Knowledge-aware Multi-modal Adaptive Graph Convolutional Networks for Fake News Detection

SHENGSHENG QIAN, JUN HU, and QUAN FANG, National Lab of Pattern Recognition, Institute of Automation, CAS, China, and University of Chinese Academy of Sciences, China
CHANGSHENG XU, National Lab of Pattern Recognition, Institute of Automation, CAS, China, School of Artificial Intelligence, University of Chinese Academy of Sciences, China, and Peng Cheng Laboratory, China

In this article, we focus on fake news detection task and aim to automatically identify the fake news from vast amount of social media posts. To date, many approaches have been proposed to detect fake news, which includes traditional learning methods and deep learning-based models. However, there are three existing challenges: (i) How to represent social media posts effectively, since the post content is various and highly complicated; (ii) how to propose a data-driven method to increase the flexibility of the model to deal with the samples in different contexts and news backgrounds; and (iii) how to fully utilize the additional auxiliary information (the background knowledge and multi-modal information) of posts for better representation learning. To tackle the above challenges, we propose a novel Knowledge-aware Multi-modal Adaptive Graph Convolutional Networks (KMAGCN) to capture the semantic representations by jointly modeling the textual information, knowledge concepts, and visual information into a unified framework for fake news detection. We model posts as graphs and use a knowledge-aware multi-modal adaptive graph learning principal for the effective feature learning. Compared with existing methods, the proposed KMAGCN addresses challenges from three aspects: (1) It models posts as graphs to capture the non-consecutive and long-range semantic relations; (2) it proposes a novel adaptive graph convolutional network to handle the variability of graph data; and (3) it leverages textual information, knowledge concepts and visual information jointly for model learning. We have conducted extensive experiments on three public real-world datasets and superior results demonstrate the effectiveness of KMAGCN compared with other state-of-the-art algorithms.

CCS Concepts: • **Information systems → Data mining;**

Additional Key Words and Phrases: Fake news detection, graph convolutional network, multi-modal learning

98

This work was supported by National Key Research and Development Program of China (No. 2017YFB1002804), National Natural Science Foundation of China (No. 62036012, 61721004, 61720106006, 61802405, 62072456, 61832002, 61936005 and U1705262), the Key Research Program of Frontier Sciences, CAS, Grant NO. QYZDJSSWJSC039, and the K.C.Wong Education Foundation.
Authors' addresses: S. Qian, J. Hu, and Q. Fang, National Lab of Pattern Recognition, Institute of Automation, CAS, China, and University of Chinese Academy of Sciences, 95, ZhongGuanChun East Road, Beijing, China, 100190; emails: shengsheng.qian@nlpr.ia.ac.cn, hujunxianligong@gmail.com, qfang@nlpr.ia.ac.cn; C. Xu, National Lab of Pattern Recognition, Institute of Automation, CAS, China, School of Artificial Intelligence, University of Chinese Academy of Sciences, China, and Peng Cheng Laboratory, 95, ZhongGuanChun East Road, Beijing, China, 100190; email: csxu@nlpr.ia.ac.cn.

1551-6857/2021/07-ART98 $15.00
https://doi.org/10.1145/3451215

ACM Reference format:
Shengsheng Qian, Jun Hu, Quan Fang, and Changsheng Xu. 2021. Knowledge-aware Multi-modal Adaptive Graph Convolutional Networks for Fake News Detection. *ACM Trans. Multimedia Comput. Commun. Appl.* 17, 3, Article 98 (July 2021), 23 pages.
https://doi.org/10.1145/3451215

1 INTRODUCTION

Nowadays, social media websites have become a convenient platform that offers different kinds of multimedia information in our daily life all around the world. With easy accessibility, more and more people choose to acquire knowledge through social media, where they can share information, express, and exchange opinions. Unfortunately, due to a large number of users, various fake news has been fostered on social media websites. These widespread fake news are utilized by some guys to mislead readers, which could do serious harm to our society and may cause great economic loss. Therefore, in this scenario, it is necessary and urgent to detect fake news on social media and ensure users receive truthful information.

To date, many approaches [27, 28, 39, 60] have been proposed to detect fake news, which mainly include traditional learning methods and deep learning-based techniques. Traditional methods [28] such as **Support Vector Machine (SVM)** and Decision Tree heavily rely on hand-craft features to track and debunk fake news, which is time-consuming and labor-intensive. For example, SVM-TS [28] utilizes the linear SVM [45] and heuristic rules to detect fake news on Twitter. Recently, to alleviate the heavy manual efforts, deep neural networks are emerging tools to automatically identify fake news and have achieved great improvements over traditional methods. For instance, the **recurrent neural network (RNN)** [27] is proposed to obtain the hidden representations and sequential features from the propagation of fake news. The **convolutional neural networks (CNN)** [60] is introduced to learn high-level representations to identify fake news.

Compared with traditional methods, deep learning models have shown outstanding performance due to their superior capability of extracting features and flexibility to deal with sequential information. However, most of them ignore the long-range semantic relationships among words in feature representations and only focus on local semantics in small sliding windows (short messages or word-level syntactics). For example, the left panel of Figure 1 is the relatively long post text, in which the keywords are "Roger Stone," "Donald Trump," "guilty," "2016 campaign," and so on. However, these keywords are not grouped together and distributed throughout the whole post. It is relatively difficult to capture the dependency of semantics and structure information among them only with small sliding windows. Therefore, how to effectively capture the structural dependency information of posts in feature representations becomes more and more important for fake news detection. To better capture the structural information, many efforts have been made to model text as graphs [36]. For example, in Figure 1, the post is modeled as a graph (the right panel) with nodes representing words and edges denoting dependencies between words. However, obtaining representations for graphs is a challenging task, which may result in many subgraph features. Recently, the **graph convolutional network (GCN)** [21, 47, 54] has been proposed to leverage the content and structure information of graphs and represent them as feature vectors, which effectively avoids relying on subgraph features. As GCNs have shown the superb capability of feature learning, they are utilized as an alternative solution to capture the long-range dependency among words in the text. For example, **Text Graph Convolutional Network (TextGCN)** [58] is proposed to utilize a single text graph and learn word and text embeddings through a graph convolutional network. Although GCNs have shown impressive performance in graph learning,

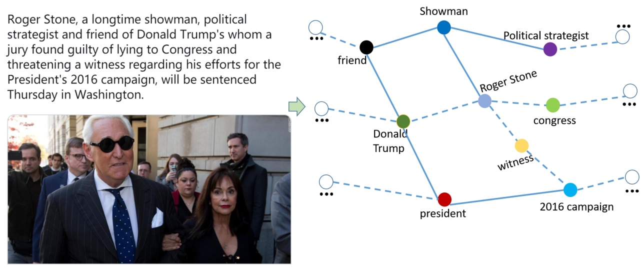


Fig. 1. An example of building a graph-structured post. The vertices represent words, and the edges denote dependency between words in a post text. The solid lines indicate direct relationships between two words, and the dashed line represents long-range semantic relationships.

to the best of our knowledge, no existing work has explored modeling each post as a graph and using GCNs to learn feature representations in the fake news detection task.

Furthermore, the graph structure in traditional GCN methods is often heuristically predefined and is not guaranteed to be optimal for fake news detection. For example, in Figure 1, the relationships among "Roger Stone," "guilty," and "threatening" are vital information for recognizing whether this particular post is fake news. However, it is difficult for GCNs to capture the dependencies among these keywords, since their relationships may not exist in the predefined graph. Therefore, the graph structure should be data dependent and dynamically updated together with other parameters of the network.

As we known, posts often contain various types of information, which is quite different from the text-only form and brings challenges for fake news detection tasks. However, most existing methods focus on extracting features from the post text only and ignore the background knowledge and visual information, which are often used as the auxiliary information to judge the credibility of posts in reality. For example, Figure 1 shows various types of information of posts, such as the textual content (i.e., the post text), visual information (i.e., the news image), and the background knowledge (i.e., Roger Stone is close to Donald Trump and Donald Trump won the President's 2016 campaign). This indicates that posts on social media platforms contain a large number of multi-model properties and knowledge-level connections, which can be leveraged to judge whether the news is fake. Therefore, how to obtain the background knowledge of the post and systematically integrate the textual information, knowledge background, and visual information is the key for fake news detection.

To build an effective framework for fake news detection, we have to address the following challenges:

- *Challenge 1:* How to effectively represent social media posts and preserve non-consecutive and long-range dependencies among words for capturing structural information in instance level?
- *Challenge 2:* How to propose a data-driven approach to increase the model flexibility for graph construction and bring more generality to adapt to various data samples in feature level?
- *Challenge 3:* How to make full use of the additional auxiliary information (the background knowledge and multi-modal information) of posts for fake news detection task?

