Fake News Detection on Social Media using Geometric Deep Learning

Mudit Wadhwa
Dept. of CSIS
BITS Pilani
Rajasthan, INDIA
f20180330@pilani.bits-pilani.ac.in

Hritwik Goklaney
Dept. of CSIS
BITS Pilani
Rajasthan, INDIA
f20180238@pilani.bits-pilani.ac.in

Yash Gupta
Dept. of CSIS
BITS Pilani
Rajasthan, INDIA
f20180262@pilani.bits-pilani.ac.in

Ayush Sharma
Dept. of EEE
BITS Pilani
Rajasthan, INDIA
f20180326@pilani.bits-pilani.ac.in

Vikram Singh Chundawat
Dept. of CE
BITS Pilani
Rajasthan, INDIA
f20180128@pilani.bits-pilani.ac.in

I. Introduction

A. Background

Due to rapidness and low cost of news dissemination, Social media has now become one of the main channels for people to access and consume news. However using social media as a platform for news updates can make it a hotbed of fake news dissemination. Fake news brings negative impacts on individuals political, economic and social well being. Fake news can become extremely influential and has proven to travel fast at an impressive rate.

With the increase in number of people using social media, the community is at a high risk of being exposed to new misinformation everyday which may be difficult to correct and may have lasting applications. Therefore, automatically detecting fake news has now become a global concern attracting tremendous research efforts.

B. Problem Statement

Detecting fake news via content-based analysis approaches face many challenges. One of the main reasons is that the interpretation of the news as fake or real requires some political context or "common sense" which the current day algorithms are still missing.

Many recent studies have shown that fake and real news spread differently on social media forming propagation patterns which can be used in the detection of fake news. Moreover, propagation-based techniques are flexible to handle adversarial attacks and are language independent. In the current paper, a model based on geometric deep learning is proposed which can automatically detect fake news using the propagation patterns.

C. Motivation

Detecting fake news has now become one of the crucial problems. Fake news has the capability to break the authenticity balance of the news ecosystem. For example, extensive spread of fake news during the U.S. 2016 presidential elections. Fake news is generally manipulated to persuade consumers to accept false or biased beliefs. Moreover fake news also has the power to change the way people interpret and respond to real news. Hence detecting fake news in the early phase of its spread becomes very important.

II. Literature Review

Approaches used for fake news detection typically make use of three types of information which involve- content-based, user-based and propagation-based approaches. Content-based approaches [9] rely on linguistic, semantic, sentiment, writing style and visual-based features of the message being spread to discriminate between real and fake information on social media. However, most content-based approaches are language dependent which limits the generality of these approaches.

User-based features are related to analysis of news distribution and diffusion patterns through social network structure. This analysis is helpful for the identification of fake and unreal accounts in social networks. User-based features use user profiling feature analysis, temporal and posting behaviour analysis.

As compared to user-based and content-based approaches, considerably less research has been dedicated to applying propagation features to fake news detection on social media. Propagation based approaches utilise the fact that fake news travels faster as compared to real news. [3] Several studies have demonstrated that fake and real news present significantly different propagation patterns on Twitter. [4]

However, a more novel approach for fake news detection that aims to optimize the use of propagation features makes use of the recent advances in geometric deep learning. The field of Geometric deep learning refers to the methods that aim to generalize neural network models to non-Euclidean domains

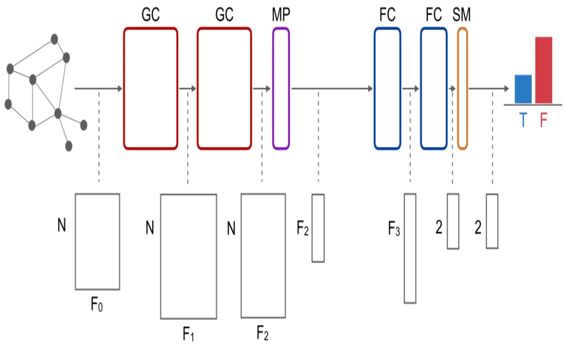

Fig. 1. Proposed Architecture. GC = Graph Convolution, MP = Mean Pooling, FC = Fully Connected, SM = SoftMax layer

such as graphs. Geometric deep learning approaches counter the limitation of most neural network models including CNN that work only on Euclidean data. [2]

