Multi-Depth Graph Convolutional Networks for Fake News Detection

Guoyong Hu[1], Ye Ding[2], Shuhan Qi[1], Xuan Wang[1], and Qing Liao[1,3](✉)

[1] Harbin Institute of Technology (Shenzhen)
huguoyong@stu.hit.edu.cn,liaoqing@hit.edu.cn,
{shuhanqi,wangxuan}@cs.hitsz.edu.cn
[2] Dongguan University of Technology
dingye@dgut.edu.cn
[3] Peng Cheng Laboratory

Abstract. Fake news arouses great concern owing to its political and social impacts in recent years. One of the significant challenges of fake news detection is to automatically identify fake news based on limited information. Existing works show that only considering news content and its linguistic features cannot achieve satisfactory performance when the news is short. To improve detection performance with limited information, we focus on incorporating the similarity of news to discriminate different degrees of fakeness. Specifically, we propose a multi-depth graph convolutional networks framework(M-GCN) to (1) acquire the representation of each news node via graph embedding; and (2) use multi-depth GCN blocks to capture multi-scale information of neighbours and combine them by attention mechanism. Experiment results on one of the largest real-world public fake news dataset LIAR demonstrate that the proposed M-GCN outperforms the latest five methods.

Keywords: Fake news detection · Graph convolutional networks · Graph embedding.

1 Introduction

With the rapid development of social media, millions of information flood into our lives every day, since one could easily post messages on microblogging websites. A study about the spread of fake content shows that false news diffused significantly faster, deeper and more broadly than the truth [23]. For instance, within the final three months of the 2016 U.S. presidential election, the fake news generated to favour either of the two nominees was believed by many people and was shared by more than 37 million times on Facebook [1]. Such large amount of false information causes the serious adverse effects on both individuals and society. Therefore, automatic fake news discriminator is meaningful to detect fake news and lessen the negative impact.

Fake news detection aims to determine the truthfulness of a given claim. Traditional approaches either designed a range of hand-crafted feature from text content, speaker profiles and diffusion patterns of the post to establish supervised machine learning model [4,6,27], or exploited rules and regular expressions to discover unusual patterns from tweets [29]. However, it is not easy to design all appropriate artificial features, since fake news is usually written across different topics, writing styles and social media platforms [20].

Deep learning methods [9,12,19,26] were proposed to alleviate manual effort and learn the pattern from news contents and propagation paths. These works improve the performance of detection but the accuracy drops quickly when processing the short text. For example, the accuracy of Hybrid-CNN [25] on the LIAR dataset is only 27.4%. Besides, most of these works directly fed all features to learn the representation instead of exploring the relationship among news samples. It is worth noting that the three democrats, Barack Obama, Charlie Crist and Tim Kaine, share similar credit history distribution collected from their previous statements by Wang [25]. Thus, we aim to acquire this kind of similarity to benefit the detection performance via graph embedding methods.

Recent years have seen a growth in network embedding approaches [8,16,22,24], wherein they aim to map the nodes in a network to a low-dimensional vector space preserving the network structure and node feature. The simplified Graph Convolutional Networks (GCN) [10] look at the complete 1-hop neighbourhood around the node for aggregation, but it fails to capture information beyond the second-order neighborhood instead of stacking the convolution layers. Besides, GCN iteratively propagates neighborhood features to the node, which makes information morph at each step, i.e. higher-depth information is propagated via nodes at lower-depth [21]. Therefore, the way of propagation makes the high-order information over-smoothing.

To tackle the major challenges, we propose the Multi-depth Graph Convolutional Networks(M-GCN) to classify news with speaker profiles, including the information of party of the speaker, the topic of news, home state, and so on. M-GCN preserves the multi-order information in explicit way, which makes nodes from different categories become more recognizable. Specifically, instead of directly encoding the original speaker profiles, we view each news as a node and employ their speaker profiles to construct graphs. Each graph presents a specific relationship network transformed from a kind of relationship, i.e. job-title. To take advantage of neighbors information at various depths, we expand Graph Convolutional Networks to capture the multi-scale information of neighbours, and then the nodes feature and the outputs of multi-depth GCNs blocks are integrated by the attention mechanism to obtain the final representation for fake news detection. The main contributions of the paper can be summarized as follows:

- We use graph networks to represent the speaker profiles on the LIAR dataset and capture the intrinsic correlation between two news. The correlation is exploited to enhance the performance of fake news detection.
- We expand GCN to acquire multi-scale information of neighbours based on a certain graph. The multi-depth GCN preserves the multi-granularity

information in explicit way, which improves the diversity of representation for each node.
- By using multi-depth information of neighbourhood and integrating the node feature, text representation and credit history, the proposed model outperforms the existing methods.

