Geometric Data Analysis: Fake News Detection on Social Media using Geometric Deep Learning

Niloo Sahebalzamany
Columbia University
ns2434@caa.columbia.edu

Clément Rouvroy
Ecole normale supérieure
clement.rouvroy@ens.psl.eu

Grégoire Le Corre
Ecole normale supérieure
gregoire.le.corre@ens.psl.eu

Abstract

Social media platforms have revolutionized news consumption but have simultaneously amplified the spread of fake news, posing critical challenges to information integrity. Traditional detection methods relying on content analysis struggle with nuanced misinformation requiring contextual or common-sense understanding. This paper introduces a novel geometric deep learning framework leveraging graph-based propagation features for fake news detection. By analyzing propagation dynamics, user behaviors, and network structures, our model achieves language independence and resistance to adversarial attacks. Tested on verified Twitter datasets spanning 2013–2018, their method attains 92.7% ROC AUC in URL-wise classification and demonstrates reliable early detection within hours of news dissemination. This study highlights the potential of propagation-based detection as a robust alternative or complement to content-focused approaches, with applications extending to scalable, real-time solutions for combating misinformation on social media.

1 Introduction

Social media platforms are prominent sources of news today, offering rapid dissemination but posing significant risks such as the spread of fake news. Fake news often relies on misinformation and manipulation, challenging traditional content-based detection methods. Geometric deep learning provides a promising alternative, analyzing how news spreads within social networks rather than its content. This method leverages graph-based data structures to integrate content, user profiles, and propagation patterns, allowing early and accurate fake news detection. This report examines these methods and evaluates their effectiveness using real-world datasets, expanding upon the work of Monti et al. [1]

2 Literature Review

2.1 Previous Approaches to Fake News Detection

Content-based methods focus on linguistic features to identify deceptive cues, while social context-based approaches consider user demographics and interactions. Propagation-based methods leverage the spread patterns of news across social networks, offering greater resilience to adversarial attacks.

2.2 Geometric Deep Learning in Social Media Analysis

Geometric deep learning has been applied to tasks like social network analysis, misinformation detection, and community detection. By generalizing neural networks to non-Euclidean data, these methods efficiently model graph-structured information.

3 Methodology

Fake news detection is formulated as a graph classification problem, where nodes represent users and edges capture social interactions or news propagation paths.

3.1 Data

The dataset consists of annotated fake and real news stories spread on Twitter between 2013 and 2018, verified by professional fact-checking organizations such as Snopes, PolitiFact, and Buzzfeed. Each story is linked to a URL and propagated as a cascade, defined by a source tweet referencing the URL and its subsequent retweets. The dataset includes the following:

- **Claims**: 1,084 labeled claims, each linked to an article and classified as true or false.
- **Cascades**: 158,951 cascades representing propagation patterns of news on Twitter.
- **Users and Connections**: 202,375 unique users connected by 2,443,996 edges in the Twitter social graph.

Features. The dataset provides rich feature sets categorized as:

- **User profile:** Includes geolocalization, language, profile settings, verification status, word embeddings of self-descriptions, and account creation dates.
- **User activity:** Metrics such as the number of favorites, lists, and statuses.
- **Network and spreading:** Social connections (followers and friends), cascade tree structures, retweet timestamps, source devices, and counts of replies, quotes, favorites, and retweets for the source tweet.
- **Content:** Word embeddings of tweet content and hashtags.

Cascade Size. The average cascade size is 2.79 tweets, with a significant portion of cascades being small. To focus on meaningful propagation patterns, cascades with fewer than six tweets were excluded from some experiments.

Credibility and Polarization. Users were assigned a credibility score ranging from -1 to +1, reflecting the proportion of true versus fake news they propagated. The social network demonstrated significant polarization, with credible (blue) and non-credible (red) users forming distinct communities.

3.2   Model Architecture

The proposed model is a Graph Convolutional Network (GCN) , [2] designed for fake news classification using propagation patterns [3]. It integrates heterogeneous data, including user profiles, tweet content, and social interactions, into a unified framework. The architecture consists of the following components:

Input Graph Representation: Each news cascade is represented as a directed graph. The nodes represent tweets or users, and are characterized by several features: - *User Activity:* Features such as retweet count, account age, and the number of friends and followers. - *Content Embeddings:* 200-dimensional GloVe embeddings pre-trained on Twitter data. - *Metadata:* Includes timestamps, verified status, and geolocation.

