Who Drives Misinformation? Key Node Detection with Heterogeneous Graph Neural Networks

Liliana Martirano(✉), Francesco Scala, Carmela Comito, and Luigi Pontieri

Institute of High Performance Computing and Networking (ICAR-CNR), Via P. Bucci, 87036 Rende, CS, Italy
{liliana.martirano,francesco.scala,carmela.comito, luigi.pontieri}@icar.cnr.it

Abstract. Misinformation propagation in online networks involves multifaceted interactions between users, contents, and engagement mechanisms (likes, shares, comments). Addressing this issue entails both understanding how information spreads and identifying influential users driving the dissemination process. To tackle these challenges, this paper proposes a framework based on a Graph Attention Network model, applied to a heterogeneous graph representing social interactions and context-aware dynamics. Targeting the binary classification of real vs fake news, it offers insights into both propagation patterns and influential users in the dissemination process. A core contribution is the adoption of two post-hoc mechanisms for uncovering such users: uncertainty-based Active learning-like and GNN-Explainer. A detailed comparative analysis reveals that nodes where the model exhibits the highest confidence often lack rich content information; nevertheless, combining both high-confidence and content-rich nodes grasps complementary aspects and better aligns with influential users in information propagation. The framework is benchmarked against traditional centrality measures, widely used to identify influential users in social networks. A comparative evaluation on two heterogeneous, real-world, social networks confirms that the proposed method both achieves compelling accuracy in finding influential nodes and shows a potential to scale-up to densely-connected graphs on which classic approaches may fail.

Keywords: Graph Neural Networks · GNN-Explainer · Active Learning · Misinformation · Influential nodes

1 Introduction

The proliferation of misinformation in online social networks presents a significant challenge, influencing public opinion, shaping societal discourse, and even affecting democratic processes. With the increasing reliance on digital platforms for information consumption, the rapid and widespread dissemination of misleading content has become a pressing issue. This phenomenon is driven by

S. Džeroski et al. (Eds.): DS 2025, LNAI 16090, pp. 47–62, 2025.
https://doi.org/10.1007/978-3-032-05461-6_4

complex interactions among users, content, and engagement mechanisms such as likes, shares, and comments. Consequently, understanding how misinformation spreads and identifying the key users responsible for its propagation are crucial for developing effective mitigation strategies.

Traditional approaches to influence analysis in social networks often rely on graph-based centrality measures such as degree centrality, closeness centrality, betweenness centrality [16] and PageRank [2], or influence-spreading heuristics related to centrality like VoteRank [17]. While these methods provide valuable insights into network structures, they are not directly applicable to heterogeneous networks, i.e., graphs with multiple node and/or edge types. Moreover, computational constraints make them less suitable for large-scale social media datasets, necessitating the development of more scalable and adaptive solutions. Finally, existing methods often assume uniform importance among neighboring nodes and fail to account for the structural and semantic diversity of relationships in heterogeneous networks.

To address these limitations, we propose a meta-path embedding-based Graph Attention Network (GAT) model [6] to identify influential nodes in heterogeneous networks. The framework leverages GAT and meta-paths applied to heterogeneous social graphs to capture multi-typed dynamics and model misinformation propagation patterns. The proposed model performs binary classification of real and fake news while simultaneously uncovering key users driving information dissemination. A key contribution of this work is the integration of post-hoc techniques to identify influential nodes in the network. Specifically, we explore two complementary approaches: (1) Uncertainty-based Active Learning-like (AL-like) approaches, which prioritize nodes where the model exhibits high confidence, and (2) GNN-Explainer, which identifies critical nodes for the classification task.

To validate our approach, we conduct an extensive comparative analysis against traditional centrality measures, evaluating the efficacy and efficiency of our model in identifying influential users.

Our findings demonstrate that the proposed framework not only achieves competitive accuracy but also demonstrates promising scalability to densely connected graphs, making it well-suited for real-world, large-scale misinformation detection tasks. Experimental validation on two heterogeneous, real-world social graphs exhibiting dissimilar topologies further substantiates the framework's robustness, offering a scalable solution for influence analysis in online networks.

By bridging graph-based deep learning with explainability techniques, our work contributes to a deeper understanding of misinformation propagation, paving the way for more effective countermeasures against the spread of false information in digital environments.

The rest of the paper is structured as follows. Section 2 discusses major related approaches. Section 3 introduces the proposed methodology and describes its base components. Section 4 illustrates the experimental investigation. Finally, Sect. 5 concludes the work and provides pointers for future research.

2 Related Work

Methods for identifying influential nodes fall into two categories: traditional and deep learning-based.

Traditional methods rely on centrality metrics, such as local (e.g., degree), global (e.g., betweenness), and hybrid (e.g., PageRank, VoteRank). Examples include DC [10], WSLC [12], and LGC [11], which use various combinations of edge weighting, subgraph extraction, and path-based analysis. However, they often neglect influence variability, node attributes, and heterogeneity—key limitations in complex or heterogeneous networks.

