FAKE NEWS DETECTION WITHOUT TEXTUAL ANALYSIS

Raphael Faure, Maria Roman

université PARIS-SACLAY
CentraleSupélec

Introduction

The propagation of fake news on social media remains a significant issue, affecting public trust and society at large. This study focuses on using Graph Neural Networks (GNNs) to detect fake news by analyzing their unique propagation patterns without relying on textual information like content or user descriptions, and tackles the issue of handling unseen data efficiently without retraining from scratch.

Method & model used

Data Generation

- **Dataset:** FakeNewsNet, labeled data from PolitiFacts & GossipCop
- **Propagation Graphs:** nodes are tweets and retweets, and edges signify information flow

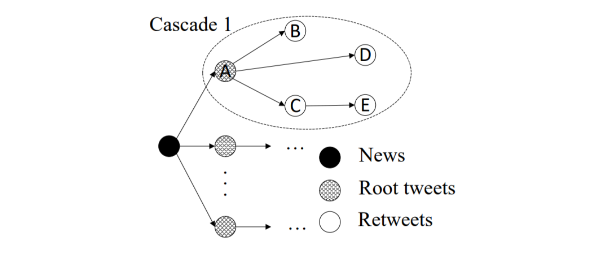

Figure 1: An illustration of the graph for each item of news.

Each graph is associated with an **adjacency matrix** that represents the structure of the graph, and a **feature matrix** that represents the features of each node.

Features: caracteristics of the tweet except from the text (whether the user is verified, when the user was created, number of followers, number of friends, timestamp of the tweet, ...)

Model architecture

The final model used for fake news detection is a DiffPool model built on top of GraphSage. The choice of model architecture varied depending on the dataset:

- PolitiFact Dataset: A four-layer GNN with 64 hidden dimensions and 64 embedding dimensions was selected.
- GossipCop Dataset: A two-layer GNN was chosen due to the larger size of this dataset, ensuring better performance.

Dealing with new data

Challenge: GNNs trained on one dataset often fail on another (e.g., PolitiFact vs. GossipCop).

Solution:

- **Incremental training:** further train the model obtained from one dataset on the other dataset
- **Continual learning methods** to prevent catastrophic forgetting
  - Gradient Episodic Memory (**GEM**): Retains key samples from old tasks to prevent knowledge loss.
  - Elastic Weight Consolidation (**EWC**): Penalizes significant changes to parameters critical to previous tasks.

Optimization problem under GEM

$$\min_{\theta} \sum_{(G_i,y_i)\in\mathcal{D}_2} \text{loss}\left(f\left(A_i^{(k)}, H_i^{(k)}; \theta^{(k)}\right), y_i\right),$$
$$\text{subject to} \sum_{(G_j,y_j)\in\mathcal{M}} \text{loss}\left(f\left(A_j^{(k)}, H_j^{(k)}; \theta^{(k)}\right), y_j\right) \leq \gamma$$

Loss function under EWC

$$\sum_{(G_i,y_i)\in\mathcal{D}_2} \text{loss}\left(f\left(A_i^{(k)}, H_i^{(k)}; \theta^{(k)}\right), y_i\right) + \frac{\lambda}{2} F\left(\theta - \theta^*_{\mathcal{D}_1}\right)^2$$

Alternative methods

Existing studies can be broadly categorized into three approaches:

- **Content-based Approaches**: these methods analyze news headlines and body content to verify validity.
- **Context-based Approaches**: these rely on user interactions (e.g., retweets, replies) to assess credibility. Thus, **propagation-based methods** fall into this category.
- **Mixed Approaches**: these integrate both content features and social interactions.

Results comparison

Performance on Single Datasets

- GNNs achieve accuracy **comparable or superior** to text-based methods.
- **Early-stage fake news detectio**n using limited propagation data is effective.
- Models trained on one dataset **perform poorly** on the other

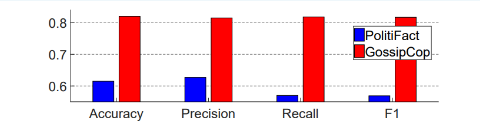

Figure 5: Models first trained on the clipped dataset of PolitiFact and then on GossipCop only perform well on the latter dataset on which it is trained, *i.e.*, GossipCop.