The edges in the graph capture propagation relationships such as retweets or follows, with each edge enriched with features indicating the type and direction of interaction.

Graph Convolutional Layers: The model employs two graph convolutional layers to compute node embeddings: - Each layer outputs 64-dimensional feature vectors for the nodes. - A single-head attention mechanism is applied to assign weights to edges based on their significance. - Scaled Exponential Linear Unit (SELU) activation is applied after each convolution operation.

Pooling Layer: Mean pooling aggregates the node-level embeddings into a single graph-level feature vector that summarizes the cascade.

Fully Connected Layers: The graph-level representation is passed through the following layers: - A dense layer reduces the feature dimension to 32. - A final output layer with 2 units performs binary classification (fake/true news). **Loss and Optimization:** - The model is trained using hinge loss, as it outperformed cross-entropy in early experiments. - AMSGrad is used as the optimizer, with a learning rate of 5×10^{-4}. - Training is performed for 25,000 iterations for URL-wise classification and 50,000 iterations for cascade-wise classification, with a mini-batch size of 1.

Early Propagation Settings: The model can classify news based on early propagation patterns by considering truncated cascades: - Cascades are truncated at intervals (e.g., 2 hours, 7 hours, and 24 hours) to analyze detection performance over time. - A minimum cascade size of 6 tweets is required for classification, as smaller cascades lack sufficient propagation patterns.

This architecture leverages propagation dynamics, graph structure, and multi-modal features to provide robust and early detection of fake news.

3.3   Mathematical Framework of Geometric Deep Learning

Geometric deep learning generalizes classical deep learning techniques [4] to non-Euclidean domains such as graphs and manifolds [5]. This paper leverages these techniques to model social media propagation patterns as graph-structured data, enabling the detection of fake news.

Graph Representation. A social media cascade is represented as a graph $G = (V, E)$, where V denotes the nodes (tweets or users), and E denotes the edges (retweets, follows, or interactions). The adjacency matrix $\boldsymbol{A}$ encodes these connections, while the degree matrix $\boldsymbol{D}$ represents node degrees. The normalized graph Laplacian $\boldsymbol{L}_{\mathrm{norm}}$ is:

$$\boldsymbol{L}_{\mathrm{norm}} = \boldsymbol{I} - \boldsymbol{D}^{-1/2} \boldsymbol{A} \boldsymbol{D}^{-1/2}.$$

Graph Convolution. The model utilizes spatial graph convolutions, which iteratively aggregate features from neighboring nodes. For a node i, the feature update at layer $l + 1$ is:

$$\boldsymbol{h}_i^{(l+1)} = \sigma \left(\sum_{j \in \mathcal{N}(i)} \frac{1}{c_{ij}} \boldsymbol{W}^{(l)} \boldsymbol{h}_j^{(l)} \right),$$

where $\boldsymbol{h}_j^{(l)}$ is the feature vector of node j at layer l, $\boldsymbol{W}^{(l)}$ is the learnable weight matrix, c_{ij} is a normalization constant, and σ is the activation function (e.g., Scaled Exponential Linear Unit, SELU).

Attention Mechanism. Graph Attention Networks (GATs) enhance traditional convolution by dynamically weighting edges. The attention coefficient α_{ij} for nodes i and j is computed as:

$$\alpha_{ij} = \frac{\exp(\mathrm{LeakyReLU}(\boldsymbol{a}^\top [\boldsymbol{W}\boldsymbol{h}_i || \boldsymbol{W}\boldsymbol{h}_j]))}{\sum_{k \in \mathcal{N}(i)} \exp(\mathrm{LeakyReLU}(\boldsymbol{a}^\top [\boldsymbol{W}\boldsymbol{h}_i || \boldsymbol{W}\boldsymbol{h}_k]))},$$

where $\boldsymbol{a}$ is a learnable vector, and $[\cdot || \cdot]$ denotes concatenation.

Propagation Modeling. News propagation paths are constructed by combining retweet timestamps and social graph structures. These paths define the graph G, where node and edge features encode user activity, content, and propagation dynamics. This graph is processed iteratively by the geometric deep learning model to learn propagation-specific patterns indicative of fake or real news.

3.4   Model Settings

The model operates in two distinct settings:

- **URL-wise Classification**: In this setting, the model aggregates information from all cascades linked to a specific URL to determine its veracity. On average, each URL has 141 cascades.
- **Cascade-wise Classification**: Here, the model evaluates a single cascade independently to predict the label of its associated URL. Cascades with fewer than six tweets are excluded to ensure sufficient propagation data.