Deep learning-based approaches better handle high-dimensional data and complex structures but often focus on homogeneous graphs or rely on rigid meta-paths in heterogeneous ones. Models such as those by Yu et al. [15], Kou et al. [5], Zhao et al. [18], Keikha et al. [4], and Ahmad et al. [1] use CNNs, GCNs, or embeddings for influence prediction. MEGA [13], the most relevant to our work, targets academic networks using meta-path-based GAT aggregation. Unlike MEGA, our method dynamically combines active learning and GNN-Explainer to rank nodes based on learned influence patterns, not just predefined structures. It also offers interpretability and is tailored for misinformation detection, broadening its applicability across real-world contexts.

3 Proposed Methodology

The methodology presented in this study is designed to detect key actors, such as influencers, within complex social media misinformation networks. At its core, the approach involves constructing a heterogeneous graph that captures the interactional and structural dynamics of the network, focusing on entities such as news articles, users, tweets, and hashtags, and the relationships among them. This graph serves as the foundation for applying a Graph Attention Network, which leverages attention mechanisms to classify news articles as fake or real while learning from the contextual and social interaction signals embedded in the graph. The final step of our framework involves extracting model-driven insights to identify the most influential (top-k) nodes driving misinformation propagation. In the following, we delve into the details of the proposed methodology (see Fig. 1 for the workflow overview).

3.1 Network Modeling

In the first step we propose two feature-rich network models able to capture the complex semantics underlying the news propagation process within the social media ecosystem. We first construct a heterogeneous attributed graph for the fake news detection task and then derive a homogeneous user-centric meta-path-based weighted graph enabling the conventional application of well-known network analysis techniques while preserving the prominent information.

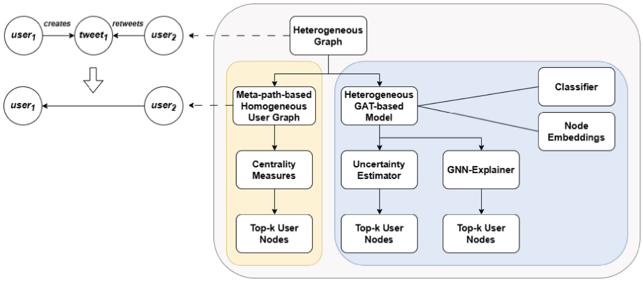


Fig. 1. Logical schemes for detecting influential nodes in a given heterogeneous graph: baseline approach based on computing centrality measures on a homogeneous-node view of the graph (yellow) vs the proposed heterogeneous GAT-based approach (blue). For the yellow block, we extract the (homogeneous) meta-path based graph of user nodes on which we compute the traditional centrality measures to extract top-k nodes. For the blue block, we apply on the original (heterogeneous) graph a GAT-based model, trained at classifying news/claim nodes (e.g., using labels obtained via fact checking), which learns embeddings for all node types; from the trained model we extract top-k user nodes based on uncertainty estimation (AL-like approaches) and explainability techniques.(Color figure online)

The heterogeneous graph captures the intricate relationships among different types of entities involved in the propagation process, like news articles, social media posts, users and hashtags. Each edge type represents a specific type of interaction or association, including retweets, mentions and co-occurrence within posts. The user-centric weighted graph focuses on a single target node type for further analysis using as-is traditional techniques, e.g., to detect the most influential users leading the spread of information.

Formally, we define a heterogeneous attributed network as $G = \langle \mathcal{V}, \mathcal{E}, A, R, \phi, \varphi, \mathcal{X} \rangle$, where $\mathcal{V}$ is the set of nodes, $\mathcal{E} = \bigcup_{r \in R} \mathcal{E}_r \subseteq \mathcal{V} \times \mathcal{V}$ is the set of (directed) edges, A is the set of node types, R is the set of relation types, $\mathcal{E}_r$ is the subset of edges associated with any relation $r \in R$, $\phi : \mathcal{V} \rightarrow A$ is the node-type mapping function, $\varphi : \mathcal{E} \rightarrow R$ is the edge-type mapping function, and $\mathcal{X}$ is a set of matrices $\mathcal{X} = \{\mathbf{X}^{(a)} \mid a \in A\}$ storing node attributes.

More complex interactions in a heterogeneous graphs can be modeled via meta-paths. A *meta-path type* (or simply *meta-path*) in a heterogeneous network is a composite relation modeling high-order proximity induced by a path $a_1 \xrightarrow{r_1} a_2 \xrightarrow{r_2} \ldots \xrightarrow{r_x} a_{x+1}$ between two node types, $a_1 \in A$ and $a_{x+1} \in A$, which are expected to share information – see Fig. 1 for a concrete meta-path example. Let σ_m be a meta-path of type m, and let $\mathcal{M}$ be the set of all the chosen meta-path types in the heterogeneous graph. Notation $\sigma_m(v, v')$ will denote the fact that nodes v and v' (such that v and v' belong to the starting node type and final node type in meta-path σ_m) are connected via σ_m.

