

# FANG: Leveraging Social Context for Fake News Detection Using Graph Representation

Van-Hoang Nguyen

vhnguyen@u.nus.edu

National University of Singapore  
Singapore

Preslav Nakov

pnakov@hbku.edu.qa

Qatar Computing Research Institute, HBKU  
Doha, Qatar

Kazunari Sugiyama

kaz.sugiyama@i.kyoto-u.ac.jp

Kyoto University  
Kyoto, Japan

Min-Yen Kan

kanmy@comp.nus.edu.sg

National University of Singapore  
Singapore

## ABSTRACT

We propose Factual News Graph (FANG), a novel graphical social context representation and learning framework for fake news detection. Unlike previous contextual models that have targeted performance, our focus is on representation learning. Compared to transductive models, FANG is scalable in training as it does not have to maintain all nodes, and it is efficient at inference time, without the need to re-process the entire graph. Our experimental results show that FANG is better at capturing the social context into a high fidelity representation, compared to recent graphical and non-graphical models. In particular, FANG yields significant improvements for the task of fake news detection, and it is robust in the case of limited training data. We further demonstrate that the representations learned by FANG generalize to related tasks, such as predicting the factuality of reporting of a news medium.

## CCS CONCEPTS

• **Information systems** → **Social networks**; • **Computer systems organization** → **Neural networks**; • **Theory of computation** → *Semi-supervised learning*; • **Computing methodologies** → *Natural language processing*.

## KEYWORDS

Disinformation, Fake News, Social Networks, Graph Neural Networks, Representation Learning

### ACM Reference Format:

Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. 2020. FANG: Leveraging Social Context for Fake News Detection Using Graph Representation. In *The 29th ACM International Conference on Information and Knowledge Management (CIKM '20)*, October 19–23, 2020, Virtual Event, Ireland. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3340531.3412046>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

CIKM '20, October 19–23, 2020, Virtual Event, Ireland

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-6859-9/20/10...\$15.00

<https://doi.org/10.1145/3340531.3412046>

## 1 INTRODUCTION

Social media have emerged as an important source of information for many worldwide. Unfortunately, not all information they publish is true. During critical events such as a political election or a pandemic outbreak, disinformation with malicious intent [38], commonly known as “fake news”, can disturb social behavior, public fairness, and rationality. As part of the fight against COVID-19, the World Health Organization also addressed the *infodemic* caused by fatal disinformation related to infections and cures [41].

Many sites and social media have devoted efforts to identify disinformation. For example, Facebook encourages users to report non-credible posts and employs professional fact-checkers to expose questionable news. Manual fact-checking is also used by fact-checking websites such as Snopes, FactCheck, PolitiFact, and Full Fact. In order to scale with the increasing amount of information, automated news verification systems consider external knowledge databases as evidence [13, 34, 42]. Evidence-based approaches achieve high accuracy and offer potential explainability, but they also take considerable human effort. Moreover, fact-checking approaches for textual claims based on textual evidence are not easily applicable to claims about images or videos.

Some recent work has taken another turn and has explored contextual features of the news dissemination process. They observed distinctive engagement patterns when social users face fake versus factual news [17, 25]. For example, the fake news shown in Table 1 had many engagements shortly after its publication. These are mainly verbatim re-circulations with negative sentiment of the original post explained by the typically appalling content of fake news. After that short time window, we see denial posts questioning the validity of the news, and the stance distribution stabilizes afterwards with virtually no support. In contrast, the real news example in Table 1 invokes moderate engagement, mainly comprised of supportive posts with neutral sentiment that stabilize quickly. Such temporal shifts in user perception serve as important signals for distinguishing fake from real news.

Previous work proposed partial representations of social context with (i) news, sources and users as major entities, and (ii) stances, friendship, and publication as major interactions [16, 32, 33, 39]. However, they did not put much emphasis on the quality of representation, modeling of entities and their interactions, and minimally supervised settings at all.

**Table 1: Engagement of social media users with respect to fake and real news articles. Column 2 shows the time since publication, and columns 4–7 show the distribution of stances (S: Support, D: Deny, C: Comment, and R: Report).**

News title (Label)	Time	# Posts	S	D	C	R	Noticeable responses
Virginia Republican Wants Schools To Check Children’s Genitals Before Using Bathroom ( <b>Fake</b> )	3h	38	0.00	0.03	0.19	0.78	“DISGUSED SO TRASNPHOBIC”, “FOR GODS SAKE GET REAL GOP”, “You cant make this up folks”
	3h - 6h	21	0.00	0.10	0.10	0.80	“Ok This cant be real”, “WTF IS THIS BS”, “Rediculous RT”
	6h+	31	0.00	0.10	0.14	0.76	“Cant make this shit up”, “how is this real”, “small government”, “GOP Cray Cray Occupy Democrats”
1,100,000 people have been killed by guns in the U.S.A. since John Lennon was shot and killed on December 8, 1980 ( <b>Real</b> )	3h	9	0.56	0.00	0.00	0.44	“#StopGunViolence”, “guns r the problem”
	3h+	36	0.50	0.00	0.11	0.39	“Some 1.15 million people have been killed by firearms in the United States since Lennon was gunned down”, “#StopGunViolence”

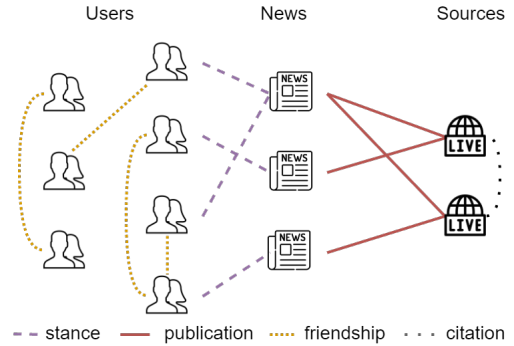
Naturally, the social context of news dissemination can be represented as a heterogeneous network where nodes and edges represent the social entities and the interactions between them, respectively. Network representations have several advantages over some existing Euclidean-based methods [23, 35] in terms of structural modeling capability for several phenomena such as echo chambers of users or polarized networks of news media. Graphical models also allow entities to exchange information, via (i) homogeneous edges, *i.e.*, user–user relationship, source–source citations, (ii) heterogeneous edges, *i.e.*, user–news stance expression, source–news publication, as well as (iii) high-order proximity (*i.e.*, between users who consistently support or deny certain sources, as illustrated in Figure 1). This allows the representation of heterogeneous entities to be dependent, leveraging not only fake news detection but also related social analysis tasks such as malicious user detection [7] and source factuality prediction [3].

Our work focuses on improving contextual fake news detection by enhancing representations of social entities. Our main contributions can be summarized as follows:

- (1) We propose a novel graph representation that models all major social actors and their interactions (Figure 1).
- (2) We propose the Factual News Graph (FANG), an inductive graph learning framework that effectively captures social structure and engagement patterns, thus improving representation quality.
- (3) We report significant improvement in fake news detection when using FANG and further show that our model is robust in the case of limited training data.
- (4) We show that the representations learned by FANG generalize to related tasks such as predicting the factuality of reporting of a news medium.
- (5) We demonstrate FANG’s explainability thanks to the attention mechanism of its recurrent aggregator.

