Incorporating Relational Knowledge in Explainable Fake News Detection

Kun Wu,[1] Xu Yuan,[2] Yue Ning[1]

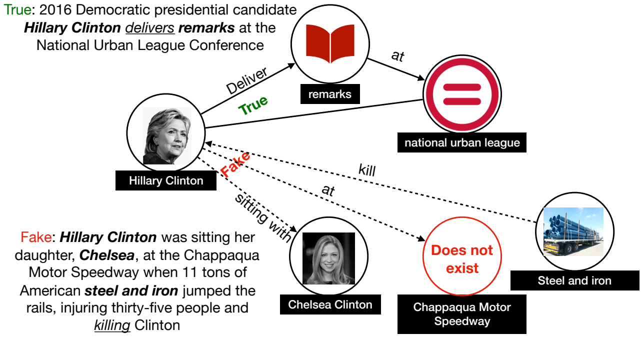
[1] Stevens Institute of Technology, Hoboken, NJ, USA
[2] University of Louisiana at Lafayette, Lafayette, LA, USA
kwu14, yue.ning@stevens.edu, xu.yuan@louisiana.edu

Abstract. The greater public has become aware of the rising prevalence of untrustworthy information in online media. Extensive adaptive detection methods have been proposed for mitigating the adverse effect of fake news. Computational methods for detecting fake news based on the news content have several limitations, such as: 1) Encoding semantics from original texts is limited to the structure of the language in the text, making both bag-of-words and embedding-based features deceptive in the representation of a fake news, and 2) Explainable methods often neglect relational contexts in fake news detection. In this paper, we design a knowledge graph enhanced framework for effectively detecting fake news while providing relational explanation. We first build a credential-based multi-relation knowledge graph by extracting entity relation tuples from our training data and then apply a compositional graph convolutional network to learn the node and relation embeddings accordingly. The pre-trained graph embeddings are then incorporated into a graph convolutional network for fake news detection. Through extensive experiments on three real-world datasets, we demonstrate the proposed knowledge graph enhanced framework has significant improvement in terms of fake news detection as well as structured explainability.

Keywords: Fake News Detection · Knowledge Graphs · Explainable Machine Learning.

1 Introduction

Misinformation in online media has become a menace, from being a public concern [12, 7] to causing major financial loss and security risks. Existing work on content-based fake news detection focuses on semantic content using statistical or deep learning models [22] while neglecting rich relational information among entities (names, organizations, etc). In this paper, we propose to investigate a self-discovered knowledge graph method to enhance the representation learning of entities and relations in fake news detection. While knowledge-based fact checking approaches [15, 28, 4] have been studied, they often suffer from issues such as reliability or incompleteness of web knowledge. In contrast to these previous approaches, we extract a credential based multi-relation knowledge graph


True: 2016 Democratic presidential candidate ***Hillary Clinton*** *delivers* ***remarks*** at the National Urban League Conference

Fake: ***Hillary Clinton*** was sitting her daughter, ***Chelsea***, at the Chappaqua Motor Speedway when 11 tons of American ***steel and iron*** jumped the rails, injuring thirty-five people and *killing* Clinton

Fig. 1: An example of a knowledge graph extracted from news articles.

from our corpus without external domain knowledge. We do not involve external data considering 1) accessibility: extracting knowledge from given news corpora is more flexible and scalable compared to external knowledge base; 2) dynamic context: knowledge is dynamic and updates over time. Keeping external data up to date requires excessive human labor efforts; and 3) relevance: external knowledge often contain global noise rather than useful information.

In this work, we follow the broad definition of fake news [31] as "false news" where news includes false information related to public figures and organizations in articles, statements, and speeches . The veracity of news articles can be discovered from multiple aspects such as writing styles, languages, and focused stories. From the presented example in Figure 1, we observe one important factor that distinguishes fake news from real news is the involved entities and their relations. Given the significant roles of entities and their relations in news content, we create a knowledge graph of entities and relations that appear in existing data to represent a few aspects of structured knowledge: credentials, relations, and contexts. In which, each node represents an entity (e.g., persons, organizations, locations) and each edge/relation between a pair of nodes indicates the action (e.g., predicate) among them.