To deal with the above challenges, we propose a **knowledge-aware multi-modal adaptive graph convolutional networks (KMAGCN)** for fake news detection by modeling posts as graphs and combining the textual information, knowledge concept, and visual information into a unified deep model. (1) For *Challenge 1*, we model each post content as a graph rather than word sequences to accurately capture word relations in instance level. (2) For *Challenge 2*, we propose a novel adaptive graph convolutional network to adaptively learn the topology of the graph in feature level. It is composed of three types of graphs, the first is a global graph by the **pointwise mutual information (PMI)**, which represents the common patterns for all data. The second is a parameterized graph that is dynamically updated together with other parameters of the network in an end-to-end learning manner. The third is an individual graph by utilizing word features of each post, which denotes the individual pattern of each sample. These three graphs are optimized individually for each layer, which effectively improves the flexibility of the model. (3) For *Challenge 3*, we propose a novel knowledge conceptualization to provide complementary background knowledge, and visual information is also utilized to provide supplementary semantic information by the feature-level attention mechanism.

In summary, the contributions of this work are as follows:

- We propose an end-to-end KMAGCN for fake news detection by modeling posts as graph structures and combining the textual information, knowledge concept and visual information into a unified deep model.
- We model each post as a graph rather than the word sequences to obtain the non-consecutive and long-range information. Furthermore, a novel adaptive graph convolutional network is proposed to increase the flexibility of the model for graph construction.
- Knowledge conceptualization is used to provide supplementary background knowledge, which generalizes well for the newly emerged posts. In addition, the feature-level attention mechanism is utilized to provide visual information as complementary information to facilitate fake news detection.
- We evaluate our method on three real-world datasets, and experimental results demonstrate our KMAGCN approach outperforms the baseline methods.

2 RELATED WORK

2.1 Fake News Detection

With the rapid growth of social media content in Internet, recognizing and detecting fake news become increasingly challenging. Researchers have made a lot of efforts on fake news detection and proposed many different and effective methods [3, 16, 23, 39, 46, 48, 49], which can be summarily reviewed from two perspectives: single-modal (e.g., text or images) fake news detection and multi-modal fake news detection.

In single-modal analysis, existing methods [3, 11, 23, 39] for fake news detection typically concentrate on extracting textual features or visual features from the text content or image information of the posts. For example, Castillo et al. [3] utilize the decision-tree to classify the post by learning the topic-based features from the text content. Yu et al. [60] obtain high-level interactions and key features of related posts by convolutional neural networks. Ma et al. [27] learn latent features from the relevant textual posts by recurrent neural networks. In Reference [35], the authors only exploit the rich visual information with different pixel domains and utilize a multi-domain visual neural network to detect fake news. However, the social media platforms contain rich multimodal information, e.g., images, texts, and videos, which can complement each other and contribute to social media analysis [15, 18, 53, 56].

Recently, multi-modal fake news detection has received considerable attentions. Some approaches are based on the simple features of images in the posts [12, 17, 33]. However, such features are handcrafted and difficult to effectively capture the complicated distributions of the image detail. As deep neural networks have achieved extraordinary performance on nonlinear representation learning [26, 62, 63], many multi-modal representation methods [16, 48] utilize deep schemes to learn the representative features, and obtain superior performance for fake news detection. Jin et al. [16] leverages a fake news detection method based on deep learning, which can learn the multi-modal and social context information and use the attention to merge them. In Reference [48], an event-invariant feature representation is learnt to obtain the multi-modal features of each post for fake news detection through an adversarial network aggregated with a multi-modal feature extractor. In Reference [19], Khattar et al. utilize a multi-modal variational autoencoder for detecting fake news, which obtains the multi-modal representations by feeding the multi-modal features into a bimodal variational autoencoder. Cui et al. [6] propose an end-to-end deep embedding framework (SAME) for detecting fake news, where users' latent sentiments are utilized to help distinguish fake news. In Reference [43], Shivangi et al. make use of the pre-trained BERT to learn text feature, and apply VGG-19 pre-trained on ImageNet dataset to learn image feature. SpotFake+ [42] is an advanced version of SpotFake [43], and the textual features are extracted by pre-trained XLNet model [57]. In Reference [64], the representations of news textual and visual information along with their relationship are jointly learned and used to predict fake news. In Reference [13], Hu et al. focus on incorporating the similarity of news to discriminate against different degrees of fake news, which uses multi-depth GCN blocks to capture multi-scale information of neighbors and combine them by attention mechanism. Compared with our previous work [61], the differences are as follows: (1) The tasks are different. The goal of the proposed model is to identify whether a claim is fake news at the post-level, where a claim is a single post on social media. The task of Reference [61] is to identify whether a claim is fake news at the event-level, where each claim consists of a sequence of correlative posts (reposts and comments) and each post is associated with a timestamp. (2) The methods are different. First, the difference is in learning textual feature. In this article, an adaptive graph convolutional network is utilized to obtain the textual feature of each post. However, in Reference [61], the authors feed a sequence of word vectors into a bidirectional GRU to capture the contextual information of the text sequence [5]. Second, the difference is in knowledge conceptualization. For knowledge concepts, we add these knowledge concepts from knowledge graphs into the text of the post, all of which are taken as inputs to feed into the adaptive graph convolutional network to get the textual feature. However, Reference [61] employs concept attention to reduce the negative effects of some inappropriate concepts, and then maps the knowledge embeddings from their original space to the word space as a channel of the multi-channel input for fake news detector. Third, the difference is in multi-modal fusion. In this work, we apply a feature-level attention mechanism to characterize certain correlations between visual and textual content, which can be used to obtain the final visual feature representation for each post. Finally, the textual feature representation and visual feature representation are concatenated to form a multi-modal feature representation for each post. However, in Reference [61], the visual feature is obtained by a word-guide visual attention module, and then Reference [61] produces a multi-channel CNN for combining all the information of a post.

Although these methods have made fairly good performance, most of existing methods are only capable of capturing the local semantic information based on narrow sliding windows for post text, while ignoring the long-range relationships between words in feature representations. In our work, we propose the KMAGCN for fake news detection by modeling posts as graph data structure, and combining the textual information, knowledge concept and visual information into a unified learning framework.

2.2 Entity Linking

Entity linking [37] is defined as assigning different entity mentions in texts to their corresponding items in the knowledge graph. Nevertheless, an entity mention may represent different named entities, and each named entity may take various forms, e.g., its abbreviations, aliases, or other alternate spellings. Therefore, a good entity linking system should not only disambiguate entity mentions in the text context but also identify the mapped entities for every entity mention. In recent years, local information [24, 55] and global information [8–10] have drawn research community's great attention. Here, local information is utilized based on the word occurrences in a context window near an entity mention, while global information is obtained by using the document-level consistency of referenced entities. In this article, different from the traditional entity linking methods, which conduct text analysis based on entity linking with a knowledge base, we focus on fake news detection via knowledge-aware multi-modal adaptive graph convolutional network learning.

2.3 Graph Neural Networks

In recent years, graph neural networks have attracted high attention in various research areas, which utilizes deep learning architectures to learn latent representations on graph-structured data [21, 50, 51, 54]. GCN [21] is a deep convolutional learning paradigm for graph-structured data which integrates the graph topology structure and local node features in convolutional layers. GAT [47] improves GCN by applying the attention method to aggregate features from the neighbors of a node with discrimination. Recently, many **Natural Language Processing (NLP)** tasks [2, 29] try to employ graph neural networks for text representation learning. For instance, in Reference [29], GCNs are utilized to incorporate the predicate-argument structure of source sentences into sentence encoders. Recently, TextGCN [58] is proposed to learn word and document embeddings by the corpus-based word co-occurrence and document-word relationships based on corpus. However, the above methods mainly focus on employing the model for textual corpora and may not be appropriate for multi-modal posts in social media. Therefore, to tackle the limitations of previous methods, in our work, we propose an end-to-end KMAGCN by modeling posts as graph data structure, and combining the textual information, knowledge concept and visual information into a unified deep model.

3 THE PROPOSED ALGORITHM

This section defines the problem to be addressed and introduces the key notations used in the article as summarized in Table 1. Then we present the overall framework for the problem.

3.1 Problem Statement

Fake news detection task can be defined as a binary classification problem, which focuses on whether posts on social media are fake news or not. Assuming that the set $P = \{p_1, \ldots, p_N\}$ contains N multimedia posts from social media where p_i is a post which consists of text messages and corresponding images. We will learn a model $\mathcal{F} : P \rightarrow Y$, which can classify each post p_i into the binary categories $Y = \{0, 1\}$, which is the ground truth of the post (1 denotes Fake news, while 0 denotes Real news).

3.2 Overall Framework

Our purpose is to identify whether a post is fake news or not. To this end, we present a KMAGCN to capture the underlying representation of the posts by jointly modeling the textual information, knowledge concepts and visual information into a unified framework for fake news detection. Our framework, as shown in Figure 2, is mainly composed of four components:

- **Knowledge Conceptualization:** The knowledge conceptualization from a real-word knowledge graph can complement the semantic representations of short texts of posts. Furthermore, the conceptual information extracted from entities can provide additional

Table 1. The Main Notations of Our Proposed Model

Notation	Description
P	the multimedia posts set
P_T	the training set
N	the number of posts
M	the size of training set
p_i	a post consists of a set of words and the attached image
Y	the ground-truth label set of posts
C	the concept set
T	the entity set
$G = (V, E)$	an undirected graph
V	the node set
E	the edge set
X	the feature matrix
A	the adjacency matrix
I	the identify matrix
$Z^{(j+1)}$	the output feature matrix of the jth GCN layer
n	the total length of each post which includes the textual content and the knowledge concepts
$\mathcal{G}$	the words relations over the global corpus
H	the global graph
U	the parameterized graph
Q	the individual graph
R_w	the representation of words in each post text
R_t	the textual feature representation of each post
R_v	the visual feature representation of each post
R_{vgg}	the feature of attached image of each post extracted by VGG-19
att^i	the attention weight of the ith words for the visual feature R_{vgg}
R_p	the multi-modal feature representation of each post
$\mathcal{F}$	the binary classifier

evidence to improve the performance of fake news detection. Specifically, this module aims to retrieve relevant knowledge for each post from Knowledge Graphs.