## 2 RELATED WORK

In this section, we first review the existing work on contextual fake news detection and the way the social context of news is represented in such work. We then detail recent advances in the Graph Neural Network (GNN) formalism, forming the premise of our proposed graph learning framework.



**Figure 1: Graph representation of social context.**

### 2.1 Contextual Fake News Detection

Previous work on contextual fake news detection can be categorized based on the approach used to represent and learn the social context.

*Euclidean approaches* represent the social context as a flat vector or a matrix of real numbers. They typically learn a Euclidean transformation of the social entity features that best approximates the fake news prediction [32]. The complexity of such transformation varies from the traditional shallow (as opposed to “deep”) models, *i.e.*, Random Forest or Support Vector Machines (SVM) [6, 44] to probabilistic graphical models [33] and deep neural networks such as Long Short-Term Memory (LSTM) [15] that model engagement temporality [35]. However, given our formulation of social context as a heterogeneous network, Euclidean representations are less expressive [5]. Although pioneering work used user attributes such as demographics, news preferences, and social features, *e.g.*, the number of follower and friends [26, 38], they do not capture the user interaction landscape, *i.e.*, what kind of social figures they follow, which news topics they favor or oppose, etc. Moreover, in graphical representation, node variables are no longer constrained by the independent and identically distributed assumption, and thus they can reinforce each other’s representation via edge interactions.

Having acknowledged the above limitations, some researchers have started exploring non-Euclidean or *geometric approaches*. They generalized the idea of using social context by modeling an underlying user or the news source network and by developing representations that capture structural features about the entity.

**Table 2: Comparison between representation learning frameworks for social entities (1. users, 2. news, 3. sources) and interactions (4. user-user friendship, 5. user-news engagement, 6. source-news publication, 7. source-source citation) on whether they consider engagement time, graph modelling of social context, deep learning, inductiveness, and representation learning.**

Approach	Social Entities & Interactions	Temporal	Graphical	Deep	Inductive	Representative
Feature engineering [6, 26, 32, 44]	1, 2				✓	
Popat [33]	2, 3, 6	✓				
CSI [35]	1, 2, 4, 5	✓		✓	✓	
TriFN [39]	1, 2, 3, 4, 5, 6		✓			✓
MVDAM [21]	2, 3, 6, 7		✓	✓		
Monti [29]	1, 2, 4, 5	✓	✓			
GLAN [45]	1, 2, 5		✓	✓		
<b>FANG</b> (Our proposed approach)	1, 2, 3, 4, 5, 6, 7	✓	✓	✓	✓	✓

The *Capture, Score, and Integrate* (CSI) model [35] used linear dimensionality reduction on the user co-sharing adjacency matrix, combining it with news engagement features obtained from a recurrent neural network (RNN).

The *Tri-Relationship Fake News* (TriFN) detection framework [39] – although similar to our approach – neither differentiated user engagements in terms of stance and temporal patterns, nor modeled source–source citations. Also, matrix decomposition approaches, including CSI [35], can be expensive in terms of graph node counts and ineffective for modeling high-order proximity.

Other work on citation source network [21], propagation network [29], and rumor detection [10, 45] used recent advances in GNNs and multi-head attention to learn both local and global structural representations. These models optimized solely for the objective of fake news detection, without accounting for representation quality. As a result, they are not robust when presented with limited training data and cannot be generalized to other downstream tasks, as we show in Section 5. Table 2 compares these approaches.

## 2.2 Graph Neural Networks (GNNs)

GNNs have successfully generalized deep learning methods to model complex relationships and inter-dependencies on graphs and manifolds. Graph Convolutional Networks (GCNs) are among the first methods that effectively approximate convolutional filters [19]. However, GCNs impose a substantial memory footprint in storing the entire adjacency matrix. They are also not easily adaptable to our heterogeneous graph, where nodes and edges with different labels exhibit different information propagation patterns. Furthermore, GCNs do not guarantee generalizable representations, and are transductive, requiring the inferred nodes to be present at training time. This is especially challenging for contextual fake news detection or general social network analysis, as their structure is constantly evolving.

With these points in mind, we build our work on GraphSage that generates embeddings by sampling and aggregating features from a node’s local neighborhood [12]. GraphSage provides significant flexibility in defining the information propagation pattern with parameterized random walks and recurrent aggregators. It is well-suited for representation learning with unsupervised node proximity loss, and generalizes well in minimal supervision settings. Moreover, it uses a dynamic inductive algorithm that allows the creation of unseen nodes and edges at inference time.

## 3 METHODOLOGY

We first introduce the notation, and then formally define the problem of fake news detection. Afterwards, we discuss our methodology, namely the process of construction of our social context graph – FANG – as well as its underlying rationale. Finally, we describe the process of feature extraction from social entities as well as the modeling of their interactions.

### 3.1 Fake News Detection Using Social Context

Let us first define the social context graph  $G$  with its entities and interactions shown in Figure 1:

- (1)  $A = \{a_1, a_2, \dots\}$  is the list of questionable **news articles**, where each  $a_i$  ( $i = 1, 2, \dots$ ) is modeled as a feature vector  $\mathbf{x}_a$ .
- (2)  $S = \{s_1, s_2, \dots\}$  is the list of **news sources**, where each source  $s_j$  ( $j = 1, 2, \dots$ ) has published at least one article in  $A$ , and is modeled as a feature vector  $\mathbf{x}_s$ .
- (3)  $U = \{u_1, u_2, \dots\}$  is the list of **social users**, where each user  $u_k$  ( $k = 1, 2, \dots$ ) has engaged in spreading an article in  $A$ , or is connected with another user;  $u_k$  is a feature vector  $\mathbf{x}_u$ .
- (4)  $E = \{e_1, e_2, \dots\}$  is the list of **interactions**, and each interaction  $e = \{v_1, v_2, t, \mathbf{x}_e\}$  is modeled as a relation between two entities  $v_1, v_2 \in A \cup S \cup U$  at time  $t$ ;  $t$  is absent in time-insensitive interactions. The interaction type of  $e$  is defined as the label  $\mathbf{x}_e$ .

Table 3 summarizes the characteristics of different types of interactions, both homogeneous and heterogeneous. Stances are special types of interaction, as they are not only characterized by edge labels and source/destination nodes, but also by temporality as shown in earlier examples in Table 1. Recent work has highlighted the importance of incorporating temporality not only for fake news detection [26, 35], but also for modeling online information dissemination [14]. We use the following stance labels: *neutral support*, *negative support*, *deny*, *report*. The major *support* and *deny* stances are consistent with the prior work (e.g., [28]), whereas the two types of *support* – *neutral support* and *negative support* – are based on reported correlation between news factuality and invoked sentiment [1]. We assign the *report* stance label to a user–news engagement when the user simply spreads the news article without expressing any opinion. Overall, we use stances to characterize news articles based on opinions about them as well as social users by their view on various news articles.

