Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Hybrid Model for Fake News Detection in Persian Language: a Feature Fusion Based on Hierarchical Propagation Networks

Sara Bourbour Hosseinbeigi1

s.bourbour@modares.ac.ir

Tarbiat Modares University

Mohammad Javad Fatemi

Tarbiat Modares University

Research Article

Keywords: Persian Fake News Detection, Hierarchical Propagation Networks, Social Network Analysis, Social Context, Hybrid Features Fusion

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11507132/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Hybrid Model for Fake News Detection in Persian Language: a Feature Fusion Based on Hierarchical Propagation Networks

Sara Bourbour Hosseinbeigi[1*] and Mohammad Javad Fatemi[1]

[1]Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran.

*Corresponding author(s). E-mail(s): s.bourbour@modares.ac.ir;
Contributing authors: mohammadjavad.fatemi@modares.ac.ir;

Abstract

The rapid proliferation of fake news on social media has emerged as a significant threat to information credibility and public trust. Fake news detection approaches can be broadly categorized into two main groups: text-based and structure-based methods.Detecting fake news solely based on textual content is often insufficient, because fake news often exhibits linguistic patterns and writing styles similar to those of genuine news.Moreover, the scarcity of datasets capturing hierarchical information propagation patterns has hindered progress in low-resource languages such as Persian. To address these limitations, we introduce P-HNet, a novel Persian fake news dataset collected from X (formerly Twitter) that captures information propagation at both macro and micro levels. The macro level represents how news spreads from original publishers to users who repost it, while the micro level captures conversations and interactions surrounding the news, including user replies, community responses, and engagement patterns. Building on this hierarchical representation, we propose PGT, a hybrid feature-fusion framework that combines features extracted from news content and propagation networks. Experimental results on P-HNet demonstrate the effectiveness of PGT, achieving an F1-score of 74.51% and an accuracy of 82.19%, outperforming baseline methods such as DSS, HPFN, and conventional machine-learning models. Ablation studies further highlight the importance of textual representations, micro-level propagation patterns, and emotion-related features in detecting Persian fake news.

Keywords: Persian Fake News Detection, Hierarchical Propagation Networks, Social Network Analysis, Social Context, Hybrid Features Fusion

1 Introduction

Traditional media, such as newspapers and television, were once the main sources of news. However, the rise of online social networks (OSNs) and digital media platforms has changed how people access, share, and consume news. In the United States, for example, the proportion of users obtaining news from social networks increased from 49% in 2012 to 62% in 2016 (Alghamdi et al. 2024). This transition toward digital platforms has created favorable conditions for the rapid dissemination of misinformation. A study conducted by Stanford University reported that 72.3% of false information is propagated through established news sources and social media channels (Allcott and Gentzkow 2017). In the current post-AI era, the rapid expansion of social media has transformed platforms such as X and Facebook into major channels for information dissemination. Although these platforms facilitate global participation and instantaneous access to information, they also create substantial challenges for information integrity. In particular, the rapid spread of fake news can erode public trust in the media, promote conspiracy theories, and pose serious social and security risks. Misinformation may also contribute to practical threats such as phishing and social engineering, potentially affecting social stability and democratic processes. Given the vast amount of information shared online and its rapid spread, manual verification alone is insufficient, highlighting the growing need for automated fake news detection.(Ebiwonjumi et al. 2026; Kuntur et al. 2026; Phan et al. 2023; Soga et al. 2024).

To address this growing challenge, researchers have increasingly focused on developing artificial intelligence-based methods for fake news detection. (Zheng 2026), (Dair and Sadirmekova 2026).

Many early approaches to automated fake news detection primarily focused on analyzing textual content (Oshikawa et al. 2020). However, recent studies have increasingly explored social context, as the way news spreads and users interact with it can provide valuable insights into its veracity (Lv et al. 2025). These approaches go beyond textual features by incorporating user behavior, social interactions, and information propagation patterns.

Social context generally refers to the interactive environment in which news is disseminated and is commonly investigated from two complementary perspectives: user-based and propagation-based analysis. User-based approaches assess news credibility by examining the characteristics and behaviors of users who share or interact with news, as well as their relationships within social networks items (Zhou and Zafarani 2019). In contrast, propagation-based approaches focus on the structural, temporal, and linguistic characteristics of information diffusion, typically modeling the hierarchical propagation patterns through which news spreads across social networks (Shu et al. 2020). More recently, hybrid approaches have been proposed to combine multiple forms of social-context information. For instance, models that jointly leverage propagation and stance networks derived from user interactions have shown improved performance in fake news detection (Davoudi et al. 2022).

Despite these advances, most existing datasets and propagation-based fake news detection studies are designed for high-resource languages, particularly English. Consequently, research on low-resource languages such as Persian remains relatively limited.

Publicly available datasets such as FakeNewsNet (Shu et al. 2020), which includes the PolitiFact and GossipCop collections, primarily target English-language content and do not provide suitable resources for Persian fake news detection. In particular, there is currently a lack of publicly available Persian datasets that capture hierarchical propagation structures and user interaction patterns. Nevertheless, efforts have been made to address resource scarcity in Persian NLP. Among these, the Matina corpus (Hosseinbeigi et al. 2025) is particularly valuable due to its large-scale collection of Persian texts, including substantial content from news sources and social media platforms.