III. Proposed Methodology

A. Dataset Details

In the current study authors have collected each 'story' which has an underlying article published on the web. In order to classify the articles as real or fake, they relied on professionals from fact-checking organizations such as Snopes, PolitiFact and Buzzfeed. In total they collected 1084 labeled claims spread on Twitter in 158951 cascades [5].
Features describing users, news, and their activity were organised into 4 categories: User Profile, User activity, Network and spreading and Content. A significant amount of polarization was observed in the social network collected, as the credible and non-credible users tend to form two different clusters. The formation of two distinct groups suggests the fact that fake and true news have different propagation patterns.

B. Model Architecture

One of the most popular deep learning models such as convolutional neural networks (CNNs) are based on classical signal processing theory with an assumption of Euclidean data. Graph CNNs generalise deep learning techniques to non-Euclidean data such as graphs and manifolds.
The proposed learning framework for detecting fake news shown in Fig. 1 consists of a four-layer Graph CNN with two convolutional layers and two fully connected layers. Graph attention is used with every convolutional layer to implement the filters along with mean-pooling for dimensionality reduction. The activation function used is Scaled Exponential Linear Unit (SELU) and Hinge loss is used for loss computation to train the neural network.

C. Input Generation

In Fig. 2, Figure A shows the general connectivity graph for a URL spread over a social platform which tracks the news spread. The red nodes are the local central points of cascade of URL. Grey nodes represent the users that (re)tweet the URL following the cascade. The blue lines denote social relationships of the different users that (re)tweet i.e., follower/followee or friendship.

In Fig. 2, Figure B takes a closer look at a local cascade of the URL '**u**' being tweeted in $T_u = \{ t_0, t_1, t_2 .... t_{n-2}, t_{n-1} \}$ tweets which are authored by $A_u = \{ a_0, a_1, a_2 ...... a_{n-2}, a_{n-1} \}$ respectively. The solid blue arrow denotes a person following another while a dotted arrow indicates an indirect social connection between two people. The graph is drawn with the purpose of estimating news diffusion on a social platform. Hence, we timestamp a set of tweets (T_u), record their authors (A_u) and then attempt to connect the very next tweet t_n and its author a_n to the existing set of nodes and tweets present in the graph. There are two scenarios for a_n (both illustrated in the figure) :

I) The author an follows one of the members of A_u between a_1 and a_n (refer to the figure).
II) The author an does not follow anyone from the A_u directly and thus we consider indirect news spread to an from the member having the maximum number of followers. Consequently, a dotted blue line is drawn to connect a_{n-1} and a_n .

The graph generated in this manner is used as input for the proposed geometric deep learning algorithm.[Fig. 2]

IV. Results and Discussion

A. Findings of the Paper

The model is evaluated in two different settings of fake news detection: URL-wise and cascade-wise. In the first setting, the output label of a URL containing a news story from all twitter cascades is predicted. On the other hand, in the second setting, a label associated with a URL, given only one cascade arising from that URL, is predicted. The proposed method achieved mean ROC AUC of 92.70 ± 1.80% and 88.30 ± 2.74% in the URL- and cascade-wise settings, respectively.
Also, out of the four groups of features defined, user-profile and network/spreading appear as the two most important feature groups, and allow achieving around 90% ROC AUC with the proposed model. This study suggests that just a few (approximately 2) hours of news spread are sufficient to achieve above 90% mean ROC AUC in URL-wise fake news classification.