2 Related Work

Automatic Fake News Detection. Detecting fake news is a vital research topic and has been studied in various methods [14]. Supervised classification was widely used to identify fake news in social media posts. Castillo et al. [4] provided well designed hand-crafted features from the post contents, user profiles and propagation patterns. Feng et al. [6] utilised a wide range of linguistic features such as n-gram, part-of-speech tags and production rules based on the probabilistic context-free grammar. The main concern of this approach is to define useful features for training classifiers.

Since the ability of deep learning in automatically extracting features, many researchers focus on detecting fake news by deep neural network. Based on post content and user interactions at different times, Ma et al. [12] and Rath et al. [18] proposed deep neural network model that used RNN to learn the representations of fake news and its spreaders. Ma et al. [13] optimized rumor detection and stance classification at the same time so that more textual character can be captured. Unfortunately, the above methods for specific participants bring plenty of noise and cannot extract valid information from newly emerged events [26].

Recently, a party of studies are turning to hybrid neural network methods. Wang [25] presented the first large scale fake news detection benchmark LIAR dataset, in which each statement only contains 17.9 tokens in average. In addition to lexical features, this dataset includes speakers' information and draw plenty of attention from relevant researchers. Gottipati et al. [7] had demonstrated that speaker profiles information can be used to indicate the credibility of a piece of news. Long et al. [11] adopted speaker profiles as attention factors to propose a hybrid LSTM model to detect fake news and Karimi et al. [9] combined information from multiple sources and to discriminate between different degrees of fakeness by attention mechanism. However, most existing works aim to make good use of the additional speaker profiles to improve the performance of fake new detection but ignore the relationship between news.

Therefore, we regard the speaker profiles as multiple relationships and construct graphs to describe the similarity between two nodes(news). With the help of proposed framework, the nodes feature and multiple relationship can be merged perfectly to return coherent representation.

Graph Covolution Networks. Motivated by the successful attempt of Convolutional Neural Networks in dealing with Euclidean data to model graph-structured data, the topic of Graph Neural Networks has received growing attention. Some studies generalized well-defined neural network models to work on structured graphs. These convolution-based approaches for network embedding not only leverage the feature information of a node and its neighbourhood but also preserve global structure information in graph embedding.

Graph Convolutional Networks have shown significant improvements in semi-supervised learning on graph-structured data. In their pioneering work, Kipf and Welling [10] presented a simplified graph neural network model, graph convolutional networks, which integrated the connectivity patterns and node features. Though the model achieved state-of-the-art classification results on a large of benchmarks, it still has two limitations. On the one hand, GCN requires expensive computation to intergate high-order information by stacking convolutional layers. On the other hand, GCN iteratively propagates neighbourhood features to the node, i.e. higher depth information is propagated via nodes at lower-depth, which makes information morph at each step [21].

3 Formal Problem Definition

Let $\mathcal{D} = \{d_1, d_2, \ldots, d_{|\mathcal{D}|}\}$ be the news set with $|\mathcal{D}|$ news, where each news i contains text content representation x_i and news side information q_i. q_i^t represents a kind of speaker profiles t. Additionally, let $Y = \{y_1, y_2, \ldots, y_c\}$ donates a set of class labels. We can build adjacency matrix A_t based on $Q^t = \{q_1^t, q_2^t, \ldots, q_{|\mathcal{D}|}^t\}$. Each news is viewed as one unique node. For a given graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $N = |\mathcal{V}| = |\mathcal{D}|$ node and the edge set $\mathcal{E}$, $A_t \in \mathbb{R}^{N \times N}$ is the adjacency matrix(binary or weighted) and $X \in \mathbb{R}^{N \times F}$ represents feature matrix. Label for a subset of nodes $\mathcal{V}_L \subset \mathcal{V}$ are observed.

