

Nothing Stands Alone: Relational Fake News Detection with Hypergraph Neural Networks

Ujun Jeong*, Kaize Ding*, Lu Cheng[†], Ruocheng Guo[‡], Kai Shu[§] and Huan Liu*

*School of Computing and Augmented Intelligence, Arizona State University

[†]Department of Computer Science, University of Illinois Chicago

[‡]Bytedance AI Lab, London

[§]Department of Computer Science, Illinois Institute of Technology

Email: *{ujeong1, kding9, huanliu}@asu.edu, [†]lucheng@uic.edu, [‡]rguo.asu@gmail.com, [§]kshu@iit.edu

Abstract—Nowadays, fake news easily propagates through online social networks and becomes a grand threat to individuals and society. Assessing the authenticity of news is challenging due to its elaborately fabricated contents, making it difficult to obtain large-scale annotations for fake news data. Due to such data scarcity issues, detecting fake news tends to fail and overfit in the supervised setting. Recently, graph neural networks (GNNs) have been adopted to leverage the richer relational information among both labeled and unlabeled instances. Despite their promising results, they are inherently focused on pairwise relations between news, which can limit the expressive power for capturing fake news that spreads in a group-level. For example, detecting fake news can be more effective when we better understand relations between news pieces shared among susceptible users. To address those issues, we propose to leverage a hypergraph to represent group-wise interaction among news, while focusing on important news relations with its dual-level attention mechanism. Experiments based on two benchmark datasets show that our approach yields remarkable performance and maintains the high performance even with a small subset of labeled news data.

Index Terms—Fake News Detection, Social Media Mining, Hypergraph, Graph Neural Networks

I. INTRODUCTION

With the growing amount of information that spreads through online social networks, people are at high risk of being exposed to fake news [1]. Online social networks often facilitate the spread of fake news as people tend to search for, recall, and share information that supports one’s prior beliefs due to *confirmation bias* [2]. Further, the *echo chamber effect* [3] reinforces this prism of reality among people who share the same opinion. Consequently, fake news jeopardizes the public trust in government and accelerates polarization in society [4].

Nevertheless, assessing the authenticity of news is difficult, as fake news is written to intentionally mislead people with elaborately manipulated facts and narratives. The contents and topics covered by fake news are also diverse, which makes it extremely laborious to obtain large amounts of annotated fake news. Existing works in fake news detection mainly focus on the supervised setting where abundant labeled data is given [5]–[7]. In this case, the detection models would easily overfit and fail to make accurate predictions when only limited labeled data is available. Therefore, how to solve the problem of fake news detection under the low-resource scenario has become a challenging task that requires urgent research efforts.

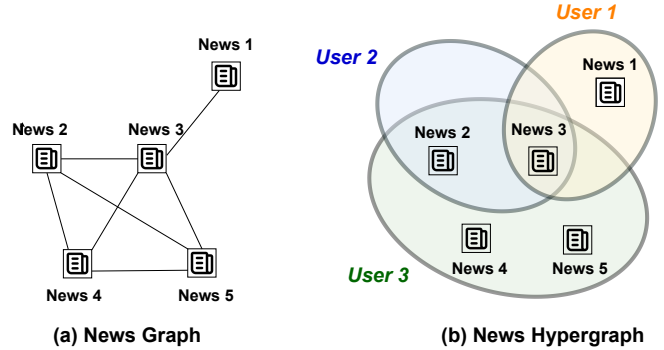


Fig. 1. Examples of a news graph (a) and a news hypergraph (b) representing users’ news sharing behavior on a social network. Each node denotes a news piece, and nodes are connected when news are shared by the same user.

In this paper, we aim to investigate whether capturing meaningful interactions between news can improve the performance of fake news detection under the low-resource setting. However, it is a non-trivial task due to the following reasons: (1) Though graph neural networks (GNNs) have been applied to fake news detection [7], [8] to capture the relational information, this line of work assumes that the news interactions are based on pairwise relations, which may omit the important group-wise interactions among different news pieces. In fact, many fake news pieces spring up organically [9], that being said, fake news is often shared by a group of similar users. This means that fake news can be better understood by capturing such group-wise interactions; (2) Most graph-based fake news detection methods merely consider the relations for the labeled news pieces that are naturally formed during the proliferation process of news on a social network [10]–[12]. However, the information from unlabeled news pieces and their interactions with labeled news are largely ignored. Hence, how to incorporate the unlabeled data that can potentially benefit the propagation of label information is another key challenge to solve.

To address the aforementioned challenges, we first propose to use a *hypergraph* as the modeling basis. A hypergraph can naturally represent group-wise interactions by connecting multiple (≥ 2) nodes to its special edge called a *hyperedge*. For example, Fig. 1 shows users’ news sharing behavior on a social network. The news hypergraph (Fig. 1(b)) can fuse

the crucial group-wise interaction that *News 3*, *News 4*, and *News 5* are shared by the same user. In contrast, the news graph (Fig. 1(a)) only captures the pairwise relations using the following edges: (*News 3*, *News 4*), (*News 3*, *News 5*), and (*News 4*, *News 5*), which can lead to the misinterpretation that the three news pieces are shared by three different users. Moreover, we design the problem of fake news detection as a semi-supervised node classification task in the hypergraph, where nodes denote news pieces and hyperedges denote any group-wise interactions. This way we are able to leverage both labeled and unlabeled news based on different hyperedges and propagate the feature patterns of labeled news in the news hypergraph. To further detect fake news on the constructed news hypergraph, we propose a new hypergraph neural network model – HGFND (**H**yper**G**raph for **F**ake **N**ews **D**etection). Considering the fact that noisy and task-irrelevant relations in the hypergraph (e.g., news shared mistakenly [13]) may jeopardize the learning process, HGFND adopts a dual-level attention mechanism to highlight useful news relations by measuring their importance in both node-level and hyperedge-level. Experiments on two real-world benchmark datasets for fake news detection (i.e., Politifact, Gossipcop) show that the proposed model outperforms the state-of-the-art methods by a large margin with both full and limited number of labeled data. Our main contributions are summarized as follows:

- To the best of our knowledge, this is the first work that incorporates group-wise interactions among news pieces by leveraging a hypergraph for fake news detection.
- We propose to construct the hypergraph based on social context and news content. We also define three types of hyperedges based on user, time, and entity shared in news.
- We propose a novel hypergraph neural network, which outperforms graph neural networks for fake news detection with both full and limited number of labeled data.
- Our model can jointly capture propagation patterns of news while leveraging the information from unlabeled news and their relations with labeled news.

II. RELATED WORK

In this chapter, we focus on introducing various methods used in fake news detection and reviewing a series of studies on graph neural networks in relational machine learning.

A. Fake News Detection

The most basic characteristic of news is its textual content. Early researches about fake news detection mostly relied on the content in the body or headline of news to understand its linguistic features and compare it with relevant evidences retrieved from Wikipedia [14], [15]. For example, a series of work has shown promising result in encoding the news textual information such as Riedel et al. [16] and other BERT-based models [17]. On the other hand, social context-based approaches include user-driven social engagements of news consumption on social network platforms. For example, CSI [18] incorporates the behavior of both parties, users and articles, and the group behavior of users who spread news.

