DETERRENT: Knowledge Guided Graph Attention Network for Detecting Healthcare Misinformation

Limeng Cui, Haeseung Seo, Maryam Tabar, Fenglong Ma, Suhang Wang, and Dongwon Lee
The Pennsylvania State University, PA, USA
{lzc334,hxs378,mfg5544,fenglong,szw494,dongwon}@psu.edu

ABSTRACT

To provide accurate and explainable misinformation detection, it is often useful to take an auxiliary source (e.g., social context and knowledge base) into consideration. Existing methods use social contexts such as users' engagements as complementary information to improve detection performance and derive explanations. However, due to the lack of sufficient professional knowledge, users seldom respond to healthcare information, which makes these methods less applicable. In this work, to address these shortcomings, we propose a novel knowledge guided graph attention network for detecting health misinformation better. Our proposal, named as DETERRENT, leverages on the additional information from medical knowledge graph by propagating information along with the network, incorporates a *Medical Knowledge Graph* and an *Article-Entity Bipartite Graph*, and propagates the node embeddings through *Knowledge Paths*. In addition, an attention mechanism is applied to calculate the importance of entities to each article, and the knowledge guided article embeddings are used for misinformation detection. DETERRENT addresses the limitation on social contexts in the healthcare domain and is capable of providing useful explanations for the results of detection. Empirical validation using two real-world datasets demonstrated the effectiveness of DETERRENT. Comparing with the best results of eight competing methods, in terms of F1 Score, DETERRENT outperforms all methods by at least 4.78% on the diabetes dataset and 12.79% on cancer dataset.

KEYWORDS

Healthcare misinformation, graph neural networks, medical knowledge graph

1 INTRODUCTION

The popularity of online social networks has promoted the growth of various applications and information, which also enables users to browse and publish such information more freely. In the healthcare domain, patients often browse the Internet looking for information about illnesses and symptoms. For example, nearly 65% of Internet users use the Internet to search for related topics in healthcare [25]. However, the quality of online healthcare information is questionable. Many studies [11, 33] have confirmed the existence and the spread of healthcare misinformation. For example, a study of three health social networking websites found that 54% of posts contained medical claims that are inaccurate or incomplete [39].

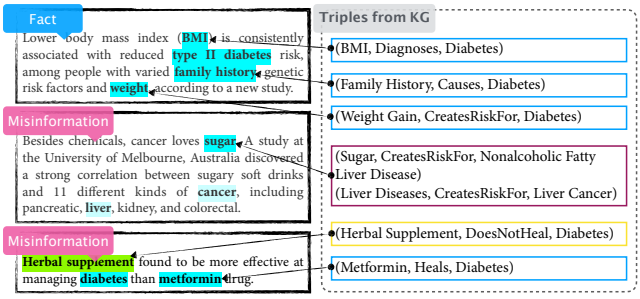


Figure 1: Healthcare article examples and related triples from a medical knowledge graph(KG). The triples can either enhance or weaken the augments in the articles.

Healthcare misinformation has detrimental societal effects. First, community's trust and support for public health agencies is undermined by misinformation, which could hinder public health control. For example, the rapid spread of misinformation is undermining trust in vaccines crucial to public health[1]. Second, health rumors that circulate on social media could directly threaten public health. During the 2014 Ebola outbreak, the World Health Organization (WHO) noted that some misinformation on social media about certain products that could prevent or cure the Ebola virus disease has led to deaths[2]. Thus, detecting healthcare misinformation is critically important.

Though misinformation detection in other domains such as politics and gossips have been extensively studied [1, 26, 30], healthcare misinformation detection has its unique properties and challenges. *First*, as non-health professionals can easily rely on given health information, it is difficult for them to discern information correctly, especially when the misinformation was intentionally made to target such people. Existing misinformation detection for domains such as politics and gossips usually adopt social contexts such as user comments to provide auxiliary information for detection[12, 16, 36, 40]. However, in healthcare domain, social context information is not always available and may not be useful because of users without professional knowledge seldom respond to healthcare information and cannot give accurate comments. *Second*, despite good performance of existing misinformation detection methods [43], the majority of them cannot explain why a piece of information is classified as misinformation. Without proper explanation, users who have no

KDD '20, August 23–27, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7998-4/20/08…$15.00
https://doi.org/10.1145/3394486.3403092

[1] https://www.nature.com/articles/d41586-018-07034-4
[2] https://www.who.int/mediacentre/news/ebola/15-august-2014/en/

health expertise might not be able to accept the result of the detection. To convince them, it is necessary to offer an understandable explanation why certain information is unreliable. Therefore, we need some auxiliary information that can (1) help detect healthcare misinformation; and (2) provide easy to understand professional knowledge for an explanation.

Medical knowledge graph, which is constructed from research papers and reports can be used as an effective auxiliary for healthcare misinformation detection, to find the inherent relations between entities in texts to improve detection performance and provide explanations. In particular, we take the article-entity bipartite graph and medical knowledge graph as complementary information, into consideration to facilitate a detection model (See Figure 1). First, article contents contain linguistic features that could be used to verify the truthfulness of an article. Misinformation (including hoaxes, rumors and fake news) is intentionally written to mislead readers by using exaggeration and sensationalization verbally. For example, we can infer from a medical knowledge graph that *Sugar* is not directly linked to *Liver Cancer*, however, the misinformation indicates that there is a "strong correlation" between the two entities. Second, the relation triples from a medical knowledge graph can add/remove the credibility of certain information, and provide explanations to the detection results. For example, in Figure 1, we can see that the triple (*BMI*, *Diagnoses*, *Diabetes*) and two more triples can directly verify that the article is real, while the triple (*Herbal Supplement*, *DoesNotHeal*, *Diabetes*) can prove that the saying in an article is wrong. Above all, it is beneficial to explore the medical graph for healthcare misinformation detection. And to our best knowledge, there is no prior attempt to detect healthcare misinformation by exploiting the knowledge graph.

Therefore in this paper, we study a novel problem of explainable healthcare misinformation detection by leveraging the medical knowledge graph. Modeling the medical knowledge graph with healthcare articles is a non-trivial task. On the one hand, healthcare information/texts and medical knowledge graph cannot be directly integrated, as they have different data structures. On the other hand, social network analysis techniques are not applied to the medical knowledge graph. For example, recommendation systems would recommend movies to users who watched a similar set of movies. However, in the healthcare domain, two medications are not necessarily related even if they can heal the same disease. To address the above two issues, we propose a knowledge guided graph attention network that can better capture the crucial entities in news articles and guide the article embedding. We incorporate the *Article-Entity Bipartite Graph* and a *Medical Knowledge Graph* into a unified relational graph and compute node embeddings along the graph. We use the *Node-level Attention* and *BPR loss* [31] to tackle the positive and negative relations in the graph. The main contributions of the paper include:

- We study a novel problem of explainable healthcare misinformation detection by leveraging medical knowledge graph to better capture the high-order relations between entities;
- We propose a novel method DETERRENT (knowleDgE guided graph aTtention nEtwoRks foR hEalthcare misiNformation deTection), which characterizes multiple positive and negative relations in the medical knowledge graph under a relational graph attention network; and
- We manually build two healthcare misinformation datasets on diabetes and cancer. Extensive experiments have demonstrated the effectiveness of DETERRENT. The reported results show that DETERRENT achieves a relative improvement of 1.05%, 4.78% on Diabetes dataset and 6.30%, 12.79% on Cancer dataset comparing with the best results in terms of Accuracy and F1 Score. The case study shows the interpretability of DETERRENT.