A *meta-path instance* is a sequence of connected nodes matching the node and edge types in the meta-path, able to make two distant nodes in the network reachable. A *meta-path-based graph* is a graph comprised of nodes connected via meta-path instances of one or more meta-path types. Given a target node type $\overline{a} \in A$ and a meta-path $\sigma_m \in \mathcal{M}$, with terminal nodes of the same type $\overline{a}$, the resulting meta-path-based graph is a homogeneous weighted graph with one node type $\overline{a}$ and one relation type σ_m, obtained from all meta-path instances of σ_m by removing the intermediate nodes and establishing a direct link between the terminal nodes, weighted on the number of meta-path instances connecting them. Many meta-paths connecting nodes of type $\overline{a}$ can be combined into a single meta-path-based graph.

3.2 Fake News Detection

The second step applies a GAT-based architecture enhanced by meta-paths to the constructed heterogeneous graph to classify news articles as fake or real by relying on both the network topology and the node attributes as in [6]. The GAT operates on the heterogeneous graph at node-level by computing attention scores for each incident edge and meta-path instance[1], dynamically weighing the contributions of neighbors. The attention mechanism allows to prioritize influential relationships, such as a widely retweeted tweet or a frequently mentioned user or frequently co-occurring hashtags.

The model architecture includes an embedding layer to encode the initial features of each node, followed by attention layers to refine the embeddings based on the influence of neighboring nodes. The final output layer classifies nodes by their likelihood of being fake news. The graph is processed in a type-aware manner, with distinct attention mechanisms applied to different edge types (e.g., users retweeting tweets vs. hashtags used in tweets). This ensures that the model captures the semantics of each relationship. Information from neighboring nodes is aggregated based on the computed attention scores, updating the node embeddings to reflect their local context. Please note that, as well as to tell fake and real news apart, the learned embeddings can be used for different downstream tasks, at node, edge and (sub)graph level. By making the GAT-based classifier learn all node embeddings within a shared latent space, we propose to leverage it for identifying influential user nodes.

3.3 Key Node Identification

Identifying key nodes within an online misinformation network is critical for understanding how false information propagates and how it can be mitigated. In the third step, node rankings are determined and key (top-k) nodes are detected based on their roles in either amplifying or suppressing misinformation. Influential users act as amplifiers by driving the spread of misinformation, while suppressors counter false narratives and help curb their dissemination. Understanding these roles is essential for designing effective intervention strategies.

[1] We treat meta-paths as additional relationships between nodes of a given type.

Traditional approaches to identify influential nodes rely on centrality measures such as degree centrality, which highlights well-connected nodes; betweenness centrality, which identifies nodes serving as network bridges; closeness centrality, which measures how efficiently a node can reach others; PageRank, which evaluates node importance based on both the quantity and quality of incoming connections. Additionally, influence-spreading heuristics like VoteRank iteratively select key nodes to maximize diffusion in a network. While these measures offer valuable structural insights, they may fail to account for the contextual aspects of misinformation spread and suffer from high computational costs, hence struggling in dense graphs where high connectivity complicates the distinction of truly influential nodes and becoming impractical for large-scale networks.

To address these limitations, our framework integrates model-driven techniques by leveraging the trained GAT and advanced post-hoc methods. Specifically, two complementary approaches are employed to uncover influential nodes: uncertainty-based AL-like techniques and GNN-Explainer. The AL-like approach prioritizes nodes where the model exhibits high confidence, ensuring that the most informative and structurally significant nodes are analyzed. On the other hand, a GNN-Explainer module provides a score for each instance allowing to identify which nodes and edges contribute the most to the model's predictions. This dual approach ensures that both structural influence and contextual significance are accounted for, approximating the traditional techniques while remaining scalable as the graph grows. Each technique applied on the heterogeneous GAT deal with the multiplicity of types in the graph and results in a ranking of nodes, allowing top-k nodes of a given type to be selected as most relevant.

Uncertainty-Based Active Learning-Like (AL-Like) Approach. Active Learning (AL) techniques aim to reduce labeling costs by iteratively selecting a small number of informative instances from a large unlabeled dataset [8]. In this work, AL-like instance ranking criteria are exploited to identify key user nodes, which deem higher attention in the analysis and management of fake news propagation.

This method prioritizes nodes for further analysis by selecting those where the GAT model exhibits the highest prediction confidence. The rationale is that user nodes associated with highly confident predictions are likely to feature some properties that contribute significantly to (mis-)information propagation. These techniques were selected based on their energy efficiency, as they require only a single inference pass per data sample. The employed methods are simple measures computed directly from the model's output.

The AL-like strategy begins by sampling nodes with high classification certainty. These nodes are examined in greater detail to assess their structural and behavioral significance in the network. Unlike traditional AL involving iterative querying of labels, this AL-like method focuses on identifying nodes that the model deems most influential without requiring additional labeling efforts.