Our goal is to learn the model $\mathcal{M}$ assigning labels to all unlabeled nodes $\mathcal{V}_U = \mathcal{V} - \mathcal{V}_L$ by using feature matrix X and known labels for nodes in $\mathcal{V}_L$. Many researches have shown that leveraging unlabelled data in training can improve learning accuracy significantly if appropriately used [30]. In this work, we encode the graph structure by neural network $f(X, A)$ and train on the labels target, which is able to learn representations of nodes both with and without labels.

4 Proposed Method

4.1 Model Overview

Multi-Depth Graph Convolution Networks (M-GCN) is an end-to-end framework illustrated by Fig 1 and consist of three parts: node feature(text representation and credit history), multi-depth input matrix generated by one kind of relationship among nodes and the output components. To be specific, one row of text content matrix stands for text embedding vector. For each news, we used the word embedding technique to fetch the low-dimension representation of a single word. Using sum-pooling for the preceding matrix, we get the fixed-length representation vector for each news. Since credit history is well-arranged data, it can be directly used as input of the neural network. Multi-depth input matrix is multiple powers of the normalized adjacent matrix produced under one kind of graph among nodes.

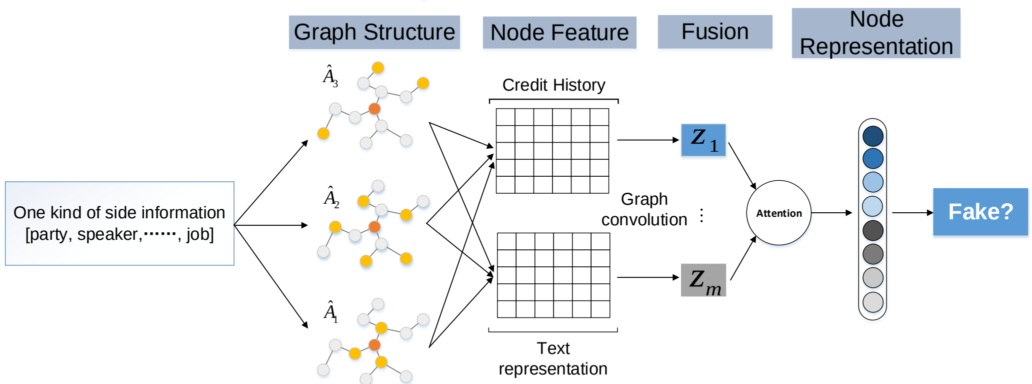


Fig. 1. An overview of the proposed model M-GCN. Given a certain kind of graph, every yellow node means one neighbour of the red node in different distances.

After integrating relational information in mutli-depth way, the textual feature and the credit history information will be fused by attention mechanism to form the final representation fed into the classifier.

4.2 Multi-Depth Graph Convolution Networks

First of all, we build edges among nodes based on whether they are in the same group. Taking job-title information for example, the sparse adjacency matrix A is defined as:

$$A_{ij} = \begin{cases} 1 \text{ if } i, j \text{ have the same job-title} \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

where i, j is the different news entity.

There are many graph convolutional methods to model the relation matrix and node feature. Spectral GCN [2] defines the convolution by decomposing a graph signal $x \in \mathbb{R}^n$ on the spectral domain and then applying a spectral filter g_θ on spectral components. Defferrard et al. [5] approximated the speactral filter with Chebyshev polynomials up to K^{th} order by building a K-localized ChebNet, where the convolution is defined as:

$$g_\theta \star x \approx \sum_{k=0}^{K} \theta'_k T_k(L_{sym})x \tag{2}$$

where $x \in \mathbb{R}^n$ is the signal on graph, g_θ is a spectral filter and $\star$ denotes the convolution operator, T_k is the Chebyshev polynomials, $\theta' \in \mathbb{R}^K$ is Chebyshev coefficients, L_{sym} is the symmetric Laplacian. Futhermore, Kipf and Welling [10] moved forward and simplified this model by limiting $K = 1$ and approximating the largest eigenvalue λ_{max} of L_{sym} by 2. The convolution becomes:

$$g_\theta \star x \approx \theta(I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}})x \tag{3}$$

where θ is the only Chebyshev coefficient. They also introduce the renormalization trick to the convolution matrix: $\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}$ with $\widetilde{A} = A + I$, which is the adjacency matrix of the undirected graph $\mathcal{G}$ with self-connections, and $\widetilde{D}_{ii} = \sum_j \widetilde{A}_{ij}$. If generalizling the definition to signal $X \in \mathbb{R}^{N\times F}$ with F input channels, which equal to a F dimensional feature vector for every node. The layer-wise propagation rule of this simplified model is:

$$H^{(l+1)} = \sigma(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{4}$$

where $H^{(l)} \in \mathbb{R}^{N\times F}$ is the matrix of activatioins in the l-th layer. $H^{(0)} = X$ is the node input features. $\sigma(\cdot)$ is an activation function, such as the $ReLU(\cdot) = max(0, \cdot)$.

The normalized adjacency matrix and Laplacian matrix $L = D - A$ describe the first-order proximity which models the local pairwise similarity between nodes. But it's not enough to model all pairwise similarity because of the sparsity.

To handle the problem of information morphing, we follow the idea of Cao et al. [3] to generalize it to k-order proximity. The normalized adjacency matrix $\widehat{A} = \widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}$ is the transition probability matrix of a single step random walk. Instead of stacking the GCN layer to merge the long distant information, we calculate the different distance proximity matrixes to describe the correlation between nodes and preserve the multi-granularity information in explicit way, which improves the diversity of representation for each node. Besides, the k-order proximity matrix can be calculated before modeling. The k-order proximity matrix $\widehat{A}^k_{ij}$ is the k-step proximity between node v_i and v_j:

$$\widehat{A}^k = \underbrace{\widehat{A} \times \widehat{A} \cdots \widehat{A}}_{k} \tag{5}$$

Each kind of step proximity matrix contain the multi-scale information of neighbours. Based on the proximity matrixes, we can use them to model the GCN layers directly and speed up the training process. Specifically, we feed different depth proximity matrixes to GCN and follow the update rule (4). For each step we get the output:

$$z_k = \widehat{A}^k ReLU(\widehat{A}^k X W_k^{(0)}) W_k^{(1)}, k = 1, 2, 3 \cdots \tag{6}$$

where X is the node features. $W_k^{(0)} \in \mathbb{R}^{F\times H}$ denotes the weight matrix for one hidden layer and $W_k^{(1)} \in \mathbb{R}^{H\times C}$ is output weight matrix. Not all outputs z_k contribute equally to detect fake news. Inspired by the great success of attention mechanism in document classification [28], we aggregate multi-depth information to form the final representation by attention mechanism. We first feed all the multi-depth outputs through a non-linear project to acquire corresponding attention score $u_i (1 \leq i \leq m)$, then each attention score is normalized by the softmax function. The final representation for each news is P_j, which is the weighted average of all z_i.

$$u_i = tanh(W_i z_i + b_i) \tag{7}$$

$$\alpha_i = \frac{\exp(u_i)}{\sum_{l=1}^{m} \exp(u_l)} \tag{8}$$

$$P_j = \sum_{i=1}^{m} \alpha_i z_i \tag{9}$$

After that, we evaluate the cross-entropy error over labeled samples:

$$\mathcal{L} = -\sum_{l \in \mathcal{Y}_L} \sum_{f=1}^{C} Y_{lf} \ln P_{lf} \tag{10}$$

where $\mathcal{Y}_L$ is the set of labels nodes and C is the dimension of final representation. We aim to minimize the loss function $\mathcal{L}$ for fake news detection.

5 Experiments

5.1 Dataset

To measure the effectiveness of the proposed approach, we evaluate the performance of M-GCN on LIAR dataset [25], which is one of the largest real-world public news datasets and famous benchmark of fake news detection. It contains 12,836 labelled short statements with 17.9 tokens in average and six fine-grained labels for the truthfulness ratings: pants-fire, false, barely-true, half-true, mostly-true and true. As a benchmark, it is divided into three set, training(80%), validation(10%) and testing(10%), in advance. The distribution of labels is relatively well-balanced. Besides, the dataset also contains a large number of speaker profiles, such as speaker name, party affiliations, job title, home state, location of speech, topics and credit history. Table 1 gives an example of the LIAR dataset.