2 RELATED WORK

In this section, we briefly review two related topics: misinformation detection and graph neural networks.

Misinformation Detection. Misinformation detection methods generally focus on using article contents and external sources. Article contents contain linguistic clues and visual factors that can differentiate the fake and real information. Linguistic features based methods check the consistency between the headlines and contents [4], or capture specific writing styles and sensational headlines that commonly occur in fake content [29]. Visual-based features can work with linguistic features to to identify fake images [43], and help to detect misinformation collectively [8, 12].

For external sources based approaches, the features are mainly context-based. Context-based features represent the information of users' engagements from online social media. Users' responses in terms of credibility [32], viewpoints [36] and emotional signals [8] are beneficial to detect misinformation. The diffusion network constructed from users' posts can evaluate the differences in the spread of truth and falsity [42]. However, users' engagements are not always available when a news article is just released, or users lack professional knowledge of relevant fields such as medicine. Knowledge graph (KG) can address the disadvantages of current methods relying on social context and derive explanations to the detection results. Some researchers use knowledge graph based methods to decide and explain whether a (Subject, Predicate, Object) triple is fake or not [7, 15, 17]. These methods use the score function to measure the relevance of the vector embedding of subject and vector embedding of object with the embedding representation of predicate. For example, KG-Miner exploits frequent predicate paths between a pair of entities [35]. Other researchers use news streams to update the knowledge graph [38].

Hence in this paper, we study the novel problem of knowledge guided misinformation detection, aiming to improve misinformation detection performance in healthcare, and provide a possible interpretation on the result of detection simultaneously.

Graph Neural Networks. Graph Neural Networks (GNNs) refer to the neural network models that are applied to graph-structured data and aim to learn node embeddings by aggregating local neighborhood information. Several variants of GNN have been proposed to improve its representation capability and efficiency. GCN [20] tries to learn node embeddings in a semi-supervised fashion using per-neighbor normalization, instead of simply averaging all the neighborhood information. GAT [41] extends GNN by incorporating the attention mechanism; thus, each neighboring node can have

a different level of contribution to the central node. R-GCN [34] is also an extension of GCN which is suitable for large-scale relational data. It is an entity encoder model that uses a new propagation model in the forward-pass update of entities to be able to handle relational data. RGAT [6] takes advantage of both the attention mechanism and R-GCN to build an efficient graph classification model suitable for relational input data. Signed Networks [9, 14, 22, 23] are variants of GCNs applicable to the signed graph domain, in which each edge has a positive or negative sign. These methods benefit from the balance theory in social psychology to be able to correctly captures negative and positive links in the aggregating process and propagate information across layers.

However, existing methods are not suitable for modeling the positive and negative relations in the medical knowledge graph, as mentioned in the introduction. In this work, we model the medical knowledge graph under a relational graph attention network, and use BPR loss to capture positive and negative relations.

3 PROBLEM FORMULATION

In this section, we describe the notations and formulate medical knowledge graph guided misinformation detection problem. The medical knowledge graph describes the entities collected from the medical literature, as well as positive/negative relations (e.g., *Heals*/*DoesNotHeal*) among entities. For example, (*Calcium Chloride*, *Heals*, *Hypocalcemia*) contains a positive relationship, but (*Actonel*, *DoesNotHeal*, *Hypocalcemia*) has a negative relationship.

Definition 1. ***Medical Knowledge Graph***: *Let $\mathcal{G}_m = \{\mathcal{E}, \mathcal{R}, \mathcal{R}^-, \mathcal{T}, \mathcal{T}^-\}$ be a knowledge graph, where $\mathcal{E}$, $\mathcal{R}$, $\mathcal{R}^-$, $\mathcal{T}$ and $\mathcal{T}^-$ are the entity set, positive relation set, negative relation set, positive subject-relation-object triple set and negative triple set, respectively. The positive triples are presented as $\{(e_i, r, e_j)|e_i, e_j \in \mathcal{E}, r \in \mathcal{R}\}$, which describes a relationship r from the head node e_i to the tail node e_j. Similarly, negative triples are represented as $\{(e_i, r, e_j)|e_i, e_j \in \mathcal{E}, r \in \mathcal{R}^-\}$.*

We denote $\mathcal{D}$ as the health-related article set. Each article $S \in \mathcal{D}$ contains $|S|$ words, $S = \{w_1, w_2, \ldots, w_{|S|}\}$. We perform entity linking to build the word-entity alignment set $\{(w, e)|w \in \mathcal{V}, e \in \mathcal{E}\}$, where (w, e) means that word w in the vocabulary $\mathcal{V}$ can be linked with an entity e in the entity set. To capture the co-relationships of articles and entities in a medical knowledge graph, we define the article-entity bipartite graphs as follow.

Definition 2. ***Article-Entity Bipartite Graph***: *The article-entity bipartite graph is denoted as $\mathcal{G}_{ae} = (\mathcal{D} \cup \mathcal{E}, \mathcal{L})$, where $\mathcal{L}$ is the set of links. The link is denoted as $\{(S, Has, e)|S \in \mathcal{D}, e \in \mathcal{E}\}$. If an article S contains a word that can be linked to entity e, there will be a link "Has" between them, otherwise none.*

Exploiting the knowledge path between entities is of great importance. Here we formally define the knowledge path.

Definition 3. ***Knowledge Path***: *A knowledge path between entity e_1 and e_k is denoted as $e_1, r_1, e_2, r_2 \ldots, r_{k-1}, e_k$, where $e_k \in \mathcal{E}$, $r_k \in \mathcal{R}$ and $(e_{k-1}, r_{k-1}, e_k) \in \mathcal{T}$.*

Consider such a knowledge path: e_1, r_1, e_2, r_2, e_3, of which the two relations are $(Diabetes, CreatesRiskFor, Kidney\ Disease)$ and $(Kidney\ Disease, Causes, Edema)$. The two relations build a path between "diabetes" and "edema", which implies a potential link between two disorders. Such a knowledge path can add credibility to the article mentioning these two disorders. Conversely, if two words are not reachable in a knowledge graph, such two words are largely irrelevant, which reduces the credibility of related articles. For example, although "bipolar disorder" and "fenofibrate" may be the causes of "diabetes", there is no strong connection between two entities themselves from a medical perspective. However, existing text classification methods regard 'bipolar disorder" and "fenofibrate" as related as they both co-occur with "diabetes" a lot. Hence, we argue that considering knowledge paths between words through a knowledge graph can provide medical evidence in healthcare misinformation detection, which yields higher detection accuracy.