To address this gap, we constructed a new dataset consisting of 1,093 Persian news instances, including 385 fake and 708 real items. The dataset captures information diffusion at multiple levels and is specifically designed to support the analysis of propagation-based and social-contextual features.

In addition to introducing the dataset, we propose PGT, a hybrid model for Persian fake news detection. The proposed framework integrates several complementary feature categories, including macro-level propagation network features, stance network features, emotion-related features extracted from news content, semantic representations of news text, and micro-level propagation representations. These heterogeneous features are then fused and fed into a multilayer perceptron (MLP) classifier for final prediction. By combining textual, emotional, and propagation-related information, the framework aims to capture both the semantic characteristics of news content and the social dynamics underlying its dissemination.

The main contributions of this study can be summarized as follows. First, we introduce a new Persian dataset for fake news detection that explicitly captures hierarchical propagation structures and social interactions. Second, we propose PGT, a hybrid feature-fusion model tailored to the characteristics of this dataset and to the challenges of fake news detection in Persian. The experimental results show that the proposed approach achieves superior performance compared with several baseline methods, including models previously evaluated on widely used fake news benchmark datasets. The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the dataset collection and preparation process. Section 4 introduces the proposed model architecture. Section 5 presents the experimental setup, implementation details, hyperparameter settings, and evaluation metrics. Section 6 reports the experimental results, including ablation studies and comparisons with baseline and traditional machine learning methods. Finally, Section 7 concludes the paper and discusses directions for future research.

2 Related Works

Research on fake news detection has expanded beyond traditional content-based approaches to incorporate social context, propagation patterns, and user interactions. This section reviews the major methodological developments in the field and discusses their limitations.More broadly, fake news detection can be examined through three complementary dimensions: the textual content of news, user reactions and interactions, and the characteristics of the sources involved in its publication and

dissemination(Ruchansky et al. 2017). Beyond textual content, social context provides valuable insights into news credibility by capturing how information propagates through social networks and how users interact with it (Zhang and Ghorbani 2020).

Traditionally, fake news detection methods have been broadly classified into content-based and social-context-based approaches. While this distinction provides a useful overview, it does not fully capture the diversity of methods that integrate multiple sources of information. Recent studies have increasingly explored combinations of textual and visual features, user behavior, network structure, and propagation patterns to improve fake news detection. (Meel and Vishwakarma 2020). To provide a more fine-grained classification, some studies distinguish among three main categories of fake news detection methods: content-based, source-based, and propagation-based approaches. (Albahar 2021).

Content-based methods assess news credibility primarily by analyzing the content of news items. One prominent branch of this category is knowledge-based detection, which assesses factual claims extracted from news by matching them with previously fact-checked claims or verifying them against external knowledge sources, such as knowledge graphs. (Zhou and Zafarani 2020). In practice, natural language processing techniques, often combined with named entity recognition, can be used to detect semantic inconsistencies between news claims and external knowledge sources.(Mayank et al. 2022; Hu et al. 2021).In this context, previous studies have shown that even incomplete knowledge graphs can provide useful evidence for detecting misinformation.(Pan et al. 2018). Another major line of research focuses on stylistic and linguistic indicators, including syntactic patterns, lexical characteristics, and emotional cues. This approach is motivated by the observation that deceptive content often exhibits distinctive writing styles and emotionally charged expressions that may attract attention and influence readers (Alonso et al. 2021; Zhou et al. 2020; Choudhary and Arora 2021).

Deep learning models have also been widely adopted for content-based fake news detection. Architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been extensively used to learn high-level textual representations and improve classification performance (Nasir et al. 2021; Umer et al. 2020). In the context of Persian fake news detection, a hybrid transformer-based model incorporating sentiment analysis, hate speech detection, and topic identification was proposed and evaluated on a social media dataset containing 2,500 COVID-19-related instances (Hosseini and Jafari 2024).In a related study, augmenting semantic text representations with complementary content-based features was found to improve detection performance. Specifically, semantic representations extracted using CNNs were combined with handcrafted content-based features, leading to improved classification performance on an English COVID-19 fake news dataset.(Samadi and Momtazi 2025).

With the increasing prevalence of multimedia content on social media, visual information has also become an important source of evidence of fake news detection. Images and videos can reinforce misleading narratives and increase the apparent credibility of false claims. A number of studies have therefore employed CNN-based architectures to extract visual patterns and identify manipulated or misleading images (Qi et al. 2019;

Uppada et al. 2022). The emergence of deepfake technology has further complicated this problem by enabling the generation of highly realistic synthetic media (Yu et al. 2021). Since social media posts often contain multiple types of information, including text, images, and user-related metadata, multimodal approaches have been developed to jointly model these heterogeneous signals. Existing studies generally indicate that combining textual and visual information can yield better performance than relying on a single modality.(Khattar et al. 2019; Segura-Bedmar and Alonso-Bartolome 2022).

Beyond feature-level fusion, multi-expert architectures offer an alternative strategy for integrating complementary modeling capabilities. For instance, Hosseinbeigi et al. (2025) introduced Matina, a Persian language model that integrates multiple domain-specialized LoRA experts through a classification-based selection mechanism. Although developed for language modeling rather than fake news detection, this work demonstrates how specialized models can be coordinated within a unified framework. A similar principle could potentially be extended to fake news detection, where specialized models assess different aspects of news credibility, such as textual content, source reliability, user behavior, and propagation patterns.