DEFEND [19] utilizes user and news comments based on hierarchical attention network [20]. Han et al. [11] utilize user features such as the number of followers and friends without any textual information related to news to detect. Cheng et al. [21] show causal relationships between user’s attributes and susceptibility of users. A series of works using GCN [22] have been used for modeling news propagation tree such as GCNFN [6], BiGCN [12] and GNN-CL [11]. UPFD [10] benchmarks FakenewsNet [23] by jointly capturing the news contents and surrounding exogenous context of news through propagation tree. GTN [24] and TGNF [25] consider time difference between tweets/retweets in propagation trees using time-aware message aggregation and attention networks. Silva et al. [26] focus on predicting propagation patterns of fake news in its early stage by reconstructing news propagation trees. Various methods have been applied to leverage auxiliary information related to news. HGAT [7] leverages a heterogeneous graph to model pairwise relations between news, creators, and subjects through node and semantic-level attention mechanism. A hypergraph has recently drawn attention in fake news detection by clustering the news content sharing patterns of users [27], or jointly modeling topic relations between news [28] with hierarchical attention network.

B. Graph Neural Networks

Graph Neural Network (GNN) models aim to obtain better node representations via message passing among local neighbors in the graph using neighborhood aggregation [29], [30]. GCN [22] can be explained as a mean-pooling neighborhood aggregation. GraphSAGE [31] uses node features with max pooling or LSTM based aggregation, which can be used in inductive representation learning. GAT [32] incorporates trainable attention weights to learn weights on neighbors when aggregating neighboring information about a node. Furthermore, a heterogeneous graph has been adopted to represent various types of nodes and edges [33] to enhance expressive power of the graph. However, existing graph neural networks only consider pairwise relationships between objects, and cannot apply to non-pairwise relation learning. Recently, hypergraph-based methods have drawn much attention due to its expressive power to represent high-order relations. Convolutional hypergraph neural networks [34] has shown promising results by applying the convolution operation on graph to hypergraph in semi-supervised learning setting. For instance, HGNN [35] proposes to build hyperedges by connecting nodes that are semantically similar based on a distance metric. While convolutional hypergraph neural networks expand the convolution operation on a graph to a hypergraph, hypergraph attention further enhances the capacity of representation learning by leveraging an attention module [34]. For example, HyperGAT [36] uses dual-level attention mechanism to capture high-order information between words and sentences for document classification in an inductive setting. Additionally, hypergraph neural networks have been widely applied in many other applications, such as recommendation system [37], [38], visual object detection [35], and advertisement classification [39].

Compared to past approaches in fake news detection, our approach has a clear distinction in that it can represent group-wise interactions among news with hypergraph neural networks, which leverages the hypergraph constructed from both social context and news content. The proposed model helps to identify important group-wise interactions for fake news detection with dual-level attention mechanism for nodes (news) and hyperedges in the stage of message aggregation.

III. PRELIMINARIES

In this section, we introduce the problem of relational fake news detection. To this end, we compare the definition of a propagation tree and a hypergraph and define the problem statement based on these structures. Foremost, the formal definition of a propagation tree is as follows:

a) Propagation Tree: Given total N number of news as $\mathcal{X} = \{x_i\}_{i=1}^N$ and user engagement associated to news x_i as $\mathcal{U}_i = \{u_{t_j}\}_{j=1}^{|\mathcal{U}_i|}$ where t_j denotes the sequence of timestamp of a user's engagement such as tweet and retweet. A propagation tree is defined as $P_i = (\{x_i\} \cup \mathcal{U}_i, \mathbf{A})$ where \mathbf{A} is an adjacency matrix estimated by time-inferred diffusion [1].

A propagation tree is designed to represent the cascade of news that spreads on a social network through user engagements. While a propagation tree is used for a single news piece, we use a hypergraph to represent the relations among multiple news pieces. The formal definition of a hypergraph is as follows:

b) Hypergraph: A hypergraph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_i\}_{i=1}^N$ denotes a set of nodes in the graph, and $\mathcal{E} = \{e_j\}_{j=1}^M$ represents a set of hyperedges. Different from a graph, a hypergraph can connect two or more nodes by a hyperedge. A hypergraph \mathcal{G} can be represented by an incidence matrix $\mathbf{H} \in \mathbb{R}^{N \times M}$, with entries defined as:

$$\mathbf{H}_{i,j} = \begin{cases} 1 & \text{if } v_i \in e_j \\ 0 & \text{if } v_i \notin e_j \end{cases} \quad (1)$$

The existing approaches to detecting fake news mainly focus on either to learn the characteristics of fake news individually and independently [16], [19] or use a graph to represent news interactions as pairwise relations [6], [7]. However, a graph can be limited in representing the group-wise interactions when the relations between news pieces are higher-order [40]. To bridge this gap, we propose to define the problem of relational fake news detection as follows:

c) Problem Statement: Given total N number of news, the goal of relational fake news detection is to predict the labels of news into two labels $\mathcal{Y} = \{y_i\}_{i=1}^N \subset \{0, 1\}^N$ where news corresponds to a node in the hypergraph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. The hypergraph connects nodes (news) through a hyperedge based on the news contents $\mathcal{X} = \{x_i\}_{i=1}^N$ and their social context obtained from the propagation trees $\mathcal{P} = \{P_i\}_{i=1}^N$.

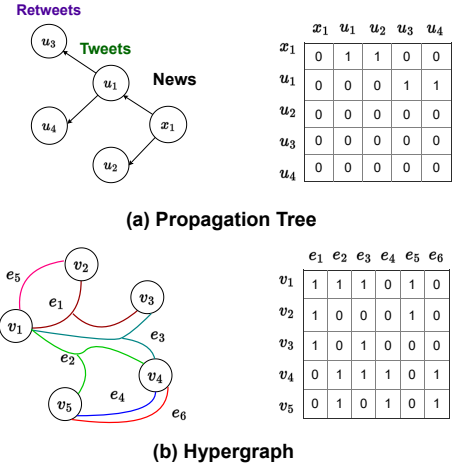


Fig. 2. A comparison between (a) a propagation tree and (b) a hypergraph

IV. METHODOLOGY

A. Constructing Hypergraph for News Relations

To construct the hypergraph for news relations, we assume there is a dataset that has already collected news content and user engagements, such as FakeNewsNet [23]. Based on this, we propose a hypergraph construction method for fake news detection using three types of hyperedges, as follows:

1) *User (Shared User ID between News Pieces):* This type of hyperedge implies the news sharing behavior of a user. This means a hyperedge corresponds to a *User ID*. To this end, we retrieve the unique *User IDs* from tweet/retweet in the propagation tree of news. In order to Then, if a piece of news shares the same *User ID* as other news pieces, we connect them to a hyperedge.