However, several challenges are encountered in learning KG-based representations. First, multiple relations may exist between pairs of entities when two entities appear in different news articles. For instance, [Obama, approves, nuclear deals] (*fake*) and [Obama, plans, nuclear policy changes] (*real*) are two relations between the same pair of entities from different contexts. Thus, relational information embeds credential values while most existing works ignore these structured knowledge. Second, relations between entities can be complex and changing over time. Figure 1 shows an example in the PolitiFact dataset [23], where same entities appear in both fake news and true news. However, the relation between entity "Hilary Clinton" and entity "Steel and iron" is not trustworthy given that they only appear in fake news. Third, integrating and fusing heterogeneous information is challenging. Relational representations into semantic encoding has been proven effective in several natural language processing tasks [30]. However, discovering relational indicators for fake news remains open.

To address aforementioned challenges, we summarize the main contributions of this work as below:

- We build a credential-based multi-relation knowledge graph from existing fake news corpora. Each link between a pair of entities indicates the relation/action from the source entity to the target entity.
- We apply a compositional graph convolutional network to pre-train relational representations of entities and relations simultaneously from the discovered knowledge graph. Thus, the representations of new entities or relations can be inferred and updated.
- We design a new framework to enhance semantic embeddings with structured knowledge in order to predict the trustworthiness of a news article (fake or real). The knowledge embeddings include both relations and entities information. In addition, the proposed framework is able to provide explainable relational evidence for predictions of fake news.

2 Related Work

This section reviews the state-of-the-art methods in the context of fake new detection and knowledge graph learning, and discuss the advantage of our work over them.

Content-based Fake News Detection. Content-based solutions have attracted wide attention, which mainly focus on extracting the semantics or writing styles of the news articles [9, 27, 22]. For example, an attention-based deep learning approach (i.e., dEFEND [22]) was proposed for jointly capturing explainable top-k sentences and user comments for fake news detection using a sentence-comment co-attention sub-network. Wang *et al.* [27] presented an event adversarial network in multi-task learning to derive event-invariant features, which can benefit the detection of fake news on newly arrived events. They considered event types along with an adversarial network to better learn the representation of news. Additionally, the images in news articles are also encoded with a CNN model for combining the image features with text features. Levi *et al.* [13] designed a machine learning model using semantic and linguistic features to distinguish fake news from satire stories. Recently, Nguyen *et al.* [16] presented a Markov random field (MRF) model to study the correlation association among documents to assist fake news detection. The most recent approach [14] takes into account short texts (e.g, tweets, users' credits, and propagation patterns) in a Graph-aware Co-Attention Network (GCAN) where the representations of the corresponding source text, user features, and propagation graphs are learned first. Then, a dual co-attention model is developed for prediction. However, most of the content-based detection methods face a few limitations: 1) the leveraged auxiliary features (e.g. user comments, images, etc) are tailored to the specific domains, which thus cannot be scaled or generalized to a different domain; 2) explainability is limited: attention mechanisms focus on existing features (e.g., words), failing to capture relational dependencies.

Knowledge-based Fact Checking. Information retrieval methods based on knowledge have been proposed to determine the veracity of news articles. For instance, Magdy *et al.* [15] identified the trustworthiness of a claim by using query results from the web. Wu *et al.* [28] presented a method through "perturbing" a claim from querying knowledge bases and using the result variations as an indicator for fact checking. In addition, Ciampaglia *et al.* [4] considered the shortest path between concepts in a knowledge graph and [21] employed a link prediction algorithm with discriminative meta paths for fact checking. However, these approaches encounter problems of determining the trustworthiness and reliability of the external knowledge (web or knowledge base). In addition, they are deficient when the corresponding entries do not exist in a knowledge base or the knowledge base is compromised.

Knowledge Graphs (KG) that organize relations of entities in directed graphs are widely used in many fields, such as link prediction and question answering. KGs are constructed from triples, e.g., (head, relation, tail) or (subject, predicate, object), to provide rich and strong facts to enhance the understanding of natural languages [26]. Knowledge graph embedding focuses on learning hidden representations of nodes and/or relations. A few state-of-the-art approaches include: TransE[2], DistMult[29], and ConvE[6]. A recent development for multi-relation representation learning in KGs, CompGCN [25], jointly learns the embeddings of nodes and relations using a graph convolutional network. A knowledge graph based fake news detection [18] utilizes the ability of link prediction in KGs to detect fake news. It extracts triples from news and employs TransE to present entities and relations into a vector space. By measuring the distances between subjects combined with relations and objects extracted from news, they can predict the veracity of the news. However, this approach only considers the features of knowledge graphs, omitting global semantic features of news which also provide critical information for fake news detection.