- **Adaptive Graph Convolutional Network:** We model posts as graphs to effectively capture the structural dependency information of posts. Based on the graph representation, we design a novel adaptive graph convolutional network which can adaptively learn the graph topology in the feature level. Specifically, it is composed of three types of graphs, a global graph H, a parameterized graph U, and an individual graph Q. These three types of graphs are optimized individually for each layer to effectively improve the model flexibility.
- **Textual and Visual Feature Representation:** Based on the output of adaptive graph convolution operation, our model can obtain representation of words in textual content for each post, which can be used to get the textual representation of each post. For visual feature representation, we utilize the attention mechanism to match textual and visual semantic concepts. A feature-level attention mechanism is proposed to find certain correlations between visual and textual content, which can obtain the final visual feature representation for each

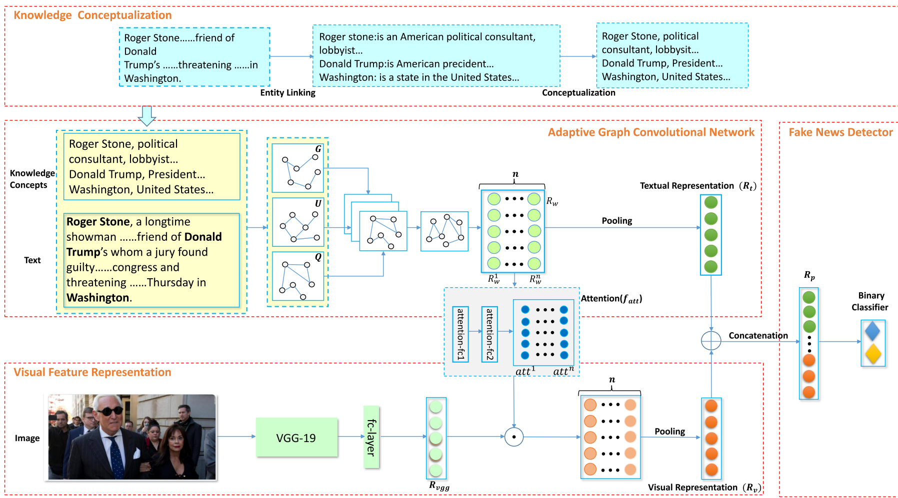


Fig. 2. The overall architecture of KMAGCN. The input includes the text content of a post, the corresponding knowledge concepts and the attached image. First, our model obtains background knowledge by knowledge conceptualization. Then we leverage a novel adaptive graph convolutional network to obtain the textual representation R_t. Next, a feature-level attention is used to get the visual representation R_v. Finally, both R_t and R_v are concatenated as the multi-modal representation R_p, which is fed into binary classifier to identify whether the post is true or not.

post. Finally, the concatenation of the textual feature representation and the visual feature representation constitutes a multi-modal feature representation for each post.

- **Fake News Detector:** The purpose of Fake News Detector is to recognize fake posts. The input is the multi-modal feature representation and the output is the predicted probability, which is used to determine whether the post is fake or not. It is composed of a fully connected layer activated by a non-linear function.

4 METHODOLOGY

This section presents our knowledge-aware multi-modal adaptive graph convolutional networks for fake news detection.

4.1 Knowledge Conceptualization

Given a post text, we hope to find a concept set C relevant to it. The process of knowledge conceptualization consists of four steps. Given the short text content of posts, many entity linking methods, such as Rel-Norm [25], Link Detector [31], EDEL [22], and STEL [4], can be utilized to link the ambiguous entity mentions M in a text to the correct entities T in the knowledge graph that they refer to. Note that, we choose the STEL in our experiment. We obtain conceptual information of each identified entity $t \in T$ from an existing knowledge graph, such as YAGO [44] and Probase [52] by conceptualization. In this article, isA relation is chosen as an example. For instance, given a short text "Thank you @realDonaldTrump for appointing me Chairman of the Minnesota Trump campaign! We will make Minnesota great again!," we obtain the entity set $T = \{realDonaldTrump, Minnesota\}$ through entity linking. Then, the entities in T are conceptualized to obtain corresponding concept set, $C_{\text{realDonaldTrump}} =$ (*Donald Trump, American president, politician, Republican*), $C_{\text{Minnesota}} =$ (*state, the United States*), from external knowledge graphs. Given a post p_i, we can conduct the knowledge conceptualization from knowledge graph and get a set of concepts for every entity contained in p_i. For every concept set $C_\mathrm{t} = (c_1, c_2, \ldots, c_\mathrm{m})$ where t is the original entity, c_i is the ith concept, and m is the set size, we aim at producing the concept knowledge for each post.

4.2 Adaptive Graph Convolutional Network

4.2.1 Graph Convolutional Network. The traditional GCN [21] is a multi-layer neural network model that computes on a pre-defined graph. An undirected graph $G = (V, E)$, where V are sets of nodes and E are sets of edges, is inputed into the GCN which outputs embedding vectors of nodes based on the propagation information of their neighborhoods. Here, each node is assumed to be self-connected, i.e., $(v, v) \in \mathbb{E}$ for every v. We use n ($n = |V|$) to denote the number of nodes, which includes the textual content and knowledge concepts of each post, and each node is associated with a d-dimensional feature vector. We introduce a feature matrix $X \in \mathbb{R}^{n \times l^{(0)}}$ to represent the features of all nodes n, in which the ith row corresponds to the $l^{(0)}$-dimensional feature vector of the ith node. Here, $l^{(0)} = d$. An adjacency matrix $A \in \mathbb{R}^{n \times n}$ is utilized to represent the edge set E, where A_{ij} denotes the weight value of the edge between the ith node and the jth node. D is the diagonal degree matrix and $D_{ii} = \sum_j A_{ij}$. Each GCN layer inputs the feature matrix $Z^{(j)} \in \mathbb{R}^{n \times l^{(j)}}$ (the input feature matrix of the first layer is $X \in \mathbb{R}^{n \times l^{(0)}}$) and outputs a higher-level feature matrix $Z^{(j+1)} \in \mathbb{R}^{n \times l^{(j+1)}}$ for nodes:

$$Z^{(0)} = X, \tag{1}$$

$$Z^{(j+1)} = \sigma\left(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} Z^{(j)} W\right), \tag{2}$$

where $\hat{A} = A + I$, I is the identify matrix, $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$ and $W \in \mathbb{R}^{l^{(j)} \times l^{(j+1)}}$ is the transformation matrix for jth layer and σ is an non-linear function.

4.2.2 Adaptive Graph Convolutional Network. Given a post, the proposed model creates an undirected graph to capture its content information, where each content word is taken as a graph node. However, the graph structure in traditional GCN methods is often heuristically predefined, which is lack of flexibility and may not be the best choice. For better dynamically updating and enriching the representation of each post, we utilize an adaptive graph structure to replace the predefined graph structure in GCN. The adaptive graph structure is constructed by three types of graphs: the global graph $H \in \mathbb{R}^{n \times n}$, the parameterized graph $U \in \mathbb{R}^{n \times n}$, and individual graph $Q \in \mathbb{R}^{n \times n}$. Specifically, Equation (2) can be transformed into:

$$Z^{(j+1)} = \sigma\left(\hat{D}^{-\frac{1}{2}} (H + U + Q) \hat{D}^{-\frac{1}{2}} Z^{(j)} W\right). \tag{3}$$

The first part H represents the common patterns of all the data, and is built by the PMI [59]. Specifically, a fixed-size window is employed on all post content to gather the word co-occurrence frequency. Then, we calculate the PMI of word pairs as follows:

$$p(w_i) = \frac{W(w_i)}{|W|}, \tag{4}$$

$$p(w_i, w_{i'}) = \frac{W(w_i, w_{i'})}{|W|}, \tag{5}$$

$$PMI(w_i, w_{i'}) = \log \frac{p(w_i, w_{i'})}{p(w_i) p(w_{i'})}, \tag{6}$$

where $|W(w_i)|$ is the statistics of sliding windows containing the word w_i, $|W(w_i, w_{i'})|$ is the frequency of sliding windows containing both the word w_i and $w_{i'}$, and $|W|$ is the total statistics of sliding windows. The PMI score reflects the correlation between words. Note that the frequency is based on the global corpus instead of a concrete post content. In the article, we merely preserve edges with positive PMI scores, since positive PMI scores imply high semantic correlations:

$$\mathcal{G}_{ii'} = \begin{cases} PMI(w_i, w_{i'}) & PMI(w_i, w_{i'}) > 0 \\ 0 & PMI(w_i, w_{i'}) \leq 0 \end{cases}, \tag{7}$$

where $\mathcal{G}_{ii'}$ is the relation between the word w_i and w_i'. After this process, we obtain the word relations $\mathcal{G}$ over the global corpus, and the global graph H is the subset of $\mathcal{G}$ for each post p_i.