With the above notations and definitions, we formulate the knowledge guided misinformation detection task as follows:

Problem 1 (Medical Knowledge Graph Guided Misinformation Detection) *Given a set of healthcare articles $\mathcal{D}$, their corresponding label set $\mathcal{Y}$, and the medical knowledge graph $\mathcal{G}$, the goal is to learn a prediction function f to distinguish if a news is fake.*

4 METHODOLOGY

Our proposed framework consists of three components, which is shown in Figure 2: 1) an information propagation net, which propagates the knowledge between articles and nodes by preserving the structure of KG; 2) knowledge aware attention, which learns the weights of a node's neighbors in KG and aggregates the information from the neighbors and an article's contextual information to update its representation; 3) a prediction layer, which takes an article's representation as input and outputs a predicted label. Next, we introduce the details of each component.

4.1 Information Propagation Net

The medical knowledge graph can provide medical evidence in healthcare misinformation detection. To fully utilize the medical knowledge graph for healthcare misinformation detection, motivated by previous work [34, 38], we leverage inherent directional structure of the medical database to learn the entity embedding. To propagate the information from knowledge graph to the article, we incorporate the Article-Entity Bipartite Graph and Medical Knowledge Graph into a unified relational graph, and add a set of self-loops (edge type 0) denoted as $\mathcal{A} = \{(e_i, 0, e_i)|e_i \in \mathcal{E}\}$, which allows the state of a node to be kept. Hence, the new graph is defined as $\mathcal{G} = \{\mathcal{E}', \mathcal{R}', \mathcal{R}^-, \mathcal{T}', \mathcal{T}^-\}$, where $\mathcal{E}' = \mathcal{E} \cup \mathcal{D}$, $\mathcal{R}' = \mathcal{R} \cup \mathcal{R}^- \cup \{Has, 0\}$ and $\mathcal{T}' = \mathcal{T} \cup \mathcal{T}^- \cup \mathcal{L} \cup \mathcal{A}$.

Information Propagation: As there are multiple relations in a graph, we use R-GCN [34] to model the relational data, which is very effective in modeling multi-relational graph data. In R-GCN, each node is assigned to an initial representation $\mathbf{h}_i^{(0)}$. The layer-wise propagation rule updates the node representation using the representations of its neighbors in the graph in the $(l+1)$-th layer, yielding the representation $\mathbf{h}_i^{(l+1)}$ as follows:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\sum_{r \in \mathcal{R}'} \sum_{(j,r,i) \in \mathcal{T}'} \frac{1}{c_{i,r}} \mathbf{W}^r \mathbf{h}_j^{(l)}\right), \tag{1}$$

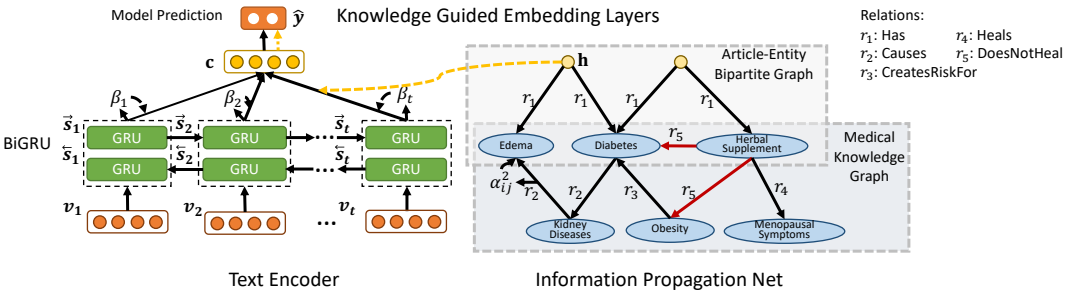


Figure 2: Illustration of the proposed DETERRENT **model. The left subfigure shows the Knowledge Guided Embedding Layers of** DETERRENT**, and the right subfigure presents the Information Propagation Net of** DETERRENT**. The Information Propagation Net is performed on the unified graph of Article-Entity Bipartite Graph and Medical Knowledge Graph, which has positive (in black) and negative (in red) relations.**

where $c_{i,r}$ is a normalization factor which is usually set to the number of neighbors of node $i \in \mathcal{E}'$ under relation $r \in \mathcal{R}'$, $\mathbf{W}^r$ is a learnable edge-type-dependent weight parameter and $\sigma(\cdot)$ denotes an activation function (we use LeakyReLU in this paper).

Node-level Attention: Each entity has relations with multiple entities. Not all relations are equally important for the healthcare misinformation detection problem. However, each neighbor has different importance to the node representation. Thus, we introduce the attention mechanism into the Information Propagation in Eq. (1) to assign more weights to important neighboring nodes, and the node representation is computed as the weighted sum of neighbors':

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}'}\sum_{(j,r,i)\in\mathcal{T}'} \alpha_{ij}^r \mathbf{W}^r \mathbf{h}_j^{(l)}\right) \tag{2}$$

where α_{ij}^r measures the importance of node i for a neighbor j, which is calculated as follows:

$$\begin{aligned} \mathbf{u}_{ij}^r &= \mathbf{W}^r(\mathbf{h}_i^{(l)} \parallel \mathbf{h}_j^{(l)}), \\ \alpha_{ij}^r &= \frac{\exp(\mathbf{a}^r \mathbf{u}_{ij}^r)}{\sum_{(k,r,i)\in\mathcal{T}'} \exp(\mathbf{a}^r \mathbf{u}_{ik}^r)} \end{aligned} \tag{3}$$

where $\mathbf{a}^r$ is the learnable parameter that weighs different feature dimensions of the node representation.

An issue of Eq. 2 is that, with the increasing number of relation types, the model will be quickly over-parameterized. To alleviate this problem, we apply Basis Decomposition [34] for regularization. This approach decomposes the weight matrix into a linear combination of several basic matrices, which largely decreases the number of model parameters.

Modeling Negative Relations: Since negative relations have different effects on the target node compared with positive relations, they should be treated separately. For example, the following three positive triples between four entities in a medical knowledge graph: 1) *Calcitriol* can heal *Calcium Deficiency*; 2) *Actonel* can heal *Calcium Deficiency*; and 3) *Calcitriol* can alleviate *Hypocalcemia*. Intuitively, we can infer that *Actonel* is a potential treatment for *Hypocalcemia*. However, a negative triple in a medical knowledge graph indicates that *Actonel* does not heal *Hypocalcemia*. Although the fact overrides our guess, it is explainable medically: Both *Calcitriol* and *Actonel* can treat *Calcium Deficiency*. However, the active ingredients in them are Vitamin D and Risedronate, respectively. Furthermore, the Vitamin D in *Calcitriol* can alleviate *Hypocalcemia* while Risedronate cannot. Thus, when we are modeling the graph, we hope the discrepancy between two entities in a negative triple is larger than in a positive triple. To achieve this goal, we choose BPR loss [31]. It is commonly used in recommendation systems, to maximize the difference between the scores of the positive and negative samples. Hence, we first conduct inner product of entity representations as the matching score:

$$m_{ij} = (\mathbf{W}^r \mathbf{h}_j)^{\mathrm{T}} \tanh(\mathbf{W}^r \mathbf{h}_i) \tag{4}$$

where $\mathbf{h}_i$ and $\mathbf{h}_j$ are the representations for entity e_i and e_j under relation r in each layer. Then we use BPR loss to penalize the scores of two entities in a negative triple:

$$\mathcal{L}_k = \sum_{\substack{(e_j,r,e_i)\in\mathcal{T}' \\ (e_k,r,e_i)\in\mathcal{T}^-}} -\ln\sigma\left(m_{ij} - m_{ik}\right) \tag{5}$$

where $\sigma(\cdot)$ is the Sigmoid function.