**Table 3: Interactions in FANG’s social context network.**

Interaction	Linking Entities	Link Type	Description	Temporal
Followership	User–user	Unweighted, undirected	Whether a user follows another user on social media	No
Citation	Source–source	Unweighted, undirected	Whether sources refers to another source via a hyperlink	No
Publication	Source–news	Unweighted, undirected	Whether the source published the target news	Yes
Stance	User–news	Multi-label, undirected	The stance of the user with respect to the news	Yes

**Table 4: Some examples from the stance-annotated dataset, all concerning the same event.**

Text	Type	Annotated stance
Greta Thunberg tops annual list of highest-paid Activists!	reference headline	-
Greta Thunberg is the ‘Highest Paid Activist’	related headline	support (neutral)
No, Greta Thunberg not highest paid activist	related headline	deny
Can’t speak for the rest of ‘em, but as far as I know, Greta’s just a schoolgirl and has no source of income.	related post	deny
The cover describes Greta Thunberg to be the highest paid activist in the world	related tweet	support (neutral)
A very wealthy 16yo Fascist at that!	related post	support (negative)

**Table 5: Statistics about our stance-annotated dataset.**

	# Samples	# Supports	# Denies
Train	2,089	931	1,158
Test	438	207	231

We can now formally define our task as follows:

**Definition 3.1.** *Context-based fake news detection:* Given a social context  $G = (A, S, U, E)$  constructed from news articles  $A$ , news sources  $S$ , social users  $U$ , and social engagements  $E$ , context-based fake news detection is defined as the binary classification task to predict whether a news article  $a \in A$  is fake or real, in other words,  $F_C : a \rightarrow \{0, 1\}$  such that,

$$F_C(a) = \begin{cases} 0 & \text{if } a \text{ is a fake article} \\ 1 & \text{otherwise.} \end{cases}$$

### 3.2 Graph Construction from Social Context

**News Articles.** Textual [6, 34, 39, 44] and visual [18, 43] features have been widely used to model news article contents, either by feature extraction, unsupervised semantics encoding, or learned representation. We use unsupervised textual representations as they are relatively efficient to construct and optimize. For each article  $a \in A$ , we construct a TF.IDF [36] vector from the text body of the article. We enrich the representation of news by weighting the pre-trained embeddings from GloVe [30] of each word with its TF.IDF value, forming a semantic vector. Finally, we concatenate the TF.IDF and semantic vector to form the news article feature vector  $\mathbf{x}_a$ .

**News Sources.** We focus on characterizing news media sources using the textual content of their websites [3, 21]. Similarly to article representations, for each source  $s$ , we construct the source feature vector  $\mathbf{x}_s$  as the concatenation of its TF.IDF vector and its semantic vector derived from the words in the *Homepage* and the *About Us* section, as some fake news websites openly declare their content to be satirical or sarcastic.

**Social Users.** Online users have been studied extensively as the main propagator of fake news and rumors in social media. As in Section 2, previous work [6, 44] used attributes such as demographics, information preferences, social activity, and network structure such as the number of followers or friends. Shu *et al.* [39] conducted feature analysis of user profiles and pointed out the importance of signals derived from profile description and timeline content. A text description such as “*American mom fed up with anti american leftists and corruption. I believe in US constitution, free enterprise, strong military and Donald Trump #maga*” strongly indicates the user’s political bias and suggests the tendency to promote certain narratives. We calculate the user vector  $\mathbf{x}_u$  as the concatenation of a pair consisting of a TF.IDF vector and a semantic vector derived from the user profile’s text description.

**Social Interactions.** For each pair of social actors  $(v_i, v_j) \in A \cap S \cap U$ , we add an edge  $e = \{v_i, v_j, t, x_e\}$  to the list of social interactions  $E$  if they are linked via interaction type  $x_e$ . Specifically, for following, we examine whether user  $u_i$  follows user  $u_j$ ; for publication, we check whether news article  $a_i$  was published by source  $s_j$ ; for citation, we examine whether the *Homepage* of source  $s_i$  contains a hyperlink to source  $s_j$ . In the case of time-sensitive interactions, *i.e.*, *publication* and *stance*, we record their relative timestamp with respect to the article’s earliest time of publication.

**Stance Detection.** The task of obtaining a viewpoint for a piece of text with respect to another one is known as *stance detection*. In the context of fake news detection, we are interested in the stance of a user reply with respect to the title of a questionable news article. We consider four stances: support with neutral sentiment or *neutral support*, support with negative sentiment or *negative support*, *deny*, and *report*. We classify a post as verbatim reporting of the news article if it matches the article title after cleaning the text from emojis, punctuation, stop words, and URLs. We train a stance detector to classify the remaining posts as *support* or *deny*. Popular stance detection datasets either do not explicitly describe the target text [8], have a limited number of targets [27, 40], or define the source/target texts differently, as in the *Fake News Challenge*.<sup>1</sup>

<sup>1</sup><http://www.fakenewschallenge.org/>

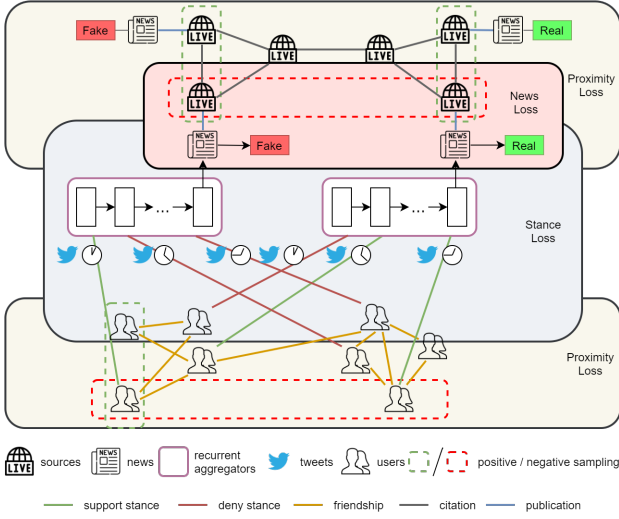


Figure 2: Overview of our FANG framework.

In order to overcome this difficulty, we constructed our own dataset for stance detection between social media posts and news articles, which contains 2,527 labeled source–target sentence pairs from 31 news events. For each event with a reference headline, the annotators were given a list of related headlines and posts. They labeled whether each related headline or post supports or denies the claim made by the reference headline. Aside from the *reference headline–related headline* or the *headline–related post* sentence pairs, we further made second-order inferences for *related headline–related post* sentence pairs. If such a pair expressed a similar stance with respect to the reference headline, we inferred a *support* stance for the *related headline–related post*, and *deny*, otherwise. Tables 4 and 5 show example annotations and statistics about the dataset. The inter-annotator agreement evaluated with Cohen’s Kappa is 0.78, indicating a substantial agreement. In order to choose the best stance classifier, we fine-tuned the model on our dataset using various pre-trained large-scale Transformers [9, 22]. RoBERTa [22] turned out to work best, achieving *Accuracy* of 0.8857, *F<sub>1</sub> score* of 0.8379, *Precision* of 0.8365, and *Recall* of 0.8395, and thus we chose it for our stance classifier.

In order to further classify *support* posts into such with neutral and with negative sentiment, we fine-tuned a similar architecture on the Yelp Review Polarity dataset to obtain a sentiment classifier. Altogether, the stance prediction of a user–article engagement  $e$  is given as  $stance(e)$ .