Table 1. An example of the LIAR dataset

Statement	Our real unemployment is anywhere from 18 to 20 percent. Don't believe the 5.6. Don't believe it.
Home State	New York
Speaker	Donald Trump
Political Party	Republican
News Topic	Economy, Jobs
Current Job	President-Elect
Credit history	(63, 114, 51, 37, 61, 14)
Location of speech	his presidential announcement speech
Label	FALSE

5.2 Experimental Settings

The 300-dimensional Glove [15] word embedding was applied for each cleaned word. Out-Of-Vocabulary(OOV) words are initialized from a uniform distribution with range $[-0.25, 0.25]$. We utilise the validation set to tune the hyperparameters with grid search over ranges of different values. The hidden unit in GCN is set to [128, 64] and learning rate is 0.001. The dropout rate is set to 0.5. We use Adam optimizer to train all the parameters with weight decay strategy for 200 epochs. Since the dataset is fairly balanced and consistent with Wang [25], we use accuracy as the performance metric and calculate the average accuracy with ten trials to reduce the influence of random.

5.3 Results

To evaluate the fake news detection performance on LIAR dataset, we compared our method with the latest fake news detection methods:

- **Hybrid-CNN** [25] A hybrid CNN that integrates text and contextual information together to detect fake news.
- **LSTM-Attention** [11] A hybrid LSTM that takes the importance of words by attention mechanism into account.
- **Memory-Network** [17] A memory network that uses contextual information as attention factors to detect fake news.
- **MMFD** [9] A multi-source multi-class fake news detection model to detect multi-class fake news with multiple sources information.
- **GCN** [10] A semi-supervised method for classification that uses text content or credit history as node features and one kind of speaker profiles to build a graph, which is limited by its fundamental.

Table 2 shows the fake news classification result on LIAR dataset according to different methods. Since compared methods use different parts of contextual information, we choose the best result of the state-of-the-art models to compare. LSTM-Attention uses party, location, job position and credit history. Memory-Network only uses party and credit history. Though Hybrid-CNN had adopted all features, its accuracy is just 27.4%. This phenomenon inspires us to make good use of the features because some features may bring noise and weaken the performance. For MMFD, it outperformed the Hybrid-CNN by introducing additional information, *Report*, which is collected from *politifact.com* and longer than the statements. For the attention-based models, LSTM-Attention and Memory-Network, they manually filtered some features and weighted the importance of factors by attention mechanism. Therefore, their accuracy greatly improved. Howerver, they ignore the relationship between nodes with rich interactive information, which may be helpful for classification. Our model surpasses their results by more than 7% and 2% respectively, which indicates the effective of information integration in M-GCN. Compare to MMFD, our method gets the 10% improvement even without using the extra report information. Using the same features and side information, *Speaker*, our model also exceeds the result of original GCN by 4%. And we get similar gaps in other speaker profiles. The fact shows that the extension of GCN can better utilise the similarity among samples and preserve the rich multi-granularity information to improve fake news detection performance.

To figure out the effect of each speaker profiles information, *Speaker*, *Party*, *Topic*, *State* and *Job*. Table 3 shows the result of the proposed model with one

Table 2. Performance of fake news detection models on the LIAR dataset

Model	Detection Accuracy(%)
Hybrid-CNN [25]	27.4
MMFD [9]	38.8
LSTM-Attention [11]	41.5
Memory-Network [17]	46.7
GCN [10]	45.3
M-GCN	**49.2**

kind of speaker profiles under the condition of depth $K = 2$. We found that the other confusion matrix statistics perform similarly as accuracy. Compare with others, the *Speaker* perform better in test dataset and return the highest accuracy.