In the current implementation of the approach, we use the following three uncertainty-based ranking criteria, in a opposite way w.r.t. typical AL settings (i.e., the lower the uncertainty, the higher the rank):

- *Least Confidence Sampling* (for short, *AL LCS*): Let p be the probability of the most likely class for a data instance x. Then the least confidence score assigned to x is simply computed as $1 - p$;
- *Margin Sampling* (for short, *AL Margin*): This criterion focuses on the difference between the probability of the most likely class and the second most likely class. If, for a data instance x, p_{top1} and p_{top2} are the probability of the most likely class and of the second most likely class, respectively, then the margin score of x is computed as $p_{top1} - p_{top2}$;
- *Entropy Sampling* (for short, *AL Entropy*): Entropy measures the overall uncertainty across all classes. A high entropy value means the model is unsure about the correct class. For a data instance x, if there are C classes and p_i is the probability of the i-th class, then it is calculated as $-\sum_{i=1}^{C} p_i \log p_i$.

GNN Explainer Approach. GNN-Explainer [14] is a post-hoc interpretability method designed to uncover the most important structural and feature-related factors influencing Graph Neural Network's predictions. It can be extended to heterogeneous graphs with minimal effort. We use a Captum-based explainer which optimizes a mask over the input graph, assigning importance scores to node features and relations based on their impact on the gradient, highlighting key edges and node features that drive the GAT classification outcomes.

To identify highly central nodes responsible for amplifying fake news, we compute the node-level score according to two complementary strategies:

- *feature-based score* (for short, *Expl. Feats.*), computed by summing along the feature dimension, i.e., by aggregating the contributions of all its features;
- *relation-based score* (for short, *Expl. Rel.*), computed by summing along the incident edges, i.e., by aggregating the contributions of all its neighbors.

The disjunction of the two node-level scores enable the understanding of the predominance of one contribution over the other in analyzing the most critical actors in the network. For user nodes, we assign an importance score based on node features —such as engagement metrics, credibility scores, and content characteristics —as well as an importance score based on edges representing interactions between users, allowing for a better understanding of how misinformation spreads and which users play central roles in its propagation.

Unlike the AL-like approach, which focuses on confident nodes, GNN-Explainer provides a broader view by identifying nodes that play a pivotal role in shaping misinformation spread, regardless of their classification confidence.

4 Experimental Evaluation

The goal of this evaluation is to systematically assess the efficacy of the proposed approach by analyzing its ability to identify influential nodes and maximize information diffusion. The objective is to generate a ranking of nodes such that selecting the top-k nodes ensures the highest possible spread of influence on the remaining nodes in few steps.

The evaluation focuses on two performance dimensions: (i) Influence Spread Capacity, i.e., evaluating how many nodes are influenced when selecting top-ranked nodes, and (ii) Computational Efficiency, i.e., comparing the execution time of each method to determine scalability.

4.1 Baselines

To assess the effectiveness of the proposed approach in detecting the most influential nodes, we compared its performance against traditional centrality measures, which evaluate structural significance within the network while offering the advantage of being highly interpretable:

- *Degree Centrality.* Reflects the number of direct connections a user has, capturing their immediate influence;
- *Betweenness Centrality.* Measures how often a user acts as a bridge between other nodes, highlighting their role in spreading information by facilitating communication between otherwise disconnected groups;
- *Closeness Centrality.* Indicates how quickly a user can reach others in the network, capturing their ability to spread information across the network;
- *Pagerank.* Quantifies a user's overall importance within the graph, based on the connectivity with other influential nodes;
- *Voterank.* Identifies influential nodes in a network based on an iterative voting process by selecting influential nodes based on their ability to influence others while avoiding redundancy.

4.2 Performance Metrics

Influence Spread Metrics. To evaluate the influence spread effectiveness, in terms of the extent to which selected influential nodes propagate information, two widely used metrics are considered. Given N the total number of nodes in the network, K the number of selected top-k influential nodes (seed nodes) and R the number of nodes reached (influenced) by the K selected nodes using Breadth-First Search (BFS) algorithm within a fixed small, empirically determined depth, we define our reachability and coverage metrics as follows:

- $Coverage(K) = \frac{R+K}{N}$: measures the proportion of covered nodes, including both seed nodes and influenced nodes, out of the total number of nodes in the graph;
- $Reachability(K) = \frac{R}{N-K}$: evaluates how well the selected seed nodes propagate influence among non-seed nodes, by quantifying the total number of influenced nodes w.r.t the maximum number of nodes that can be influenced.

Table 1. MuMiN statistics, in terms of no. of nodes, edges and meta-paths. In bold the ranked node type and the relations held in the meta-path-based graph.

# Nodes	Claim (C)	2168
	Tweet (T)	4340
	Reply (R)	195459
	User (U)	153168
	Hashtag (H)	28091
	Image (I)	1020
	Article (A)	1453
# Meta-paths	C-T-U-T-C	28867
	C-T-H-T-C	21577
	C-T-R-T-C_r	2859
	C-T-R-T-C_q	3042
	U-T-U	11412
	U-R-U	146056

# Edges	T *discusses* C (C-T)	5081
	R *reply_to* T (T-R_r)	90196
	R *quote_of* T (T-R_q)	101216
	T *has_hashtag* H (T-H)	2289
	T *has_article* A (T-A)	1898
	T *has_image* I (T-I)	1028
	T *mentions* U (U-T_m)	1119
	U *posted* T (U-T_p)	4091
	U *posted* R (U-R)	179247
	U *retweeted* T (U-T_r)	13402
	U *follows* U (U-U_f)	18379
	U *mentions* U (U-U_m)	2797
	U *has_hashtag* H (U-H)	50451

As K increases, i.e., the number of starting influencial nodes grows, Coverage(K) is expected to increase, as it accounts for both the K influential nodes selected at start (based on centrality-like rankings) and those that they can potentially influence; however, this may not necessarily hold for Reachability(K), since a higher K reduces the set of reachable nodes.