### 3.3 Factual News Graph (FANG) Framework

We now describe our FANG learning framework on the social context graph described in Section 3.2. Figure 2 shows an overview of FANG. While optimizing for the fake news detection objective, FANG also learns generalizable representations for the social entities. This is achieved by optimizing three concurrent losses: (i) unsupervised *Proximity Loss*, (ii) self-supervised *Stance Loss*, and (iii) self-supervised *Fake News Detection Loss*.

**Representation Learning.** We first discuss how FANG derives the representation of each social entity. Previous representation learning frameworks such as Deep Walk [31] and node2vec [11] compute a node embedding by sampling its neighborhood, and then optimizing for the proximity loss similarly to word2vec. However, the neighborhood is defined by the graph structure. These methods use the neighborhood structure only, and they are suitable when the node auxiliary features are unavailable or incomplete, *i.e.*, when optimizing for each entity’s structural representation separately. Recently, GraphSage [12] was proposed to overcome this limitation by allowing auxiliary node features to be used jointly with proximity sampling as part of the representation learning.

Let  $GraphSage(\cdot)$  be GraphSage’s node encoding function. Thus, we can now obtain the structural representation  $z_u \in \mathbb{R}^d$  of any user and source node  $r$  as  $z_r = GraphSage(r)$ , where  $d$  is the structural embedding dimension. For news nodes, we further enrich their structural representation with user engagement temporality, which we showed to be distinctive for fake news detection in Section 1 above. This can be formulated as learning an aggregation function  $F(a, U)$  that maps a questionable news  $a$ , and its engaged users  $U$  to a temporal representation  $v_a^{temp}$  that captures  $a$ ’s engagement pattern. Therefore, the aggregating model (*i.e.*, the aggregator) has to be time-sensitive. RNNs fulfill this requirement: specifically, the Bidirectional LSTM (Bi-LSTM) can capture a long-term dependency in information sequence in both the forward and the backward directions [15]. On top of the Bi-LSTM, we further incorporate an attention mechanism that focuses on essential engagement during the encoding process. Attention is not only expected to improve the model quality but also its explainability [9, 24]. By examining the model’s attention, we learn which social profiles influence the decision, mimicking human analytic capability.

Our proposed LSTM input is a user–article engagement sequence  $\{e_1, e_2, \dots, e_{|U|}\}$ . Let  $meta(e_i) \in \mathbb{R}^l = (time(e_i), stance(e_i))$  be the concatenation of  $e_i$ ’s elapsed time since the news publication and a one-hot stance vector. Each engagement  $e_i$  has its representation  $x_{e_i} = (z_{U_i}, meta(e_i))$ , where  $z_{U_i} = GraphSage(U_i)$ . A Bi-LSTM encodes the engagement sequence and outputs two sequences of hidden states: (i) a forward sequence,  $H^f = h_1^f, h_2^f, \dots, h_n^f$ , which starts from the beginning of the engagement sequence, and (ii) a backward sequence,  $H^b = h_1^b, h_2^b, \dots, h_n^b$ , which starts from the end of the engagement sequence.

Let  $w_i$  be the attention weight paid by our Bi-LSTM encoder to the forward ( $h_i^f$ ) and to the backward ( $h_i^b$ ) hidden states. This attention should be derived from the similarity of the hidden state and the news features, *i.e.*, how relevant the engaging users are to the discussed content, and the particular time and stance of the engagement. Therefore, we formulate the attention weight  $w_i$  as:

$$w_i = \frac{\exp(z_a M_e h_i + meta(e_i) M_m)}{\sum_{j=1}^n \exp(z_a M_e h_j + meta(e_j) M_m)}. \quad (1)$$

where  $l$  is the meta dimension,  $e$  is the encoder dimension, and  $M_e \in \mathbb{R}^{d \times e}$  and  $M_m \in \mathbb{R}^{l \times 1}$  are the optimizable projection matrices for engagement and meta features shared across all engagements.  $w_i$  is then used to compute the forward and the backward weighted feature vectors as  $h^f = \sum_i^n w_i h_i^f$  and  $h^b = \sum_i^n w_i h_i^b$ , respectively.

Finally, we concatenated the forward and the backward vectors to obtain the temporal representation  $v_a^{temp} \in \mathbb{R}^{2e}$  for article  $a$ . By explicitly setting  $2e = d$ , we can combine the temporal and the structural representations of a news  $a$  into a single representation:

$$z_a = v_a^{temp} + \text{GraphSage}(a). \quad (2)$$

**Unsupervised Proximity Loss.** We derive the Proximity Loss from the hypothesis that closely connected social entities often behave similarly. This is motivated by the echo chamber phenomenon, where social entities tend to interact with other entities of common interest to reinforce and to promote their narrative. This echo chamber phenomenon encompasses inter-cited news media sources publishing news of similar content or factuality, as well as social friends expressing similar stance with respect to news article(s) of similar content. Therefore, FANG should assign such nearby entities to a set of proximal vectors in the embedding space. We also hypothesize that loosely-connected social entities often behave differently from our observation that social entities are highly polarized, especially in left-right politics [4]. FANG should enforce that the representations of these disparate entities are distinctive.

The social interactions that define the above characteristics the most are user-user friendship, source-source citation, and news-source publication. As these interactions are either (a) between sources and news or (b) between news, we divide the social context graph into two sub-graphs, namely *news-source sub-graph* and *user sub-graph*. Within each sub-graph  $G'$ , we formulate the following Proximity Loss function:

$$\mathcal{L}_{prox} = - \sum_{u \in G'} \sum_{r_p \in P_r} \log(\sigma(z_r^\top z_{r_p})) + Q \cdot \sum_{r_n \in N_r} \log(\sigma(-z_r^\top z_{r_n})), \quad (3)$$

where  $z_r \in \mathbb{R}^d$  is the representation of entity  $r$ ,  $P_r$  is the set of nearby nodes or *positive set* of  $r$ , and  $N_r$  is the set of disparate nodes or *negative set* of  $r$ .  $P_r$  is obtained using our fixed-length random walk, and  $N_r$  is derived using negative sampling [12].

**Self-supervised Stance Loss.** We also propose an analogous hypothesis for the user-news interaction, in terms of stance. If a user expresses a stance with respect to a news article, their respective representations should be close. For each stance  $c$ , we first learn a user projection function  $\alpha_c(u) = \mathbf{A}_c z_u$  and a news article projection function  $\beta_c(a) = \mathbf{B}_c z_a$  that map a node representation of  $\mathbb{R}^d$  to a representation in the stance space  $c$  of  $\mathbb{R}^{d_c}$ . Given a user  $u$  and a news article  $a$ , we compute their similarity score in the stance space  $c$  as  $\alpha(u)^\top \beta(a)$ . If  $u$  expresses stance  $c$  with respect to  $a$ , we maximize this score, and we minimize it otherwise. This is the stance classification objective, optimized using the Stance Loss:

$$\mathcal{L}_{stance} = - \sum_{u,a} \sum_c y_{u,a,c} \log(f(u, a, c)), \quad (4)$$

where  $f(u, a, c) = \text{softmax}(\alpha_c(u)^\top \beta_c(a))$  and

$$y_{u,a,c} = \begin{cases} 1 & \text{if } u \text{ expresses stance } c \text{ over } a, \\ 0 & \text{otherwise.} \end{cases}$$

**Supervised Fake News Loss.** We directly optimize the main learning objective of fake news detection via the supervised Fake News Loss. In order to predict whether article  $a$  is false, we obtain its contextual representation as the concatenation of its representation and the structural representation of its source, i.e.,  $v_a = (z_a, z_s)$ .