2) *Time (News Shared in Proximal Time Range):* This type of hyperedge assumes that similar news can emerge and be shared in the same time. This means a hyperedge corresponds to the creation timestamp of tweet/retweet in the propagation tree of news. Since the timestamp is in seconds, there may not be many news shared at the exact same time. To solve this problem, we round the timestamp to the nearest day or hour when connecting news. The rounding method can be opted flexibly based on the performance on the task.

3) *Entity (Shared Entity between News Contents):* This type of hyperedge aims to connect news that deal with similar contents by checking the entity shared between news pieces. For example, we connect news by a hyperedge if they share the entity ‘COVID-19’. To extract entities from news content, we employ an entity recognition tool from spacy [41]. Here, we adopt seven types of entities provided in spacy: organizations (*ORG*), people (*PER*), products (*PROC*), title of arts such as books and songs (*WORK_OF_ART*), nationalities or religious or political groups (*NORP*), non-GPE locations (*LOC*), and named events such as hurricanes, wars, and sports (*EVENT*).

TABLE I
NOTATIONS AND DESCRIPTIONS.

Notation	Description
\mathcal{X}	set of features for news contents
\mathcal{U}_i	set of features for user engagements on news x_i
\mathcal{P}	set of propagation trees for \mathcal{X}
\mathcal{V}	set of nodes in the hypergraph
\mathcal{E}	set of hyperedges in the hypergraph
\bar{x}_i	representation of the propagation on news x_i
v_i^l, e_j^l	representation of v_i and e_j in the l -th layer
α_i, β_j	attention coefficient for v_i and e_j
h_i, r_j	hidden representation for v_i and e_j
a_1, a_2	context vector for h_i and r_j
y_i, \hat{y}_i	ground-truth label and predicted logit of v_i

To construct the proposed hypergraph, we simply combine the incidence matrix of \mathbf{H}^{user} , \mathbf{H}^{time} , and \mathbf{H}^{entity} from three construction methods by $\mathbf{H} = \mathbf{H}^{user} \oplus \mathbf{H}^{time} \oplus \mathbf{H}^{entity}$ where \oplus denotes concatenation operation. Kindly note that the hypergraph needs to be constructed before training the model.

B. Leveraging News Relations for Fake News Detection

In this section, we propose our hypergraph neural network, well-designed for fake news detection. Fig. 3 shows how the proposed method can learn important correlations in the hypergraph while jointly modeling the propagation patterns of news. In the following, we explain about three major modules: (1) encoding propagation tree for node representation in the hypergraph, (2) learning relational information between nodes and hyperedge using dual-level attention mechanism [36], and (3) the final layer for news classification.

1) *Encoding Propagation Tree*: Given a set of propagation trees $\mathcal{P} = \{P_i\}_{i=1}^N$, we use Graph Neural Networks (GNNs) to encode each propagation tree. To obtain the propagation pattern of news, we extract the root representation of the propagation tree $\bar{x}_i \in \mathbb{R}^F$ where F is the size of the input feature matrix. Next, we concatenate \bar{x}_i with the original news features $x_i \in \mathbb{R}^F$ by skip-connection to generate a news representation enhanced with its propagation pattern. Then, we project it into $v_i^0 \in \mathbb{R}^d$ for the initial node representation of the hypergraph with hidden dimension size d as follows:

$$\begin{aligned} \bar{x}_i &= \text{ROOT}(GNN(P_i)), \\ v_i^0 &= f(\sigma(x_i \oplus \bar{x}_i)), \end{aligned} \quad (2)$$

where σ is the nonlinearity such as ReLU, f is a fully connected layer mapping $\mathbb{R}^{2 \times d} \rightarrow \mathbb{R}^d$, and GNN denotes a multi-layer graph neural network encoder. Specifically, we use GraphSAGE [31] as the backbone of the GNN module.

2) *Node-level Attention for Hyperedge Representation*: Given all the node representation $v^{l-1} \in \mathbb{R}^{|\mathcal{V}| \times d}$ in the layer $l-1$, the first layer of the hypergraph attention network learns the representations of hyperedges \mathcal{E} . To reflect the different importance of nodes to hyperedge representation, we use a node-level attention mechanism to highlight those nodes that are important to the meaning of the hyperedge. Formally,

nodes are aggregated to compute the hyperedge representation $e_j^l \in \mathbb{R}^d$ in the l -th layer as follows:

$$e_j^l = \sigma \left(\sum_{v_k \in e_j} \alpha_{jk} W_1 v_k^{l-1} \right), \quad (3)$$

where v_k is a node connected to the hyperedge $e_j \in \mathcal{E}$, $W_1 \in \mathbb{R}^{d \times d}$ is a trainable weight matrix, and $a_1 \in \mathbb{R}^{|\mathcal{E}| \times |\mathcal{V}|}$ is a context vector. The attention coefficient α_{jk} for node representation v_k^{l-1} in the l -th layer is computed as follows:

$$\begin{aligned} \alpha_{jk} &= \frac{\exp(a_1^T h_k)}{\sum_{v_p \in e_j} \exp(a_1^T h_p)}, \\ h_k &= \text{LeakyReLU}(W_1 v_k^{l-1}), \end{aligned} \quad (4)$$

3) *Hyperedge-level Attention for Node Representation*: With all the hyperedge representations $e^l \in \mathbb{R}^{|\mathcal{E}| \times d}$ in the l -th layer, we apply an edge-level attention mechanism to focus on the useful hyperedges for learning the next-layer representation of node v_i . This process can be expressed as:

$$v_i^l = \sigma \left(\sum_{e_j \in \mathcal{E}_i} \beta_{ij} W_2 e_j^l \right), \quad (5)$$

where v_i^l is the representation of node v_i in the l -th layer, $W_2 \in \mathbb{R}^{d \times d}$ is a weight matrix, and \mathcal{E}_i is a set of hyperedges connected to node v_i . Then, the attention coefficient β_{ij} for hyperedge representation e_j^l is computed as follows:

$$\begin{aligned} \beta_{ij} &= \frac{\exp(a_2^T r_j)}{\sum_{v_i \in e_j, e_j \in \mathcal{E}_i} \exp(a_2^T r_i)}, \\ r_j &= \text{LeakyReLU}([W_2 e_j^l \oplus W_1 v_i^{l-1}]), \end{aligned} \quad (6)$$

where $a_2 \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{E}|}$ is another context vector adopted to measure the importance of the hyperedges when aggregating the messages to the node representation.

4) *News (Node) Classification*: The node representation $v^L \in \mathbb{R}^{|\mathcal{V}| \times d}$ is a high-level representation resulted from node-level aggregation, where L is the last layer of the hypergraph attention network. For the news classification layer, we use a fully connected layer with a trainable weight matrix $W_3 \in \mathbb{R}^{d \times 2}$ and bias term b to classify a node into two labels. To this end, we use Negative Log Likelihood to optimize the loss between the predicted logit of the i -th node \hat{y}_i and its ground-truth label, y_i where \mathcal{Y}_{train} is labels for training data.

$$\begin{aligned} \hat{y}_i &= \text{Softmax}(f(W_3 v_i^L + b)), \\ \mathcal{L} &= - \sum_{y_i \in \mathcal{Y}_{train}} (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \end{aligned} \quad (7)$$

Note that the proposed framework works in the semi-supervised setting for transductive node classification in the hypergraph. This means the incidence matrix and features of all nodes in the hypergraph are accessible for training. However, there is no label leakage issues, since the labels of other nodes apart from the training data are unknown during the training phase [42].