Table 3. M-GCN using different speaker profiles for 6-label classification(%)

metric	Speaker		Party		Topic		State		Job	
	Valid	Test	Valid	Test	Valid	Test	Valid	Test	Valid	Test
Accuracy	47.8	**49.2**	48.5	48.8	49.0	48.6	49.1	49.1	48.6	48.6
Precision	49.7	51.7	51.4	**51.9**	50.9	50.6	51.2	51.7	50.2	50.6
Recall	47.6	**49.9**	47.8	48.7	48.8	49.3	48.7	49.6	48.5	49.6
F1 Score	47.9	**49.7**	48.3	48.7	48.9	48.5	49.1	49.2	48.7	48.5

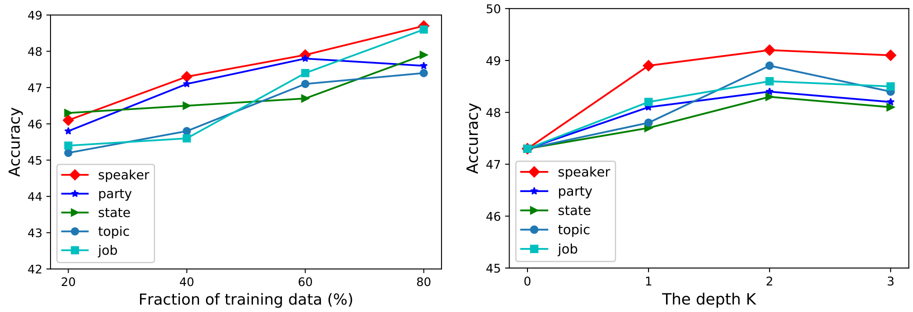
To discriminate the relative importance of Text, Credit history and speaker profiles, we also calculate the average attention scores with single speaker information *Speaker* on test dataset during the experiment. The credit history is the most informative factor in detecting fake news, which is consistent with the finding by Long [11]. The speaker profiles play a positive role in improving the performance, and the normalized attention score of text, 0.19, is far lower than the sorce of credit history's 0.59, which also makes sense because of the very short textual content. And the score of *Speaker*, 0.22, demonstrates the benefit of introducing other information.

Table 4. M-GCN and GCN using different speaker profiles without credit history(%)

Model	Speaker	Party	Topic	State	Job
GCN	22.3	20.2	20.7	22.1	21.9
M-GCN	25.7	25.3	24.9	24.6	23.4

It's worth noting that credit history is statistical data collected by previous statements of speakers and not commonly available. We also conduct experiment on using these speaker profiles without credit history. Table 4 shows the classification performance between the GCN [10] and our model. Although the overall classification performance has been significantly reduced, our model is still better than the original GCN.

To investigate whether the higher ratio of training data can improve the performance of classification, LIAR dataset is split by different percentage with the



(a) (b)

Fig. 2. (a) Performance with different fraction of training data. (b) Performance with different depth K.

random sampling. To this end, we plot the learning curves of different relationships, which include party, topic, speaker, state and job. The test data is set to fixed 10% of the total data in advance, and there is not overlap between test set and training set. The training data is randomly selected from the rest of data at each fraction. Fig 2(a) shows that the accuracy curves rise steadily in general and four features bring a little bit different performance. The *speaker* information is more remarkable in improving performance. Even when less training information is available, the performance of our M-GCN model doesn't drop quickly.

We also explore the influence of the depth of neighbour K for model performance. The Fig 2(b) presents the variation in accuracy. With the depth K ranging from zero to three, most of the features reach the peak while $K = 2$. The best reason may lie in the lack of rich relational information. Most of the relationships just contain two or three variables in LIAR dataset, and then most neighbour nodes can be reached with one or two hops, so it may introduces the noise information while $K > 2$. What we want to stress is that our M-GCN can better use node fetaures and relation graphs while receiving the same inputs under the same order. If a certain relationship has more variables, it may return better performance by adding rich information of dense graph structure.

6 Conclusion

In this paper, different from the methods directly encoding the speaker profiles or attention-based methods, we acquire the representation of each news node with graph structure information converted from speaker profiles. To make good use of neighbours' features, we expand original GCN to capture the multi-scale information of neighbours and preserves the rich multi-granularity information for each node. The experiments results on the LIAR dataset show that multi-depth graph convolution networks(M-GCN) can utilise the similarity of news nodes to improve classification performance and identify the authenticity of news more effectively than the existing five methods.