---

#### Algorithm 1: FANG Learning Algorithm

---

**Input :** The social context graph  $G = (A, S, U, E)$   
The news labels  $Y_A$ , and the stance labels  $Y_{U,A,C}$   
**Output:** FANG-optimized parameters  $\theta$   
Initialize  $\theta$ ;  
**while**  $\theta$  has not converged **do**  
  **for** each news batch  $A_i \subset A$  **do**  
    **for** each news  $a \in A_i$  **do**  
       $U_a \leftarrow$  users who have engaged with  $a$ ;  
       $z_a \leftarrow$  Equation (2);  
       $z_s \leftarrow \text{GraphSage}(s)$ ;  
      **for** each user  $u \in U_a$  **do**  
         $z_u \leftarrow \text{GraphSage}(u)$ ;  
         $\mathcal{L}'_{stance} \leftarrow$  Equation (4);  
      **end**  
       $\mathcal{L}'_{news} \leftarrow$  Equation (5);  
    **end**  
  **for** each news-source or user sub-graph  $G'$  **do**  
    **for** each entity  $r \in G'$  **do**  
       $P_r \leftarrow$  positive samples of  $r$  in  $G'$ ;  
       $N_r \leftarrow$  negative samples of  $r$  in  $G'$ ;  
       $\mathcal{L}'_{prox} \leftarrow$  Equation (3);  
    **end**  
  **end**  
   $\mathcal{L}_{total} \leftarrow \text{SUM}(\mathcal{L}'_{stance}, \mathcal{L}'_{news}, \mathcal{L}'_{prox})$ ;  
   $\theta \leftarrow \text{Backpropagate}(\mathcal{L}_{total})$ ;  
**end**  
**return**  $\theta$

---

This contextual representation is then input into a fully connected layer whose outputs are computed as  $o_a = \mathbf{W}v_a + b$ , where  $\mathbf{W} \in \mathbb{R}^{2d \times 1}$  and  $b \in \mathbb{R}$  are the weights and the biases of the layer. The output value  $o_a \in \mathbb{R}$  is finally passed through a sigmoid activation function  $\sigma(\cdot)$ , and trained using the following cross-entropy Fake News Loss  $\mathcal{L}_{news}$ , defined as follows:

$$\mathcal{L}_{news} = \frac{1}{T} \sum_a \{y_a \cdot \log(\sigma(o_a)) + (1 - y_a) \cdot \log(1 - \sigma(o_a))\}, \quad (5)$$

where  $T$  is the batch size,  $y_a = 0$  if  $a$  is a fake article, and  $y_a = 1$  otherwise.

We define the total loss by linearly combining these three component losses:  $\mathcal{L}_{total} = \mathcal{L}_{prox} + \mathcal{L}_{stance} + \mathcal{L}_{news}$ . We provide detailed instructions for training FANG in Algorithm 1.

## 4 EXPERIMENTS

### 4.1 Data

We conducted our experiments on a Twitter dataset collected by related work on rumor classification [20, 25] and fake news detection [37]. For each article, we collected its source, a list of engaged users, and their tweets if they were not already available in the previous dataset. This dataset also includes Twitter profile description and the list of Twitter profiles each user follows.

**Table 6: Statistics about our dataset.**

Fake	448	Publications / source	2.38	Cites / source	8.38
Real	606	Engagements / news	71.9	Friends / user	58.25
Sources	442	Neu. support / news	19.07	Deny / news	5.27
Users	54461	Neg. support / news	10.83	Report / news	36.73

We further crawled additional data about media sources, including the content of their *Homepage* and their *About us* page, together with their frequently cited sources on their *Homepage*.

The truth value of the articles, namely, whether they are fake or real news, is based on two fact-checking websites: Snopes and Politifact. We release the source code of FANG and the stance detection dataset.<sup>2</sup> Table 6 shows some statistics about our dataset.

## 4.2 Fake News Detection Results

We benchmark the performance of FANG on fake news detection against several competitive models: (i) a content-only model, (ii) a Euclidean contextual model, and (iii) another graph learning model. In order to compare our FANG with the content-only model, we use a Support Vector Machine (SVM) model on TF.IDF feature vectors constructed from the news content (see Section 3.2). We also compare our approach with a current Euclidean model, CSI [35], a fundamental yet effective recurrent encoder that aggregates the user features, the news content, and the user-news engagements. We re-implement the CSI with source features by concatenating the overall score for the users and the article representation with our formulated source description to obtain the result vector for CSI’s Integrated module mentioned in the original paper. Lastly, we compare against the GCN graph learning framework [19]. First, we represent each of  $k$  social interactions in a separated adjacency matrix. We then concatenate GCN’s output on  $k$  adjacency matrices as the final representation of each node, before passing the representation through a linear layer for classification.

We also verify the importance of modeling temporality by experimenting on two variants of CSI and FANG: (i) temporal-insensitive CSI(- $t$ ) and FANG(- $t$ ) without  $time(e)$  in the engagement  $e$ ’s representation  $\mathbf{x}_e$ , and (ii) temporal sensitive CSI and FANG with  $time(e)$ . Table 7 shows the macroscopic results. For evaluation, we use the area under the Receiver Operating Characteristic curve or AUC score as standard. All context-aware models (*i.e.*, CSI(- $t$ ), CSI, GCN, and FANG(- $t$ )) and FANG improve over the context-unaware baseline by 0.1153 absolute with CSI(- $t$ ) and by 0.1993 absolute with FANG in terms of AUC score. This demonstrates that considering social context is helpful for fake news detection. We further observe that both time-sensitive CSI and FANG improve over their time-insensitive variants, CSI(- $t$ ) and FANG(- $t$ ) by 0.0233 and 0.0339, respectively. These results demonstrate the importance of modeling the temporality of news spreading. Finally, two graph-based models, FANG(- $t$ ) and GCN are consistently better than the Euclidean CSI(- $t$ ) by 0.0501 and 0.0386, respectively: this demonstrates the effectiveness of our social graph representation described in Section 3.2. Overall, we can observe that FANG outperforms the other context-aware, temporally-aware, and graph-based models.

<sup>2</sup><https://github.com/nguyenvanhoang7398/FANG>

**Table 7: Comparison between FANG and baseline models on fake news detection, evaluated with AUC score.**

Model	Contextual	Temporal	Graphical	AUC
Feature SVM				0.5525
CSI(- $t$ ) (without $time(e)$ )	✓			0.6678
CSI	✓	✓		0.6911
GCN	✓		✓	0.7064
FANG(- $t$ ) (without $time(e)$ )	✓		✓	0.7179
<b>FANG</b>	✓	✓	✓	<b>0.7518</b>

## 5 DISCUSSION

We now answer the following research questions (RQs) to better understand FANG’s performance under different scenarios:

- RQ1: Does FANG work well with limited training data?
- RQ2: Does FANG differentiate between fake and real news based on their contrastive engagement temporality?
- RQ3: How effective is FANG’s representation learning?

