Han, Shanku Karanasoker, and Christopher Leckie. 2020. Graph neural networks with continual learning for fake news detection from social media. arXiv preprint arXiv:2007.03316 (2020).
[24] Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. Multi: Multimodal variational autoencoder for fake news detection. In The World Wide Web Conference 2915–2921.
[25] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016).
[26] Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. PHEME dataset for Rumour Detection and Veracity Classification. https://doi.org/10.6084/m9. figshare.6392078.v1
[27] Sejong Kwon, Meeyoung Cha, Ryomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent features of rumor propagation in online social media. In 2013 IEEE 13th international conference on data mining, IEEE, 1103–1108.
[28] Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In International conference on machine learning. PMLR, 1188–1196.
[29] Qin Ma, Wei Gao, Prasenjit Mitra, Sejong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. (2016).
[30] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. Journal of machine learning research, 9, Nov (2008), 2579–2605.
[31] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013).
[32] Van-Hoang Nguyen, Kazarnani Sugiyama, P尔斯lav 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. 1165–1174.
[33] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of LIWC2015. Technical Report.
[34] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefèvre, and Rada Mihalcea. 2018. Automatic Detection of Fake News. In Proceedings of the 27th International Conference on Computational Linguistics. 3391–3401.
[35] Martin Potthast, Johannes Kiesel, Kevin Reimann, Janek Bevendorff, and Benno Stein. 2018. A Stylistic Inquiry into Hyperpartisan and Fake News. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 231–240.
[36] Shengsheng Qian, Jinguang Wang, Jun Hu, Quan Fang, and Changsheng Xu. 2021. Hierarchical multi-modal contextual attention network for fake news detection. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval 153–162.
[37] Yuxiang Ren, Bo Wang, Jiawei Zhang, and Yi Chang. 2020. Adversarial active learning based heterogeneous graph neural network for fake news detection. In 2020 IEEE International Conference on Data Mining (ICDM). IEEE, 452–461.
[38] Victoria L Rubin and Tatiana Lukosanovia. 2015. Truth and deception at the rhetorical structure level. Journal of the Association for Information Science and Technology 66, 5 (2015), 905–917.
[39] Nataša Ruchansky, Sunggyoung Seo, and Yan Liu. 2017. Csi: A hybrid deep model for fake news detection. In Proceedings of the 2017 ACM on Conference on Management of Information Knowledge Management. 797–806.
[40] 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. Springer, 593–607.
[41] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2018. FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media. arXiv preprint arXiv:1809.01286 (2018).
[42] Kai Shu, Suhang Wang, and Huan Liu. 2019. 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. 312–320.
[43] Kai Shu, Xinyi Zhou, Suhang Wang, Reza Zafarani, and Huan Liu. 2019. 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. 436–439.
[44] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by rotational relaxation in complex space. arXiv preprint arXiv:1902.10197 (2019).
[45] Matej Velicković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. arXiv preprint arXiv:1710.10903 (2017).
[46] Nguyen Vo and Kyumin Lee. 2021. Hierarchical Multi-head Attentive Network for Evidence-aware Fake News Detection. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main, 965–975.
[47] Xiaowei Wang, Huiyue Ji, Chuan Shi, Bai Wang, Yinfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous graph attention network. In The World Wide Web Conference.
A NOTATIONS

Table 8: Notations used in the paper.

| Notation | Meaning |
|----------|---------|
| $G = (V,E)$ | Heterogeneous graph of news |
| $V$ | A set of nodes in the graph $G$ |
| $E$ | A set of edges in the graph $G$ |
| $R$ | A set of relations between two nodes (type of edge) |
| $A$ | A set of types of nodes $A = \{A_p, A_n, A_u\}$ |
| $A_p$ | Node type: publisher |
| $A_n$ | Node type: news |
| $A_u$ | Node type: user |
| $\mathcal{P}$ | A set of Meta-Path |
| $P_U$ | Meta-Path: $News \rightarrow User \rightarrow News$ |
| $P_S$ | Meta-Path: $News \rightarrow Publisher \rightarrow News$ |
| $p$ | A Meta-Path instance |
| $W_A$ | Type-specific transformation matrix |
| $x^A_v$ | Initial feature of the node $v$ of type $A$ |
| $h^A_v$ | Transformed feature of the node $v$ of type $A$ |
| $f_{enc}$ | Meta-Path instance encoding function |
| $\parallel$ | Concatenation operator |
| $\odot$ | Element-wise product |
| $\tilde{v}$ | Reshape the vector $v$ in a 2D form |

B DESCRIPTION OF EXISTING METHODS

The bench-marked fake news detection methods can be categorized into text-based approaches and graph-based approaches. Hetero-SCAN is also compared with representative graph embedding methods made for the homogeneous and heterogeneous graph. The detail of the methods we compared is listed below.

Text-based Methods:

- **TF-IDF + SVM**: TF-IDF is short for term frequency-inverse document frequency. It is intended to represent the importance of a word in a document. Feature vectors were extracted based on news article contents with TF-IDF, and SVM is applied to it.
- **LIWC [33] + SVM**: LIWC stands for Linguistic Inquiry and Word Count. It is widely used to extract words falling into psychologically meaningful categories, and these words can be used to compose a feature vector.
- **Doc2Vec [28] + SVM**: Doc2Vec is a paragraph embedding technique based on Word2Vec [31]. It uses skip-gram and CBOW model to learn the representation vector. Doc2Vec is considered as an unsupervised learning approach to learn the latent representation of a document.