The second part U is also a n x n adjacency matrix for each post. In contrast to the fixed H matrix, U is parameterized and optimized along with other parameters during training. In addition, no limitations act on the value of U, denoting that the graph is completely learned in a data-driven manner.

The third part Q learns an individual graph for each post. To know whether two words are connected, the normalized embedded Gaussian function is utilized to measure the similarity between two words in an embedding space as follows:

$$Q_{ii'} = sim(w_i, w_{i'}) = \frac{exp(vec(w_i)^T vec(w_{i'}))}{\sum_{i'=1}^{N} exp(vec(w_i)^T vec(w_{i'}))}, \tag{8}$$

where $Q_{ii'}$ is the similarity between the word w_i and word $w_{i'}$ in a post, $vec(w_i)$ is the embedding vector of word w_i. N denotes the length of sentence in the text of posts. The value of $sim(w_i, w_{i'})$ is normalized to the range $(0, 1)$, which is used as the complementary information between two words.

Specifically, given a post, we can conduct the knowledge conceptualization from the knowledge graph and get a set of concepts for every entity contained in the text of the post, and then we add these knowledge concepts into the text of the post before building H, U, and Q. The words in the text are taken as the graph nodes and relationships are taken as edges. Finally, we add H, U, and Q together as the final adaptive adjacency matrix, and propose a novel adaptive graph convolutional network to exploit the rich semantic relations between words in a data-driven manner, which can bring more generality for fake news detection. Here, in the implementation, we first define a Pytorch parameter u (a tensor with length $\frac{n \times (n+1)}{2}$). The u contains the elements of the upper triangular matrix of the $n \times n$ adjacency graph matrix. Then, we construct graph U from u by projecting $U[i, j]$ and $U[j, i]$ to the same $u[k]$. In this way, we guarantee that the u is trainable and the graph U is undirected. A pre-trained Word2Vec method [30] is utilized to obtain the feature matrix X for each post.

4.3 Textual and Visual Feature Representation

After performing the adaptive graph convolution operation on text content and corresponding knowledge concepts for each post, we obtain the representation of words of each post text denoted as $R_w = [R_w^1, \ldots, R_w^i, \ldots R_w^n]$ by GCN layers. The representation of each post text content denoted as R_t can be obtained by the average pooling-layer as follows:

$$R_t = \frac{1}{n} \sum_{i=1}^{n} (R_w^i), \tag{9}$$

where n is the total length of each post which includes the textual content and the knowledge concepts.

Visual features R_{vgg} of each post can be gained by feeding the image into VGG-19 [41] and a fully connected layer. Inspired by the idea of the attention mechanism that attempts to learn shared representations in a multi-modal setting [16], we use a feature-level attention mechanism to characterize certain correlations between visual and textual content. The words in the text content are assumed to be related with some semantic concepts in the image, and that more weights should be given to the visual features of each image that can be semantically similar to that word. Specifically, we use $att = [att^1, \ldots, att^i, \ldots, att^n]$ to weight the contributions of different visual features of each image for different words. The ith word embedding $R_w^i \in R_w$ is connected into a function f_{att} to obtain the attention weight att^i, we can calculate it as follows:

$$att^i = f_{att}(R_w^i), \tag{10}$$

where att^i has the same dimension as the visual features R_{vgg} of each image, f_{att} consists of a fully connected layer activated by ReLU function and another fully connected layer activated by softmax function.

We then perform attention weight att^i to compute the correlation between the ith word in the text and the visual features R_{vgg} of each image as follows:

$$v_i = att^i \cdot R_{vgg}, \tag{11}$$

where att^i is the feature-level attention weight of the ith word for the visual features R_{vgg}. According to this attention mechanism, the attention vector att^i computed by the network in the text can determine which visual feature deserves more attention. The final visual representation R_v of each image can be generated as follows:

$$R_v = \frac{1}{n}\sum_{i=1}^{n}(v_i). \tag{12}$$

In the end, the concatenation of text feature representation R_t and visual feature representation R_v constitutes the multi-modal feature representation denoted as $R_p = R_t \bigoplus R_v$, which is the final representation of each post. Note that the textual feature representation R_t contains original textual information of each post and knowledge concepts about entities.

4.4 Fake News Detector

Fake News Detector inputs multimodal representation R_p and aims at classifying the post as fake news or not. It is composed of a fully connected layer activated by a non-linear function denoted as $\mathcal{F}$. We feed the multi-modal representation R_p of each post into a binary classifier $\mathcal{F}$ and get the predictor r_i:

$$r_i = \mathcal{F}(R_p). \tag{13}$$

We leverage cross-entropy to measure the classification loss as follows:

$$\mathcal{L} = \sum_{\mathrm{i}=1}^{N} - \left[y_{\mathrm{i}} * log(r_{\mathrm{i}}) + (1 - y_{\mathrm{i}}) * log(1 - r_{\mathrm{i}})\right], \tag{14}$$

where N is the number of posts, y_i is the ground-truth label of the ith post.

4.5 Algorithm Description

Our algorithm is illustrated in Algorithm 1. Given a training set $P_T = \{p_1, \ldots., p_M\}$ and a global graph H by the PMI, where p_i consists of the textual content, corresponding knowledge concepts and visual information. M is the training size. Our goal is to learn a model $\mathcal{F} : P \rightarrow Y$, to classify every post p_{i} into the binary categories $Y = \{0, 1\}$, in which 0 denotes Real news, and 1 denotes

Fake news. First, we randomly initialize parameterized graph U and dynamically update it with other parameters of the network (Step 1). Next we can obtain the adaptive adjacency matrix by adding H_i, U_i, with Q_i together (Steps 6 and 7). Then we feed them into a graph convolutional network to output representation of words for each post denoted as R_w (Step 8). Later, we use the feature-level attention to compute the correlation between words in the text and the visual features R_{vgg} to obtain final visual feature representation R_v (Steps 9–11). Afterwards we concatenate R_t and R_v to obtain the multi-modal feature representation R_p of each post (Step 12). Finally, we can calculate the objective function by Equation (14), back propagate gradients and update network parameter (Step 14).

ALGORITHM 1: KMAGCN Algorithm

Input: Training set $P_T = \{p_1, \ldots, p_M\}$, p_i is each post which consists of the textual content, visual information and corresponding knowledge concepts, M is the size of training set; Global graph $\mathbf{H} = \{H_1, \ldots, H_M\}$; mini-batch size B;
Output: A binary classifier model, $\mathcal{F} : P \rightarrow Y$.
1: Initialize epoch t = 0; Randomly initialize the parameterized graph $\mathbf{U} = \{U_1, \ldots, U_M\}$ for P_T;
2: **repeat**
3: t = t+1;
4: **for** $\lfloor \frac{M}{B} \rfloor$ iterations **do**
5: Randomly sample b instances from P_T to construct a mini-batch;
6: Calculate the third graph $\mathbf{Q}_i$ for each post p_i in Equation (8);
7: Obtain the adaptive adjacency matrix by adding $\mathbf{H}_i, \mathbf{U}_i$ with $\mathbf{Q}_i$ together;
8: Obtain the representation of words of each post text R_w through Equation (3) and the representation of each post text R_t through Equation (9);
9: Obtain the visual features of each post R_{vgg} by feeding the image into VGG-19 and a fully connected layer;
10: Obtain the attention weight att^i of the ith word for the visual features R_{vgg} through Equation (10);
11: Compute the correlation between the ith word in the text and the visual features R_{vgg} through Equation (11) and obtain the final visual feature representation R_v through Equation (12);
12: Obtain the multi-modal feature representation of each post $R_p = R_t \bigoplus R_v$;
13: Calculate the objective function by Equation (14), back propagate gradients and update network parameters;
14: **end for**
15: **until** convergence;

5 EXPERIMENTAL RESULTS

We present the results of our experimental in this section. We would like to answer two questions:

(1) Does the proposed algorithm, KMAGCN, outperforms existing algorithms for fake news detection (Quantitative Results in Section 5)?
(2) How important is each component, including visual information, knowledge concepts and the three different types of graphs (H, U, Q) for the adaptive graph convolutional network module (Analysis of KMAGCN Components in Section 5)?

Table 2. The Statistics of Three Real-world Datasets

News	WEIBO	TWITTER	PHEME
# of Fake News	4,749	7,898	1,972
# of Real News	4,779	6,026	3,830
# of Images	9,528	514	3,670

5.1 Dataset

We compare our KMAGCN with state-of-the-art baselines on three public datasets: WEIBO [16], TWITTER [16, 19], and PHEME [66]. The WEIBO dataset is collected from XinHua News Agency[1] and Weibo.[2] Each post contains tweet id, text, and image. The TWITTER dataset [19] is composed of tweets containing textual information, visual information, and social context information associated with it. The PHEME dataset [66] is collected based on five breaking news, each containing a set of posts. Each dataset includes a large number of texts and images with labels. Following the [19], the WEIBO and PHEME dataset are divided into training and testing set with an tweet ratio of 8:2, and the TWITTER dataset is utilized with the development data for training set and the test data for testing set. Table 2 shows the statistics of the three datasets.