### 5.1 Limited Training Data (RQ1)

In order to address RQ1, we conducted the experiments described in Section 4.2 using different sizes of the training dataset. We found consistent improvements over the baselines under both limited and sufficient data conditions. Table 8 shows the experimental results and Figure 3 (left) further visualizes them. We can see that FANG consistently outperforms the two baselines for all training sizes: 10%, 30%, 50%, 70%, and 90% of the data. In terms of AUC score at decreasing training size, among the graph-based models, GCN’s performance drops by 16.22% from 0.7064 at 90% to 0.5918 at 10%, while FANG’s performance drops by 11.11% from 0.7518 at 90% to 0.6683 at 10%. We further observe that CSI’s performance drops the least by only 7.93% from 0.6911 at 90% to 0.6363 at 10%. Another result from an ablated baseline, FANG(- $s$ ), where we removed the stance loss, highlights the importance of this self-supervised objective. When training on 90% of the data, the relative underperforming margin of FANG(- $s$ ) compared to FANG is only 1.42% in terms of AUC. However, this relative margin increases as the availability of training data decreases, to at most 6.39% at 30% training data. Overall, the experimental results emphasize our model’s effectiveness even at low training data availability compared to the ablated version, GNN and Euclidean, which confirms RQ1.

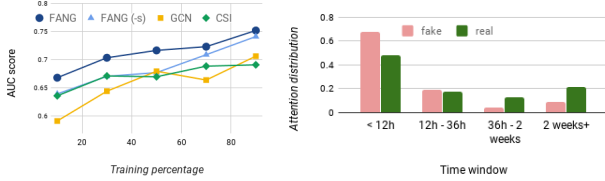
### 5.2 Engagement Temporality Study (RQ2)

To address RQ2 and to verify whether our model makes its decisions based on the distinctive temporal patterns between fake and real news, we examined FANG’s attention mechanism. We accumulated the attention weights produced by FANG within each time window and we compared them across time windows. Figure 3 (right) shows the attention distribution over time for fake and for real news.

We can see that FANG pays 68.08% of its attention to the user engagement that occurred in the first 12 hours after a news article has been published to decide whether it is fake. Its attention then sharply decreases to 18.83% for the next 24 hours, then to 4.14% from 36 hours to two weeks after publication, and finally to approximately 9.04% from the second week onward.

**Table 8: Comparison of FANG against baselines (AUC score) by varying the size of the training data.**

Systems	AUC score at different training percentages				
	10%	30%	50%	70%	90%
CSI	0.6363	0.6714	0.6700	0.6887	0.6911
GCN	0.5918	0.6445	0.6797	0.6642	0.7064
FANG(-s) (without stance loss)	0.6396	0.6708	0.6773	0.7090	0.7411
<b>FANG</b>	<b>0.6683</b>	<b>0.7036</b>	<b>0.7166</b>	<b>0.7232</b>	<b>0.7518</b>



**Figure 3: FANG’s performance against baselines (AUC score) for varying training data sizes (left), and attention distribution across time windows for fake vs. real news (right).**

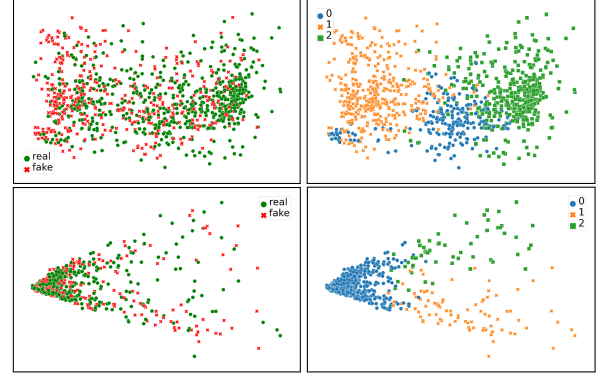
On the other hand, for real news, FANG places only 48.01% of its attention on the first 12 hours, which then decreases to 17.59% and to 12.85% in the time windows of 12 to 36 hours and 36 hours to two weeks, respectively. We also observe that FANG maintains 21.53% attention even when the real news has been published after two weeks.

Our model’s characteristics are consistent with the general observation that the appalling nature of fake news generates the most engagements within a short period of time after its publication. Therefore, it is reasonable that the model places much emphasis on these crucial engagements. On the other hand, genuine news attracts fewer engagements, but it is circulated for a longer period of time, which explains FANG’s persistent attention even after two weeks after publication. Overall, the temporality study here highlights the transparency of our model’s decision, largely thanks to the incorporated attention mechanism.

### 5.3 Representation Learning (RQ3)

Our core claim is to improve the quality of representation with FANG, and we verify it in intrinsic and extrinsic evaluations.

In the intrinsic evaluation, we verify how generalizable the minimally supervised news representations are for the fake news detection task. We first optimize both GCN and FANG on 30% of the training data to obtain news representations. We then cluster these representations using an unsupervised clustering algorithm, OPTICS [2], and we measure the homogeneity score — the extent to which clusters contain a single class. The higher the homogeneity score, the more likely the news articles of the same factuality label (i.e., fake or real) are to be close to each other, which yields higher quality representation. Figure 4 visualizes the representations obtained from two approaches with factuality labels and OPTICS clustering labels.



**Figure 4: 2D PCA plot of FANG’s representations with factuality labels (top left) and OPTICS clustering labels (top right), and GCN’s news representations with factuality labels (bottom left) and OPTICS clustering labels (bottom right).**

In the extrinsic evaluation, we verify how generalizable the supervised source representations are for a new task: source factuality prediction. We first train FANG on 90% of the training data to obtain all source  $s$  representations as  $z_s = \text{GraphSage}(s)$ , and the total representation as  $v_s = (z_s, x_s, \sum_{a \in \text{publish}(s)} x_a)$ , where  $x_s$ ,  $\text{publish}(s)$ , and  $x_a$  denote the source  $s$  content representation, the list of all articles published by  $s$ , and their content representations.

We propose two baseline representations that do not consider the source  $s$  content,  $v'_s = (z_s, x_s)$ . Finally, we train two separate SVM models for  $v_s$  and  $v'_s$  on the source factuality dataset, consisting of 129 sources of high factuality and 103 sources of low factuality, obtained from Media Bias/Fact Check<sup>3</sup> and PolitiFact.<sup>4</sup>

For intrinsic evaluation, the Principal Component Analysis (PCA) plot of labeled FANG representation (see Figure 4 top left) shows moderate collocation for the groups of fake and real news, while the PCA plot of labeled GCN representation (Figure 4 bottom left) shows little collocation within either the fake or the real news groups. Quantitatively, FANG’s OPTICS clusters (shown in Figure 4 top right) achieve a homogeneity score of 0.051 based on news factuality labels, compared to 0.0006 homogeneity score for the GCN OPTICS clusters. This intrinsic evaluation demonstrates FANG’s strong representation closeness within both the fake and the real news groups, indicating that FANG yields improved representation learning over another fully supervised graph neural framework.