B. Our Implementation

The dataset used by the authors in the current paper was a commercial dataset which was not available for use, therefore we have used PROTEINS dataset [6] in the current report.
The PROTEINS dataset is used to evaluate and train the proposed architecture. The nodes in protein data represent secondary structure elements and two nodes will be considered as connected, or there will be an edge between two nodes, when they are neighbours in 3D space or amino acid sequence. Each node contains a type attribute and a vector which contains physical and chemical properties of the protein. The

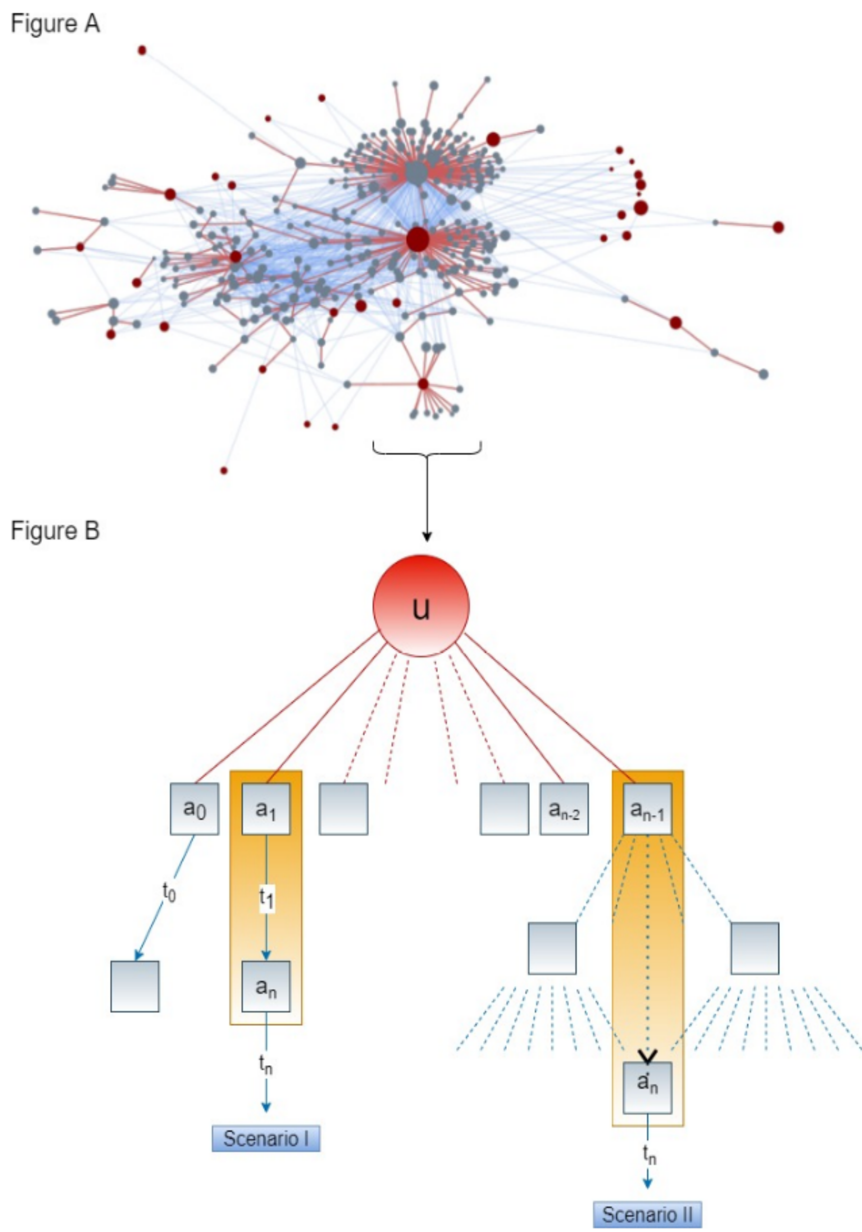

Fig. 2. Input Generation of a single news story spreading on a subset of the twitter social network. Social connections between users are marked by light-blue edges. Circle size represents number of followers. A news URL is tweeted by multiple users each producing a cascade propagating over a subset of the social graph.

type attribute contains whether the protein is helix, sheet or turn.

The task here is to predict whether the given protein is enzyme or non-enzyme. The data is available at torch-geometric dataset module, the dataset was directly fetched from the module and further used for training. Fig. 3 shows the learned graph representations in PROTEINS dataset and also highlights two classes with different colors.

Various hyperparameters were tried for proposed architecture. The best results were observed with the following hyperparameters: learning rate = 0.0005, learning rate decay = 0.1, and batch size = 128. Adam optimizer was used for training. Figure 4, 5 and 6 shows Model loss, F1 Score and Model accuracy on training and validation data respectively.