5.2 Baselines

We compare our model with two types of baseline models: single-modal and multi-modal models.

5.2.1 Single-modal Models. As against such multi-modal approach, we compare with four single-modal models described below.

- *SVM-TS* [28]: SVM-TS utilizes heuristic rules and a linear SVM classifier to detect fake news.
- *CNN* [60]: CNN employs a convolutional neural network to learn the feature representation by framing relative posts into fixed-length sequence.
- *GRU* [27]: GRU uses the multilayer GRU network to consider the post as a variable-length time series.
- *TextGCN* [59]: TextGCN is an algorithm that uses the graph convolutional network to learn words and document embeddings. The whole corpus is modeled as a heterogeneous graph.

5.2.2 Multi-modal Models. Multi-modal models utilize information from both textual and visual data for the fake news detection task. We also compare with six multi-modal approaches described below.

- *EANN* [48]: EANN is composed of the multimodal feature extractor, the fake news detection, and the event discriminator. For fairness of the comparison, we conduct experiments with a simplified version of EANN that excludes the event discriminator.
- *att_RNN* [16]: att_RNN uses an attention mechanism to combine textual, visual and social content infomation. To make a fair comparison, we remove the component processing social context information in our experiments.
- *MVAE* [19]: MVAE learns the multi-modal representations by feeding the multi-modal features into a bimodal variational autoencoder. Then the latent multi-modal representations produced by bimodal VAE are classified through a binary classifier.

[1] http://www.xinhuanet.com/.
[2] https://weibo.com/.

- *Spotfake* [43]: Spotfake utilizes the pre-trained BERT to learn the textual information, and the image feature is learned from VGG-19 pre-trained on ImageNet [7] dataset to classify a tweet as true or false.
- *Spotfake+* [42]: Spotfake+ is an advanced version of Spotfake, and the textual feature is extracted by pre-trained XLNet [57] model.
- *SAFE* [65]: SAEF is a similarity-aware multi-modal method for fake news detection, which extracts both textual and visual features from news content, and investigates their relationships to obtain the final representation.

Additionally, we have conducted several control experiments to show the effectiveness of each component in the proposed model. Details of the variants will be introduced in the analysis of KMAGCN components in Section 5.

5.3 Experimental Setting

In the process of knowledge conceptualization, we obtain an entity set T about textual content in each post through utilizing the existing entity linking solutions [4]. We then retrieve the entities in Probase [52] and YAGO [44], while merely considering the isA relation. We obtain the first type of graph, the global graph, through the PMI, in which the size of the sliding window is set to 20. For each dataset, we pre-train the Word2Vec model with an unsupervised manner by default parameter settings. We set the dimension of the post text feature to 128. The adaptive graph convolutional networks contain two hidden layers with structure as 128-128. We generate visual features via the second-to-last layer of a 19-layer VGGNet pre-trained on ImageNet [40], where the dimension is set to $4,096$. Note that, we do not finetune the weights of VGG, and the dimension of the visual representation is set to 128. We use all the news to feed in our proposed model. For posts without images attached, we generate dummy images for data alignment. To make a fair comparison, when we compare KMAGCN with SpotFake and SpotFake+, we replace the pre-trained Word2Vec model with the pre-trained BERT-base ($KMAGCN_{bert}$) and XLNet-base ($KMAGCN_{xlnet}$) available on tfhub.[3] In the SAFE model, the visual features are obtained by the pre-trained image2sentence model.[4] However, the author did not provide the pre-trained model, so we replace the image2sentence with the VGG-19 model. Our algorithms are implemented on Pytorch deep learning framework [14, 32] and are trained with Adaptive Moment Estimation [20] optimizer. The batch size is set as 128 in the training process which takes 300 epochs with a learning rate of 0.01. For fairness of comparison, we use Accuracy and F1 score as evaluation metrics [19, 48].

5.4 Quantitative Results

Table 3 displays the experimental results of KMAGCN and all baseline approaches. From Table 3, we have the following observations:

(1) SVM-TS performs the worst among all models in three datasets, indicating that hand-crafted features are weak and not enough to identify fake news.
(2) The deep learning approaches (CNN, GRU) have better performance than SVM-TS, which shows that deep learning methods have competitive advantages over traditional models. In TWITTER dataset, we also observe that CNN only outperforms SVM-TS, and is worse than other five baselines and our proposed model. It is probably because CNN ignores the long-range semantic relationships between words, which is important in fake news detection.

[3] https://tfhub.dev/.
[4] https://github.com/nikhilmaram/Show_and_Tell.

(3) TextGCN has better performances than CNN and SVM-TS on the three datasets, which shows that the graph structure can effectively capture word co-occurrences and document-word relations by the flexible graph convolutional network.
(4) As multi-modal models, att_RNN has superior performance than GRU, indicating that the attention mechanism can consider the parts of the image related to the text and improve the performance of the model. In addition, MVAE has better performance than single-modal models, which indicates that the additional visual information can be used as complementary information to facilitate fake news detection.
(5) SAFE performs better than CNN on three datasets due to jointly exploiting multi-modal (textual and visual) and relational information to learn the representation of posts. In addition, Spotfake and Spotfake+ achieve better results in all baselines on TWITTER and WEIBO datasets, which shows that the pre-trained BERT and pre-trained XLNet can better obtain contextual information for text features to improve the performance of the model.
(6) $KMAGCN_{bert}/KMAGCN_{xlnet}$ performs better than SpotFake/SpotFake+ on three datasets, which demonstrates the effectiveness of our proposed method for improving the performance of fake news detection. The proposed KMAGCN ($KMAGCN$, $KMAGCN_{bert}$, and $KMAGCN_{xlnet}$) consistently outperforms all the baselines on the three datasets. It demonstrates that the proposed model can better capture the underlying representation of the posts by jointly modeling the textual information, knowledge concepts and visual information into a unified framework.

5.5 Analysis of KMAGCN Components

Because the proposed KMAGCN contains multiple key components, we additionally compare variants of KMAGCN with respect to the following perspectives to demonstrate the effectiveness of KMAGCN—(1) the effect of the knowledge concept component, (2) the effect of the visual information, and (3) the impact of the three different types of graphs (H, U, Q) for the adaptive graph convolutional network module. The following KMAGCN variants are designed for comparison.

- KMAGCN¬k: A variant of KMAGCN with the knowledge concept component being removed.
- KMAGCN¬v: A variant of KMAGCN with the visual information being removed.
- KMAGCN¬kv: A variant of KMAGCN with the knowledge concept component and visual information being removed, and only using the post text information.
- KMAGCN¬a: A variant of KMAGCN with the attention mechanism being removed, and only using a simple concatenation for textual and visual features.
- KMAGCN¬h: A variant of KMAGCN¬kv with the global graph being removed.
- KMAGCN¬u: A variant of KMAGCN¬kv with the parameterized graph being removed.
- KMAGCN¬q: A variant of KMAGCN¬kv with the individual graph being removed.

The ablation study results are shown in Table 4 and Table 5.

5.5.1 Effects of the Knowledge CConcept Component. We compare the performance of KMAGCN with KMAGCN¬k on the three datasets (WEIBO, TWITTER, and PHEME) to investigate the effectiveness of knowledge concept component. According to the result, we observe that the proposed KMAGCN performs better than KMAGCN¬k which confirms the superiority of introducing the knowledge concept to our model.

5.5.2 Effects of the Visual Information. We compare the performance of KMAGCN with KMAGCN¬v on the three datasets (WEIBO, TWITTER, and PHEME) to investigate the

Table 3. Results of Comparison among Different Models on WEIBO, TWITTER, and PHEME Datasets