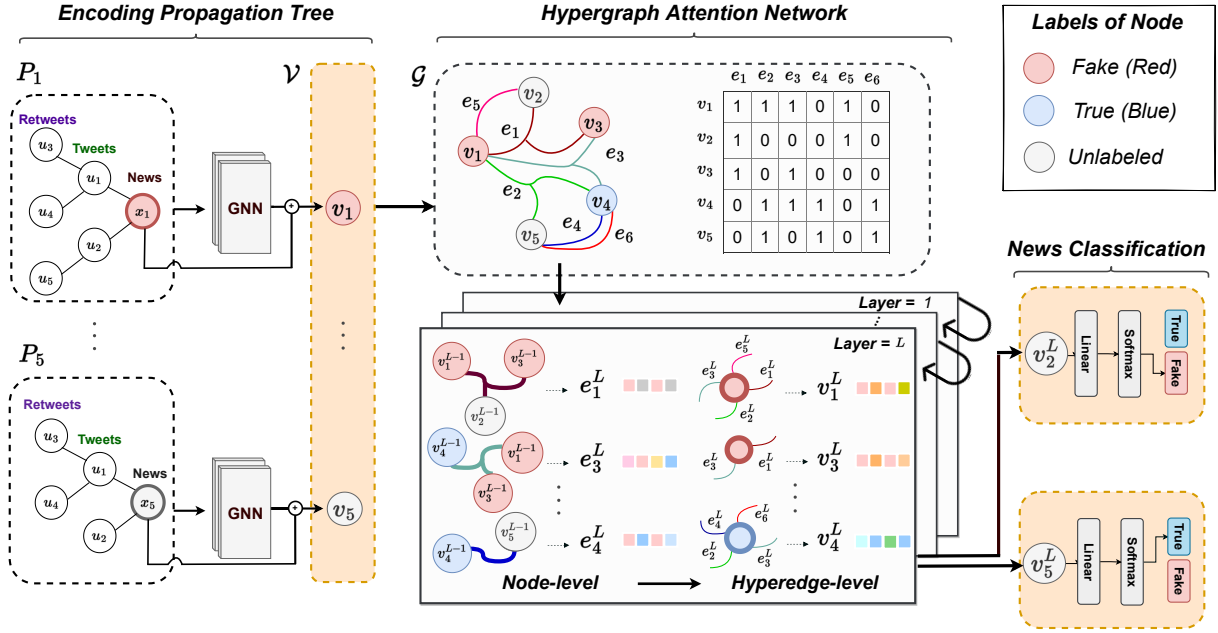


Fig. 3. Illustration of the proposed framework, which jointly models propagation trees and the hypergraph for classifying unlabeled news. The figure shows an example of transductive node classification where the number of news is 5 and the number of hyperedge is 6. The color of node in the hypergraph represents the label of news. i.e., fake (red), true (blue), and unlabeled (gray). The figure is best shown in color.

Algorithm 1 The overall process of HGFND

Input: The news feature matrix \mathcal{X} ,
The propagation trees \mathcal{P} .

Output: The optimized node representation on \mathcal{V} ,
The node-level attention weight α ,
The hyperedge-level attention weight β .

- 1: Construct incidence matrix for the hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
 $\mathbf{H} = \mathbf{H}^{user} \oplus \mathbf{H}^{time} \oplus \mathbf{H}^{entity}$;
- 2: **for** $P_i \in \mathcal{P}$ **do**
- 3: Encode the propagation tree associated with news
 $\bar{x}_i \leftarrow \text{ROOT}(GNN(P_i))$;
- 4: Create node (news) representations for the hypergraph
 $v_i^0 \leftarrow f(\sigma(x_i \oplus \bar{x}_i))$;
- 5: **end for**
- 6: **for** $l = 1 \dots L$ **do**
- 7: **for** $e_j \in \mathcal{E}$ **do**
- 8: Calculate hidden state h_k ;
- 9: Calculate weight coefficient α_{jk} ;
- 10: Uptime the hyperedge representation e_j^l
 $e_j^l \leftarrow \sigma\left(\sum_{v_k \in e_j} \alpha_{jk} W_1 v_k^{l-1}\right)$;
- 11: **end for**
- 12: **for** $v_i \in \mathcal{V}$ **do**
- 13: Calculate hidden state r_k ;
- 14: Calculate weight coefficient β_{ij} ;
- 15: Uptime the node (news) representation v_i^l
 $v_i^l \leftarrow \sigma\left(\sum_{e_j \in \mathcal{E}_i} \beta_{ij} W_2 e_j^l\right)$;
- 16: **end for**
- 17: **end for**
- 18: **return** $\{v_i^L\}_{i=1}^N, \alpha, \beta$.

V. EXPERIMENTS

In this section, we aim to answer the following questions:

- **RQ1:** Can our proposed approach using hypergraph neural network improve fake news detection performance?
- **RQ2:** Does the proposed approach have a greater effect when only a limited number of training labels is given?
- **RQ3:** Is it more effective to use the hypergraph to represent news relations over a normal graph?
- **RQ4:** What types and combinations of hyperedges contribute most to performance of fake news detection?
- **RQ5:** Would the proposed method be useful for detecting users who tend to spread fake and true news?

1) *Infrastructures and Implementations:* All experiments are performed on a Linux machine with a GeForce RTX 6000 GPU (24 GB of VRAM), 64 GB of system memory, and Intel i7 @ 3.60GHz processors. Our models are implemented using PyTorch 1.9.0 [43] and PytorchGeometric 2.0.2 [44]

2) *Datasets:* We use FakeNewsNet [23] which contains news from two fact-checking platforms (Politifact, Gossipcop) with user's tweet and retweet associated to each news piece. However, since many Twitter accounts in FakeNewsNet are no longer accessible, we use the news and user information collected in UPFD [10]. In addition, UPFD provides news and user information in a form of pre-computed features with two different options: (1) news content, user's recent 200 tweets, and user's comments, which are encoded either in BERT [45] or word2vec [41]; and (2) user's 10 profile attributes such as the number of followers, timestamp, status verification, and so on. The full list of 10 profile features can be found in [11]. UPFD randomly splits data to train-val-test (20%-10%-70%).

TABLE II
STATISTICS OF NEWS PROPAGATION TREES IN UPFD [10].