Table 1: Dataset Statistics

Dataset	#Graphs (#Fake)	#Total Nodes	#Total Edges	#Avg. Nodes per Graph
Politifact (POL)	314 (157)	41,054	40,740	131
Gossipcop (GOS)	5464 (2732)	314,262	308,798	58

Each setting is trained using supervised learning, with hinge loss as the objective function. The URL-wise setting benefits from aggregated data, while the cascade-wise approach focuses on individual propagation patterns, making it more challenging but also more granular.

Summary of Results. The model demonstrated strong performance across both settings, as measured by the area under the ROC curve (ROC AUC):

- **URL-wise Classification**: Achieved an average ROC AUC of 92.70%, indicating high accuracy when aggregating information from all cascades linked to a URL.
- **Cascade-wise Classification**: Despite the increased difficulty of evaluating single cascades, the model achieved an impressive average ROC AUC of 88.30%.

Further analysis revealed that cascade size significantly impacts classification performance. Larger cascades provide more reliable propagation patterns, with a minimum threshold of six tweets per cascade optimizing results. Moreover, the model exhibited robust early detection capabilities, achieving above 90% accuracy within a few hours of news propagation in the URL-wise setting.

4 Implementation

Our implementation is publicly accessible here. We employed the model architecture described in Monti et al.[1] as detailed in Section 3.2, with a modification to the batch size, which we set to 128 instead of 1 to improve training efficiency and handle larger data batches.

In this study, we leverage URL-wise classification to determine the veracity of a news story by aggregating information from all Twitter cascades associated with a given URL. This approach aligns naturally with the organic structure of our dataset, which is inherently organized at the URL level. When utilizing all available features, the model achieved a robust Area Under the Receiver Operating Characteristic curve (AUC-ROC) score of 96%, underscoring the efficacy of feature integration in enhancing classification performance.

Data: The dataset used by Monti et al. [1] is not publicly available. Instead, we leveraged the publicly accessible dataset introduced by Dou et al.[6], which differs slightly in its construction. Specifically, each news diffusion on Twitter is modeled as an undirected graph, with the news itself serving as the root node (rather than the tweet citing it). Nodes directly connected to the root represent tweets that reference the news URL, while the edges between these nodes capture the retweet relationships of those initial tweets.

Node Features: In addition to the graph structure, we incorporated various node features derived from the User
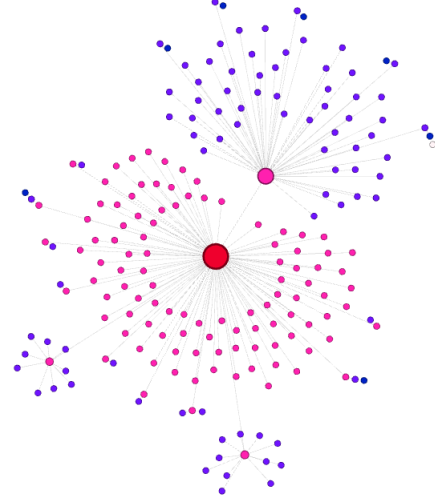

Figure 1: Graph representing an URL cascade. Color depends on depth (i.e. number of reweet) and size represents degree. The big node red is the URL, node it pink tweets containing URL, node in blue first retweet, ...

Preference-aware Fake News Detection dataset (UPFD). These features include:

- **Spacy:** A 300-dimensional word2vec embedding representing the historical content of the user's profile.
- **BERT:** A 768-dimensional embedding capturing the semantic content of the user's profile using BERT.
- **Profile:** A 10-dimensional vector encapsulating key profile attributes, including the username, Twitter handle, and follower count.
- **Content:** A concatenated vector combining:
  (1) A 300-dimensional word2vec embedding of the tweet's content.
 (2) The 10-dimensional profile vector described above.

Figure 1 provides a visualization of an example URL cascade from our dataset, highlighting its graph representation. Table 1 and Figure 2 summarizes the dataset statistics, offering a detailed overview of its composition and key characteristics.