Graph-based Methods:

- **SAFER [13]**: SAFER uses GCN and pre-trained RoBERTa model to embed news nodes in the heterogeneous graph. They concatenate two vectors and apply Logistic Regression to classify the news embeddings.
- **CSI [39]**: CSI is a hybrid deep learning based framework that aims to model the response, text, and user engagement of the news. The representation of response and text is concatenated with the user vector and score.
Algorithm 1 Learning Algorithm

**Input:** Heterogeneous Graph of News $G = (V, E)$, node feature $(x_v, y_v \in V)$, node types $\mathcal{A} = \{A_p, A_n, A_u\}$, label $y$

**Output:** Learn-able parameters $\theta$

for each epoch do
  for each Meta-Path schema $P \in \mathcal{P}$ do
    # Node feature transformation
    $h^0_v = W_A \cdot x^0_v$
    end for
    # Calculate the weight coefficient $a_{P_k}$ for each Meta-Path
    instance.
    $h^0_P = \sum_{i=1}^{K} \sigma(\sum_{k \in \mathcal{A}} a_{P_k} [h^0_k] \cdot h^0_v)$
    end if
    if $P = \mathcal{P}_U$ then
      All $h^0_{P_k}$ are sorted chronologically
      $h^0_{P_k} = \text{GRU}(h_p_1, h_p_2, ..., h_p_n)$
      $P \in \mathcal{P}_U$
    end if
    $s_P = \frac{1}{|V|} \sum_{v \in V} \text{tanh}(M_A \cdot h^0_v + b_A)$
    end for
  Calculate the weight coefficient $b_P$ for each Meta-Path.
  $h_v = \sum_{P \in \mathcal{P}} b_P \cdot h^0_{P_k}$
  # a fully connected layer for new classification.
  $z_v = W_c \cdot h_v$
  $[P_{\text{real}}, P_{\text{fake}}] = \text{softmax}(z_v)$
  $L = - \sum_{i} (y \log(P_{\text{fake}}) + (1 - y) \log(P_{\text{real}}))$
  $\theta \leftarrow \text{Backpropagate}(L)$
end for

- **FANG** [32]: FANG divides the detection task into several sub-tasks, such as textual encoding and stance detection. The final detection object is optimized by defining loss functions for those sub-tasks.
- **AA-HGNN** [37]: AA-HGNN uses active learning to tackle the limited training data problem and extends GAT [45] to learn the news representation in the graph.

**GNN baselines:**
- **GCN** [25]: GCN is a deep learning based method on a graph-structured data. Each node is learned by aggregating the feature information from its neighbors and the feature of itself.
- **GAT** [45]: GAT is similar to GCN, but it introduces the attention mechanism to replace the statically normalized convolution operation in GCN.
- **GraphSAGE** [22]: GraphSAGE is a general inductive framework that learns a node representation by sampling its neighbors and aggregating features of sampled nodes.
- **R-GCN** [40]: R-GCN is an application of the GCN framework for modeling relational data. In R-GCN, edges can represent different relations.
- **HAN** [47]: HAN is an extension of GAT on the heterogeneous graph. Meta-Path extraction strategy and attention mechanism are adopted to learn the representation of a node.

C VALIDATION LOSS DURING TRAINING

In Section 5.3, to see the impact of temporal information in Hetero-SCAN, we replace the RNN with attention mechanism. In other words, we compare the Hetero-SCAN trained with and without temporal information. These two Hetero-SCAN are trained with two dataset, and corresponding validation loss during the training is shown in Figure 6. The Hetero-SCAN trained with temporal information has faster convergence speed than Hetero-SCAN trained without temporal information in FANG dataset; In the HealthStory dataset, however, two models have no significant difference. Considering that fake news in FANG dataset is disinformation, and fake news in HealthStory is misinformation, temporal information can accelerates the convergence speed of training when identifying disinformation.

D ABLATION STUDY ON META-PATH INSTANCE ENCODING METHODS

In Section 4.3.2, we propose to use knowledge triple embedding methods to encode Meta-Path instances, and we adopt TransE in Hetero-SCAN. We wanted to examine the performance differences by changing the Meta-Path encoding method to other knowledge triple embedding methods, RotatE and ConvE. Descriptions of the three encoding methods are introduced below.

- **TransE** [49]: The TransE model represents relations as translations and aims to model the inversion and composition patterns. It defines each relation as a translation from the subject entity to the object entity.
- **RotatE** [44]: The RotatE model maps the entities and relations to the complex vector space and defines each relation as a rotation from the subject entity to the object entity.
- **ConvE** [17]: The ConvE model uses 2D convolution over embedding and multiple layers of nonlinear features to model knowledge graphs. They reshape the embedding of subject and predicates in a 2D form and apply convolution calculations on it.

To show the performance differences when different knowledge triple embedding methods are applied, F1 score, Accuracy, and AUC score were measured on two datasets: FANG and HealthStory. Table 9 indicates that TransE gives better results than the others. This reason can be drawn from the fact that TransE requires fewer parameters and operations than RotatE and ConvE. With limited training data, complex models are easy to suffer from over-fitting, which will cause performance degradation.
Table 9: Performance of detection result when apply different Meta-Path encoding method. Bold texts indicate the best encoding method in Hetero-SCAN.

|          | F1 Score | Accuracy | AUC        |
|----------|----------|----------|------------|
| TransE   | 0.831±0.037 | 0.831±0.037 | 0.900±0.057 |
| 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 |

E T-SNE VISUALIZATION

To show that the news representation produced by Hetero-SCAN is better than the existing methods, t-SNE was adopted to visualize news representation in a two-dimensional plane (Figure 7). The t-SNE technique is a well-known method to visualize the high-dimensional data in a two-dimensional plane [30]. As can be seen in Figure 7, the representations of Hetero-SCAN are clustered tighter than the other methods, implying a significant improvement over existing methods.

Figure 7: t-SNE visualization of news representations.