It is worth noting that the signed GCNs [9, 14] use balance theory [13] in social psychology to deal with the negative relations in GCN. The balance theory suggests a positive relationship between two nodes, if there exists a knowledge path between the nodes that have an even number of negative relations (e.g., "The enemy of my enemy is my friend"). However, these methods cannot be used in modeling the medical knowledge graph due to the complexity of entities (medications and diagnoses). Distinct from the existing methods, our model uses a soft assumption on the negative relations, which does not require the graph to be balanced.

4.2 Knowledge Guided Embedding Layers

After going through the Information Propagation Net, we can get the neighboring attention weights of nodes (including articles). In this section, we propose Knowledge Guided Embedding Layers to use the relevance scores of entities to an article to guide the embedding of the article.

Text Encoder: To fully capture the contextual information of an article, we use BiGRU [3] to encode word sequences from both directions of words. To be specific, given the word embeddings

$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{|S|}\}$ of an article S, the article embedding is computed as below:

$$\begin{aligned} \overrightarrow{\mathbf{s}}_t &= \text{GRU}(\overrightarrow{\mathbf{s}}_{t-1}, \mathbf{v}_t) \\ \overleftarrow{\mathbf{s}}_t &= \text{GRU}(\overleftarrow{\mathbf{s}}_{t-1}, \mathbf{v}_t) \end{aligned} \tag{6}$$

We concatenate the forward hidden state $\overrightarrow{\mathbf{s}}_t$ and the backward hidden state $\overleftarrow{\mathbf{s}}_t$ as $\mathbf{s}_t = [\overrightarrow{\mathbf{s}}_t, \overleftarrow{\mathbf{s}}_t]$, which captures the contextual information of the article centered around word $\mathbf{v}_t$.

Since not all words equally contribute to the semantic representation of the article, we leverage the attention mechanism to learn the weights to measure the importance of each word, and compute the article representation vector as follows:

$$\mathbf{c} = \sum_{t=1}^{|S|} \beta_t \mathbf{s}_t \tag{7}$$

where β_t measures the importance of the t-th word for the article, which is calculated as follows:

$$\begin{aligned} \mathbf{u}_t &= \tanh\left(\mathbf{W}_c \mathbf{s}_t + \mathbf{b}_c\right) \\ \beta_t &= \frac{\exp(\mathbf{u}_t^{\mathrm{T}} \mathbf{g})}{\sum_{k=1}^{|S|} \exp(\mathbf{u}_k^{\mathrm{T}} \mathbf{g})} \end{aligned} \tag{8}$$

where $\mathbf{u}_t$ is a hidden representation of $\mathbf{v}_t$ obtained by feeding the hidden state $\mathbf{v}_t$ to a fully embedding layer, and $\mathbf{g}$ is a trainable parameter to guide the extraction of the context.

Knowledge Guided Attention: To incorporate the knowledge guidance into the textual information, we update the $\mathbf{g}$ in Eq. 8 by $\mathbf{g}'$ to get the final attention function:

$$\mathbf{g}' = \gamma \mathbf{g} + (1 - \gamma) \mathbf{W}_k \mathbf{h}^s \tag{9}$$

where $\mathbf{h}^s$ is the node embedding of the article S obtained from the Information Propagation Net, $\mathbf{W}_k$ is a learnable transformation matrix and $\gamma \in [0, 1]$ is a trade-off parameter that controls the relative importance of the two terms. If we set $\gamma = 1$, then $\mathbf{g}'$ degenerates to $\mathbf{g}$ and our framework degenerates to a text classifier without the information from the medical knowledge graph. It makes it easy to pre-train the model to get good word embeddings for misinformation detection. The updated context vector $\mathbf{g}'$ takes both linguistic features from BiGRU and knowledge guidance into consideration. The Information Propagation Net propagates more information among similar entities and articles through the knowledge paths. We further use the attention score β_t to compute the articles representation vector $\mathbf{c}$ by Eq. 7.

4.3 Model Prediction

We have introduced how we can encode article contents through knowledge guidance. We further feed the embeddings to a softmax layer for misinformation classification as follows:

$$\hat{y} = \text{Softmax}(\mathbf{W}_f \mathbf{c} + \mathbf{b}_f) \tag{10}$$

where $\hat{y}$ is the predicted value which indicates the probability of the article being fake. For each article, our goal is to minimize the cross-entropy loss:

$$\mathcal{L}_d = -y \log \hat{y} - (1 - y) \log(1 - \hat{y}) \tag{11}$$

where $y \in \{0, 1\}$ is the ground truth label being 0 (fact) and 1 (misinformation), respectively.

Table 1: Statistics of datasets

Disease	Diabetes	Cancer
# Misinformation	608	1,476
# Fact	1,661	4,623
# Entities	11,572	22,301
# Relations	39,110	60,481

4.4 Training and Inference with DETERRENT

Finally, we combine the detection goal with BPR loss to form the final objective function as follows:

$$\mathcal{L}_{final} = \mathcal{L}_k + \mathcal{L}_d + \eta \left\|\Theta\right\|_2^2 \tag{12}$$

where Θ is the model parameters, and η is a regularization factor.

During the training, we optimize $\mathcal{L}_k$ and $\mathcal{L}_d$ alternatively. We use Adam [19] to optimize the embedding loss and the prediction loss. Adam is a widely used optimizer, which can compute individual adaptive learning rates for different parameters w.r.t. the absolute value of gradient.

5 EXPERIMENTS

In this section, we present the experiments to evaluate the effectiveness of DETERRENT. Specifically, we aim to answer the following evaluation questions:

- **RQ1**: Is DETERRENT able to improve misinformation classification performance by incorporating the medical knowledge graph?
- **RQ2**: How effective are knowledge graph and knowledge aware attention, respectively, in improving the misinformation detection performance of DETERRENT?
- **RQ3**: Can DETERRENT provide reasonable explanations about misinformation detection results?

Next, we first introduce the datasets and baselines, followed by experiments to answer these questions.

5.1 Datasets

As the medical knowledge graph, we use a public medical knowledge graph KnowLife[3] [10] with 25,334 entity names and 591,171 triples. We extract six positive relations including *Causes*, *Heals*, *CreatesRiskFor*, *ReducesRiskFor*, *Alleviates*, *Aggravates* and four negative relations including *DoesNotCause*, *DoesNotHeal*, *DoesNotCreateRiskFor*, *DoesNotReduceRiskFor*.