Ranking Similarity Metrics. To compare two rankings over the same set, multiple rank correlation coefficients measuring the overall similarity and overlap measures are considered:

- *Spearman's Rank Correlation Coefficient* (ρ): measures how well the rankings match through the correlation between the rank positions of elements in both lists. Given d_i the difference between the ranks of element i in both lists: $\rho = \frac{6 \sum d_i^2}{N(N^2-1)}$;
- *Kendall's Tau* (τ): counts the number of concordant (C) and discordant (D) pairs between the two lists: $\tau = \frac{C-D}{N(N-1)/2}$;
- *Jaccard Similarity Index:* measures the percentage of common nodes in the two top-K rankings A_K and B_K: $Jaccard@K = \frac{|A_K \cap B_K|}{|A_K \cup B_K|}$;
- *Precision@K*: calculates how many of the top-K items in one list appear in the other list: $Precision@K = \frac{|A_K \cap B_K|}{K}$;
- *Normalized Discounted Cumulative Gain* (NDCG@K): measures ranking quality by considering the relevance and order of ranked nodes: $NDCG@K = \frac{DCG@K}{IDCG@K}$ where DCG is the Discounted Cumulative Gain and measures the effectiveness of a ranking method by considering both relevance scores and positioning of ranked elements, while $IDCG$ is the ideal discounted cumulative gain, where nodes are perfectly ranked.

4.3 Datasets

The experimental evaluation has been conducted on two real-world datasets, i.e., MuMiN and PolitiFact datasets. Both are modeled as a heterogeneous information network, with multiple node and edge types and external information

Table 2. PolitiFact statistics, in terms of no. of nodes, edges and meta-paths. In bold the ranked node type and the relations held in the meta-path-based graph.

# Nodes	News (N)	696
	Tweet (T)	268306
	User (U)	169106
	Hashtag (H)	18631
# Meta-paths	N-T-H-T-N	44682
	N-T-U-T-N	46025
	N-T-U-U-T-N	20056
	U-T-U	533

# Edges	T *discusses* N (N-T)	276676
	T *has_hashtag* H (T-H)	59782
	U *posted* T (U-T_p)	285124
	U *retweeted* T (U-T_r)	539
	U *mentions* U (U-U)	84093

associated with nodes available as a set of attributes (for detailed information on node features, please refer to the corresponding publication). The former is multi-topic; a complete description is provided in [7]. The latter is extracted from the FakeNewsNet data repository [9], which fact-checks news pertaining to the US political system.

In the following, we use the terms Claim and News interchangeably, as the MuMiN dataset refers to news items as Claims, while the PolitiFact dataset uses the term News. Classification is performed on nodes of type Claim (C) in the MuMiN dataset and nodes of type News (N) in the PolitiFact dataset.

To enhance classification performance by capturing richer structural and semantic relationships, we handcrafted some *meta-paths* toward the Claim/News node type. Two types are common to both datasets, including connections between pairs of claims (news) discussed in tweets by the same user or associated with the same hashtag. Additionally, dataset-specific meta-paths are defined. For MuMiN, we identify pairs of claims belonging to the same conversation thread through reply or quote relationships, respectively. For PolitiFact, we include pairs of news discussed in tweets posted by users who mention each other.

To construct the meta-path-based user graph we look at all the edges between users and additionaly integrate the U-T-U meta-path for both datasets, which connects users if one has retweeted the other's post, and the U-R-U meta-path for MuMiN, which connects users who have commented on the same tweet.

Tables 1 and 2 show the statistics in terms of number of nodes, number of edges, and number of meta-path instances for each type, for the two datasets.

4.4 Results

To assess the effectiveness of the proposed methodology in identifying influential user nodes, we compare individual and combined rankings versus the baselines using the two influence spread metrics defined in Sect. 4.2. Before delving into this comparative study, let us first show the results of some preliminary analyses.

Table 3. Results of the Fake News Detection task averaged over 5 runs. Execution time is expressed in seconds.