Source-based approaches assess news credibility by analyzing the reliability of the publisher or the users involved in disseminating the information. Although such approaches often infer veracity indirectly, they can be particularly useful for early-stage fake news detection (Zhou and Zafarani 2020). One representative example is MESE, a source-oriented model that leverages user comments to capture the evolution of emotional and semantic information, as well as variations in user attention, thereby improving prediction performance (Xu et al. 2026). Other studies have investigated author-level attributes, such as historical credibility, behavioral patterns, and previous involvement in spreading misinformation (Yuan et al. 2020). Along similar lines, Park and Chai (2023) proposed a multidimensional framework that integrates user attributes, news content, and network properties based on social capital theory. Their results showed that a Random Forest classifier trained on prioritized features achieved an accuracy of approximately 94%. In parallel, several studies have focused on identifying malicious or automated entities, including social bots and deceptive user profiles, which are frequently associated with large-scale misinformation propagation (Ramalingam and Chinnaiah 2018; Kudugunta and Ferrara 2018; Rostami and Karbasi 2020; Gao et al. 2020). More recent findings also suggest that user credibility is neither static nor universal across topics, highlighting the importance of topic-aware credibility modeling in fake news verification systems (Bazmi et al. 2023).

Although source-based approaches provide useful signals for estimating news credibility, they assess veracity indirectly through the reliability of publishers or users. A complementary direction is claim-level verification, which focuses on evaluating the factual claims themselves rather than relying solely on source credibility. In the context of Persian, Bourbour et al. (Bourbour et al. 2026) proposed a hybrid claim-verification pipeline combining large language models (LLMs) with rationale-quality classification. This work represents a relevant direction for Persian-language claim verification and complements source-based approaches by addressing the assessment of individual claims.

Propagation-based methods shift the focus from the content or source of a news item to the way information spreads through social networks. Prior studies have demonstrated that fake and real news often exhibit distinct diffusion patterns in terms of propagation speed, structural organization, and cascade characteristics (Monti et al. 2019). Propagation-based approaches typically model either the diffusion cascade itself or additional contextual information associated with propagation, such as user comments and interactions.

Recent studies have further suggested that information propagation can be more effectively modeled from a hierarchical perspective. At the macro level, propagation captures the dissemination path from the original publisher to subsequent reposters, whereas the micro level reflects local user interactions, discussions, replies, and reactions surrounding the news item. The HPFN model, for example, adopts this hierarchical perspective and achieved an accuracy of 84.3% on the PolitiFact dataset (Shu et al. 2020). Such multi-level representations provide richer contextual information than approaches relying exclusively on textual features.

More recent propagation-oriented methods have also incorporated stance-related information into the detection process. In these approaches, propagation networks are combined with stance networks constructed from user responses, enabling the model to capture both diffusion structure and the attitudes expressed by users toward the news item. The DSS model is one example of this strategy, where stance information extracted from user replies is represented through network-based features. Under specific parameter settings, DSS achieved an accuracy of 99.1% on the PolitiFact dataset (Davoudi et al. 2022). These findings highlight the potential of jointly modeling propagation structure and user stance for improving fake news detection performance.

In the context of multimodal fake news detection, Bourbour Hosseinbeygi et al. (Bourbour Hosseinbeygi et al. 2026) investigated an approach incorporating three complementary sources of information: news text, user history, and propagation networks. This study highlights the relevance of combining content-based and social-contextual information for fake news detection.

3 Dataset

3.1 P-HNet: Hierarchical Propagation Network Dataset

In this study, the TAJ dataset introduced in (Samadi et al. 2021) was used as the primary source for constructing the proposed P-HNet hierarchical propagation network dataset. The overall dataset construction pipeline is illustrated in Figure 1.

Our data collection process focuses on tweets published by influential users on the X platform, where an influential user is defined as an account with more than 3,000 followers. Following the methodology adopted in (Shu et al. 2020), the objective was to reconstruct the propagation structure associated with each news item and subsequently analyze how the corresponding information diffused through user interactions, replies, and reposting behavior.

To construct the propagation dataset, we first retrieved a set of candidate tweets potentially related to each news article. The retrieval process consisted of several stages. First, Part-of-Speech (POS) tagging was applied to the news text using the