Dataset	Methods	Accuracy	Fake news			Real news		
			Precision	Recall	F1	Precision	Recall	F1
WEIBO	SVM-TS	0.640	0.741	0.573	0.646	0.651	0.798	0.711
	GRU	0.702	0.671	0.794	0.727	0.747	0.609	0.671
	CNN	0.740	0.736	0.756	0.744	0.747	0.723	0.735
	SAFE	0.763	0.833	0.659	0.736	0.717	0.868	0.785
	att_RNN	0.772	0.854	0.656	0.742	0.720	0.889	0.795
	EANN	0.782	0.827	0.697	0.756	0.752	0.863	0.804
	TextGCN	0.787	0.975	0.573	0.727	0.712	0.985	0.827
	MVAE	0.824	0.854	0.769	0.809	0.802	0.875	0.837
	KMAGCN	0.849	0.925	0.762	0.836	0.797	0.938	0.862
	SpotFake	0.869	0.8765	0.859	0.868	0.861	0.879	0.870
	SpotFake+	0.870	0.887	0.849	0.868	0.855	0.892	0.873
	SpotFake*	0.892	0.902	0.964	0.932	0.847	0.656	0.739
	$KMAGCN_{bert}$	**0.922**	0.993	0.851	0.917	0.869	0.994	**0.927**
	$KMAGCN_{xlnet}$	**0.944**	0.991	0.897	**0.942**	0.906	0.992	**0.947**
TWITTER	SVM-TS	0.529	0.488	0.497	0.496	0.565	0.556	0.561
	CNN	0.549	0.508	0.597	0.549	0.598	0.509	0.550
	GRU	0.634	0.581	0.812	0.677	0.758	0.502	0.604
	EANN	0.648	0.810	0.498	0.617	0.584	0.759	0.660
	att_RNN	0.664	0.749	0.615	0.676	0.589	0.728	0.651
	TextGCN	0.703	0.808	0.365	0.503	0.680	0.939	0.779
	MVAE	0.745	0.801	0.719	0.758	0.689	0.777	0.730
	SAFE	0.766	0.777	0.795	0.786	0.752	0.731	0.742
	SpotFake	0.771	0.784	0.744	0.764	0.769	0.807	0.787
	SpotFake*	0.777	0.751	0.900	0.820	0.832	0.606	0.701
	KMAGCN	0.787	0.799	0.724	0.760	0.782	0.845	0.812
	SpotFake+	0.790	0.793	0.827	**0.810**	0.786	0.747	0.766
	$KMAGCN_{bert}$	**0.804**	0.787	0.784	0.785	0.817	0.819	**0.818**
	$KMAGCN_{xlnet}$	**0.827**	0.816	0.805	**0.810**	0.836	0.846	**0.841**
PHEME	SVM-TS	0.639	0.546	0.576	0.560	0.729	0.705	0.717
	EANN	0.681	0.685	0.664	0.694	0.701	0.750	0.747
	CNN	0.779	0.732	0.606	0.663	0.799	0.875	0.835
	SpotFake+	0.800	0.730	0.668	0.697	0.832	0.869	0.850
	SAFE	0.811	0.827	0.559	0.667	0.806	0.940	0.866
	SpotFake	0.823	0.743	0.745	0.744	0.864	0.863	0.863
	TextGCN	0.828	0.775	0.735	0.737	0.827	0.828	0.828
	GRU	0.832	0.782	0.712	0.745	0.855	0.896	0.865
	att_RNN	0.850	0.791	0.749	0.770	0.876	0.899	0.888
	MVAE	0.852	0.806	0.719	0.760	0.871	0.917	0.893
	KMAGCN	**0.864**	0.857	0.763	**0.791**	0.877	0.939	**0.907**
	$KMAGCN_{xlnet}$	**0.865**	0.817	0.783	**0.800**	0.889	0.908	**0.898**
	$KMAGCN_{bert}$	**0.867**	0.830	0.775	**0.800**	0.885	0.914	**0.900**

*means the results are from the article.

Table 4. Results of Comparison among Different Variants in KMAGCN on WEIBO, TWITTER, and PHEME Dataset

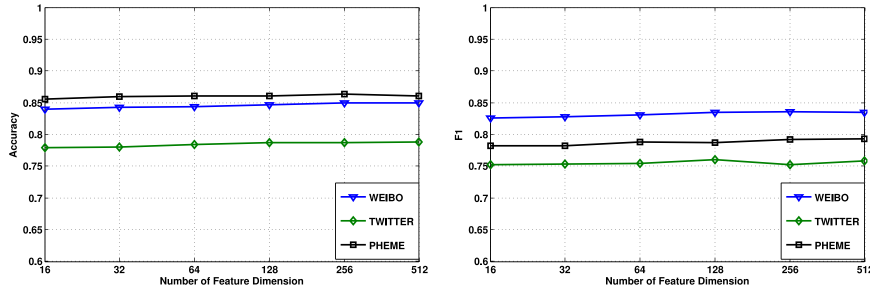
Dataset	Methods	Accuracy	Fake news			Real news		
			Precision	Recall	F1	Precision	Recall	F1
WEIBO	KMAGCN¬kv	0.834	0.923	0.729	0.815	0.775	0.939	0.849
	KMAGCN¬v	0.839	0.935	0.730	0.820	0.778	0.949	0.855
	KMAGCN¬k	0.842	0.902	0.769	0.831	0.798	0.916	0.853
	KMAGCN¬a	0.842	0.887	0.785	0.832	0.806	0.899	0.850
	KMAGCN	**0.849**	0.925	0.762	**0.836**	0.797	0.938	**0.862**
TWITTER	KMAGCN¬kv	0.748	0.838	0.561	0.672	0.708	0.907	0.796
	KMAGCN¬v	0.761	0.832	0.601	0.698	0.725	0.896	0.802
	KMAGCN¬k	0.771	0.782	0.693	0.736	0.763	0.838	0.798
	KMAGCN¬a	0.785	0.839	0.659	0.738	0.754	0.892	0.808
	KMAGCN	**0.787**	0.799	0.724	**0.760**	0.782	0.845	**0.812**
PHEME	KMAGCN¬kv	0.849	0.801	0.761	0.781	0.873	0.897	0.885
	KMAGCN¬v	0.855	0.796	0.774	0.786	0.884	0.896	0.890
	KMAGCN¬k	0.854	0.785	0.807	0.796	0.897	0.879	0.886
	KMAGCN¬a	0.855	0.828	0.714	0.787	0.866	0.925	0.894
	KMAGCN	**0.864**	0.804	0.822	**0.813**	0.899	0.888	**0.894**

Table 5. Results of Comparison when Adding Adaptive Graph Convolution Operation with or without H, U, Q on WEIBO, TWITTER, and PHEME Dataset

Dataset	Methods	Accuracy	Fake news			Real news		
			Precision	Recall	F1	Precision	Recall	F1
WEIBO	KMAGCN¬h	0.635	0.604	0.786	0.683	0.692	0.483	0.569
	KMAGCN¬u	0.818	0.979	0.651	0.782	0.738	0.986	0.844
	KMAGCN¬q	0.810	0.970	0.641	0.772	0.731	0.979	0.837
	KMAGCN*¬*kv	**0.834**	0.923	0.729	**0.815**	0.775	0.939	**0.849**
TWITTER	KMAGCN¬h	0.597	0.534	0.953	0.685	0.882	0.293	0.440
	KMAGCN¬u	0.745	0.824	0.568	0.672	0.709	0.897	0.792
	KMAGCN¬q	0.723	0.843	0.488	0.618	0.679	0.913	0.782
	KMAGCN*¬*kv	**0.748**	0.838	0.561	**0.672**	0.708	0.907	**0.796**
PHEME	KMAGCN¬h	0.720	0.502	0.005	0.010	0.651	0.985	0.794
	KMAGCN¬u	0.839	0.701	0.746	0.723	0.868	0.840	0.854
	KMAGCN¬q	0.834	0.709	0.760	0.733	0.873	0.842	0.857
	KMAGCN*¬*kv	**0.849**	0.801	0.761	**0.781**	0.873	0.897	**0.885**

effectiveness of the visual information. From the result, we can observe that the proposed KMAGCN outperforms KMAGCN¬v, which shows that the visual information can consistently provide supplementary information to benefit our model.

5.5.3 Effects of the Attention Mechanism. We compare the performance of KMAGCN with KMAGCN¬a on the three datasets (WEIBO, TWITTER, and PHEME) to investigate the effectiveness of the attention mechanism. From the result, we can observe that the proposed KMAGCN outperforms KMAGCN¬a, which shows the effectiveness of introducing the attention mechanism to our model.



(a) Accuracy (b) F1 score of fake news

Fig. 3. Impact of feature dimensions of post text output embeddings for the accuracy and F1 score of fake news on three datasets.

5.5.4 Influence of the Three Types of Graphs for the Adaptive Graph Convolutional Network Module of KMAGCN. To validate the effectiveness of each graph for the adaptive graph convolutional network, we design four variants models KMAGCN$\neg kv$, KMAGCN$\neg h$, KMAGCN$\neg u$, and KMAGCN$\neg q$. Note that KMAGCN$\neg kv$ is a variant with three types of graphs (H, U, Q) in our model. From Table 5, we can easily observe that the KMAGCN$\neg kv$ model significantly outperforms KMAGCN$\neg h$, KMAGCN$\neg u$, and KMAGCN$\neg q$, which can indicate the effectiveness of each graph for the adaptive graph convolutional network module. We can observe that the global graph H has the greatest impact on KMAGCN$\neg kv$ in three types of graphs on the three datasets, which shows that the usage of the common pattern built by the PMI can learn a superior feature representation for fake news detection.

5.6 Impact of Feature Dimensions of Post Text

KMAGCN adopts two-layer GCNs with output dimensionality of both 128, and the feature dimension d of post text output embeddings is 128. We vary d from 16 to 512, and test the impact for the accuracy and F1 score of fake news on the three datasets, respectively, as shown in Figure 3(a) and Figure 3(b). When d grows from 16 to 256, the accuracy of our model increase slightly on three dataets for fake news detection. We also observe that with d increasing from 256 to 512, the accuracy is stable on TWITTER, and slightly decreased on PHEME, while is slightly improved on WEIBO. However, for F1 score, we can find that the increase of d (from 128 to 256) does not necessarily result in the F1 score improvements on TWITTER. Therefore, we set each post textual feature dimension $d = 128$, and KMAGCN can achieve better and stable performance.