The results obtained are not par with the state of the art techniques. We just implemented the model proposed for Fake News Detection on a different dataset. In future hyperparamter tuning can be done to get good results.

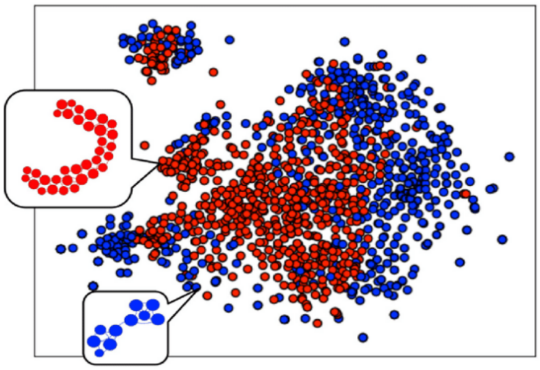

Fig. 3. Learned graph representations in Proteins dataset [1], which has two classes: enzyme (red) and non-enzyme (blue). [8]

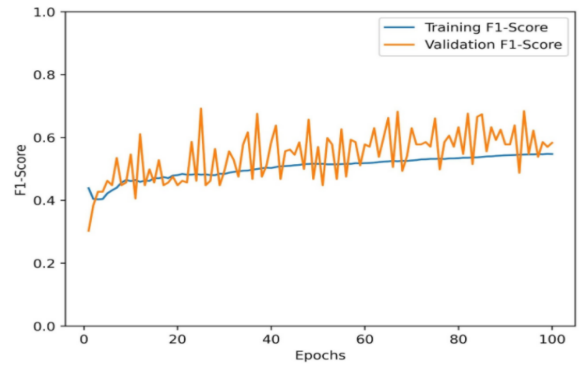

Fig. 5. F1 Score

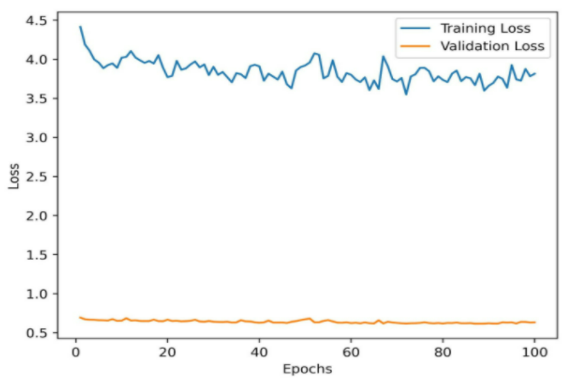

Fig. 4. Model Loss

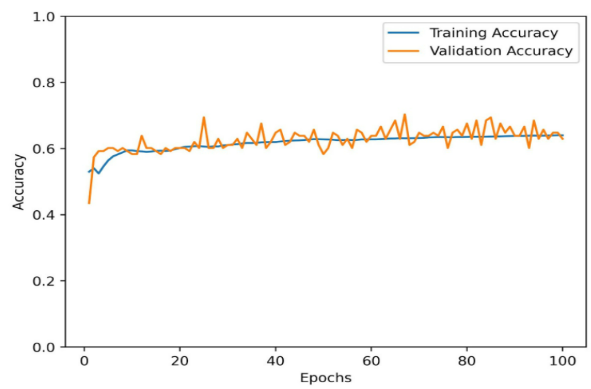

Fig. 6. Model Accuracy

V. Limitations

Because the proposed network architecture classifies a given piece of news as fake or not altogether, if some part of it is not fake, we would not be able to tell by using a propagation based approach.
If a particular article is posted with a comment denying the facts in the article, then understanding the comment also becomes important.
So, using propagation based approaches we might miss out on details given in the content of the tweet which might result in wrong labelling. For example : If a fact checker posts a fake news article, to mark it as fake, that tweet could itself be marked as fake because of the same URL.
Moreover, proper regularization (e.g. dropout or weight decay) should thus be introduced to prevent the model from overfitting and to improve performance at test time.
The model only accounts the sharing/tweeting of URL as the process of news diffusion. However many cascades result from an alternative way to spread the same news i.e., a screenshot spreading or quoting the headlines with just the name of the source not the URL. Hence it loses out on quite a large amount of graph data and wrong news diffusion paths for some users.
The model also does not account for cross platform news diffusion. Every user who may have read information on some other social platform would be treated as a new user if he/she decides to share the information on the platform we are studying. Moreover, it may indirectly affect the propagation pattern on the platform we currently are studying too.