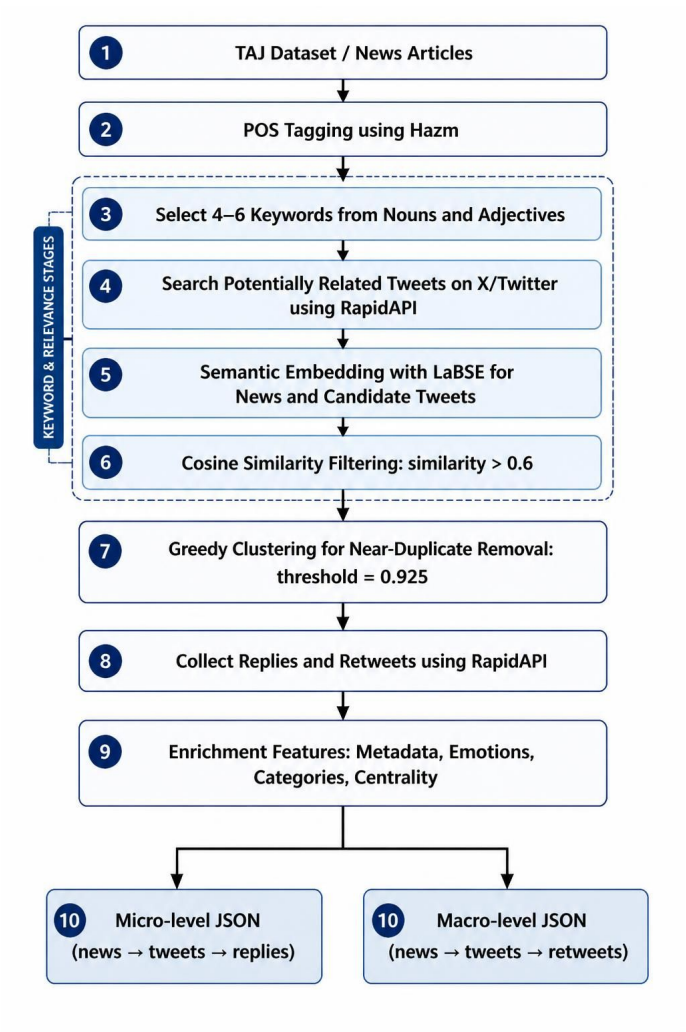


Fig. 1: P-HNet dataset construction pipeline.

Hazm library. Next, four to six keywords were randomly selected from the nouns and adjectives identified in the article. These keywords were then used as search queries to retrieve potentially relevant tweets through the X API provided via RapidAPI[1].

Since lexical overlap alone does not necessarily indicate semantic relevance, an additional semantic filtering stage was applied to the retrieved tweets. Specifically, both the original news article and each candidate tweet were encoded into dense semantic representations using the Language-agnostic BERT Sentence Embedding (LaBSE) model introduced in (Feng et al. 2022). Cosine similarity was subsequently computed

[1] https://rapidapi.com

between the corresponding embeddings, and only tweets with a similarity score greater than 0.6 were retained. To further reduce redundancy, near-duplicate tweets were identified using a greedy clustering procedure with a cosine-similarity threshold of 0.925. For each cluster of highly similar tweets, a single representative instance was preserved. This procedure was designed to retain semantically relevant and diverse samples while limiting redundancy and reducing noise in the resulting dataset. After identifying the relevant tweets, their corresponding replies and retweets were collected through RapidAPI. Each tweet, reply, and retweet was stored together with a set of associated metadata attributes, including id, type, screen_name, published_at, like_count, retweet_count, quote_count, reply_count, follower_count, following_count, and is_verified.

In addition to the raw textual content and metadata, several categories of derived features were extracted for each news article and its associated social media interactions:

1. **Text emotions** (6 binary features): Inspired by the model proposed in (Sarlakifar et al. 2024), we extracted the following binary features: Anger, Fear, Happiness, Disgust, Sadness, Surprise.
2. **Content categories:** The 11-class text classification model proposed in (Eghbalzadeh et al. 2012) was used to extract the following binary features for each news article, tweet, reply, and retweet: Educational, Social, Historical, Economic, Health-related, Scientific, Political, Cultural, Jurisprudence and Law, Religious, and Sports-related.
3. **Centrality measures**: To capture the structural properties of the propagation network, centrality-based features were computed using the NetworkX library[2]. For each news article, tweet, reply, and retweet, we calculated five centrality measures: degree_centrality, closeness_centrality, betweenness_centrality, PageRank_centrality, and Katz_centrality.

The resulting propagation information was organized into two JSON files, each corresponding to a different level of information diffusion:

1. **Micro-level propagation**: The file contains the news article, its associated tweets, and the replies linked to each tweet.
2. **Macro-level propagation**: The file contains the news article, its associated tweets, and the retweets linked to each tweet.

Together, the micro- and macro-level representations provide complementary views of information diffusion on X. The micro-level structure reflects conversational dynamics and user engagement, whereas the macro-level structure captures large-scale dissemination patterns. These hierarchical propagation representations serve as the basis for the network-aware models evaluated in the subsequent experiments.

[2] https://networkx.org

3.2 Data Pre-processing

Data preprocessing constitutes the initial stage of the proposed modeling pipeline. First, missing values were identified and handled appropriately, and timestamp information associated with different interaction events was converted into a consistent format. The dataset was also examined to identify news articles with no associated social interactions. Since every article in the constructed dataset contained at least one relevant interaction, no instances were removed on this basis. Finally, the numerical feature values were standardized using `StandardScaler`[3]. Standardization was applied to transform the features onto comparable scales, thereby preventing variables with larger numerical ranges from disproportionately influencing the learning process and ensuring that the dataset was appropriately prepared for model training.

4 PGT Model

The proposed method comprises six stages, as illustrated in Figure 2. In the first stage, news articles are preprocessed as described in Subsection 3.2. Subsequently, once a news item reaches a propagation depth of two within the structure at the end of a predefined time window, three components are constructed: a micro-level propagation network, a macro-level propagation network, and a stance network. The third stage involves extracting textual emotion features for each news item. Specifically, emotions such as anger, fear, happiness, disgust, sadness, and surprise are extracted using the dataset described in Section 3. In the fourth stage, the constructed propagation and stance networks are utilized to derive network-based and stance-related features. These features capture the structural dissemination dynamics of news and user interaction patterns, which are subsequently utilized in later stages of the model. In the fifth stage, LaBSE is employed to generate news content representations. Concurrently, aligned with (Khoshraftar and An 2024), a Message Passing Neural Network (MPNN), a recognized graph embedding technique, is employed to generate representations of the micro-level propagation network. Finally, all features, including news content emotion features, network-based propagation features, stance network features, as well as the representations of both news content and the micro-level propagation network, are concatenated into a unified feature vector. This vector is then fed into an MLP to predict news veracity and classify it as real or fake.