5.7 Qualitative Results

To illustrate the importance of textual features extracted by adaptive graph convolutional network, knowledge concepts obtained by knowledge conceptualization, and visual features refined by the feature-level attention mechanism for fake news detection, we show the examples of the recognized fake news to demonstrate the effectiveness of the three parts in our model.

Figure 4(a) and Figure 4(b) show two fake news examples detected by KMAGCN. In these two pieces of news, the attached images look quite regular, but the corresponding textual description is surprising and appears to be impossible, which may mislead the readers. This shows that Our adaptive GCN can effectively identify the fake news by merely using textual features when the visual features convey only a small amount of information. In Figure 4(c) and Figure 4(d), "Adolf Hitler," "Donald Trump," and "White House" are key information for readers to identify the truth of the news. However, not all readers are familiar with all of these key words. Our model can

(a) NASA Confirms Earth Will Experience 15 Days Of Darkness In Nov 2015 november blackout.

(b) Russian Laboratory Invents Chemical That Peels Off Black skin changing it to White.

(c) Adolf Hitler is Donald Trumps grandfather–Mugabe puts Trump in his place.

(d) The White House Just DISQUALIFIED Donald Trump From Being President Seriously.

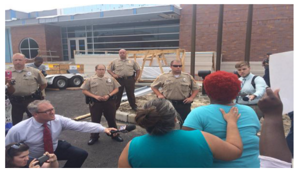
(e) A Guy Creating An Amazing Fried Rice Wave.

(f) #Believe me# This weird animals would definitely have their place in the top 10.

Fig. 4. Illustration of some fake news detected by KMAGCN.

utilize knowledge conceptualization to obtain knowledge concepts about "Adolf Hitler," "Donald Trump," and "White House," which are helpful to detect whether the news is fake. In Figure 4(e) and Figure 4(f), the two attached images look quite suspicious and are doubtless to be fake illustrations, while their textual content is normal and cannot show any evidence to discriminate that the news are fake. Thus, it is difficult for normal readers to judge whether they are fake. By inputting the visual and textual content into the proposed KMAGCN, both news are classified as fake with high confidence scores. Therefore, through combining the textual and visual content of news, the proposed KMAGCN can accurately recognize that they are fake news, which indicates the visual features can provide complementary information for normal textual content for fake news detection. The qualitative results verify that our proposed model can effectively utilize multi-modal information of posts to detect fake news. In addition, background knowledge information is utilized as an auxiliary information to judge the credibility of posts by knowledge conceptualization.

To further analyze the proposed method, we have added the error analysis. Figure 5(a) and Figure 5(b) are two fake news examples that the proposed method failed to detect. In Figure 5(a), the attached image in the news is complicated, which makes the model difficult to extract the contained semantic information and fail to detect it as a fake news. In Figure 5(b), the text content of the post is very short, leading to poor performance of the proposed adaptive graph convolutional network.



(a) Lots of angry mothers here in Ferguson this morning.

(b) SolarEclipse.

Fig. 5. Illustration of some fake news which cannot be detected by KMAGCN.

6 DISCUSSION

Fake news detection is a significant challenge, since it involves feature learning, prediction loss, and classification confidence. In our proposed design, we learn a good fake news detector that captures information in all available modality data (textual information, visual information, extended knowledge concept) and experimental results on the given benchmark datasets demonstrate the effectiveness of our model.

This challenging task has attracted more and more attention and the definition of fake news has changed over time. The most commonly used definition of fake news is: "news articles that are intentionally and verifiably false, and could mislead readers" [1]. We can observe that different motivations may lead to different types of fake news, which calls for carring out the fine-grained fake news detection. As a data-driven effort, detecting the fine-grained fake news needs the datasets that classify the fake news into fine-grain categories. However, in our article, we only employ the dataset that coarsely divides news into true and false. Therefore, our model may be biased when dealing with various types of fake news, and will divide all the fake news into false, rather than subdividing them according to motivations. In future work, we will apply our model to fine-grained fake news datasets.

In addition, as an area of great concern, there are various types of branches in fake news detection, including content-based fake news detection (ours) [42], fact-based fake news detection (the fact-checking methods) [38], and so on. The fact-based methods first extract facts from the target news and then compare them with the facts on the public knowledge base, while the content-based approaches aim to find cues from the target news content. Usually, the fact-based methods and the content-based methods use different datasets and task settings for evaluation. The content-based methods are the vast majority in the fake news detection task, and many content-based models have been proposed. However, most content-based methods using deep learning models have obvious deficiencies in interpretability. It is difficult to say what the proposed method sees, or what factors it relies on. Although it may be limited in principle, it still improves the ability to detect fake news. For example, in our daily life, most news can be distinguished based on its content, such as the wording, the writing style and the photograph [34], through the deep learning approaches. As for the interpretability of fake news detection, we have realized that this is an urgent problem to be addressed and we will study this issue in our future work.

7 CONCLUSIONS

In this article, we propose a novel KMAGCN for fake news detection task. We argue that existing methods mainly treat post texts as word sequences and are difficult to capture the non-consecutive and long-range semantic relations between words. Additionally, existing methods mainly ignore the variability of word relations in different background along with different types of information appeared in posts. To address these limitations, KMAGCN is proposed to use three technical innovations: (1) models posts as graphs to capture long-range semantic relationships, (2) proposes an adaptive graph convolutional network to increase the model flexibility for the data-driven feature learning, and (3) utilizes multi-modal information and knowledge conceptualization to jointly enhance the feature representation and model learning. The three approaches are combined to form an end-to-end optimization framework. Experiments and comparisons demonstrate that the proposed KMAGNN performs more effective and robust than state-of-the-art baselines on three public datasets for fake news detection. In the future, we will explore a more effective way to extract visual information and background knowledge, which may provide better complementary content for fake news detection.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *J. Econ. Perspect.* 31, 2 (2017), 211–236.

[2] Joost Bastings, Ivan Titov, Wilker Aziz, Diego Marcheggiani, and Khalil Sima'an. 2017. Graph convolutional encoders for syntax-aware neural machine translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP'17)*. 1957–1967.

[3] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th International Conference on World Wide Web*. ACM, 675–684.

[4] Lihan Chen, Jiaqing Liang, Chenhao Xie, Yanghua Xiao. 2018. Short text entity linking with fine-grained Topics. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 457–466.

[5] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio. 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. 1724–1734.

[6] Limeng Cui, Suhang Wang, and Dongwon Lee. 2019. SAME: sentiment-aware multi-modal embedding for detecting fake news. In *Proceedings of the IEEE ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'19)*. 41–48.

[7] Jia Deng, Wei Dong, Richard Socher, Li Jia Li, and Fei Fei Li. 2009. ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition*. 248–255.

[8] Octavian-Eugen Ganea, Marina Ganea, Aurelien Lucchi, Carsten Eickhoff, and Thomas Hofmann. 2016. Probabilistic bag-of-hyperlinks model for entity linking. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 927–938.

[9] Amir Globerson, Nevena Lazic, Soumen Chakrabarti, Amarnag Subramanya, Michael Ringaard, and Fernando Pereira. 2016. Collective entity resolution with multi-focal attention. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 621–631.

[10] Zhaochen Guo and Denilson Barbosa. 2018. Robust named entity disambiguation with random walks. *Semant. Web* 9, 4 (2018), 459–479.

[11] Aditi Gupta, Ponnurangam Kumaraguru, Carlos Castillo, and Patrick Meier. 2014. Tweetcred: Real-time credibility assessment of content on twitter. In *Proceedings of the International Conference on Social Informatics*. Springer, 228–243.

[12] Manish Gupta, Peixiang Zhao, and Jiawei Han. 2012. Evaluating event credibility on twitter. In *Proceedings of the 2012 SIAM International Conference on Data Mining*. SIAM, 153–164.

[13] Guoyong Hu, Ye Ding, Shuhan Qi, Xuan Wang, and Qing Liao. 2019. Multi-depth graph convolutional networks for fake news detection. In *Proceedings of the 8th CCF International Conference on Natural Language Processing and Chinese Computing*. 698–710.

[14] Jun Hu, Shengsheng Qian, Quan Fang, Youze Wang, Quan Zhao, Huaiwen Zhang, and Changsheng Xu. 2021. Efficient graph deep learning in tensorflow with tf_geometric. *CoRR* abs/2101.11552 (2021). Retrieved from https://arxiv.org/abs/2101.11552.

[15] Rongrong Ji, Fuhai Chen, Liujuan Cao, and Yue Gao. 2019. Cross-modality microblog sentiment prediction via bi-layer multimodal hypergraph learning. *IEEE Trans. Multimedia* 21, 4 (2019), 1062–1075.

[16] Zhiwei Jin, Juan Cao, Han Guo, Yongdong Zhang, and Jiebo Luo. 2017. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In *Proceedings of the 25th ACM International Conference on Multimedia*. ACM, 795–816.