Dataset	F1-micro	F1-macro	Precision T	Recall T	Precision F	Recall F	Time
MuMiN	0.954 ± 0.007	0.788 ± 0.117	0.721±0.139	0.647 ± 0.191	0.967 ± 0.014	0.985 ± 0.009	189 ± 21
Politifact	0.859 ± 0.039	0.845 ± 0.054	0.848 ± 0.066	0.943 ± 0.044	0.909 ± 0.063	0.736 ± 0.153	332 ± 18

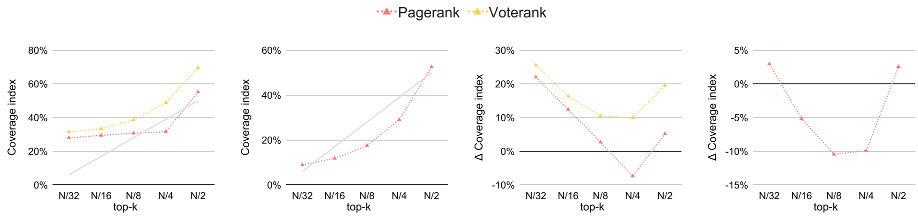
Table 4. Agreement between node importance rankings produced by pairs of post-hoc techniques, using multiple metrics.

Dataset	Pair of techniques	Spea. ρ	Kend. τ	Jacc. index		Precision		NDCG	
				@$\frac{N}{2}$	@$\frac{N}{4}$	@$\frac{N}{2}$	@$\frac{N}{4}$	@$\frac{N}{2}$	@$\frac{N}{4}$
Politifact	Expl feats - AL entr	-0.001	-0.001	0.333	0.133	0.500	0.235	0.935	0.868
	Expl feats - AL lcs	0.029	0.020	0.344	0.154	0.512	0.266	0.9400	0.880
	Expl feats - AL marg	-0.001	-0.001	0.333	0.133	0.500	0.235	0.935	0.868
	Expl rel - AL entr	-0.003	-0.002	0.331	0.140	0.497	0.246	0.936	0.870
	Expl rel - AL lcs	0.032	0.021	0.343	0.144	0.511	0.252	0.94	0.873
	Expl rel - AL marg	-0.003	-0.002	0.33	0.140	0.497	0.245	0.936	0.870
	Expl feats - Expl rel	0.638	0.531	0.622	0.502	0.767	0.669	0.978	0.959
	AL entr - AL lcs	0.319	0.218	0.431	0.206	0.602	0.341	0.938	0.884
	AL entr - AL marg	1.000	0.999	1.000	0.999	1.000	0.999	1.000	1.000
	AL lcs - AL marg	0.319	0.218	0.430	0.206	0.602	0.341	0.957	0.894
Mumin	Expl feats - AL entr	-0.031	-0.021	0.313	0.157	0.476	0.272	0.936	0.874
	Expl feats - AL lcs	-0.049	-0.034	0.289	0.143	0.449	0.25	0.934	0.87
	Expl feats - AL marg	-0.031	-0.022	0.312	0.157	0.476	0.271	0.936	0.874
	Expl rel - AL entr	-0.012	-0.008	0.323	0.15	0.489	0.260	0.936	0.873
	Expl rel - AL lcs	-0.039	-0.026	0.32	0.141	0.485	0.247	0.935	0.87
	Expl rel - AL marg	-0.008	-0.005	0.327	0.151	0.493	0.262	0.937	0.873
	Expl feats - Expl rel	0.184	0.201	0.384	0.177	0.555	0.301	0.953	0.889
	AL entr - AL lcs	0.815	0.649	0.719	0.545	0.836	0.706	0.991	0.973
	AL entr - AL marg	1.000	0.988	0.978	0.976	0.989	0.988	1.000	1.000
	AL lcs - AL marg	0.816	0.650	0.721	0.546	0.838	0.706	0.987	0.979

Preliminary tests: GAT accuracy and correlation analysis Table 3 shows a number of accuracy metrics for the fake news classification task. These results confirm that the proposed GAT-based architecture performs well on both datasets ([6]).

Here, we report the main parameters used in the experimentation. As regards the graph neural network model, we employed a l-layer GAT [3] architecture with dropout set to 0.4, hidden channels dimension set to 64 and out channel dimension set to 2 as the number of classes. The number of layers l was set to 2 for the PolitiFact dataset and to 3 for the MuMin dataset, based on the empirical performance and to balance representation capacity with over-smoothing avoidance. We employed a weighted cross entropy loss function in a fully supervised setting, with weights inversely proportional to the class frequency. We trained the model over 200 epochs and used the Adam optimization algorithm. The learning rate was set to 0.005 while the weight decay to 0.001.

We exploit AL-like techniques and GNN-Explainer on the trained model to identify high-confidence and key influential nodes, respectively. These methods



(a) $C(K)$, PolitiFact (b) $C(K)$, MuMiN (c) $\Delta C(K)$, PolitiFact (d) $\Delta C(K)$, MuMiN

Fig. 2. Qualitative analysis of the baseline approach, configured with ranking methods Pagerank and Voterank, when varying the size of the core set K, with dataset and maximum depth $d = 2$ fixed. (a) and (b) show the coverage index $C(K)$, while (c) and (d) show the increment of $C(K)$ given by the influenced nodes and measured as $\Delta C(K) = C(K) - \frac{K}{N} = \frac{R}{N}$.