4.1 Propagation Networks and Stance Network Construction

4.1.1 Construction of Propagation Networks

To model the diffusion behavior of news items, propagation networks are constructed using the NetworkX library in Python. Adopting the hierarchical propagation modeling strategy proposed in (Shu et al. 2020), two distinct graph structures are generated for each news item: a micro-level propagation network and a macro-level propagation network. Based on the dataset structure described in Section 3, both propagation networks are constrained to a maximum depth of two.

[3] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

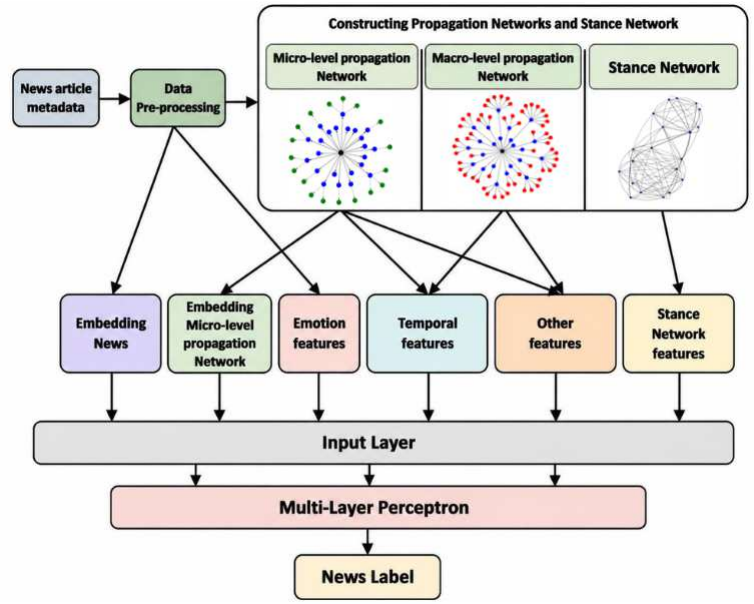
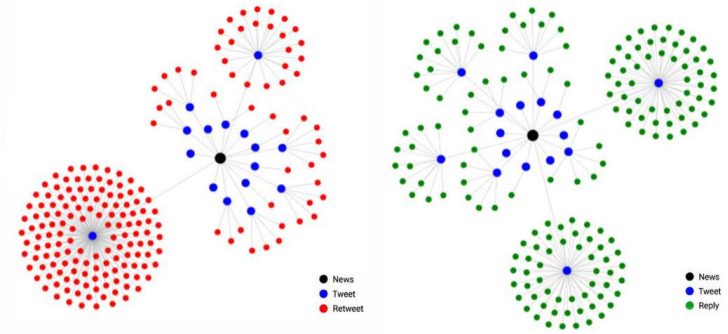


Fig. 2: The overall architecture of PGT.

The micro-level propagation network captures local conversational interactions, including replies associated with relevant tweets, whereas the macro-level propagation network represents broader dissemination patterns through retweeting behavior. Collectively, these two graph structures provide complementary perspectives on how a news item is discussed and disseminated across the platform.

Examples of the micro-level and macro-level propagation networks for the fake news article depicted in Figure 5 are presented in Figure 3a and Figure 3b, respectively.



(a) Macro-level propagation network (b) Micro-level propagation network

Fig. 3: Propagation networks for the fake news article shown in Figure 5.

4.1.2 Stance Network Construction

For each reply node in the micro-level propagation network, a sentiment-based score is computed. To estimate sentiment in Persian text, we employ a pre-trained sentiment analysis model, Behpouyan Sentiment (Behpouyan 2024), which is based on a transformer architecture and available in the Hugging Face Transformers library. The model classifies each text into three categories, positive, neutral, and negative, and yields a probability distribution across these classes. The final sentiment score for each reply node v_i, denoted as $\text{score}(v_i) \in [-1, +1]$, is defined as:

$$\text{score}(v_i) = (P_{\text{pos}}(v_i) \times 1) + (P_{\text{neu}}(v_i) \times 0) + (P_{\text{neg}}(v_i) \times -1) \tag{1}$$

where $P_{\text{pos}}(v_i)$, $P_{\text{neu}}(v_i)$, and $P_{\text{neg}}(v_i)$ represent the predicted probabilities for positive, neutral, and negative sentiment classes, respectively.

Following the formulation in (Davoudi et al. 2022), we construct an undirected stance network denoted by $G_s = (V_s, E_s)$, where V_s is the set of reply nodes in the propagation tree. An edge $e_{ij} \in E_s$ is established between nodes v_i and v_j if the following condition is satisfied:

$$|\text{score}(v_i) - \text{score}(v_j)| \leq 0.5 \tag{2}$$

For each connected pair of nodes (v_i, v_j), the corresponding edge weight w_{ij}^s is calculated based on their sentiment scores as follows:

$$w_{ij}^s = \frac{1}{|\text{score}(v_i) - \text{score}(v_j)| + \varepsilon_0} \tag{3}$$

where $\varepsilon_0 = 0.01$ is introduced to prevent division by zero when two nodes have identical sentiment scores.