Dataset	# Graph	# Fake	# True	# Node	# Edge
Politifact	314	157	157	41,054	40,740
Gossipcop	5,464	2,732	2,732	314,262	308,798

3) *Baselines*: We train the baselines with 10 different random seeds and the news samples provided in UPFD. Then, we classify the test data using the parameters of the model that has shown the best accuracy for validation. We evaluate the following representative baselines in fake news detection based on (1) a propagation tree [6], [12] [10], [11], [25], [46], (2) a heterogeneous graph [7], and (3) a hypergraph [35] for a comprehensive evaluation. The features for news and users are based on the implementation of each baseline as follows:

- **GNN-CL** [11]: We use DiffPool [47] to model structural information on a propagation tree based on user’s profile attributes associated to each tweet/retweet. Due to the difference in the problem setting, training GNN with continuous learning is not considered in this study.
- **BiGCN** [12]: It uses two GCN [22] modules in both top-down and bottom-up directions to capture propagation and dispersion patterns of news on a social network. User profile attributes are used to represent each tweet/retweet.
- **TGNF** [25]: We use TGAT [48] to model temporal evolving pattern of a propagation tree by aggregating neighbors of tweet/retweet shared in proximity in time. It uses BERT to obtain the word vectors for tweet/retweet.
- **GTN** [24]: We use Graph Transformer Network [46] to encode a propagation tree, where the weight of an edge is determined by the time difference between tweet and retweet. Each tweet/retweet is represented by the corresponding user’s recent 200 tweets, encoded by BERT.
- **HGAT** [7]: We use a heterogeneous graph with two types of nodes (user and news) by linking them together if a user shares news. Based on two level aggregations with semantic and node level attention mechanism, it learns the relations between user and news nodes. We use BERT to encode each user’s 200 recent tweets and news contents.
- **HGNN** [35]: It is a transductive node classification model using a hypergraph in a semi-supervised setting. It constructs the hypergraph based on the top 10 closest nodes according to euclidean distance for each node. Each node in the hypergraph is news content encoded by word2vec.
- **GCNFN** [6]: It uses two-layer GCN [22] to model a propagation tree with the fusion of heterogeneous data such as user’s profile, user’s comments, and news contents which are encoded by word2vec.
- **UPFD** [10]: It combines various signals such as news text features and user preference features by jointly modeling news content and a propagation tree using a skip-connection. We experiment the model with three variants using GCN [22], GAT [32], and SAGE [31] as the GNN module. We use BERT to represent news content and the user preference based on the user’s recent 200 tweets.

TABLE III

THE RESULTS INDICATE MEAN \pm STANDARD DEVIATION. OUR PROPOSED MODEL OUTPERFORMS ALL THE BASELINE WHILE MEETING STATISTICAL SIGNIFICANCE UNDER T-TESTS ($P < 0.05$). OOM DENOTES OUT OF MEMORY, AND THE BEST PERFORMANCE IS HIGHLIGHTED IN BOLDFACE.

Model	Politifact		Gossipcop	
	Accuracy	F1-score	Accuracy	F1-score
GNN-CL	65.79 \pm 8.96	65.02 \pm 9.46	94.98 \pm 0.80	94.94 \pm 0.80
BiGCN	74.16 \pm 3.57	74.16 \pm 3.57	88.04 \pm 0.48	87.95 \pm 0.49
TGNF	74.28 \pm 1.74	74.09 \pm 1.81	85.07 \pm 0.08	85.07 \pm 0.08
GTN	81.67 \pm 4.16	81.53 \pm 4.35	92.41 \pm 0.98	92.38 \pm 0.98
HGAT	81.53 \pm 1.16	80.47 \pm 1.75	OOM	OOM
HGNN	79.96 \pm 4.89	79.28 \pm 5.16	93.38 \pm 0.49	93.38 \pm 0.49
GCNFN	80.63 \pm 4.23	80.31 \pm 4.57	95.37 \pm 0.21	95.33 \pm 0.21
UPFD-GCN	80.27 \pm 4.35	80.16 \pm 4.41	95.55 \pm 0.63	95.51 \pm 0.64
UPFD-GAT	79.09 \pm 3.73	78.95 \pm 3.79	96.03 \pm 0.62	96.00 \pm 0.62
UPFD-SAGE	80.40 \pm 4.22	80.13 \pm 4.65	96.38 \pm 0.48	96.36 \pm 0.48
HGFND	91.11\pm1.89	91.11\pm1.89	97.46\pm0.30	97.46\pm0.30

4) *Hyperparameters*: We use unified size of the hidden dimension (128), batch size (128), optimizer (Adam), learning rate (0.001), dropout (0.3), and epoch (200) with early-stopping strategy. For the other parameters in the baselines, we apply the parameters suggested in each baseline. We use the pre-computed features of news and user’s recent 200 tweet/retweet encoded by BERT [45] for the proposed model, where the initial feature dimension is 768, and we fix the number of layers in the hypergraph attention network to 2.

A. RQ1: Performance on Fake News Detection

Table III shows the performance of all compared methods. HGFND achieves the best performance across all datasets w.r.t. both accuracy and F1-score. Based on the evaluation, we make three primary observations as follows:

First, in contrast to propagation-based methods, HGFND constantly shows better performance. While variations of UPFD show comparable performance in Gossipcop, they show much lower performance in Politifact compared to HGFND. Considering that the size of Gossipcop is ~ 17 times larger than Politifact, the difference in performances on two datasets shows the robustness of HGFND to available training labels.

Second, in contrast to the heterogeneous graph-based method, HGFND has advantages in terms of space usage and expressive power in capturing relational information. HGAT runs out of memory since it includes all users in the dataset as nodes and connects them to a news node, while HGFND includes users when they share more than one news. HGAT is also limited in capturing the relational information between news nodes, as its attention mechanisms only aggregates the local neighbors (i.e., user nodes) of a single news node.

Third, in comparison with the hypergraph-based method, HGFND shows greater performance than HGNN by $\sim 12\%$ on Politifact and $\sim 4\%$ on Gossipcop w.r.t both accuracy and F1-score. This result proves that relying on the semantic similarity between nodes can involve noisy and task-irrelevant information to fake news detection and finding meaningful relation for news is important for the quality of the hypergraph.

TABLE IV
TEST ACCURACY BY VARYING THE PERCENTAGE OF AVAILABLE LABELS FOR TRAINING DATA WHERE OOM DENOTES OUT OF MEMORY.