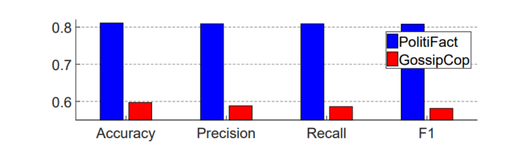

Figure 4: Models trained on the clipped dataset of PolitiFact perform poorly on the dataset of GossipCop.

Efficiency: GNNs converge quickly and handle large-scale data efficiently, and no significant difference in training time between GEM and EWC.

Performance Across Datasets

- Incremental training without continual learning leads to catastrophic forgetting.
- GEM outperforms EWC in balancing performance across datasets.

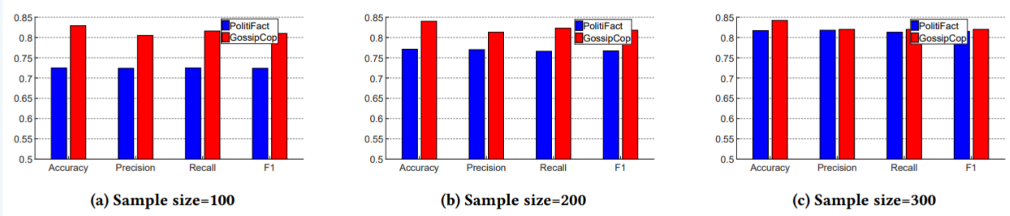

Figure 6: Performance of models first trained on the clipped dataset of PolitiFact and then on GossipCop using GEM.

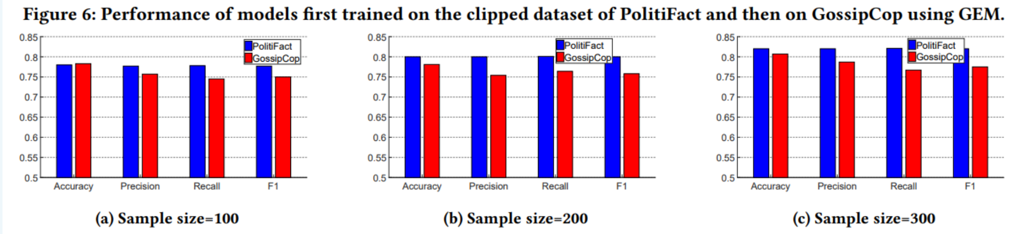

Figure 7: Performance of models first trained on the clipped dataset of GossipCop and then on PolitiFact using GEM.

Our contributions

1. We managed to reproduce the results of this study and found similar results.

2. We introduced a **new penalization method** using the **KL divergence** in order to prevent catastrophic forgetting. This way, we make sure that our probabilistic distribution on a new dataset stays close to the one we learnt on the previous dataset, and thus our model doesn't start from scratch. Indeed, we keep a subset of data from the old dataset, and using the Kullback-Leibler divergence, we force the model to ensure that the new distribution remains similar to the old one on this subset.

$$\min_{\theta} \sum_{(G_i,y_i)\in\mathcal{D}_2} \text{loss}\left(f\left(A_i^{(k)}, H_i^{(k)}; \theta^{(k)}\right), y_i\right) + \lambda \sum_{(G_j)\in\mathcal{M}} \bar{f}(A_j^{(k)}, H_j^{(k)}; \theta^{(k)}) \log\left(\frac{\bar{f}(A_j^{(k)}, H_j^{(k)}; \theta^{(k)})}{\bar{f}(A_j^{(k)}, H_j^{(k)}; \theta^{(k)}_{old})}\right)$$

We have to perform now a cross-validation to check which size of the kept subset and value for the penalty are the best.

Conclusion

This study constitutes an empirical evidence that fake news and real news spread differently online, making propagation-based methods a great alternative to textual analysis.
However, a critical limitation is the poor generalization of GNNs trained on one dataset to new datasets with different graph structures. Thus, improvements on continuous learning and feature selection seem to hold the key to achieving robust performances.

References

1. Yi Han Shanika Karunasekeran Christopher Leckie (2020), *Graph Neural Networks with Continual Learning for Fake News Detection from Social Media*
2. Huyen Trang Phan, Ngoc Thanh Nguyen, Dosam Hwang (2023), *Fake news Detection: A survey of graph neural network methods*