[17] Zhiwei Jin, Juan Cao, Yongdong Zhang, Jianshe Zhou, and Qi Tian. 2016. Novel visual and statistical image features for microblogs news verification. *IEEE Trans. Multimedia* 19, 3 (2016), 598–608.

[18] Ilias Kalamaras, Anastasios Drosou, and Dimitrios Tzovaras. 2014. Multi-objective optimization for multimodal visualization. *IEEE Trans. Multimedia* 16, 5 (2014), 1460–1472.

[19] Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. Mvae: Multimodal variational autoencoder for fake news detection. In *Proceedings of the World Wide Web Conference*. 2915–2921.

[20] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. arXiv:1412.6980. Retrieved from https://arxiv.org/abs/1412.6980.

[21] Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th International Conference on Learning Representations (ICLR'17)*. 1–14.

[22] Nikolaos Kolitsas, Octavian-Eugen Ganea, and Thomas Hofmann. 2018. End-to-end neural entity linking. arXiv:1808.07699. Retrieved from https://arxiv.org/abs/1808.07699.

[23] Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent features of rumor propagation in online social media. In *Proceedings of the 2013 IEEE 13th International Conference on Data Mining*. IEEE, 1103–1108.

[24] Nevena Lazic, Amarnag Subramanya, Michael Ringgaard, and Fernando Pereira. 2015. Plato: A selective context model for entity resolution. *Trans. Assoc. Comput. Ling.* 3 (2015), 503–515.

[25] Phong Le and Ivan Titov. 2018. Improving entity linking by modeling latent relations between mentions. arXiv:1804.10637. Retrieved from https://arxiv.org/abs/1804.10637.

[26] Song Liu, Shengsheng Qian, Yang Guan, Jiawei Zhan, and Long Ying. 2020. Joint-modal distribution-based similarity hashing for large-scale unsupervised deep cross-modal retrieval. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'20)*. 1379–1388.

[27] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'16)*. 3818–3824.

[28] Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. 2015. Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 1751–1754.

[29] Diego Marcheggiani, Joost Bastings, and Ivan Titov. 2018. Exploiting semantics in neural machine translation with graph convolutional networks. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT'18)*. 486–492.

[30] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*. 3111–3119.

[31] David Milne and Ian H Witten. 2008. Learning to link with wikipedia. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*. ACM, 509–518.

[32] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. In *Proceedings of the 30th Conference on Neural Information Processing Systems NeurIPS Workshop*. 1–4.

[33] Dong ping Tian et al. 2013. A review on image feature extraction and representation techniques. *Int. J. Multimedia Ubiq. Eng.* 8, 4 (2013), 385–396.

[34] Martin Potthast, Johannes Kiesel, K. Reinartz, Janek Bevendorff, and Benno Stein. 2018. A stylometric inquiry into hyperpartisan and fake news. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 231–240.

[35] Peng Qi, Juan Cao, Tianyun Yang, Junbo Guo, and Jintao Li. 2019. Exploiting multi-domain visual information for fake news detection. arXiv:1908.04472. Retrieved from https://arxiv.org/abs/1908.04472.

[36] François Rousseau, Emmanouil Kiagias, and Michalis Vazirgiannis. 2015. Text categorization as a graph classification problem. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL'15)*. 1702–1712.

[37] Wei Shen, Jianyong Wang, and Jiawei Han. 2015. Entity linking with a knowledge base: Issues, techniques, and solutions. *IEEE Trans. Knowl. Data Eng.* 27, 2 (2015), 443–460.

[38] Baoxu Shi and Tim Weninger. 2016. Discriminative predicate path mining for fact checking in knowledge graphs. *Know.-Based Syst.* 104, C (Jul. 2016), 123–133. https://doi.org/10.1016/j.knosys.2016.04.015

[39] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explor. Newslett.* 19, 1 (2017), 22–36.

[40] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556. Retrieved from https://arxiv.org/abs/1409.1556.

[41] Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the 3th International Conference on Learning Representations*. 1–14.

[42] Shivangi Singhal, Anubha Kabra, Mohit Sharma, Rajiv Ratn Shah, and Ponnurangam Kumaraguru. 2020. SpotFake+: A multimodal framework for fake news detection via transfer learning (student abstract). In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*. 13915–13916.

[43] Shivangi Singhal, Rajiv Ratn Shah, Tanmoy Chakraborty, Ponnurangam Kumaraguru, Shin'ichi Satoh. 2019. SpotFake: A multi-modal framework for fake news detection. In *Proceedings of the 2019 IEEE 5th International Conference on Multimedia Big Data*. 39–47.
[44] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2008. Yago: A large ontology from wikipedia and wordnet. *Web Semant.* 6, 3 (2008), 203–217.
[45] Simon Tong and Daphne Koller. 2001. Support vector machine active learning with applications to text classification. *J. Mach. Learn. Res.* 2 (Nov. 2001), 45–66.
[46] Vishaal Udandarao, Abhishek Maiti, Deepak Srivatsav, Suryatej Reddy Vyalla, Yifang Yin, and Rajiv Ratn Shah. 2020. COBRA: Contrastive Bi-Modal Representation Algorithm. arXiv:cs.LG/2005.03687. Retrieved from https://arxiv/org/abs/2005.03687.
[47] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. arXiv:1710.10903. Retrieved from https://arxiv.org/abs/1710.10903.
[48] 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*. ACM, 849–857.
[49] Youze Wang, Shengsheng Qian, Jun Hu, Quan Fang, and Changsheng Xu. 2020. Fake news detection via knowledge-driven multimodal graph convolutional networks. In *Proceedings of the 2020 on International Conference on Multimedia Retrieval (ICMR'20)*. 540–547.
[50] Man Wu, Shirui Pan, Chuan Zhou, Xiaojun Chang, and Xingquan Zhu. 2020. Unsupervised domain adaptive graph convolutional networks. In *Proceedings of the World Wide Web Conference (WWW'20)*. 1457–1467.
[51] Man Wu, Shirui Pan, and Xingquan Zhu. 2020. OpenWGL: Open-world graph learning. In *Proceedings of the 20th IEEE International Conference on Data Mining (ICDM'20)*. 681–690.
[52] Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Q Zhu. 2012. Probase: A probabilistic taxonomy for text understanding. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*. ACM, 481–492.
[53] Xiao Wu, Chong-Wah Ngo, and Alexander G. Hauptmann. 2008. Multimodal news story clustering with pairwise visual near-duplicate constraint. *IEEE Trans. Multimedia* 10, 2 (2008), 188–199.
[54] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S Yu. 2019. A comprehensive survey on graph neural networks. arXiv:1901.00596. Retrieved from https://arxiv.org/abs/1901.00596.
[55] Ikuya Yamada, Hiroyuki Shindo, Hideaki Takeda, and Yoshiyasu Takefuji. 2017. Learning distributed representations of texts and entities from knowledge base. *Trans. Assoc. Comput. Ling.* 5 (2017), 397–411.
[56] Xiaoshan Yang, Tianzhu Zhang, and Changsheng Xu. 2015. Cross-domain feature learning in multimedia. *IEEE Trans. Multimedia* 17, 1 (2015), 64–78.
[57] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. XLNet: Generalized autoregressive pretraining for language understanding. In *Proceedings of the 32th Conference on Neural Information Processing Systems*. 5754–5764.
[58] Liang Yao, Chengsheng Mao, and Yuan Luo. 2018. Graph convolutional networks for text classification. arXiv:1809.05679. Retrieved from https://arxiv.org/abs/1809.05679.
[59] Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. Graph convolutional networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 7370–7377.
[60] Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2017. A convolutional approach for misinformation identification. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. 3901–3907.
[61] Huaiwen Zhang, Quan Fang, Shengsheng Qian, and Changsheng Xu. 2019. Multi-modal knowledge-aware event memory network for social media rumor detection. In *Proceedings of the 27th ACM International Conference on Multimedia (MM'19)*. Association for Computing Machinery, New York, NY, 1942–1951. https://doi.org/10.1145/3343031.3350850
[62] Yingying Zhang, Shengsheng Qian, Quan Fang, and Changsheng Xu. 2019. Multi-modal knowledge-aware hierarchical attention network for explainable medical question answering. In *Proceedings of the 27th ACM International Conference on Multimedia (MM'19)*. 1089–1097.
[63] Lei Zhao, Qinghua Hu, and Wenwu Wang. 2015. Heterogeneous feature selection with multi-modal deep neural networks and sparse group LASSO. *IEEE Trans. Multimedia* 17, 11 (2015), 1936–1948.
[64] Xinyi Zhou, Jindi Wu, Reza Zafarani. 2020. SAFE: Similarity-Aware multi-modal fake news detection. In *Proceedings of the 24th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'20)*. 354–367.
[65] Xinyi Zhou, Jindi Wu, and R. Zafarani. 2020. SAFE: Similarity-aware multi-modal fake news detection. arXiv:abs/2003.04981. Retrieved from https://arxiv.org/abs/2003.04981.
[66] Arkaitz Zubiaga, Maria Liakata, and Rob Procter. 2017. Exploiting context for rumour detection in social media. In *Proceedings of the International Conference on Social Informatics*. Springer, 109–123.

Received August 2020; revised January 2021; accepted February 2021