To conclude, our approach will explore semantic content of news articles and enrich the semantic features with structural embeddings from knowledge graphs. Our developed KG embedding model can be compatible with other models and offer relation level explainability beyond keywords' contributions.

3 The Proposed Method

In this section, we present our design of a novel **k**nowledge **g**raph enhanced framework for **f**ake news detection, abbreviated as **KGF**, that can be applied in a variety of deep learning models to jointly predict if a news article is fake while providing explainable structured knowledge for the prediction. The overall framework is present in Figure 2. We first introduce the notations and the problem formulation and then we discuss the details of the proposed framework.

Notations and Problem Formulation. Given a collection of news articles $\mathcal{D}$, each one contains a sequence of words $\{w_1, ..., w_k, ..., w_W\}$ and its corresponding label $y \in \{0, 1\}$ indicating if the news is fake ($y = 1$) or not ($y = 0$). We extract a knowledge graph from this corpus and denote it by $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \mathcal{X}, \mathcal{Z})$ where $\mathcal{V} = \{v_1, v_2, ...v_{|\mathcal{V}|}\}$ denotes the set of vertices (i.e., entities) such as person names or locations. $\mathcal{E}$ denotes the edges between pairs of nodes where each entry $\mathcal{E}[i, j] = \{r_1, r_2, ...r_{|\mathcal{E}[i,j]|}\}$ is a set of relations between node i and node j given that entity i and j may appear in different contexts. $\mathcal{X} \in \mathbb{R}^{|\mathcal{V}| \times d_0}$ denotes the d_0-dimensional input features of each node. $\mathcal{Z} \in \mathbb{R}^{|\mathcal{R}| \times d_0^{'}}$ denotes the $d_0^{'}$-dimensional input features of each relation.

Knowledge Graph Extraction. We apply the Stanford NLP tool, OpenIE [1], to extract triples from sentences. Each triple (u, r, v) consists of a source entity u which is the subject in a clause, a target entity v which is an object in a clause, and a relation r between them. The subject and object entities are usually persons, places, organizations, or general nouns. The relation, also called predicate, is a directed action (e.g., verb) from a subject to an object. As such, we get a set of triples from each news article. During this process, we notice that the OpenIE tool generates some noisy triples. For instance, the triples extracted from the sentence: *"the American people must be able to trust that the American people government is looking out for all of us"* are: *('american people', 'must', 'must able')*. This kind of triples is noise we want to avoid. Hence, we investigate a few techniques to improve the quality of extracted entities and relations. First, We adopt a coreference resolution approach, NeuralCoref, to avoid the ambiguity of pronouns.[3] Next, we use Spacy to lemmatize the verbs in a relation given multiple tenses.[4] To reduce the number of duplicated relations, we remove the adverb in the predicate and only keep the lemmatized verb. As shown in the previous examples, we find that subject or object entities may be invalid. We assume that entities have to contain at least one noun. Therefore, we filter out all the entities that do not contain a noun.

Learning Relational Representations. After cleaning the triples, we organize all the entities into nodes and construct a multi-relational knowledge graph $\mathbf{G}$ based on all the triples in our training corpus. In this multi-relational graph, we assume each node and relation is encoded by an embedding vector. We adapt the compositional graph convolutional network (CompGCN) [25] to jointly embed both nodes and relations in a relational graph. Assuming node v is an object entity node, $N(v)$ is a set of its immediate neighbors for its incoming edges, and each edge corresponds to a specific relation, CompGCN updates the object node embedding vector as below:

$$\mathbf{h}_v^{(l+1)} = f\Big(\sum_{(u,r) \in N(v)} \mathbf{W}_q^{(l)} \phi\big(\mathbf{h}_u^{(l)}, \mathbf{o}_r^{(l)}\big)\Big) \in \mathbb{R}^{d_{l+1}}, \tag{1}$$

where $\phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is a composition operation [17] between the subject vector $\mathbf{h}_u^{(l)}$ and the relation vector $\mathbf{o}_r^{(l)}$. Layer-wise parameter matrix $\mathbf{W}_q^{(l)} \in \mathbb{R}^{d_{l+1} \times d_l}$ maps the dimension of hidden features from layer l to layer $l + 1$. The first layer embedding matrix $\mathbf{H}^{(0)}$ is initialized by $\mathcal{X}$. After the node embedding