To evaluate the performance of DETERRENT, we need a reasonably sized collection of health-related articles of several diseases with labels. Unfortunately, there is no available dataset of adequate size. For this reason, we have collected a health-related article dataset whose years range from 2014 to 2019.

To gather real articles, we crawled from 7 reliable media outlets that have been cross-checked as reliable, e.g., Healthline, ScienceDaily, NIH (National Institutes of Health), MNT (Medical News Today), Mayo Clinic, Cleveland Clinic, WebMD. For misinformation, we crawled verified health misinformation from Snopes.com and Hoaxy API, popular hoax-debunking site and web tool. The detailed statistics of the datasets are shown in Table 1.

[3] http://knowlife.mpi-inf.mpg.de/

Table 2: Performance Comparison on Diabetes and Cancer datasets. DETERRENT **outperforms all state-of-the-art baselines including knowledge graph based and article contents based methods.**

Datasets	Metric	KG-Miner	TransE	text-CNN	CSI\c	dEFEND\c	GUpdater	HGAT	DETERRENT
Diabetes	Accuracy	0.7601	0.7671	0.7566	0.8359	0.9101	0.9012	0.8888	**0.9206**
	Precision	0.5398	0.5963	0.5563	0.6847	**0.9793**	0.9687	0.7730	0.8445
	Recall	0.6333	0.4248	0.4836	0.7826	0.6597	0.6369	0.8289	**0.8503**
	F1 Score	0.5828	0.4961	0.5174	0.7304	0.7883	0.7685	0.7996	**0.8474**
Cancer	Accuracy	0.8051	0.8536	0.8812	0.8982	0.8969	0.9022	0.8608	**0.9652**
	Precision	0.5790	0.6455	0.8531	0.7900	0.8847	0.7868	0.7226	**0.9469**
	Recall	0.7365	0.8125	0.5988	0.8165	0.6538	0.8147	0.7338	**0.9153**
	F1 Score	0.6485	0.7195	0.7037	0.8030	0.7519	0.8005	0.7282	**0.9309**

5.2 Baselines

We compare DETERRENTwith representative and state-of-the-art misinformation detection algorithms, which are listed as follows:

- KG-Miner [35]: KG-Miner is a fast discriminative path mining algorithm that can predict the truthfulness of a statement. We first use OpenIE [2] to extract the relation triple of each sentence in the article. Then we compute the score of each triple when the subject, predicate, object are all in the KG, and average all the score as output label.
- TransE [5]: TranE is a knowledge graph embedding method, which embeds entities and relations into latent vectors and completes KGs based on these vectors. We use TransE on the unified relational graph. The article embeddings are used for misinformation detection.
- text-CNN [18]: text-CNN is a text classification model that utilizes convolutional neural networks to model article contents, which can capture different granularity of text features with multiple convolution filters.
- CSI\c [32]: CSI is a hybrid deep learning-based misinformation detection model that utilizes information from article content and user response. The article representation is modeled via an LSTM model with the article embedding via Doc2Vec [21] and user response. As our datasets do not have user comments, the corresponding part of the model is ignored, and termed as CSI\c.
- dEFEND\c [37]: dEFEND utilizes a hierarchical attention neural network framework on article content and co-attention mechanism between article content and user comment for misinformation detection. As our datasets do not have user comments, the corresponding part of the model is ignored, and termed as dEFEND\c.
- HGAT [24]: HGAT is a flexible heterogeneous information network framework for classifying short texts, which can integrate any type of additional information. We add *Semantic Group* to the entities as side information, such as *Procedures* and *Disorders*.
- GUpdater [38]: GUpdater can update KGs by using news. It is built upon GNNs with a text-based attention mechanism to guide the updating message passing through KG structures. Similar to TransE, we use article embeddings for misinformation detection.

Note that for a fair comparison, we choose above contrasting methods that use features from following aspects: (1) only knowledge graph, such as TransE, KG-Miner; (2) only article contents, such as text-CNN, CSI\c, dEFEND\c and (3) both knowledge graph and article contents, such as HGAT and GUpdater. For knowledge graph methods, we feed output article embeddings into several traditional machine learning methods and choose the one that achieves the best performance. The methods include Logistic Regression, Multilayer Perceptron and Random Forest. We run these methods by using scikit-learn [27] with default parameter settings.

5.3 Experimental Setup

5.3.1 Metrics. To evaluate the performance of misinformation detection algorithms, we use the following metrics, which are commonly used to evaluate classifiers in related areas: Accuracy, Precision, Recall, and F1 score.

5.3.2 Implementation Details. We implement all models with Keras. We randomly use the labels of 75% news pieces for training and predict the remaining 25%. We set the hidden dimension of our model and other neural models to 128. The word embeddings are initialized by GloVe [28] and the dimension of pre-trained word embeddings is 100. For DETERRENT, the entity embeddings and relation embeddings are pre-trained using Information Propagation Net. We tested the depth of DETERRENT $L = \{1, 2, 3, 4\}$ and learning rate $lr = \{10^{-2}, 10^{-3}, 10^{-4}\}$. We tried $\gamma = \{0.01, 0.05, 0.1, 0.5\}$ and $\gamma = 0.05$ works best. We set $\eta = 0.05$. For other methods, we follow the network architectures as shown in the papers. For all models, we use Adam with a minibatch of 50 articles on Diabetes dataset and 100 on Cancer dataset, and the training epoch is set as 10. For a fair comparison, we use cross-entropy loss.

5.4 Misinformation Detection (RQ1)

To answer **RQ1**, we first compare DETERRENT with the representative misinformation detection algorithms introduced in Section 5.2, and then investigate the performance of DETERRENT when dealing with different types of articles.

5.4.1 Overall Comparison. Table 2 summarized the detection performance of all competing methods (reporting the average of 5 runs). From the table, we make the following observations:

- For knowledge graph-based methods, TransE and KG-Miner, the performance is less satisfactory. Although they are designed for KG triple checking and they do not incorporate linguistic features in news information. TransE can capture article-entity relations to differentiate fake and real news. When detecting fake articles, KG-Miner is dependent on OpenIE to extract relation triple from

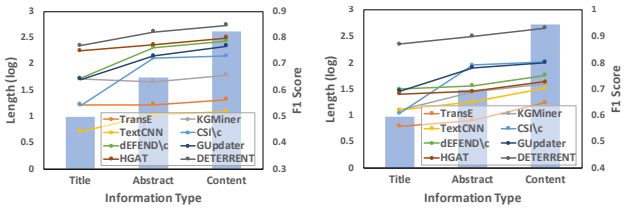
the contents, and the performance of OpenIE tends to decrease as the sentence gets longer.