5 Experiments

This study follows the methodology of Monti et al. [1] to explore the impact of different feature combinations on the performance of our model using the UPFD dataset. We performed an ablation study to evaluate how various feature
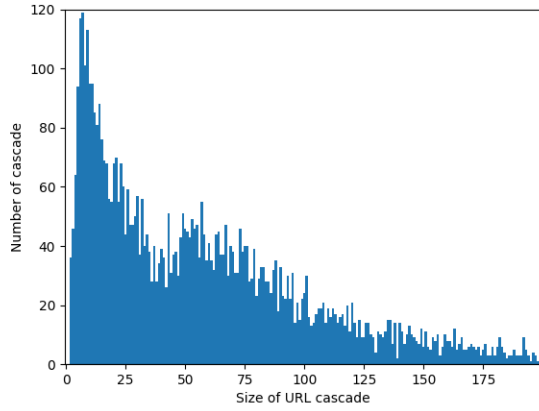
Figure 2: Breakdown of the number of cascading URLs by size (number of nodes)
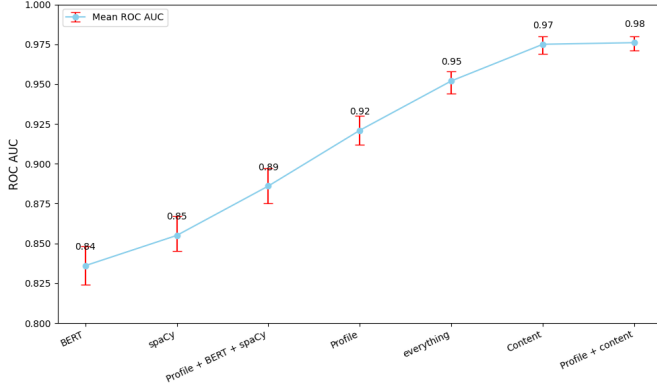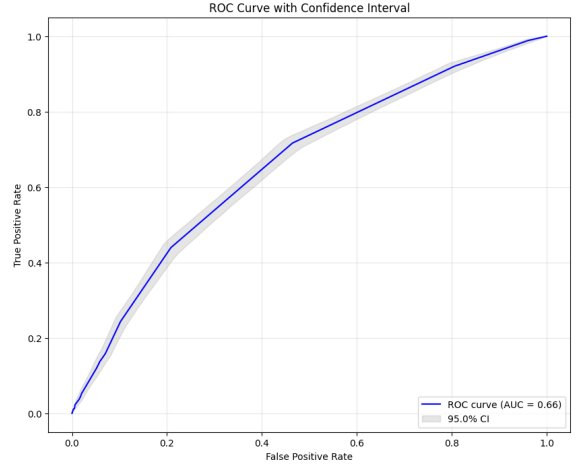

Figure 3: Ablation study on our implementation

Figure 4: ROC AUC of a model with random features on nodes.

sets contribute to model performance. The results, shown in Figure 3, reveal significant improvements as specific features are incorporated into the model.

The inclusion of user profile and preferences resulted in an ROC-AUC score of 0.88, demonstrating the importance of user-specific contextual information. Further isolating the profile features led to a substantial increase in performance to 0.93, underscoring their predictive strength. When all features—profile, preferences, content, and interactions—were combined, the model achieved a score of 0.96, showcasing the positive impact of a diverse feature set.

However, contrary to expectations, the best performance of 0.98 was observed when only the profile and content features were used together. This result suggests that while adding additional features provides some improvements, the combination of profile and content features outperforms the full set of four features. The inclusion of all four features introduced some redundancy, which may have detracted from the model's overall performance.

These findings align with the observations made by Monti et al. [1], where feature selection played a pivotal role in enhancing model accuracy. In our case, the profile and content features, when combined, provided the most effective representation of both user-specific and tweet-specific contexts, thereby achieving the highest predictive accuracy. This highlights the importance of optimizing feature combinations and avoiding unnecessary complexity, which can lead to diminishing returns.

Overall, this study emphasizes the critical role of strategic feature engineering. By carefully selecting the most informative features—particularly profile and content—we were able to achieve optimal performance without the added complexity of redundant features.

6 Extensions

6.1 Random Features and the Role of Graph Propagation in Fake News Diffusion

Building on the idea of propagation and the patterns of tweet diffusion, we explored the effect of using random features in our analysis. This approach aimed to isolate the structural properties of the graph from the influence of carefully crafted features. Notably, despite the randomness of the features, the model achieved an AUC-ROC score of 0.66, as illustrated in Figure 4, suggesting that the graph itself inherently encodes valuable information.