For the extrinsic evaluation on downstream source factuality classification, our context-aware model achieves an AUC score of 0.8049 compared to 0.5842 for the baseline. We further examined the FANG representations for sources to explain this 0.2207 absolute improvement. Figure 5 shows the source representations obtained from the textual features, GCN, and FANG with their factuality labels, i.e., high, low, mixed, and citation relationship. In the left sub-figure, we can observe that the textual features are insufficient to differentiate the factuality of media, as a fake news spreading site such as *cnsnews* could mimic factual media in terms of web design and news content.

<sup>3</sup><https://www.mediabiasfactcheck.com>

<sup>4</sup><http://politifact.com>

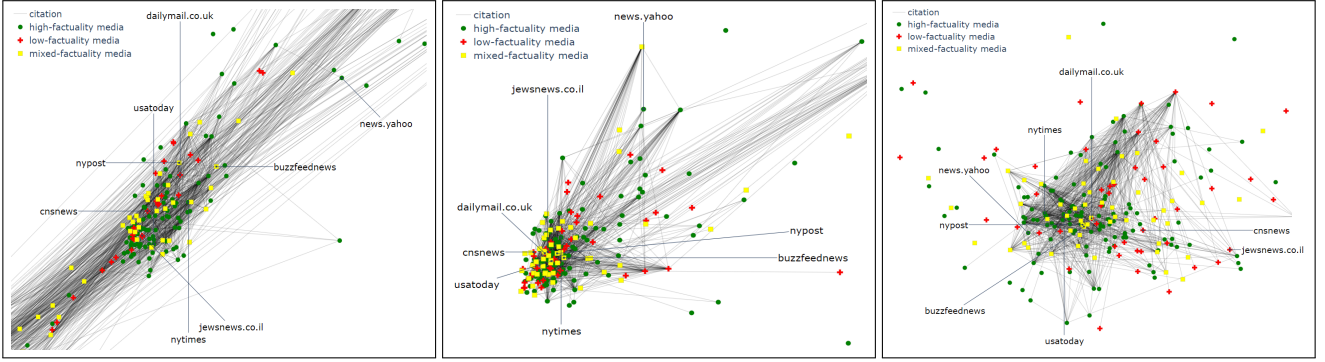


Figure 5: Plots for source representations using textual features (left), GCN (middle), and FANG (right) with factuality labels.

However, the citation between a low-factuality website and high-factuality sites would not be as high, and it is effectively used by the two graph learning frameworks: GCN and especially FANG. Yet, GCN fails to differentiate low-factuality sites with higher citations, such as *jewsnews.co.il* and *cnsnews*, from high-factuality sites. On the other hand, sources such as *news.yahoo* despite being textually different, as shown in Figure 5 (left), should still cluster with other credible media for their high inter-citation frequency. FANG, with much more emphasis on contextual representation learning, makes these sources more distinguishable. Its representation space gives us a glance into the landscape of news media, where there is a large central cluster of high-factuality inter-cited sources such as *nytimes*, *washingtonpost* and *news.yahoo*. At the periphery lie less connected media including both high- and low-factuality ones.

We also see cases where all models failed to differentiate mixed-factuality media, such as *buzzfeednews* and *nypost*, which have high citation counts with high-factuality media. Overall, the results from intrinsic and extrinsic evaluation, as well as the observations, confirm RQ3 on the improvement of FANG’s representation learning.

#### 5.4 Scalable Inductiveness

FANG overcomes the transductive limitation of previous approaches while inferring the credibility of unseen nodes. MVDAM [21] has to randomly initialize an embedding and to optimize it iteratively using node2vec [11] for any unseen node, whereas FANG directly infers the embedding with its learned feature aggregator. Other graphical approaches using matrix factorization [39] or graph convolutional layers [10, 29] learn parameters whose dimensionality is fixed to the network size  $N$ , and can be as expensive as  $O(N^3)$  [10] in terms of inference time complexity. FANG infers the embeddings of unseen nodes without reconstructing the adjacency matrix, and its inference time complexity only depends on the size of the neighborhood of unseen nodes.

#### 5.5 Microscopic Analysis

It is also helpful to analyze FANG’s predictions by examining specific test examples. The first example is shown in Figure 6, where we can see that FANG pays most of its attention to a tweet by user B. This can be explained by B’s Twitter profile description of a fact-checking organization, which indicates high reliability.

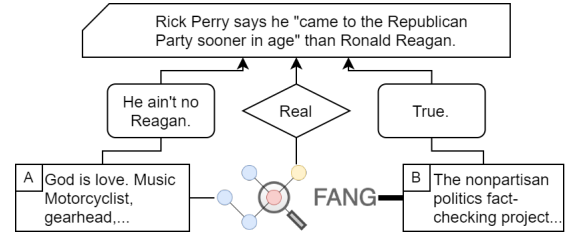


Figure 6: A test example explaining FANG’s decision.

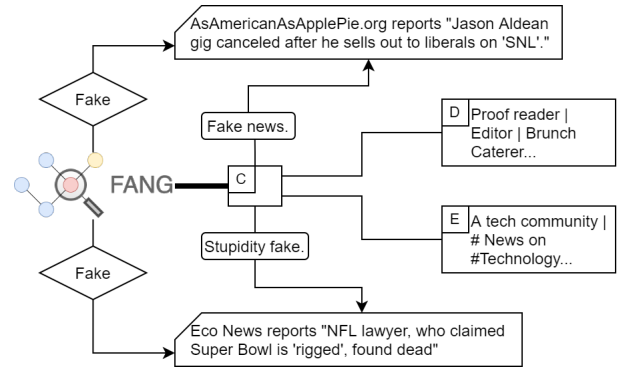


Figure 7: Another test example explaining FANG’s decision.

In contrast, a denying tweet from user A is not paid so much attention, due to the insignificant description of its author’s profile. Our model bases its prediction on the *support* stance from the fact-checker, which is indeed the correct label.

In the second example, shown in Figure 7, FANG pays most attention to a tweet by user C. Although this profile does not provide any description, it has a record of correctly denying the fake news about the dead NFL lawyer. Furthermore, the profiles that follow Twitter user C, namely user D and user E, have credible descriptions of a proof reader and of a tech community, respectively. This explains why our model bases its prediction of the news being fake thanks to the reliable denial, which is again the correct label.

## 5.6 Limitations

We note that entity and interaction features are constructed before passing to FANG, and thus errors from upstream tasks, such as textual encoding or stance detection, can propagate to FANG. Future work can address this in an end-to-end framework, where textual encoding [9] and stance detection can be jointly optimized.

Another limitation is that the dataset for contextual fake news detection can quickly become obsolete as hyperlinks and social media traces at the time of publication might no longer be retrievable.