An example of a stance network related to a fake news article is shown in Figure 4.

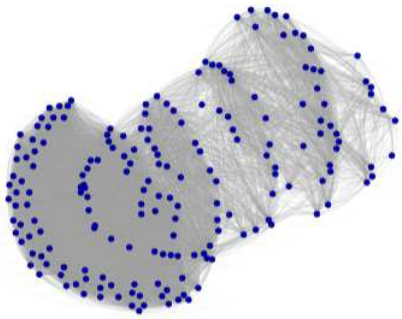


Fig. 4: Stance network for the fake news article shown in Figure 5.

4.2 Emotion features

We employ the six binary emotion features, namely Anger, Fear, Happiness, Disgust, Sadness and Surprise, introduced in the dataset described in Section 3, as content-based features for the news node text. Figure 5 illustrates an example of a fake news article.

۲ هزار ایرانی دو هفته پس از تزریق واکسن کرونا فوت کرده‌اند.

Fig. 5: An example of a fake news article in the P-HNet dataset.

Finally, the feature vector $x^{\{Emo\}}$ is constructed for each news article as follows:

For instance, consider the fake news article shown in Figure 5 Based on the dataset annotations, this article is associated with emotions of Anger and Fear. Its corresponding emotion feature vector is illustrated in Figure 6.

$$\boldsymbol{x^{\{Emo\}} = [1, 1, 0, 0, 0, 0]}$$

Fig. 6: The emotion feature vector for the fake news article shown in Figure 5.

4.3 Propagation Networks and Stance Network Features

To effectively distinguish between real and fake news, we extract a set of features from both micro-level and macro-level propagation networks, as well as the constructed stance network. Specifically, the feature extraction framework encompasses the following three main categories:

- **Temporal Features ($x^{\{\mathbf{Tem}\}}$):** Adapted from the HPFN model (Shu et al. 2020), these features characterize the temporal behavior of information diffusion across propagation networks. Given that the propagation networks in our curated dataset (described in Section 3) exhibit a maximum depth of two, certain depth-dependent features, namely T6, T12, and T13, become redundant and are therefore eliminated to avoid duplication. Consequently, the remaining distinct temporal metrics including T1 to T5 and T7 to T11 are designated for the macro-level and micro-level propagation networks.
- **Other Features ($x^{\{\mathbf{Other}\}}$):** These supplementary features, ranging from N_1 to N_{19}, capture early branching dynamics, user response variability, structural compactness, diffusion velocity, and overall network connectivity across both micro-level and macro-level networks.
- **Stance Network Features ($x^{\{\mathbf{Stance}\}}$):** Adopted from (Davoudi et al. 2022), covering features C10 through C17 and G13 through G18, these features capture

the sentiment, controversy, and topological properties of the stance network. Note that feature C16 was excluded due to the high network density in the Persian dataset making its computation exponentially complex, while features G13 and G14 were dropped as they did not positively impact model performance. Furthermore, because finding the maximum independent set is generally computationally challenging, features C17 and G18 were computed using a maximal independent set obtained via a greedy algorithm instead of the strictly maximum one.

The definitions and descriptions of all utilized network-based and stance-based features are summarized systematically in Table 1.

Finally, the respective feature vectors for each news article are formulated as follows:

$$x^{\{\text{Tem}\}} = [\text{T1 to T5 and T7 to T11}] \tag{4}$$

$$x^{\{\text{Other}\}} = [N_1, N_2, \ldots, N_{19}] \tag{5}$$

$$x^{\{\text{Stance}\}} = [\text{C10 to C15, C17, and G15 to G18}] \tag{6}$$

For example, consider the fake news article shown in Figure 5. Its temporal, other, and stance feature vector representations are illustrated in Figures 7, 8, and 9, respectively.

$$\boldsymbol{x}^{\{Tem\}} = [48663.17, 1.19e{+}07, 9.23e{+}06, 9.48e{+}06, 2.68e{+}06, 7.91e{+}06, 6572.82, 42906.06, 0, 91972]$$

Fig. 7: The temporal feature vector values in the propagation networks for the fake news article shown in Figure 5.

$$\boldsymbol{x}^{\{Other\}} = [10.12, 11.94, 5, 3, 11, 2.68e{+}06, 2.71e{+}05, 0.933, 0.933, 14.83, 17.79, 0.011, 0.010, 1.10e{-}07, 1.28e{-}07, 1.21e{+}08, 1.12e{+}08, 1.476, 1.480]$$

Fig. 8: The other feature vector values in the propagation networks for the fake news article shown in Figure 5.

$$\boldsymbol{x}^{\{Stance\}} = [-97.73, 0.881, 0.270, 4, 3.571, 5, 54, 4, 0, 0.004, -0.440, 0.026]$$

Fig. 9: The stance network feature vector values in the propagation networks for the fake news article shown in Figure 5.

Table 1: Propagation networks and stance network features used in the proposed model.