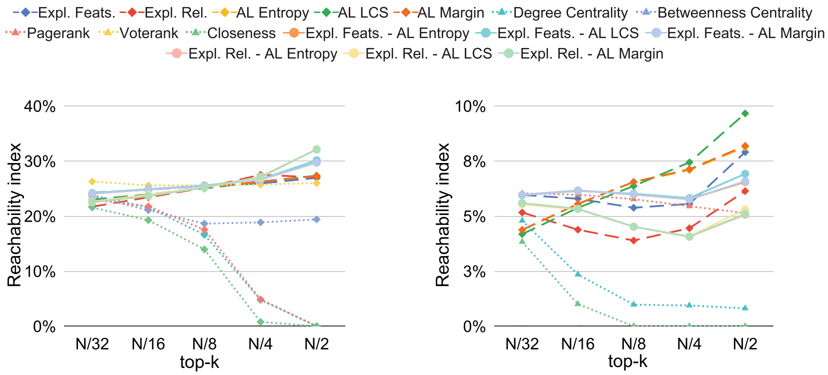
generate independent node rankings, which are compared in pairs by means of the ranking similarity metrics presented in Sect. 4.2.

Table 4 reveals low correlation across ranking techniques of different post-hoc approaches, indicating that node importance varies across methods. However, the Jaccard Index suggests moderate agreement, especially for top-half-ranked nodes. Notably, consistently high NDCG scores confirm that key influential nodes remain prominent across methods, despite variations in ranking positions.

Among variants of the AL-like approach, rankings align closely, particularly for entropy- and margin-based methods, which show near-identical orderings. Differences across datasets, particularly for the two GNN-Explainer outcomes, reflect the impact of graph sparsity, where high connectivity graphs emphasize structural importance, while more sparse ones rely more on feature-based influence. These findings led us investigate a combined approach in addition to individual rankings. To create the combined ranking, we merge top-K/2 nodes from both AL and explainability-based lists, iteratively combining them to capture complementary aspects of nodes importance.

Coverage/reachability Analysis. The top row in Fig. 3 is meant to help get a general understanding of how the coverage power of baseline methods depend on the number K of core influential nodes retrieved with these methods, on both datasets. For the sake of presentation, let us focus only on methods Pagerank and Voterank –indeed, as shown later on, these two methods were found to excel among classical ranking ones in coverage metrics. Sub-figures (c-d) show the trend of the share $\Delta C(K) = C(K) - \frac{K}{N} = \frac{R}{N}$, of $C(K)$ corresponding to the influenced nodes (i.e. the nodes reached from the initial K ones), i.e., the proportion of nodes that are successfully reached.

Notably, in denser graphs like MuMiN, VoteRank and betweenness centrality – the most competitive baselines in sparser graphs like PolitiFact – fail to compute within a reasonable time, while PageRank shows high coverage capabilities. The Coverage scores align as the core set increases, but the dependency from K



(a) PolitiFact (b) MuMiN

Fig. 3. Reachability index $R(K)$ when varying the number K of nodes in the initial core set, with maximum depth $d = 2$. Different kinds of markers (namely, diamonds, circles and triangles) and styles (namely, dashed, solid and dotted) are used to denote the proposed individual and combined post-hoc techniques and the baselines, respectively. For the sparser dataset (a), few differences are noticed as the initial core set size changes and the best performance is achieved by combining AL margin with an explainability technique, outperformed by voterank only for smaller seed size; for the denser dataset (b), for large initial core set size the individual Active Learning techniques work best, while decreasing the size the combination of AL margin with an explainability technique still achieves the best result.

is non-monotonic. Excluding the core set from the computation (Fig. 2 (c) -(d)) emphasizes the need of a proper selection of K.

Figure 3 shows the Reachability index values of individual, combined and baseline rankings as the size of the core set K varies within a maximum depth $d = 2$. This results clearly reveal that conventional measures of centrality often fail to maximize influence for short distances as the core set increases. On the more sparse graph, VoteRank remains flat while post-hoc techniques slightly increase, with predominance of combinations of explainability outcomes and AL margin. On the higher connectivity graph, no significant changes emerge, with uncertainty-based techniques prevailing.

The effectiveness of our approach is coupled with scalability in denser graphs exhibiting a higher number of connections. Table 5 shows the elapsed time for the computation of node importance scores for individual techniques. Increasing the number of edges slows down the computation and exceeds 40 h for techniques most dependent on the number of edges, such as VoteRank and betweenness centrality. An exception is closeness centrality which benefits from its implementation. Degree centrality and PageRank are efficient but fail in spreading information in few steps in more sparse graphs. Although we do not achieve the best absolute runtimes, our execution times remain reasonable while yielding superior performance results. Our analysis highlights the advantages of combining AL-like and explainability techniques for identifying influential nodes.

Table 5. Computation time of node importance scores, expressed in seconds, including training time.

Technique	politifact	mumin
GNN Explainer	44.81	51.71
AL Entropy	1183.74	3054.21
AL LCS	1352.85	3063.36
AL Margin	1183.74	3093.11
Degree Centrality	0.09	0.33
Closeness Centrality	1542.41	237.86
Betweenness Centrality	11844.87	n/a
Pagerank	1.79	2.45
VoteRank	711.75	n/a

The proposed hybrid approach achieves competitive accuracy while significantly reducing computational overhead, making it an effective solution for large-scale social media networks. Additionally, our influence spread simulations confirm that hybrid rankings outperform traditional centrality measures in identifying nodes that maximize reachability within minimal steps.