[3] https://github.com/huggingface/neuralcoref

[4] https://spacy.io/

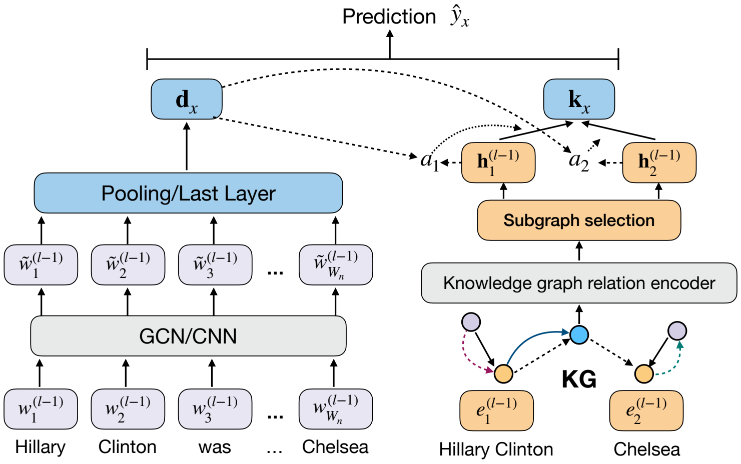


Fig. 2: The overall framework of knowledge graph enhanced fake news detection. The left module is a standard deep neural network on word embedding features. The right module learns the embeddings of entities and their relations using the self-discovered knowledge graph from the corpus. In the aggregation part, the semantic document embeddings learned from the left module are combined with the knowledge embeddings of the document.

update, the relation embeddings are also transformed as follows:

$$\mathbf{o}_r^{(l+1)} = \mathbf{W}_{\text{rel}}^{(l)} \mathbf{o}_r^{(l)} \in \mathbb{R}^{d_{l+1}}, \tag{2}$$

where $\mathbf{W}_{\text{rel}}$ is a learnable transformation matrix which projects edges to the same embedding space as nodes. The relation embedding matrix is initialized by $\mathbf{O}^{(0)} = \mathcal{Z}$.

Given a subject entity (u), a relation (r), and their anticipated object entity (v), we design a link prediction task as in ConvE [6] to estimate the embedding parameters. Both subject and relation embeddings are passed through a convolutional layer and several fully connected layers to get an estimated vector for the object (v). The score function s estimates the similarity between the estimated vector and the anticipated object entity. The loss function for the link prediction task is defined as below:

$$L_G = -\sum_{(u,r)} \sum_{e \in \mathcal{V}} \Big[y_e \log \sigma(s(u,r,e)) + (1 - y_e) \log(1 - \sigma(s(u,r,e))) \Big], \tag{3}$$

where e is a randomly sampled object entity. When e is equal to the ground truth object (v), $y_e = 1$. Otherwise, $y_e = 0$. $\mathcal{V}$ is the set of nodes. We adopt ConvE [6] as the score function $s(u,r,e) = s(\phi(\mathbf{h}_u, \mathbf{h}_r), \mathbf{h}_e)$ where ϕ denotes a composition operator for estimating the embedding vector of the object given a subject u and a relation r. All model parameters can be trained via back-propagation and optimized using the Adam algorithm.

Integration of Relational Knowledge in Detection After learning the embedding vectors of all triples in the knowledge graph, we incorporate the pre-trained embeddings in a global vector to improve prediction performance. For each node, its embedding vector is learned from its neighbors and their corresponding multi-type relations. Therefore, we utilize node embeddings instead of triple embeddings. Given a news article x, assuming there are N_x entity extracted from the article. The corresponding entity/node embedding are denoted as $N = \{n_0, n_1, ..., n_{N_x}\}$. We apply a global attention mechanism to capture the contributions of the nodes embeddings to the global semantic vector of an article. We first introduce how we learn the global vector $\mathbf{d}_x$ for each news article.