Features	Description
Category 1: Temporal Features ($x^{\{Tem\}}$)	
T1	Average time difference between the adjacent retweet nodes (macro-level)
T2	Time difference between the first tweet and the last retweets (macro-level)
T3	Time difference between the first tweet and the tweet with maximum outdegree (macro-level)
T4	Time difference between the first and last tweet posting news (macro-level)
T5	Time difference between the tweet posting news and last retweet node in deepest cascade (macro-level)
T7	Average time between tweets posting news (macro-level)
T8	Average time difference between the tweet post time and the first retweet time (macro-level)
T9	Average time difference between adjacent replies in cascade (micro-level)
T10	Time difference between the first tweet posting news and first reply node (micro-level)
T11	Time difference between the first tweet posting news and last reply node in micro network (micro-level)
Category 2: Other Features ($x^{\{Other\}}$) — Micro-level	
N1	Average degree of all tweet nodes
N3	Number of paths of length 1 originating from the root node
N5	Number of paths of length 2 originating from the root node
N7	Difference between the maximum and minimum reply time intervals
N8	Ratio of leaf nodes to the total number of nodes
N10	Average out-degree of nodes with at least one child (internal nodes)
N12	Ratio between network depth and total node count
N14	Total number of nodes divided by overall diffusion duration
N16	Standard deviation of sequential event time intervals
N18	Average shortest path length among all node pairs
Category 2: Other Features ($x^{\{Other\}}$) — Macro-level	
N2	Average degree of all tweet nodes
N4	Number of paths of length 1 originating from the root node
N6	Difference between the maximum and minimum retweet time intervals
N9	Ratio of leaf nodes to the total number of nodes
N11	Average out-degree of nodes with at least one child (internal nodes)
N13	Ratio between network depth and total node count
N15	Total number of nodes divided by overall diffusion duration
N17	Standard deviation of sequential event time intervals
N19	Average shortest path length among all node pairs
Category 3: Stance Network Features ($x^{\{Stance\}}$)	
C10	Total reply sentiment scores
C11	Local clustering coefficient
C12	Node centrality measures
C13	Stance network diameter
C14	Average stance network eccentricity
C15	Number of nodes forming the shortest path between the nodes with the maximum and minimum scores
C17	Number of nodes forming the maximal independent set of the stance network
G15	Local node connectivity
G16	Total sentiment score of nodes forming the shortest path between the nodes with the maximum and minimum scores
G17	The weighted average of sentiment scores among cliques with different sizes in the stance network
G18	The total sentiment of nodes forming the maximal independent set of the stance network

4.4 News Text Embedding

In this study, the textual content of each news item is represented as a sentence-level embedding vector. For this purpose, we use the LaBSE model. The text of each news item is fed into the LaBSE model, and the resulting 768-dimensional vector is used as its representation. Accordingly, the textual feature vector is constructed as follows:

$$x^{\{LaBSE\}} = (768 \text{ features}) \tag{7}$$

4.5 Micro-Level Propagation Network Embedding

Furthermore, a graph-based approach is employed to generate embeddings for the micro-level propagation network. In this study, we utilize the MPNN model, which was originally proposed for molecular property prediction in (Akensert 2021), to embed the micro-level propagation network into a 32-dimensional vector. In this model, each node is represented by a set of network centrality measures, including degree_centrality, closeness_centrality, betweenness_centrality, PageRank_centrality, and Katz_centrality.

These measures are computed using the NetworkX library in Python and provide information about the relative position and importance of each node within the propagation network. In addition, edge features are defined using the timestamp difference between the source and destination nodes in order to capture temporal differences between connected nodes. Finally, the propagation network feature vector is constructed as follows:

$$x^{\{MPNN\}} = (32 \text{ features}) \tag{8}$$

4.6 Integration Module

In the final step, all extracted features are concatenated to form the final representation vector $x^{\{Final\}}$, as illustrated in Figure 10.

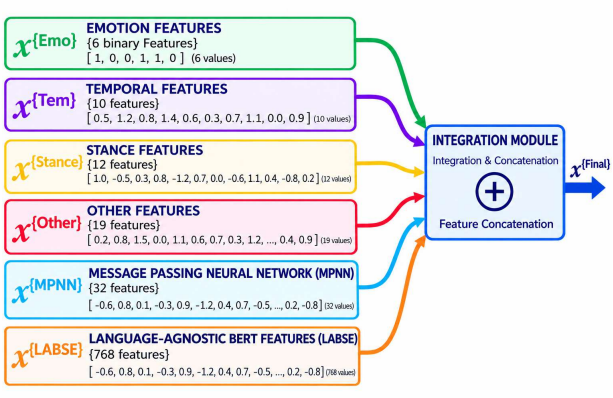


Fig. 10: Architecture of the Integration Module.

5 Evaluation

This section details the implementation environment and the infrastructure required to conduct the experiments. Furthermore, we discuss the process of hyperparameter identification and optimization, aimed at achieving the peak performance of the model. Finally, the evaluation metrics selected to assess the system's accuracy and performance are described in detail.

5.1 Implementation Environment

All experiments were conducted on a system equipped with an NVIDIA GeForce RTX 3090 GPU with 24 GB of VRAM and an AMD Ryzen 7 5800X 8-Core Processor. The machine was configured with 44 GB of RAM.

5.2 Hyperparameters and Experimental Settings

This section details the experimental settings and hyperparameter specifications for the evaluated baselines. The comparative study encompasses advanced graph-based architectures, specifically HPFN (Shu et al. 2020) and DSS (Davoudi et al. 2022). The operational parameters and training configurations for each approach are discussed in the following subsections.