Model	Politifact (small dataset)				Gossipcop (large dataset)					
	100%	75%	50%	25%	100%	75%	50%	25%	5%	2.5%
HGFND	91.13±1.8	90.89±2.6	87.91±3.0	85.61±6.0	97.46±0.3	97.32±0.3	97.13±0.4	97.05±0.3	96.69±0.4	96.45±0.4
UPFD-SAGE	80.40±4.2	77.62±2.8	75.87±2.7	68.24±5.4	96.38±0.4	96.26±0.6	95.87±0.4	94.65±0.4	92.42±1.4	87.81±3.1
UPFD-GAT	79.09±3.7	77.24±3.2	75.51±4.0	69.29±6.1	96.03±0.6	95.46±0.6	95.05±0.4	93.46±0.2	90.21±2.9	84.99±3.4
UPFD-GCN	80.21±4.3	76.86±5.3	76.70±5.3	70.25±4.0	95.55±0.6	95.38±0.4	92.93±3.1	89.22±3.2	85.77±2.3	83.14±3.6
GCNFN	80.63±4.2	78.65±6.1	77.16±6.2	71.53±8.1	95.37±0.2	95.11±0.3	94.65±0.6	91.19±1.7	89.75±2.8	88.82±3.2
HGNN	79.96±4.8	79.45±4.2	78.69±4.6	75.71±4.5	93.38±0.4	93.34±0.4	93.21±0.4	93.11±0.4	92.56±0.5	91.37±1.6
HGAT	81.53±1.1	80.00±2.0	77.51±3.6	69.04±5.7	OOM	OOM	80.95±7.6	75.73±6.0	68.06±5.2	60.12±3.9
GTN	81.67±4.1	78.85±3.4	69.21±5.4	64.28±5.0	92.41±0.9	89.54±2.3	87.38±3.4	83.81±3.9	81.34±4.4	75.82±6.8
TGNF	74.28±1.7	70.30±1.8	63.94±7.8	59.46±1.9	85.07±0.0	84.53±0.4	83.54±1.7	78.31±2.2	65.54±1.4	65.29±0.8
BiGCN	74.16±3.5	74.02±4.0	72.44±5.5	68.20±8.4	88.04±0.4	84.54±0.6	83.99±0.7	82.13±0.9	79.54±0.6	78.83±1.2
GNN-CL	65.79±8.9	64.23±6.0	57.34±9.0	56.75±5.9	94.98±0.8	93.73±0.4	91.98±0.8	91.62±1.5	84.42±1.8	80.36±2.8

B. RQ2: Is HGFND more effective with limited train labels?

To answer RQ2, we train and evaluate the models with a subset of training labels. To this end, we vary the percentage of training labels according to the two different sizes of the datasets; Gossipcop contains more labeled data than Politifact. When randomly sampling training data, we remove the unselected training data from the hypergraph, and also remove the corresponding propagation tree from the dataset.

As shown in Table IV, we find consistent improvements over the baselines regarding accuracy under both limited and sufficient data conditions. For a clear comparison, we compare the highest and lowest accuracy of the proposed model per dataset. In Politifact, HGFND’s accuracy decreases by $\sim 6\%$ ($91.13 \rightarrow 85.61$). In Gossipcop, HGFND’s accuracy decreases by $\sim 1\%$ ($97.46 \rightarrow 96.45$). While other baselines suffer a significant decrement, HGFND shows robust performance to the limited number of training labels.

Notably, another hypergraph-based method, HGNN [35], also shows robust performance to the size of training labels than the other baselines. HGNN’s accuracy decreases by $\sim 5\%$ ($79.96 \rightarrow 75.71$) in Politifact, and $\sim 2\%$ ($93.38 \rightarrow 91.37$) in Gossipcop. This experimental result supports that leveraging a hypergraph can help to better utilize training labels by learning and preserving the group-wise interaction among news.

C. RQ3: Advantage of a Hypergraph over a Normal Graph

To answer RQ3, we compare the performance between a series of GNN models, and the proposed methods based on the hypergraph constructed in Section IV-A. As it is incompatible to directly apply GNNs to a hypergraph, we first convert the hypergraph into graph structure using clique expansion [40], which replaces hyperedges with cliques to obtain a graph. In this experiment, GNNs define node features as news content encoded by BERT, and classify the unlabeled nodes in a semi-supervised setting for transductive node classification. As compared in Table V, GCN [22], GAT [32], and SAGE [31] constantly show worse performance than HGFND across all datasets w.r.t both accuracy and F1-score. The performance gap can be mainly attributed to undesired information loss that occurs when clique expansion converts group-wise interactions in the hypergraph to pairwise relations [49].

TABLE V

COMPARISON OF PERFORMANCE BETWEEN THE PROPOSED MODEL AND A SERIES OF GRAPH NEURAL NETWORKS BASED ON CLIQUE EXPANSION.

Model	Politifact		Gossipcop	
	Accuracy	F1-score	Accuracy	F1-score
HGFND	91.13±1.89	91.11±1.89	97.46±0.30	97.46±0.30
SAGE [31]	80.84±3.55	80.80±3.59	94.79±0.38	94.79±0.38
GAT [32]	80.09±4.49	79.98±4.62	94.75±0.37	94.75±0.37
GCN [22]	79.90±4.07	79.77±4.20	94.69±0.41	94.68±0.42

TABLE VI

EVALUATION OF THE PROPOSED MODEL (HGFND) ON COMBINATION OF DIFFERENT TYPES OF HYPEREDGES. U, T, AND E SYMBOLS INDICATE THE HYPEREDGE TYPES OF USER, TIME, AND ENTITY RESPECTIVELY.

Type	Politifact		Gossipcop	
	Accuracy	F1-score	Accuracy	F1-score
U T E				
✓ ✓ ✓	91.13±1.89	91.11±1.89	97.46±0.30	97.46±0.30
✓ ✓ -	89.72±3.11	89.70±3.14	97.38±0.25	97.38±0.25
✓ - ✓	89.90±2.52	89.89±2.52	97.41±0.28	97.41±0.28
- ✓ ✓	85.24±4.18	85.13±4.30	90.01±2.57	90.00±2.58
✓ - -	87.23±2.32	87.17±2.34	97.29±0.16	97.29±0.16
- ✓ -	83.61±3.28	83.49±3.38	72.15±3.26	71.85±3.39
- - ✓	76.28±3.09	76.15±3.10	84.46±0.20	84.42±0.20

D. RQ4: Contribution of Hyperedge Types on Performance

To answer RQ4, we conduct an ablation study on HGFND with combinations of the three different types of hyperedges (*User*, *Time*, and *Entity*). As shown in Table VI, combining all types of hyperedges yields the best performance. Each hyperedge type is complementary to the other, as performance always increases when adding more types of hyperedge. When comparing the contribution of each hyperedge type, *User* hyperedge shows the highest contribution to the performance on both datasets in every combination with other hyperedge types. *Time* and *Entity* hyperedges appear to be the second most contributing type of hyperedges, but the contribution is different for each dataset.

Overall, the experimental results support that users play a key role in detecting fake news. Based on such observation, we ask the following question, “How does users’ news sharing behavior help to detect fake news?”. We answer it next.

TABLE VII

STATISTICS OF THE PROPOSED NEWS HYPERGRAPH IN SECTION IV-A. NODE DEGREE IS THE NUMBER OF HYPEREDGES CONNECTED TO A NODE. HYPEREDGE SIZE IS THE NUMBER OF NODES INCLUDED IN A HYPEREDGE.