Learning Global Semantics. Assuming the word embedding matrix for a document x is represented by $\mathbf{E}_x \in \mathbb{R}^{W \times d}$, where W is the number of unique words in this document. We take advantage of a multi-layer Graph Convolutional Network to learn hidden representations of words given its effectiveness [11]. We build an adjacency matrix $\mathbf{A}$ to represent the context frequency between pairs of words. Following the setup in the Dynamic GCN model for event predictions [5], the edge weight between two words in a document is calculated as below:

$$\mathbf{A}[i,j] = \begin{cases} \mathrm{PMI}(i,j) & \mathrm{PMI}(i,j) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The PMI value of a word pair i, j is computed as $\mathrm{PMI}(i,j) = \log \frac{s(i,j)}{s(i)s(j)/S}$, where $s(i)$ and $s(j)$ are the total number of sentences in the document containing at least one occurrence of i and j, respectively. S is the total number of sentences in the document. The message passing process on the graph of words is denoted as $\mathbf{H} = f(\hat{\mathbf{A}}\mathbf{E}\mathbf{W}_g)$ where $\hat{\mathbf{A}}$ is the normalized symmetric adjacency matrix $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}}(\mathbf{A} + \mathbf{I}_W)\tilde{\mathbf{D}}^{-\frac{1}{2}}$. $\tilde{\mathbf{D}}$ is the degree matrix. $\mathbf{I}_W$ is an identity matrix with dimensions of W. Eventually the hidden features of words are updated by their neighboring word vectors. Assuming the final layer output is $\mathbf{H}^{(L)} \in \mathbb{R}^{W \times d_w}$, we adapt a pooling strategy to get the semantic embedding of the document: $\mathbf{d}_x = \mathrm{pooling}(\mathbf{H}_i) \in \mathbb{R}^{d_w}, i \in \{1, ..., W\}$.

Bridging Relational Knowledge with Global Semantics. We next learn relational representation of news based on the relational embeddings of entities. Assuming there are N_x entities in news x and each entity embedding $\mathbf{h}_i \in \mathbb{R}^{d_L} (i = 1, ...N_x)$ is learned from the previous CompGCN, the relational embedding vector of this news is calculated as $\mathbf{k}_x = \sum_i^{N_x} \alpha_i \mathbf{h}_i \in \mathbb{R}^{d_L}$ where α_i is the attention weight of each entity and is computed as follows:

$$\alpha_i = \frac{\exp(\mathbf{t}_i^\top \mathbf{d}_x)}{\sum_{n=0}^{N_x} \exp(\mathbf{t}_n^\top \mathbf{d}_x)}. \tag{5}$$

Here, $\mathbf{t}_i \in \mathbb{R}^{d_w}$ is the entity embedding after projected into the same vector space of semantic embedding by:

$$\mathbf{t}_i = \sigma(\mathbf{W}\mathbf{h_i} + \mathbf{b}) \in \mathbb{R}^{d_w}, \tag{6}$$

where σ is activation function, $\mathbf{W} \in \mathbb{R}^{d_w \times d_L}$ and $\mathbf{b} \in \mathbb{R}^{d_w}$ are trainable parameters, and $\mathbf{h}_i$ is the representation of entity i from the last layer of our multi-relational graph.

Optimization After obtaining the semantic embedding and the knowledge embedding, we apply a single layer MLP on the concatenation of these two vectors to predict the label of the document $\hat{y}_x = \sigma(\mathbf{w}^\top[\mathbf{k}_x \oplus \mathbf{d}_x])$. The ground truth labels of the news articles are binary. Thus, we adopt binary cross-entropy loss to optimize the model parameters:

$$L = -\sum_{x=1}^{D} \big(y_x \log \hat{y}_x + (1 - y_x) \log(1 - \hat{y}_x)\big), \tag{7}$$

where y is the ground truth and $\hat{y}$ is the model prediction. All model parameters can be trained via back-propagation and optimized using the Adam algorithm given its efficiency and ability to avoid overfitting.

4 Experiment Setup

In this section, we introduce the datasets, the baseline methods for comparison, and the evaluation metrics for measurement in our experiments.

Datasets. To fairly evaluate the performance of our model, we conduct the experiments on three datasets corresponding to different topics: 1) *Celebrity* dataset [19] was collected from web sources targeting rumors, hoaxes, and fake reports on celebrities. We sampled 250 fake news and 250 real news with 1670 relations, 19978 entities, and 31857 triples in total. 2) *PolitiFact* dataset [24, 23] was collected from "politifact.com" and most news are related to political campaigns. We sampled 474 real news and 369 fake news with 51918 entities, 3251 relations, and 91366 triples in total. 3) *GossipCop* dataset [24, 23] was collected from "E!Online (eonline.com)" and "GossipCop.com". We sampled 500 real news and 500 fake news with 43371 entities, 2438 relations, and 71842 triples in total.