This finding underscores the critical role of graph structure in distinguishing fake news from legitimate content. The propagation patterns unique to fake news dissemination appear to manifest in the graph topology, independent of specific node-level attributes. For instance, fake news tends to propagate through distinct clusters or patterns of re-sharing that differ significantly from genuine news diffusion. Such differences are captured by the graph's structural properties, reinforcing the
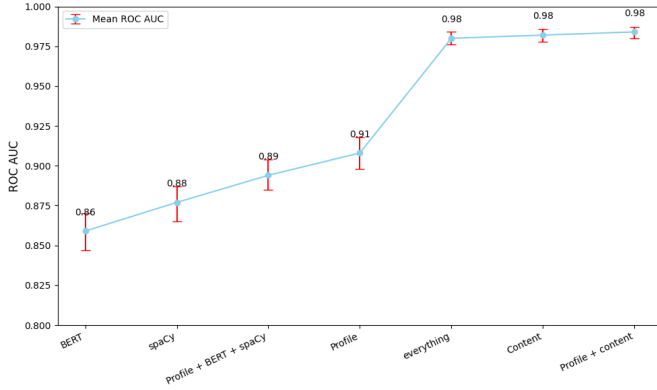
Figure 5: Ablation study with weight decay.

idea that graph-based representations can be highly effective for tasks like fake news detection.

The implications of these results are twofold:

- **Feature Agnosticism:** Even without domain-specific features, the graph's structural information contributes meaningfully to classification performance. This suggests that algorithms leveraging the graph alone, such as Graph Neural Networks (GNNs) or propagation-based heuristics, could yield robust results in detecting fake news.
- **Future Directions:** Future work could focus on integrating random features with learnable embeddings to further disentangle the contributions of structure and content. Moreover, analyzing specific subgraph patterns or motifs prevalent in fake news diffusion might help refine detection strategies further.

6.2 Impact of Regularization (Weight Decay) on Model Performance

Incorporating regularization techniques, particularly weight decay, was another avenue we explored to improve the robustness and generalization of our models. Weight decay, by penalizing large weights, aims to prevent overfitting and encourage smoother decision boundaries. However, our findings revealed that regularization did not consistently improve the AUC-ROC performance across the majority of models. In some cases, it even led to a decrease in AUC-ROC scores. Results are depicted in 5.

This counterintuitive outcome suggests that weight decay may lead to over-regularization in the context of our models. By overly constraining the parameter space, weight decay might prevent the model from capturing subtle patterns in the data that are crucial for distinguishing between fake and genuine news. These results align with the hypothesis that excessive regularization can hinder a model's ability to effectively learn from the data, especially when the underlying patterns are complex and nuanced, as is often the case with fake news detection.

The key takeaways from this observation are as follows:

- **Trade-off Between Generalization and Expressiveness:** Regularization introduces a bias toward simpler models, which can be beneficial for noisy data. However, in our case, the reduction in expressiveness outweighed the benefits, leading to suboptimal performance.
- **Sensitivity to Model Complexity and Data Characteristics:** The adverse impact of weight decay highlights the importance of tailoring regularization techniques to the specific characteristics of the data and the model. The diffusion patterns in fake news graphs, combined with potentially high-dimensional feature spaces, may require a more nuanced approach to regularization.
- **Future Work on Adaptive Regularization:** Instead of a uniform weight decay penalty, adaptive regularization strategies, such as dropout or layer-specific penalties, could be explored to better balance generalization and expressiveness. Additionally, analyzing the relationship between regularization strength and graph-based feature importance may shed further light on the interplay between model constraints and structural learning.

In summary, our extensions provide valuable insights into the interplay between graph structure, feature engineering, and regularization in fake news detection. The strong performance of random features highlights the intrinsic value of graph-based propagation patterns, while the mixed results with weight decay emphasize the need for a careful calibration of regularization strategies. These findings open avenues for further exploration into the synergy between graph representation learning and model optimization techniques in combating misinformation.

7 Limitations

This study presents several limitations that warrant consideration, particularly regarding the generalizability and contextual applicability of the proposed model:

- **Dataset Specificity**: The model was trained using a Twitter dataset annotated with fact-checking labels. Consequently, its ability to generalize to other social media platforms, particularly those with less structured or differently moderated environments, remains uncertain.
- **Dependence on Propagation Patterns**: The model heavily relies on propagation dynamics, which may not be as prominent or accessible in less networked or sparsely connected settings. This reliance could limit its utility in platforms where such patterns are difficult to observe or quantify.
- **Temporal Degradation**: Findings indicate that the model's performance degrades over time, highlighting the necessity of frequent retraining to sustain accuracy, especially in the context of rapidly evolving news ecosystems and information landscapes.