5 Conclusion

In this work, we have addressed the critical challenge of misinformation propagation in online social networks by proposing a novel attention-based framework which leverages a trained GAT model to feed advanced post-hoc techniques. Our approach enhances the identification of influential nodes within heterogeneous networks, effectively capturing the faceted nature of information dissemination. By integrating post-hoc techniques, such as uncertainty-based AL-like methods and graph neural network explanations, we provide deeper insights into the mechanisms of misinformation spread while improving model interpretability. Through extensive comparative analysis, our findings demonstrate that the proposed framework outperforms traditional graph-based centrality measures in both accuracy and scalability, . successfully identifying key users responsible for misinformation diffusion even where conventional approaches often struggle or fail. The validation on two heterogeneous, real-world social graphs further highlights the robustness and adaptability of our approach.

Building on these findings, future work will explore the extension of our framework to dynamic social networks, where the continuous evolution of user interactions and misinformation patterns poses additional challenges. Notably, integrating continual learning techniques could enable real-time adaptation to emerging trends, enhancing both the efficiency and responsiveness of misinformation mitigation strategies in large-scale, rapidly changing environments.

Test reproducibility. The code necessary to replicate our experiments is available at: https://github.com/Franco7Scala/ActiveLearningFakeNewsDetection.

Acknowledgments. This work has been partially supported by: (i) project SERICS (PE00000014) under the NRRP MUR program funded by the EU - NGEU; (ii) project MIRFAK (P2022C23K9), ERC field: PE6, funded by the EU - NGEU. (iii) project FAIR (PE00000013), funded by the EU - NGEU.

References

1. Rezaei, A.A., Munoz, J., Jalili, M., Khayyam, H.: A machine learning-based approach for vital node identification in complex networks. Expert Syst. Appl. **214**,119086 (2023)
2. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Comput. Netw. ISDN Syst. **30**(1–7), 107–117 (1998)
3. Brody, S., Alon, U., Yahav, E.: How attentive are graph attention networks? *CoRR*, abs/2105.14491 (2021)
4. Keikha, M.M., Rahgozar, M., Asadpour, M., Abdollahi, M.F.: Influence maximization across heterogeneous interconnected networks based on deep learning. Expert Syst. Appl. **140**, 112905 (2020)
5. Kou, J., Jia, P., Liu, J., Dai, J., Luo, H.: Identify influential nodes in social networks with graph multi-head attention regression model. Neurocomputing **530**, 23–36 (2023)
6. Martirano, L., Comito, C., Guarascio, M., Pisani, F.S., Zicari, P.: M3dusa: a modular multi-modal deep fusion architecture for fake news detection on social media. Soc. Netw. Anal. Min. **15**(1), 1–17 (2025)
7. Nielsen, D.S., McConville, R.: MuMiN: a large-scale multilingual multimodal fact-checked misinformation social network dataset. In: Proceedings of the 45th Internationl ACM SIGIR Conference on research and development in Information Retrival, pp. 3141–3153 (2022)
8. Settles, B: Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison (2009)
9. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: a data mining perspective. ACM SIGKDD Explor. Newsl **19**(1), 22–36 (2017)
10. Ugurlu, O.: Comparative analysis of centrality measures for identifying critical nodes in complex networks. J. Comput. Sci. **62**, 101738 (2022)
11. Ullah, A.: Identifying vital nodes from local and global perspectives in complex networks. Expert Syst. Appl. **186**, 115778 (2021)
12. Wang, X., Othman, M., Dewi, D.A., Wang, Y.: WSLC: weighted semi-local centrality to identify influential nodes in complex networks. J. King Saud Univ. Comput. Inf. Sci. **36**(1), 101906 (2024)
13. Xie, J., Yu, J., Chen, Z.: Mega: identifying influential nodes in heterogeneous networks based on meta-path and attention. J. Stat. Mech: Theory Exp. **2025**(2), 023401 (2025)
14. Ying, Z., Bourgeois, D., You, J., Zitnik, M., Leskovec, J.: GNNExplainer: generating explanations for graph neural networks. Adv. Neural Inf. Process. Syst. **32** (2019)

15. Yu, E.-Y., Wang, Y.-P., Fu, Y., Chen, D.-B., Xie, M.: Identifying critical nodes in complex networks via graph convolutional networks. Knowl. Based Syst. **198**, 105893 (2020)
16. Zafarani, R., Abbasi, M.A., Liu, H.: Social media mining: an introduction. Cambridge University Press (2014)
17. Zhang, J.-X., Chen, D.-B., Dong, Q., Zhao, Z.-D.: Identifying a set of influential spreaders in complex networks. Sci. Rep. **6**(1), 27823 (2016)
18. Zhao, G., Jia, P., Zhou, A., Zhang, B.: InfGCN: Identifying influential nodes in complex networks with graph convolutional networks. Neurocomputing **414**, 18–26 (2020)