Comparison Methods. We compare the proposed model with some common NLP models and several state-of-the-art fake news detection methods as baselines including: 1) logistic regression models with news style features by mapping the frequencies of rhetorical relations to a vector space (**RST**) [20]; 2) Recurrent neural networks (RNN) including **vanilla RNN**, Long Short-Term Memory (**LSTM**) [8], Gated Recurrent Units (**GRU**) [3]; 3) Text Convolutional Neural Networks (**Text-CNN**) [10]; 4) Graph based models such as Graph Convolutional Networks (**GCN**) [11], Compositional Graph Covluiontal Networks (**CompGCN**) [25] using pre-trained knowledge graph features; 5) Attention-based approaches such as **dEFEND**$^\diamondsuit$ [22], and 6) the Hierarchical Discourse-level Structure (**HDSF**) [9] model. We implement the **dEFEND**$^\diamondsuit$ model without news comments and use the source code of **HDSF** from its paper directly in our k-fold cross validation.

Hyperparameter Setup. In the pretraining model, the dimensions of the initial and output embeddings (for both nodes and relations) are 100 and 200, respectively. We use the combination of circular-correlation and ConvE as the operator during the training process. We introduce a 30% sparsity dropout into the ConvE layer and utilize the Adam method as the optimizer with the 0.001 learning rate. In the detection framework, we take advantage of Glove as the pretrained word embeddings with the dimension of 100. We use one layer GCN

Table 1: Performance Comparison of Fake News Prediction using Accuracy (Acc) and F1 score (%). Bold numbers are the best results and underline indicates the second best.

	Celebrity		**PolitiFact**		**GossipCop**	
	Acc	F1	Acc	F1	Acc	F1
LR+RST	54.2(±0.035)	54.7(±0.034)	57.8(±0.038)	49.3(±0.059)	53.4(±0.034)	51.6(±0.055)
RNN	53.0(±0.012)	57.1(±0.055)	68.6(±0.016)	68.1(±0.026)	63.9(±0.026)	63.1(±0.035)
LSTM	57.6(±0.047)	63.5(±0.080)	78.8(±0.024)	77.0(±0.025)	66.5(±0.045)	66.9(±0.035)
GRU	59.0(±0.081)	64.9(±0.050)	79.0(±0.027)	77.3(±0.038)	69.7(±0.025)	69.7(±0.039)
HDSF	50.0(±0.009)	66.7(±0.008)	50.4(±0.005)	66.8(±0.003)	50.7(±0.005)	67.1(±0.004)
dEFEND$^\diamond$	53.2(±0.041)	63.1(±0.056)	70.4(±0.053)	73.9(±0.039)	52.1(±0.025)	65.1(±0.025)
CompGCN	51.8(±0.053)	62.1(±0.050)	63.6(±0.035)	54.8(±0.083)	61.1(±0.067)	65.9(±0.036)
Text-CNN	64.4(±0.060)	65.0(±0.087)	77.5(±0.041)	75.3(±0.046)	69.9(±0.049)	68.4(±0.038)
GCN	62.0(±0.056)	69.1(±0.033)	79.9(±0.020)	76.7(±0.038)	65.4(±0.062)	70.0(±0.046)
KGF-CNN	68.4(±0.083)	71.7(±0.044)	81.6(±0.027)	81.1(±0.028)	71.2(±0.060)	70.8(±0.028)
KGF	**71.4**(±0.047)	**72.1**(±0.074)	**86.0**(±0.031)	**85.3**(±0.034)	**73.3**(±0.031)	**72.3**(±0.041)

model with 64 hidden units. Afterward, an average pooling layer on the output of GCN is applied to get a context vector of each news.

Evaluation. We apply 5-fold cross-validation on the datasets and compare our approach with the selected baseline methods. In each test set, we make sure the number of fake and real news are the same. Moreover, for each fold, we run all the models 10 times and average the results. To measure the performance of fake news detection, we utilize the commonly used evaluation metrics for classification problems: Accuracy and F1 score, given that our test sets are balanced over the two classes.

5 Results

Fake News Detection Performance. Table 1 exhibits the experimental results of **KGF** and other baseline methods on three datasets in terms of accuracy and F1 score. Overall, our approach outperforms all the baseline models.