- **Dataset Comparability**: Unlike Monti et al. [1], who utilized a fact-checker-verified Twitter dataset, this study employs a dataset introduced by Dou et al. [6], which represents news diffusion using undirected graphs with news URLs as root nodes. This methodological variation might constrain the direct comparability of results across different datasets, complicating broader benchmarking efforts.
- **Dependence on Graph Richness and Data Quality**: The effectiveness of the model is contingent on the availability of detailed graph structures and high-quality user profile features. This dependency could limit the model's generalizability to other datasets or platforms characterized by sparse, noisy, or incomplete data.

These limitations underscore the need for further exploration into adaptive methodologies and cross-platform evaluations to enhance the robustness and applicability of the model.

8   Results and Discussion

The effectiveness of the proposed model was evaluated using a dataset that introduced novel characteristics and advanced feature engineering. This section discusses the model's performance, the significance of individual feature sets, and insights gained through analytical experiments.

8.1   Model Performance

The model demonstrated strong predictive capabilities, achieving a 96% AUC-ROC for URL-wise classification when leveraging the full suite of features. This result underscores the robustness of integrating user profiles, content attributes, and propagation patterns. Notably, these findings align with prior studies emphasizing the efficacy of such multifaceted approaches.

Interestingly, the dataset analysis revealed nuanced interactions among the feature sets. While propagation patterns and user profiles were corroborated as critical components consistent with Monti et al. [1]'s observations, the combination of profile and content features alone yielded the highest AUC-ROC at 98%. This suggests potential redundancies in the simultaneous integration of all feature sets. These insights indicate that strategic feature selection may enhance computational efficiency without compromising accuracy.

8.2   Feature Analysis

A detailed ablation study further validated the importance of specific feature groups. Profile features, such as BERT embeddings and follower counts, emerged as pivotal for distinguishing fake news. These results resonate with prior research, reaffirming the value of user-centric metrics in addressing misinformation.

To probe the underlying discriminative power of graph-based representations, experiments using random features were conducted. Remarkably, even with the absence of domain-specific information, the graph structure achieved a 66%

AUC-ROC. This finding highlights the intrinsic ability of graph topology to capture propagation dynamics, reinforcing the hypothesis that structural characteristics alone offer significant discriminatory potential between fake and authentic news dissemination patterns.

8.3   Implications and Future Directions

The findings underscore the importance of leveraging graph-based insights and user profile attributes in detecting misinformation. However, the observed redundancies in feature integration suggest the need for further exploration into feature optimization strategies. Future work could focus on developing adaptive feature selection mechanisms to refine model efficiency while maintaining or enhancing predictive performance.

9   Conclusion and Future Work

This study demonstrates the potential of geometric deep learning in fake news detection, effectively integrating user profiles, activity patterns, propagation dynamics, and semantic content. The model achieves high predictive accuracy and adaptability across diverse datasets, showcasing its robustness and early detection capabilities.

Achievements

The proposed approach successfully generalized beyond Twitter data, maintaining strong performance while revealing the nuanced contributions of graph-based propagation and user-level features.

Challenges

Key challenges included:

- Data scarcity and structural differences across networks.
- Biased propagation patterns, which impacted model consistency.

Future Work

To address these challenges and extend the study, future directions include:

- **Cross-Platform Validation**: Extend the model to multilingual and platform-agnostic datasets (e.g., Facebook, Instagram) to evaluate generalizability.
- **Temporal Adaptation**: Implement dynamic retraining strategies to counteract model aging and adapt to evolving news ecosystems.
- **Feature Engineering**: Explore advanced embedding techniques and feature selection methods to refine performance and reduce redundancy.
- **Hybrid Approaches**: Investigate methods that combine propagation-based and content-based analyses to enhance accuracy and adversarial resilience.

- **Graph Analysis and Regularization**: Study graph motifs specific to fake news diffusion and apply adaptive regularization techniques to balance generalization and expressiveness.

This work establishes a strong foundation for advancing fake news detection through innovative graph-based methodologies, paving the way for more interpretable, resilient, and scalable solutions.

References

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

[2] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems (NIPS)*, volume 29, pages 3844–3852, 2016.

[3] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2017.

[4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[5] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2014.

[6] Yingtong Dou, Kai Shu, Congying Xia, Philip S. Yu, and Lichao Sun. User preference-aware fake news detection. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, page 2051–2055, New York, NY, USA, 2021. Association for Computing Machinery.