## 6 CONCLUSION AND FUTURE WORK

We have demonstrated the importance of modeling the social context for the task of fake news detection. We further proposed FANG, a graph learning framework that enhances representation quality by capturing the rich social interactions between users, articles, and media, thereby improving both fake news detection and source factuality prediction. We have demonstrated the efficiency of FANG with limited training data and its capability of capturing distinctive temporal patterns between fake and real news with a highly explainable attention mechanism. In future work, we plan more analysis of the representations of social users. We further plan to apply multi-task learning to jointly address the tasks of fake news detection, source factuality prediction, and echo chamber discovery.

## REFERENCES

- [1] O. Ajao, D. Bhowmik, and S. Zargari. 2019. Sentiment Aware Fake News Detection on Online Social Networks. In *ICASSP*. 2507–2511.
- [2] Mihael Ankerst, Markus M Breunig, Hans-Peter Kriegel, and Jörg Sander. 1999. OPTICS: Ordering Points to Identify the Clustering Structure. *ACM SIGMOD Record* 28, 2 (1999), 49–60.
- [3] Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. 2018. Predicting Factuality of Reporting and Bias of News Media Sources. In *EMNLP*. 3528–3539.
- [4] Levi Boxell, Matthew Gentzkow, and Jesse M. Shapiro. 2017. *Is the Internet Causing Political Polarization? Evidence from Demographics*. Technical Report NBER Working Paper No. 23258. National Bureau of Economic Research.
- [5] Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. 2017. Geometric Deep Learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine* 34, 4 (2017), 18–42.
- [6] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information Credibility on Twitter. In *WWW*. 675–684.
- [7] Kareem Darwish, Dimitar Alexandrov, Preslav Nakov, and Yelena Mejova. 2017. Seminar Users in the Arabic Twitter Sphere. In *SocInfo*. 91–108.
- [8] Leon Derczynski, Kalina Bontcheva, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Arkaitz Zubiaga. 2017. SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours. In *SemEval-2017*. 69–76.
- [9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL*. 4171–4186.
- [10] Ming Dong, Bolong Zheng, Nguyen Quoc Viet Hung, Han Su, and Guohui Li. 2019. Multiple Rumor Source Detection with Graph Convolutional Networks. In *CIKM*. 569–578.
- [11] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable Feature Learning for Networks. In *KDD*. 855–864.
- [12] William L. Hamilton, Zitao Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *NIPS*. 1024–1034.
- [13] Naeemul Hassan, Gensheng Zhang, Fatma Arslan, Josue Caraballo, Damian Jimenez, Siddhant Gawsane, Shohedul Hasan, Minumol Joseph, Aaditya Kulkarni, Anil K. Nayak, Vikas Sable, Chengkai Li, and Mark Tremayne. 2017. ClaimBuster: The First-ever End-to-end Fact-checking System. *VLDB* 10, 12 (2017), 1945–1948.
- [14] Xiangnan He, Ming Gao, Min-Yen Kan, Yiqun Liu, and Kazunari Sugiyama. 2014. Predicting the Popularity of Web 2.0 Items based on User Comments. In *SIGIR*. 233–242.
- [15] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (1997), 1735–1780.
- [16] Zhiwei Jin, Juan Cao, Yu-Gang Jiang, and Yongdong Zhang. 2014. News Credibility Evaluation on Microblog with a Hierarchical Propagation Model. In *ICDM*. 230–239.
- [17] Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. News Verification by Exploiting Conflicting Social Viewpoints in Microblogs. In *AAAI*. 2972–2978.
- [18] Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. MVAE: Multimodal Variational Autoencoder for Fake News Detection. In *WWW*. 2915–2921.
- [19] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*.
- [20] Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. PHEME dataset for Rumour Detection and Veracity Classification. (6 2018). [https://figshare.com/articles/dataset/PHEME\\_dataset\\_for\\_Rumour\\_Detection\\_and\\_Veracity\\_Classification/6392078](https://figshare.com/articles/dataset/PHEME_dataset_for_Rumour_Detection_and_Veracity_Classification/6392078)
- [21] Vivek Kulkarni, Juntong Ye, Steven Skiena, and William Yang Wang. 2018. Multi-view Models for Political Ideology Detection of News Articles. In *EMNLP*. 3518–3527.
- [22] Yinhan Liu, Mylène Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv:1907.11692* (2019).
- [23] Yang Liu and Yi-Fang Brook Wu. 2018. Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks. In *AAAI*. 354–361.
- [24] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective Approaches to Attention-based Neural Machine Translation. In *EMNLP*. 1412–1421.
- [25] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting Rumors from Microblogs with Recurrent Neural Networks. In *IJCAI*. 3818–3824.
- [26] Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. 2015. Detect Rumors Using Time Series of Social Context Information on Microblogging Websites. In *CIKM*. 1751–1754.
- [27] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 Task 6: Detecting Stance in Tweets. In *SemEval-2016*. 31–41.
- [28] Mitra Mohtarami, Ramy Baly, James Glass, Preslav Nakov, Lluís Màrquez, and Alessandro Moschitti. 2018. Automatic Stance Detection Using End-to-End Memory Networks. In *NAACL-HLT*. 767–776.
- [29] Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M Bronstein. 2019. Fake News Detection on Social Media using Geometric Deep Learning. *arXiv preprint arXiv:1902.06673* (2019).
- [30] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *EMNLP*. 1532–1543.
- [31] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online Learning of Social Representations. In *KDD*. 701–710.
- [32] Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2016. Credibility Assessment of Textual Claims on the Web. In *CIKM*. 2173–2178.
- [33] Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2017. Where the Truth Lies: Explaining the Credibility of Emerging Claims on the Web and Social Media. In *WWW*. 1003–1012.
- [34] Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2018. CredEye: A Credibility Lens for Analyzing and Explaining Misinformation. In *WWW*. 155–158.
- [35] Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. CSI: A Hybrid Deep Model for Fake News Detection. In *CIKM*. 797–806.
- [36] Gerald Salton and Michael J. McGill. 1983. *Introduction to Modern Information Retrieval*. McGraw-Hill.
- [37] 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).
- [38] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter* 19, 1 (2017), 22–36.
- [39] Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond News Contents: The Role of Social Context for Fake News Detection. In *WSDM*. 312–320.
- [40] Parinaz Sobhani, Diana Inkpen, and Xiaodan Zhu. 2017. A Dataset for Multi-Target Stance Detection. In *EACL*. 551–557.
- [41] Zoe Thomas. 2020. WHO says fake coronavirus claims causing ‘infodemic’. *BBC News* (13th Feb, 2020). <https://www.bbc.com/news/technology-51497800>
- [42] James Thorne and Andreas Vlachos. 2018. Automated Fact Checking: Task Formulations, Methods and Future Directions. In *COLING*. 3346–3359.
- [43] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. EANN: Event Adversarial Neural Networks for Multimodal Fake News Detection. In *KDD*. 849–857.
- [44] Fan Yang, Yang Liu, Xiaohui Yu, and Min Yang. 2012. Automatic Detection of Rumor on Sina Weibo. In *MDS*. 1–7.
- [45] Chunyuan Yuan, Qianwen Ma, Wei Zhou, Jizhong Han, and Songlin Hu. 2019. Jointly Embedding the Local and Global Relations of Heterogeneous Graph for Rumor Detection. In *ICDM*. 796–805.