VI. Suggestions

A deep hierarchical co-attention network to learn feature representations for fake news detection and explainable sentences/comments discovery is suggested by [7], which aims to explain why a particular piece of news is detected as fake.
This approach also ranks sentences and user comments that led to the news being detected as fake. So a plausible future scope lies in uniting propagation based approach and content based approach.
A hybrid model can be developed which incorporates both the aspects, which are propagation based approach and content based approach. This will not only improve the performance, but also ensure that the tweets from fact checkers are not marked as fake as mentioned in the above example. Our assumption is that all the cascades associated with a URL inherit the label. Hence, leaving the analysis of comments accompanying tweets/retweets as a future research direction.

Social network has a dynamic nature since the user preferences and news topics evolve in time, it is very crucial that a model trained in the past has the ability to generalise well to new circumstances. Therefore there is a need to build a model that is robust to aging and is able to learn features that are discriminative.
A simple solution to this problem is to try and use Encoder-Decoder based approach and extract out the latent features from the latent space. Then these latent features can further be used for classifying the news as real or fake. This technique is somewhat similar to feature selection wherein the important features will be automatically extracted from the latent space and further used for classification.
Since in the recent years many different social platforms, each with their unique user connection criteria, have come to existence with each having a huge user base, it becomes essential to add the ability to add more social platforms as well as including users who might not be present on all the platforms so that a comprehensive and accurate classification can take place.
Experimental validation of the conjecture that the proposed model is potentially geographic and language-independent is left for future research. Moreover additional applications of the proposed model going beyond binary classification (Fake news detection) to multi-class classification such as news topic classification can be tried out.

VII. Conclusions

This report demonstrated the potential of using geometric deep learning combined with propagation features to differentiate fake news from real news on Twitter. It has been observed that fake news travels faster as compared to real news and there are significant differences in their propagation. [3]
Additionally, most of the machine learning based approaches work only on Euclidean data. The use of geometric deep learning has facilitated us to overcome this limitation of convolutional neural network-based approaches and allowed us to adapt to higher dimensional data such as graphs. [4] The method proposed in the paper based on geometric deep learning works as a unifying framework for content, social context, and propagation-based approaches. Overall, the good performance of the proposed algorithm leads us to conclude that propagation features are very important and relevant for fake news detection on Twitter.

References

[1] Karsten M Borgwardt, Cheng Soon Ong, Stefan Schönauer, SVN Vishwanathan, Alex J Smola, and Hans-Peter Kriegel. Protein function prediction via graph kernels. *Bioinformatics*, 21(suppl_1):i47–i56, 2005.

[2] Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: Going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

[3] Adam Kucharski. Post-truth: Study epidemiology of fake news. *Nature*, 540:525–525, 12 2016.

[4] Marion Meyers, Gerhard Weiss, and Gerasimos Spanakis. *Fake News Detection on Twitter Using Propagation Structures*, pages 138–158. 10 2020.

[5] Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M. Bronstein. Fake news detection on social media using geometric deep learning, 2019.

[6] Ida Schomburg, Antje Chang, Christian Ebeling, Marion Gremse, Christian Heldt, Gregor Huhn, and Dietmar Schomburg. Brenda, the enzyme database: updates and major new developments. *Nucleic Acid Research*, 2004.

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

[8] Aynaz Taheri, Kevin Gimpel, and Tanya Berger-Wolf. Sequence-to-sequence modeling for graph representation learning. *Applied Network Science*, 4(1):1–26, 2019.

[9] Xinyi Zhou and Reza Zafarani. A survey of fake news. *ACM Computing Surveys*, 53(5):1–40, Oct 2020.