- In addition, article content-based methods, text-CNN, CSI\c and dEFEND\c perform better than those methods purely based on a knowledge graph. This indicates that the methods can utilize the semantic and syntactic clues in texts. dEFEND\c can better capture important words and sentences that can contribute to the prediction through a hierarchical attention structure.
- Moreover, methods using both article contents and knowledge graph, DETERRENT, GUpdater, and HGAT, perform comparable or better than those methods using either one of them, and those only based on the knowledge graph. This indicates that knowledge graph can provide complementary information to the linguistic features, and thus improving the detection results thereby.
- Generally, for methods based on both article contents and knowledge graph, we can see that DETERRENT consistently outperforms other methods in terms of Accuracy and F1 Score on both two datasets. DETERRENT achieves a relative improvement of 1.05%, 4.78% on Diabetes dataset and 6.30%, 12.79% on Cancer dataset, comparing against the best results in terms of Accuracy and F1 Score.
- It is worthwhile to point out that dEFEND\c and CSI\c have a relatively high Precision and low Recall, which indicates that the methods predict positive samples (misinformation) wrongly as negative (fact). Hence we can see the necessity of modeling the relations between entities, as only linguistic information is not enough to distinguish fake and real information.

5.4.2 Performance Comparison w.r.t. Article Types. Besides fake articles, misinformation also includes shorter formats such as clickbait and fake posts which can easily be posted and quickly go viral on social media. The important motivation of misinformation detection is to build a general framework to detect various types of misinformation.

Hence we investigate the performance of DETERRENT when dealing with different types of articles, including title and abstract. We evaluate DETERRENT by using articles' titles and abstracts respectively. The results in terms of F1 score on both datasets are shown in Figure 3. The bars show the word lengths of different news types in log base 10. From the results, we observe that:

- DETERRENT consistently outperforms the other models. It demonstrates the effectiveness of DETERRENT on different types of misinformation regardless of the length. It again verifies the significance of knowledge graph and knowledge guided text embedding.
- The performance of article contents based methods like CSI\c and dEFEND\c do not perform very well when the length of the information is short. This suggests that those methods rely on the linguistic features of contents and cannot avoid the disadvantages brought by limited data. Although DETERRENT leverages article contents, it also exploits the additional information of entities to address above issue. The performance of DETERRENT only slightly decreases when dealing with titles (the shortest text).
- The performance of knowledge graph-based methods, KG-Miner and TransE, is relatively stable with all types of information on the two datasets.



(a) Diabetes (b) Cancer

Figure 3: Performance comparison over the length of article types on two datasets. The background histograms indicate the length of each article; meanwhile, the lines demonstrate the performance w.r.t. F1 score.

Table 3: Effects of the network depth

Datasets	Metric	1	2	3
Diabetes	Accuracy	0.8853	0.9171	0.9206
	Precision	0.7500	0.9217	0.8445
	Recall	0.8543	0.7361	0.8503
	F1 Score	0.7987	0.8185	0.8474
Cancer	Accuracy	0.9580	0.9599	0.9652
	Precision	0.9108	0.9507	0.9469
	Recall	0.9157	0.8817	0.9153
	F1 Score	0.9132	0.9149	0.9309

5.5 Ablation Analysis (RQ2)

In order to answer **RQ2**, we explore each component of DETERRENT. We first investigate the layer number of the model, then we examine the components of knowledge graph embedding and the attention mechanisms by deriving several variants.

5.5.1 Effects of Network Depth. We vary the depth L of DETERRENT to investigate the efficiency of the usage of multiple embedding propagation layers of a knowledge graph. The larger L allows further information to propagate through the information propagation layer. In particular, we search the layer number in the set of $\{1, 2, 3, 4\}$. For $L > 3$, we did not get satisfying results on both datasets, which suggests that forth- and higher-order knowledge paths contribute little information. The results are summarized in Table 3. From this, we make the following observations:

- Increasing the depth of DETERRENT can improve the performance of DETERRENT, which demonstrates the effectiveness of modeling high-order knowledge paths.
- By analyzing Table 2 and Table 3, we can see that DETERRENT is slightly better than the article contents based methods, which indicates the effectiveness of leveraging the relations.
- Besides first-order knowledge paths, high-order paths can discover inherent relations overlooked by traditional methods.

5.5.2 Effects of Attention Mechanisms and Negative Relations. In addition to article contents, we also apply knowledge graph information and integrate it with article contents with knowledge guided attention. We further investigate the effects of these components by defining three variants of DETERRENT:

Table 4: Ablation study of DETERRENT demonstrated the advantage of the attention mechanisms and modeling both positive and negative relations.

Datasets	Metric	w/o Rel	w/o K-Att	w/o Neg
Diabetes	Accuracy	0.8412	0.9012	0.9118
	Precision	0.7164	0.8870	0.9565
	Recall	0.7988	0.7236	0.7096
	F1 Score	0.7554	0.7971	0.8148
Cancer	Accuracy	0.9022	0.9291	0.9586
	Precision	0.9291	0.9385	0.9462
	Recall	0.6569	0.7651	0.8756
	F1 Score	0.7697	0.8430	0.9096

- w/o Rel: w/o Rel is a variant of DETERRENT, which does not consider the relations in the medical knowledge graph. The Information Propagation Net is replaced by a GNN model.
- w/o K-Att: w/o K-Att is a variant of DETERRENT, which excludes the knowledge-guided attention module. Each article is represented by the concatenation of the text embedding from the text encoder and node embedding from the Information Propagation Net, and fed into the prediction module.
- w/o Neg: w/o Neg is a variant of DETERRENT, which does not specifically model the negative relations in the medical knowledge graph. The BPR loss is excluded from this variant.

When one removes a medical knowledge graph, leaving only a BiGRU text encoder, the results are far from satisfactory, and thus are omitted. We summarize the experimental results in Table 4 and have the following findings:

- When we solely use a medical knowledge graph without considering relations, the performance of DETERRENT largely degrades, which suggests the necessity of modeling relations.
- Removing knowledge guided embedding attention degrades the model's performance, as the attention mechanism will assign importance weights for words, based on the semantic clues in differentiating misinformation from fact without considering knowledge paths.
- When we do not specifically model negative relations, some entities may be embedded close in a relation wrongly through information propagation. Thus, some misinformation (label 1) may be predicted as fact (label 0), which leads to relatively high Precision and low Recall.

Through the ablation study of DETERRENT, we conclude that (1) knowledge-guided article embedding can contribute to the misinformation detection performance; (2) both positive and negative relations are necessary for effective misinformation detection.

5.6 Case Study (RQ3)

In order to illustrate the importance of knowledge graph for explaining healthcare misinformation detection results, we use an example to show the triples captured by DETERRENT in Figure 4 and the corresponding attention weight in Figure 5.

In Figure5, *Diabetes* has higher attention weights to the texts. The related triples (*PancreaticIslet*, *ReducesRiskFor*, *Diabetes*) and (*Insulin*, *DoesNotHeal*, *Diabetes*) can provide explanations about why the information is false, as the texts exaggerated the effects of *PancreaticIslet* and *Insulin*. In contrast, *Glucose* has a smaller attention weight than above two entities. We can see that DETERRENT can not only detect the given information as fake but also yields the explanations of the detection results.

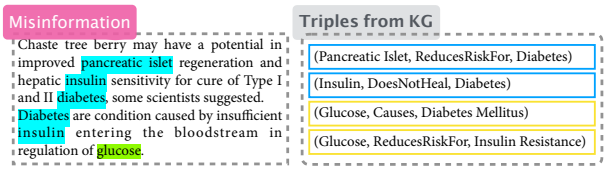


Figure 4: The explainable triples captured by DETERRENT.