Acknowledgements

This work is supported in part by National Natural Science Foundation of China under grant No.U1711261 and No.U1811463. National Key Research and Development Plan under grant No.2017YFB0802204 and the Basic Research Project of Shenzhen under grant No.JCYJ20180306174743727.

References

1. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. Journal of economic perspectives **31**(2), 211–36 (2017)
2. Bruna, J., Zaremba, W., Szlam, A., LeCun, Y.: Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203 (2013)
3. Cao, S., Lu, W., Xu, Q.: Grarep: Learning graph representations with global structural information. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. pp. 891–900. ACM (2015)
4. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on twitter. In: Proceedings of the 20th international conference on World wide web. pp. 675–684. ACM (2011)
5. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In: Advances in Neural Information Processing Systems. pp. 3844–3852 (2016)
6. Feng, S., Banerjee, R., Choi, Y.: Syntactic stylometry for deception detection. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2. pp. 171–175. Association for Computational Linguistics (2012)
7. Gottipati, S., Qiu, M., Yang, L., Zhu, F., Jiang, J.: Predicting users political party using ideological stances. In: International Conference on Social Informatics. pp. 177–191. Springer (2013)
8. Grover, A., Leskovec, J.: node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 855–864. ACM (2016)
9. Karimi, H., Roy, P., Saba-Sadiya, S., Tang, J.: Multi-source multi-class fake news detection. In: Proceedings of the 27th International Conference on Computational Linguistics. pp. 1546–1557 (2018)
10. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR) (2017)
11. Long, Y., Lu, Q., Xiang, R., Li, M., Huang, C.R.: Fake news detection through multi-perspective speaker profiles. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers). vol. 2, pp. 252–256 (2017)
12. Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B.J., Wong, K.F., Cha, M.: Detecting rumors from microblogs with recurrent neural networks. In: IJCAI. pp. 3818–3824 (2016)
13. Ma, J., Gao, W., Wong, K.F.: Detect rumor and stance jointly by neural multi-task learning. In: Companion of the The Web Conference 2018 on The Web Conference 2018. pp. 585–593. International World Wide Web Conferences Steering Committee (2018)

14. Morris, M.R., Counts, S., Roseway, A., Hoff, A., Schwarz, J.: Tweeting is believing?: understanding microblog credibility perceptions. In: Proceedings of the ACM 2012 conference on computer supported cooperative work. pp. 441–450. ACM (2012)
15. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). pp. 1532–1543 (2014)
16. Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 701–710. ACM (2014)
17. Pham, T.T.: A study on deep learning for fake news detection (2018)
18. Rath, B., Gao, W., Ma, J., Srivastava, J.: From retweet to believability: Utilizing trust to identify rumor spreaders on twitter. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017. pp. 179–186. ACM (2017)
19. Ruchansky, N., Seo, S., Liu, Y.: Csi: A hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. pp. 797–806. ACM (2017)
20. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter **19**(1), 22–36 (2017)
21. Soni, U., Bhambhani, M., Khapra, M.M.: Network embedding using hierarchical feature aggregation (2018)
22. Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q.: Line: Large-scale information network embedding. In: Proceedings of the 24th International Conference on World Wide Web. pp. 1067–1077. International World Wide Web Conferences Steering Committee (2015)
23. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018)
24. Wang, D., Cui, P., Zhu, W.: Structural deep network embedding. In: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1225–1234. ACM (2016)
25. Wang, W.Y.: " liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648 (2017)
26. Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., Gao, J.: 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. pp. 849–857. ACM (2018)
27. Yang, F., Liu, Y., Yu, X., Yang, M.: Automatic detection of rumor on sina weibo. In: Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. p. 13. ACM (2012)
28. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E.: Hierarchical attention networks for document classification. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 1480–1489 (2016)
29. Zhao, Z., Resnick, P., Mei, Q.: Enquiring minds: Early detection of rumors in social media from enquiry posts. In: Proceedings of the 24th International Conference on World Wide Web. pp. 1395–1405. International World Wide Web Conferences Steering Committee (2015)
30. Zhu, X., Goldberg, A.B.: Introduction to semi-supervised learning. Synthesis lectures on artificial intelligence and machine learning **3**(1), 1–130 (2009)