When comparing to the Logistic Regression model with Rhetorical Structure Theory(RST) features, we observe our proposed **KGF** model can improve accuracy and F1 score both by 17% on Celebrity. For PolitiFact, the proposed model outperforms LR by 30% in Accuracy and by 35% in F1. CompGCN with only KG features achieves the inferior performance compared to other baselines. From which, we believe the semantic features learned from the original text provide rich information of contexts and backgrounds in detecting fake news. Our **KGF** can also beat both HDSF and dEFEND$^\diamond$, in both performance metrics. But notably, HDSF and dEFEND$^\diamond$ models do not perform well as the reported results in the original papers. For dEFEND$^\diamond$, we think there are three reasons: (1) we used different data sampling strategies; (2) we applied K-fold cross validation for averaged results; (3) our experiments do not consider the comments of the news. For HDSF, the datasets we used in this paper cover different varieties

Table 2: Examples of selected KG entities for a fake news prediction. Lime color denotes selected entities by our model and yellow color denotes the detected relations. Cyan denotes the keywords selected by dEFEND$^\diamondsuit$ attention scores.

Police Discover Meth Lab In Back Room of Alabama Walmart DECATUR, Alabama – Police were recently tipped off to a reported meth lab that was being run by Walmart employees in what they are calling one of the biggest busts in decades. Police Chief Robert Garner said that an anonymous tip was left on their drug hotline, expressing concern about a horrible burning smell that was coming from the back of the Decatur WalMart facility. When an officer was sent to investigate, the store was instantly shut down as he discovered a meth lab that took up the entire back room. "The thing was massive, and contained enough materials to make hundreds, if not thousands, of pounds of crystal meth," said Chief Garner. "Apparently, every employee in the store was a part of it, from working with and gathering materials, to cooking, to selling it outside of the store. It was a full, massive operation."

of topics, which differs from the original paper. The GCN model achieves the best performance among all the baseline models on Celebrity and GossipCop regarding F1 scores. However, our **KGF** model can still beat GCN in both accuracy and F1 scores across all datasets, due to the use of encoded KG embeddings.

Ablation Study. In order to investigate the effectiveness of our framework, we define a variant of **KGF**: **KGF**-CNN. We substitute GCN by Text-CNN to obtain the global semantic embedding and combine it with knowledge embedding. From Table 1, we can see **KGF** outperforms **KGF**-CNN and GCN. Meanwhile, **KGF**-CNN outperforms CNN. The results show the effectiveness of relational representations in detecting fake news.

Explainability Evaluation and Case Study. We select an example from the correctly predicted fake news in the test set of PolitiFact. In Table 2, we highlight the entities and relations which received high attention weights obtained by equation 5. Meanwhile, we highlight the keywords which received high word attention scores from dEFEND with a different color. In this example, we can see that the entities (e.g., "meth lab", "Back Room of Alabama Walmart DECATUR", "Police") with relations (e.g., "tripped", "discovered") chosen by our method are the essential components to the news. Since entities and relations represent facts that the news tries to express, our **KGF** provides the facts in the news that contribute most to the predictions. We can utilize the triples with the highest attention scores to provide explanations of why the news is classified as real or fake. It is worth mentioning that the facts represented by triples in our case can be real or fake.

Model Complexity. The computational complexity of pretraining mainly relies on the number of layers of GCN, i.e., K, the dimension of entity d, the total number of relations $|R|$, and the number of basis vectors $\mathcal{B}$. CompGCN uses the basis vectors $\{v_0, v_1, ..., v_{\mathcal{B}}\}$ to initialize the relation embeddings. Thus, the computational complexity of pretraining is $O(Kd^2 + \mathcal{B}d + \mathcal{B}|R|)$.

6 Conclusion

This paper proposed a new representation learning framework for explainable fake news detection using knowledge graph enhanced embeddings. Without external databases, we first extracted and organized a knowledge graph from accessible and reliable training corpora. Then we adapt a compositional graph neural network to pre-train structured features for entities and relations. Lastly, the pre-trained relational features are incorporated with semantic features for fake news recognition. The extensive experiments on two real-world datasets demonstrated the strengths of our proposed approach in fake news detection tasks, measured by standard classification evaluation metrics. We also exhibit case studies to provide structured explanations for the prediction results. In the future, we plan to investigate meta learning approaches to extract relations from text and examine other types of news including rumor and satire news.