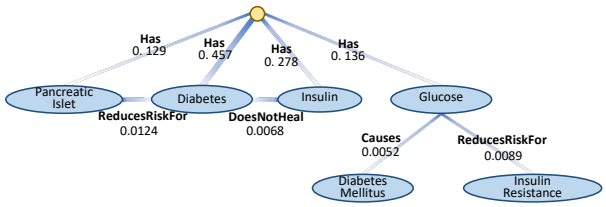
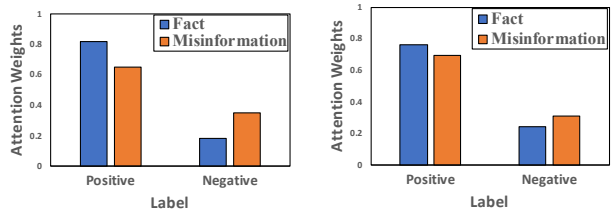


Figure 5: The visualization with attention wights.



(a) Diabetes (b) Cancer

Figure 6: The attention weight analysis indicates that positive relations contribute more to fact, and negative relations contribute more to misinformation.

We calculate the average attention weights of positive and negative relations to both misinformation and fact on two datasets. The results are shown in Figure 6. Note that positive relations have higher attention weights to fact than misinformation, while negative relations have higher attention weights to misinformation than fact. Hence, it indicates that positive relations contribute more to fact, and negative relations contribute more to misinformation.

6 CONCLUSION

In this paper, we proposed DETERRENT, a knowledge guided graph attention network for misinformation detection in healthcare. DETERRENT leverages additional information from a medical knowledge graph, to guide the article embedding with a graph attention network. The network can capture both positive and negative relations, and automatically assign more weights to important relations in differentiating misinformation from fact. The node embedding is used for guiding text encoder. Experiments on two real-world datasets demonstrate the strong performance of DETERRENT.

DETERRENT has two limitations. It only leverages a knowledge graph, instead of other complementary information. Also, it does not consider the publishing time of an article. In future, first, we can incorporate the data from medical forums to find questionable user comments. Second, other complementary information, such as doctors' remarks, can be considered. Third, time intervals between posts can be considered to model misinformation diffusion.

7 ACKNOWLEDGMENTS

We thank Patrick Ernst and Gerhard Weikum for sharing KnowLife data with us, and Jason (Jiasheng) Zhang for his valuable feedback.

This work was in part supported by NSF awards #1742702, #1820609, #1909702, #1915801, and #1934782.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives* 31, 2 (2017), 211–36.

[2] Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D Manning. 2015. Leveraging linguistic structure for open domain information extraction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 344–354.

[3] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).

[4] Gaurav Bhatt, Aman Sharma, Shivam Sharma, Ankush Nagpal, Balasubramanian Raman, and Ankush Mittal. 2018. Combining neural, statistical and external features for fake news stance identification. In *Companion Proceedings of the The Web Conference 2018*. 1353–1357.

[5] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*. 2787–2795.

[6] Dan Busbridge, Dane Sherburn, Pietro Cavallo, and Nils Y Hammerla. 2019. Relational Graph Attention Networks. *arXiv preprint arXiv:1904.05811* (2019).

[7] Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis M Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. 2015. Computational fact checking from knowledge networks. *PloS one* 10, 6 (2015).

[8] Limeng Cui, Suhang Wang, and Dongwon Lee. 2019. SAME: sentiment-aware multi-modal embedding for detecting fake news. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 41–48.

[9] Tyler Derr, Yao Ma, and Jiliang Tang. 2018. Signed graph convolutional networks. In *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 929–934.

[10] Patrick Ernst, Amy Siu, and Gerhard Weikum. 2015. KnowLife: a versatile approach for constructing a large knowledge graph for biomedical sciences. *BMC bioinformatics* 16, 1 (2015), 157.

[11] Gunther Eysenbach, John Powell, Oliver Kuss, and Eun-Ryoung Sa. 2002. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *Jama* 287, 20 (2002), 2691–2700.

[12] Han Guo, Juan Cao, Yazi Zhang, Junbo Guo, and Jintao Li. 2018. Rumor detection with hierarchical social attention network. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 943–951.

[13] Fritz Heider. 1946. Attitudes and cognitive organization. *The Journal of psychology* 21, 1 (1946), 107–112.

[14] Junjie Huang, Huawei Shen, Liang Hou, and Xueqi Cheng. 2019. Signed Graph Attention Networks. *arXiv preprint arXiv:1906.10958* (2019).

[15] Viet-Phi Huynh and Paolo Papotti. 2019. A Benchmark for Fact Checking Algorithms Built on Knowledge Bases. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 689–698.

[16] Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. News verification by exploiting conflicting social viewpoints in microblogs. In *Thirtieth AAAI conference on artificial intelligence*.

[17] Georgios Karagiannis, Immanuel Trummer, Saehan Jo, Shubham Khandelwal, Xuezhi Wang, and Cong Yu. 2019. Mining an" anti-knowledge base" from Wikipedia updates with applications to fact checking and beyond. *Proceedings of the VLDB Endowment* 13, 4 (2019), 561–573.

[18] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1746–1751.

[19] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[20] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[21] Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*. 1188–1196.

[22] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Predicting positive and negative links in online social networks. In *Proceedings of the 19th international conference on World wide web*. 641–650.

[23] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Signed networks in social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1361–1370.

[24] Hu Linmei, Tianchi Yang, Chuan Shi, Houye Ji, and Xiaoli Li. 2019. Heterogeneous graph attention networks for semi-supervised short text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 4823–4832.

[25] Amy Nguyen, Sasan Mosadeghi, and Christopher V Almario. 2017. Persistent digital divide in access to and use of the Internet as a resource for health information: Results from a California population-based study. *International journal of medical informatics* 103 (2017), 49–54.

[26] Shivam B Parikh and Pradeep K Atrey. 2018. Media-rich fake news detection: A survey. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 436–441.

[27] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12, Oct (2011), 2825–2830.

[28] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

[29] Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2017. A stylometric inquiry into hyperpartisan and fake news. *arXiv preprint arXiv:1702.05638* (2017).

[30] Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2931–2937.

[31] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*. AUAI Press, 452–461.

[32] Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 797–806.

[33] Daniel Scanfeld, Vanessa Scanfeld, and Elaine L Larson. 2010. Dissemination of health information through social networks: Twitter and antibiotics. *American journal of infection control* 38, 3 (2010), 182–188.

[34] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*. Springer, 593–607.

[35] Baoxu Shi and Tim Weninger. 2016. Discriminative predicate path mining for fact checking in knowledge graphs. *Knowledge-based systems* 104 (2016), 123–133.

[36] Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu. 2019. defend: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 395–405.

[37] Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu. 2019. DEFEND: Explainable Fake News Detection *(KDD '19)*. Association for Computing Machinery, New York, NY, USA, 395–405. https://doi.org/10.1145/3292500.3330935

[38] Jizhi Tang, Yansong Feng, and Dongyan Zhao. 2019. Learning to Update Knowledge Graphs by Reading News. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 2632–2641.