Measure	Politifact			Gossipcop		
	U	T	E	U	T	E
# hyperedges	2,953	1,717	1,040	15,646	14,586	12,673
Avg. hyperedge size	2.56	3.52	4.88	13.19	6.23	6.84
Max. hyperedge size	21	13	80	2,426	25	992
Avg. node degree	25.59	19.35	17.94	37.94	16.64	16.31
Max. node degree	153	186	147	156	124	471

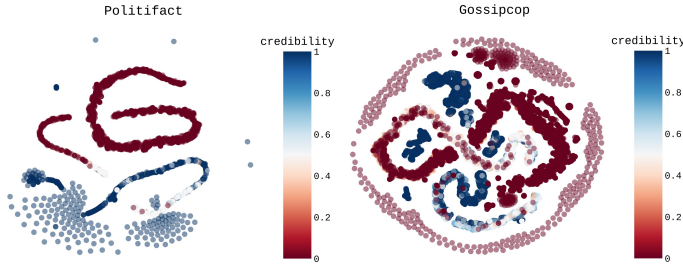


Fig. 4. t-SNE plot on the representations of the *User* hyperedge in 2D space. The credibility score (the higher, the credible) is shown by color and value.

E. RQ5: Interpreting Users' News Sharing Behavior

To answer RQ5, we examine whether HGFND can learn to characterize users who tend to spread fake or true news. To this end, we measure the credibility of a user, defined as the percentage of shared true news, as follows:

$$credibility_i = \frac{\sum_{y_j \in \mathcal{Y}_i^{user}} y_j}{|\mathcal{E}_i^{user}|}, \text{ where } \mathcal{E}_i^{user} \subset \mathcal{E}^{user} \quad (8)$$

where $credibility_i$ is the credibility score of i -th user, \mathcal{E}^{user} is a set of *User* hyperedges, and $|\mathcal{E}_i^{user}|$ is the number of nodes connected to \mathcal{E}_i^{user} . Here, \mathcal{Y}_i^{user} is a set of labels for nodes connected to \mathcal{E}_i^{user} where the label is 0 if fake and 1 if true.

To visualize the representation of \mathcal{E}^{user} , we first apply t-SNE [50] and calculate the credibility score assigned to users. As shown in Fig. 4, HGFND separates the *User* hyperedges into two distinctive groups with high and low credibility scores, respectively.

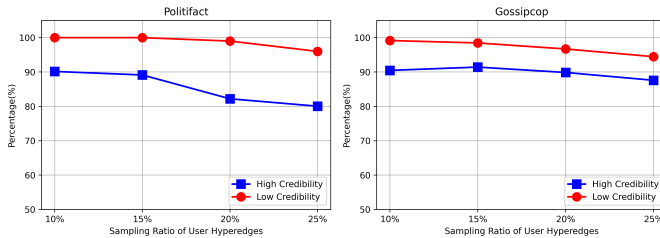


Fig. 5. The percentage of users with high and low credibility that are sampled from *User* hyperedges. Users are selected based on the sampling ratio (y-axis) according to the top/bottom attention scores of the hyperedges.

We further answer whether the attention scores assigned to *User* hyperedges can help to identify users who tend to spread true or fake news. To this end, we follow two sampling steps: (1) sampling users assigned with high and low attention scores, respectively, and (2) checking how many users are classified to have high or low credibility. Here, a user's credibility is high if the credibility score exceeds 0.9, and low if the credibility score is less than 0.1. For example, we sample the top 5% of users with the highest attention scores and check if their credibility scores are above or below the thresholds.

Fig. 5 shows the percentage of users with high and low credibility from selected users with varying sampling ratio (10%, 15%, 20%, and 25%) based on the top/bottom attention scores. For both datasets, the percentage of users with high credibility lies between 80% and 95%, and the percentage of users with low credibility is between 95% and 100%. Based on the results, we show that the proposed model can implicitly learn the credibility of users based on the attention mechanism. This finding opens a new avenue to interpret the model's decision through the lens of users' news sharing behavior.

VI. CONCLUSION

In this study, we propose a novel hypergraph-based neural network solving the problem of relational fake news detection with limited labeled news data. Our approach can better utilize the relational information between labeled and unlabeled data by capturing group-wise interactions among news pieces. Based on two real-world benchmark datasets for fake news detection, we show the superiority of the proposed model over the state-of-the-art methods on both normal and limited data settings. Experimental results demonstrate that focusing on useful subsets of the relations is important, and identifying susceptible users plays a key role in fake news detection.

VII. LIMITATIONS & FUTURE WORK

The results here are not without limitations. Some public datasets [51]–[53] are not considered in this study as they are not in line with the definition of fake news, which is an intentional and verifiable false news published by a news outlet. Furthermore, the type information of each hyperedge is not considered in the framework, which can potentially misguide the inference when more types of hyperedges are added. Additionally, we round timestamp for tweet/retweet differently for the two datasets (i.e. Politifact in days and Gossipcop in hours). The rounding method is empirically opted according to the criteria that yields the best performance.

Future work will focus on finding more types of news relations, such as news sharing similar event schema, exploitative narratives, and geolocation. Based on these comprehensive news relations, we will employ a heterogeneous hypergraph to better reflect different contribution per the type of hyperedge. Lastly, we aim to dynamically learn and exploit the structure of the hypergraph by merging and splitting existing hyperedges. The structure information will help to detect fake news when only a few news relations are known.

VIII. REPRODUCIBILITY & SOCIAL IMPACT

Existing fake news-related studies show difficulties in benchmarking their datasets, as user information is in many times not publicly available due to data sharing policies. To prevent this problem, we use pre-computed features for news and users provided by UPFD [10]. Through this, the proposed method can reproduce the result at any time without additional data collection procedure. Furthermore, as news and user features are encoded, it is expected to be safer in terms of information security. The code and data are publicly available at https://github.com/ujeong1/IEEEBigdata22_HGFND.

We acknowledge that our work provides only preliminary solutions, and it is far from objectively discerning the authenticity of news in a real-world scenario due to the complexity and diversity of fake news. Furthermore, the result of the proposed approach can be biased due to the existing bias in the dataset or the algorithm. Nevertheless, with our work, we hope to open a potential research directions in the problem of relational fake news detection while broadening discussions in mitigating the spread of fake news on online social networks.

IX. ACKNOWLEDGEMENTS

This work was supported by the Office of Naval Research under Award No. N00014-21-1-4002. Opinions, interpretations, conclusions, and recommendations are those of the authors.