Acknowledgements This work is supported in part by the US National Science Foundation under grants 1948432, 1763620 and 1948374.

Bibliography

[1] G. Angeli, M. J. Johnson Premkumar, and C. D. Manning. Leveraging linguistic structure for open domain information extraction. ACL, pages 344–354, July 2015.

[2] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. NeurIPS, pages 2787–2795. 2013.

[3] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. EMNLP, pages 1724–1734, Oct. 2014.

[4] G. L. Ciampaglia, P. Shiralkar, L. M. Rocha, J. Bollen, F. Menczer, and A. Flammini. Computational fact checking from knowledge networks. *PLOS ONE*, 10(6):1–13, 06 2015.

[5] S. Deng, H. Rangwala, and Y. Ning. Learning dynamic context graphs for predicting social events. KDD '19, pages 1007–1016, New York, NY, USA, 2019. ACM.

[6] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel. Convolutional 2d knowledge graph embeddings. In *Proceedings of the 32th AAAI Conference on Artificial Intelligence*, AAAI, pages 1811–1818, 2018.

[7] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer. Fake news on twitter during the 2016 u.s. presidential election. *Science*, 363(6425):374–378, 2019.

[8] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, Nov. 1997.

[9] H. Karimi and J. Tang. Learning hierarchical discourse-level structure for fake news detection. NAACL-HLT, pages 3432–3442, June 2019.

[10] Y. Kim. Convolutional neural networks for sentence classification. EMNLP, pages 1746–1751, Oct. 2014.

[11] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. ICLR, 2017.

[12] D. M. J. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, and J. L. Zittrain. The science of fake news. *Science*, 359(6380):1094–1096, 2018.

[13] O. Levi, P. Hosseini, M. Diab, and D. Broniatowski. Identifying nuances in fake news vs. satire: Using semantic and linguistic cues. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 31–35, Nov. 2019.

[14] Y.-J. Lu and C.-T. Li. Gcan: Graph-aware co-attention networks for explainable fake news detection on social media, 2020.

[15] A. Magdy and N. Wanas. Web-based statistical fact checking of textual documents. SMUC '10, page 103–110, 2010.

[16] D. M. Nguyen, T. H. Do, R. Calderbank, and N. Deligiannis. Fake news detection using deep Markov random fields. ACL-HLT '19, pages 1391–1400, June 2019.

[17] M. Nickel, L. Rosasco, and T. Poggio. Holographic embeddings of knowledge graphs. AAAI'16, page 1955–1961, 2016.

[18] J. Z. Pan, S. Pavlova, C. Li, N. Li, Y. Li, and J. Liu. Content based fake news detection using knowledge graphs. In *The Semantic Web – ISWC 2018*, pages 669–683, 2018.

[19] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea. Automatic detection of fake news. COLING 18, pages 3391–3401, Aug. 2018.

[20] V. Rubin, N. Conroy, and Y. Chen. Towards news verification: Deception detection methods for news discourse. 01 2015.

[21] B. Shi and T. Weninger. Fact checking in heterogeneous information networks. WWW '16 Companion, page 101–102, 2016.

[22] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu. defend: Explainable fake news detection. KDD '19, pages 395–405, 2019.

[23] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu. Fakenewsnet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media, 2018.

[24] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu. Fake news detection on social media: A data mining perspective. *SIGKDD Explor. Newsl.*, 19(1):22–36, Sept. 2017.

[25] S. Vashishth, S. Sanyal, V. Nitin, and P. Talukdar. Composition-based multi-relational graph convolutional networks. ICLR, 2020.

[26] Q. Wang, Z. Mao, B. Wang, and L. Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, 2017.

[27] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao. Eann: Event adversarial neural networks for multi-modal fake news detection. KDD 18, pages 849–857, 2018.

[28] Y. Wu, P. K. Agarwal, C. Li, J. Yang, and C. Yu. Toward computational fact-checking. *Proc. VLDB Endow.*, 7(7):589–600, Mar. 2014.

[29] B. Yang, W. tau Yih, X. He, J. Gao, and L. Deng. Embedding entities and relations for learning and inference in knowledge bases. *CoRR*, abs/1412.6575, 2015.

[30] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu. ERNIE: Enhanced language representation with informative entities. ACL, 2019.

[31] X. Zhou and R. Zafarani. Fake news: A survey of research, detection methods, and opportunities. *ArXiv*, abs/1812.00315, 2018.