[39] Christopher C Tsai, SH Tsai, Q Zeng-Treitler, and BA Liang. 2007. Patient-centered consumer health social network websites: a pilot study of quality of user-generated health information. In *AMIA Annu Symp Proc*, Vol. 1137.

[40] Sebastian Tschiatschek, Adish Singla, Manuel Gomez Rodriguez, Arpit Merchant, and Andreas Krause. 2018. Fake news detection in social networks via crowd signals. In *Companion Proceedings of the The Web Conference 2018*. 517–524.

[41] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. *International Conference on Learning Representations* (2018). https://openreview.net/forum?id=rJXMpikCZ accepted as poster.

[42] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.

[43] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*. 849–857.

A APPENDIX ON REPRODUCIBILITY

A.1 Healthcare Misinformation Detection

In this section, we provide more details of the experimental setting and configuration to enable the reproducibility of our work.

We compared the proposed framework, DETERRENT, with 7 baseline methods discussed in Section 5.2, including KG-Miner, TransE, text-CNN, CSI, dEFEND, HGAT and GUpdater. All the codes that we have implemented are available under the folder "Healthcare misinformation detection" through the following link: https://tinyurl.com/w2wxzce. Baselines were obtained as follows:

- KG-Miner: We used the implementation by the authors of [15], which is available at: https://github.com/huynhvp/Benchmark_Fact_Checking.
- TransE: We used the implementation by the authors of [15], which is available at: https://github.com/huynhvp/Benchmark_Fact_Checking.
- text-CNN: we used the publicly available implementation at: https://github.com/dennybritz/cnn-text-classification-tf.
- CSI: We used the implementation available at: https://github.com/sungyongs/CSI-Code.
- dEFEND: We used the implementation provided by the authors available at: https://tinyurl.com/ybl6gqrm.
- HGAT: We used the implementation provided by the authors of [24].
- GUpdater: We used the implementation available at: https://github.com/esddse/GUpdater.

For the health-related article dataset, we manually created a dataset on healthcare by ourselves, under the folder "Dataset" through the following link: https://tinyurl.com/w2wxzce.

For parameter settings for DETERRENT, we introduce the details of major parameter setting as shown in Table 5. The abstracts of the major parameters are as follows:

- MAX_SENTENCE_LENGTH: the threshold to control the maximum length of news sentences
- MAX_SENTENCE_COUNT: the threshold to control the maximum count of sentences
- Word Embedding: the word emending package used for initialize the word vectors
- Embedding Dimension: the dimension of embedding layer
- Vocabulary Size: the threshold to control the maximum size of vocabulary
- d: the size of hidden states for BiGRU

A.2 Medical Knowledge Graph

For a medical knowledge graph used in this paper, we use partial data from KnowLife, which is a well-known knowledge base in biomedical science. The data we used were provided by the authors of [10]. KnowLife is constructed from textual Web sources found in specialized portals and discussion forums, such as Pubmed Medline, Pubmed Central, by using information extraction (IE) techniques. The sources include both scientific publications and posts in health portals. Overall, it consists of 214k canonical entities and 78k facts for 14 relations. Example triples in KnowLife are listed in Table 7.

Left pattern phrase and right pattern phrase are entities, regarding to the head node and tail node in a medical knowledge graph, respectively. The relation indicates a directed edge from the head node to the tail node. The above three, then, forms a triple in a medical knowledge graph.

Table 5: The details of the parameters of DETERRENT

Parameter	Diabetes	Cancer
MAX_SENTENCE_LENGTH	120	120
MAX_SENTENCE_COUNT	50	50
Word embedding	GloVe	GloVe
Embedding Dimension	100	100
Vocabulary Size	20,000	20,000
Learning Rate	10^{-3}	10^{-4}
# Epochs	10	10
Minibatch Size	50	100
d	128	128
L	3	3
Adam Parameter (β_1)	0.9	0.9
Adam Parameter (β_2)	0.999	0.999

Table 6: An example of the entity name consistency

Left Fact Entity	Left Pattern Phrase
C0271650	Prediabetes
C0271650	Glucose Intolerance
C0271650	Impaired Glucose Tolerance

As there may exit multiple names for a disease/symptom, KnowLife assigns the same entity ID to all names with the same semantics to maintain the consistency of entity name, as shown in the Table 6. Left pattern phrase indicates an entity name and left fact entity indicates the corresponding entity ID. For instance, "Prediabetes", "glucose intolerance" and "impaired glucose tolerance" are several phrases that indicate the same disorder, which is characterized by the inability to properly metabolize glucose. As we can see in the example, then, they have the same left fact entity "C0271650".

A.3 Querying Examples of DETERRENT

DETERRENT can not only predict the truthfulness of a given article, but also provide related entities and triples. Hence, to show the input and output of DETERRENT more clearly, we show more examples in this section. In Table 8, we show two fake and two real snippets of information, and the detection results by DETERRENT. The related triples can help people better understand why certain information is fake (or not).

Table 7: Example triples extracted from specialized portals

Source	Sentences	Left Pattern Phrase	Relation	Right Pattern Phrase
DrugsDotCom	{"Although rare, the corticosteroid in this medicine may cause higher blood and urine sugar levels , especially if you have severe diabetes and are using large amounts of this medicine ."}	Corticosteroid	Causes	Diabetes
Wikipedia	{"One of the more serious complications of choledocholithiasis is acute pancreatitis , which may result in significant permanent pancreatic damage and brittle diabetes ."}	Acute Pancreatitis	Causes	Brittle Diabetes
pub_med_medline	{"Anemia is associated with an increased risk of cardiovascular and renal events among patients with type 2 diabetes and chronic kidney disease -LRB-CKD -RRB- ."}	Anemia	CreatesRiskFor	Diabetes
MedlinePlus	{"Diseases such as diabetes , obesity , kidney failure or alcoholism can cause high triglycerides ."}	Alcoholism	Causes	High Triglycerides

Table 8: Querying Examples of DETERRENT

Article	Ground Truth	Prediction	Related Triples
This is a detailed exploration of BME's anti-obesity effect, facilitating the rational use of this **herbal plant** to address this increasingly severe issue, **obesity**.	1	1	(Herbal Plant, DoesNotHeal, Obesity)
They contain essential fats like ALA, antioxidants like **vitamin E**…Urolithin can bind to **estrogen** receptors, making it a strong candidate for the prevention of **breast cancer**. An animal study also reported that walnuts can reduce the growth of **prostate cancer** cells.	1	1	(Vitamin E, DoesNotReduceRiskFor,Prostate Cancer) (Estrogens, DoesNotReduceRiskFor, Breast Cancer)
A study published this month found that **mediterranean diet** led to significantly lower risk of gestational **diabetes** and reduction in excess weight gain during pregnancy.	0	0	(Mediterranean Diet, ReducesRiskFor, Diabetes)
Researchers say **vitamin D** may make the body more resistant to **breast cancer**.	0	0	(Vitamin D, ReducesRiskFor, Breast Cancer)