REFERENCES

- [1] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, 2018.
- [2] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises," *Review of general psychology*, 1998.
- [3] M. Cinelli, G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, "The echo chamber effect on social media," *National Academy of Sciences*, 2021.
- [4] K. Shu, G. Zheng, Y. Li, S. Mukherjee, A. H. Awadallah, S. Ruston, and H. Liu, "Early detection of fake news with multi-source weak social supervision," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD)*. Springer, 2020.
- [5] K. Shu, S. Wang, and H. Liu, "Beyond news contents: The role of social context for fake news detection," in *ACM International Conference on Web Search and Data Mining (WSDM)*, 2019.
- [6] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," *arXiv preprint*, 2019.
- [7] Y. Ren and J. Zhang, "Fake news detection on news-oriented heterogeneous information networks through hierarchical graph attention," in *International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021.
- [8] V.-H. Nguyen, K. Sugiyama, P. Nakov, and M.-Y. Kan, "Fang: Leveraging social context for fake news detection using graph representation," in *ACM Conference on Information and Knowledge Management (CIKM)*, 2020.
- [9] D. L. Bernardi, P. H. Cheong, C. Lundry, and S. W. Ruston, "Narrative landmines," in *Narrative Landmines*. Rutgers University Press, 2012.
- [10] Y. Dou, K. Shu, C. Xia, P. S. Yu, and L. Sun, "User preference-aware fake news detection," in *ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 2021.
- [11] Y. Han, S. Karunasekera, and C. Leckie, "Graph neural networks with continual learning for fake news detection from social media," *arXiv preprint*, 2020.
- [12] T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, and J. Huang, "Rumor detection on social media with bi-directional graph convolutional networks," in *AAAI conference on artificial intelligence (AAAI)*, 2020.
- [13] X. Zhou, K. Shu, V. V. Phoha, H. Liu, and R. Zafarani, "'this is fake! shared it by mistake': Assessing the intent of fake news spreaders," in *ACM Web Conference (WWW)*, 2022.
- [14] A. Hanselowski, P. Avinesh, B. Schiller, F. Caspelherr, D. Chaudhuri, C. M. Meyer, and I. Gurevych, "A retrospective analysis of the fake news challenge stance-detection task," in *International Conference on Computational Linguistics (COLING)*, 2018.
- [15] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, "Fever: a large-scale dataset for fact extraction and verification," in *Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2018.
- [16] B. Riedel, I. Augenstein, G. P. Spithourakis, and S. Riedel, "A simple but tough-to-beat baseline for the fake news challenge stance detection task," *arXiv preprint*, 2017.
- [17] R. K. Kaliyar, A. Goswami, and P. Narang, "Fakebert: Fake news detection in social media with a bert-based deep learning approach," *Multimedia tools and applications*, 2021.
- [18] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in *ACM Conference on Information and Knowledge Management (CIKM)*, 2017.
- [19] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "defend: Explainable fake news detection," in *ACM SIGKDD international conference on Knowledge Discovery & Data Mining (SIGKDD)*, 2019.
- [20] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2016.
- [21] L. Cheng, R. Guo, K. Shu, and H. Liu, "Causal understanding of fake news dissemination on social media," in *ACM SIGKDD Conference on Knowledge Discovery & Data Mining (SIGKDD)*, 2021.
- [22] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *international conference on Learning Representations (ICLR)*, 2016.
- [23] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media," *Big data*, 2020.
- [24] H. Matsumoto, S. Yoshida, and M. Muneyasu, "Propagation-based fake news detection using graph neural networks with transformer," in *IEEE Conference on Consumer Electronics (GCCE)*. IEEE, 2021.
- [25] C. Song, K. Shu, and B. Wu, "Temporally evolving graph neural network for fake news detection," *Information Processing & Management*, 2021.
- [26] A. Silva, Y. Han, L. Luo, S. Karunasekera, and C. Leckie, "Propagation2vec: Embedding partial propagation networks for explainable fake news early detection," *Information Processing & Management*, 2021.
- [27] S. Muhuri and D. Mukhopadhyay, "A hypergraph clustering-based technique for detecting fake news from broadcasting network," in *Asian Conference on Innovation in Technology (ASIANCON)*. IEEE, 2021.
- [28] A. Borse, D. Kharate *et al.*, "Fake news prediction using hierarchical attention network and hypergraph," *International Conference on Contents, Computing Communication (ICCCC)*, 2022.
- [29] K. Ding, J. Wang, J. Li, K. Shu, C. Liu, and H. Liu, "Graph prototypical networks for few-shot learning on attributed networks," in *ACM Conference on Information and Knowledge Management (CIKM)*, 2020.
- [30] K. Ding, Z. Xu, H. Tong, and H. Liu, "Data augmentation for deep graph learning: A survey," *arXiv preprint*, 2022.
- [31] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems (NeurIPS)*, 2017.
- [32] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint*, 2017.
- [33] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, "Heterogeneous graph attention network," in *ACM Web Conference (WWW)*, 2019.
- [34] S. Bai, F. Zhang, and P. H. Torr, "Hypergraph convolution and hypergraph attention," *Pattern Recognition*, 2021.
- [35] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, "Hypergraph neural networks," in *AAAI conference on artificial intelligence (AAAI)*, 2019.
- [36] K. Ding, J. Wang, J. Li, D. Li, and H. Liu, "Be more with less: Hypergraph attention networks for inductive text classification," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.
- [37] J. Wang, K. Ding, L. Hong, H. Liu, and J. Caverlee, "Next-item recommendation with sequential hypergraphs," in *ACM SIGIR conference on research and development in Information Retrieval (SIGIR)*, 2020.

- [38] J. Wang, K. Ding, Z. Zhu, and J. Caverlee, "Session-based recommendation with hypergraph attention networks," in *SIAM International Conference on Data Mining (SDM)*. SIAM, 2021.
- [39] U. Jeong, Z. Alghamdi, K. Ding, L. Cheng, B. Li, and H. Liu, "Classifying covid-19 related meta ads using discourse representation through a hypergraph," in *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (SBP-BRIMS)*. Springer, 2022.
- [40] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," *Advances in Neural Information Processing Systems (NeurIPS)*, 2006.
- [41] M. Honnibal and I. Montani, "spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing," 2017.
- [42] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," *Advances in neural information processing systems (NeurIPS)*, 2003.
- [43] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshain, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- [44] M. Fey and J. E. Lenssen, "Fast graph representation learning with pytorch geometric," *arXiv preprint*, 2019.
- [45] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2018.
- [46] Y. Shi, Z. Huang, S. Feng, H. Zhong, W. Wang, and Y. Sun, "Masked label prediction: Unified message passing model for semi-supervised classification," *International Joint Conference on Artificial Intelligence (IJCAI)*, 2020.
- [47] Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," *Advances in neural information processing systems (NeurIPS)*, 2018.
- [48] D. Xu, C. Ruan, E. Korpeoglu, S. Kumar, and K. Achan, "Inductive representation learning on temporal graphs," *International Conference on Learning Representations (ICLR)*, 2020.
- [49] M. Hein, S. Setzer, L. Jost, and S. S. Rangapuram, "The total variation on hypergraphs-learning on hypergraphs revisited," *Advances in Neural Information Processing Systems (NeurIPS)*, 2013.
- [50] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, 2011.
- [51] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks," *International Joint Conference on Artificial Intelligence (IJCAI)*, 2016.
- [52] J. Ma, W. Gao, and K.-F. Wong, "Detect rumors in microblog posts using propagation structure via kernel learning." Association for Computational Linguistics (ACL), 2017.
- [53] E. Kochkina, M. Liakata, and I. Augenstein, "Turing at semeval-2017 task 8: Sequential approach to rumour stance classification with branch-lstm," *International Workshop on Semantic Evaluation (SemEval)*, 2017.