5.2.1 Proposed PGT Architecture

To determine the optimal configuration for the PGT model, we conducted a systematic search across 45 configurations, as detailed in Table 2. This search spanned three learning rates, three weight decay values, and five distinct network architectures. Since this space covers the entire range of potential combinations, we employed a grid search strategy, evaluating all 45 combinations in a single pass. Given the limited ranges of hyperparameters and the relatively compact network architectures considered, this exhaustive approach proved computationally feasible. The objective was to balance model capacity with overfitting prevention, while investigating how the interplay of network depth, width, dropout, and weight decay impacts overfitting control. For each configuration, we evaluated the model performance on every validation fold and ultimately selected the hyperparameters yielding the highest average OOF[4] score using stratified 5-fold cross-validation.

Based on the grid search results from the 5-fold cross-validation, the configuration producing the highest performance corresponds to Arch_1, using a learning rate of $\eta = 5 \times 10^{-4}$ and a weight decay coefficient of $\lambda = 1 \times 10^{-4}$. Consequently, the optimal MLP architecture comprises four hidden layers with 256, 128, 64, and 16 neurons, coupled with dropout rates of 0.5, 0.4, 0.3, and 0.2, respectively. Each hidden block consists of a linear transformation followed by a Batch Normalization layer, the Rectified Linear Unit (ReLU) activation function, and the corresponding dropout layer. The network is trained using the Adam optimizer with a mini-batch size of 64 for up to 30 epochs, employing a weighted binary cross-entropy loss to effectively manage class imbalance.

[4] Out-of-Fold

Table 2: Hyperparameter and Architecture Search Space

Architecture ID	Hidden Dimensions	Dropout Rates
Arch_1	$[256, 128, 64, 16]$	$[0.5, 0.4, 0.3, 0.2]$
Arch_2	$[256, 64, 16]$	$[0.5, 0.4, 0.2]$
Arch_3	$[256, 64, 16]$	$[0.3, 0.2, 0.1]$
Arch_4	$[128, 64, 32]$	$[0.4, 0.3, 0.2]$
Arch_5	$[128, 32]$	$[0.3, 0.2]$
Hyperparameters:		
Learning Rate (η): $\{1 \times 10^{-3}, 5 \times 10^{-4}, 1 \times 10^{-4}\}$		
Weight Decay (λ): $\{1 \times 10^{-2}, 1 \times 10^{-3}, 1 \times 10^{-4}\}$		

5.2.2 DSS Baseline Architecture

To facilitate a rigorous and impartial evaluation of the DSS baseline (Davoudi et al. 2022), we examined both the original hyperparameter settings of the best-performing DSS configuration and an optimized variant using the grid search space defined in Table 2. In the original study, the optimal configuration for dynamic analysis was achieved with $\alpha = 12$ and $\beta = 2$. Here, α denotes the number of constructed propagation trees and stance networks, while β represents the hourly time interval between consecutive sequences, establishing a total detection deadline of $\alpha \times \beta = 24$ hours post-publication. For this configuration, we adopted the original MLP hyperparameters: a mini-batch size of 100, a learning rate of 1×10^{-3}, a uniform dropout rate of 0.2, and an L_2 regularization parameter of $\lambda = 0.01$.

Additionally, to assess whether hyperparameter fine-tuning could improve adaptation to our dataset, we performed an exhaustive grid search over the architectures listed in Table 2. Empirical results indicated that the highest performance was achieved with Arch_5, configured with a learning rate of $\eta = 1 \times 10^{-3}$ and a weight decay coefficient of $\lambda = 1 \times 10^{-2}$. Consequently, the optimal MLP architecture comprises two hidden layers with 128 and 32 neurons, paired with dropout rates of 0.3 and 0.2, respectively. Each hidden block consists of a linear transformation, a Batch Normalization layer, the ReLU activation function, and the corresponding dropout layer. The network is trained using the Adam optimizer with a mini-batch size of 100 for up to 40 epochs. We employ a weighted binary cross-entropy loss to effectively mitigate class imbalance, alongside a learning rate reduction schedule on plateau.

5.2.3 HPFN Baseline Architecture

Following the original implementation of the HPFN baseline (Shu et al. 2020), we employed a Random Forest classifier to ensure a rigorous and fair comparison. Since the baseline study specified the use of the default parameter settings provided by the scikit-learn library[5], we adhered strictly to these default hyperparameters in our evaluation to maintain consistency and benchmark integrity.

[5] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

5.3 Evaluation Metrics

To assess the effectiveness of the proposed model, we used two widely recognized classification metrics: Accuracy and F1-score. These metrics provide complementary views of predictive performance. Accuracy measures the proportion of correctly predicted instances among all evaluated samples, reflecting the overall correctness of the classifier. It is calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

where TP represents correctly predicted positive samples, TN denotes correctly predicted negative samples, FP indicates negative samples incorrectly classified as positive, and FN refers to positive samples incorrectly predicted as negative. While accuracy provides a general view, it may not fully capture the balance between identifying positive instances and avoiding false alarms. Therefore, the F1-score is also used. The F1-score combines precision and recall into a single metric using their harmonic mean. Precision represents the proportion of predicted positive instances that are actually positive, while recall measures the proportion of true positive instances correctly identified by the model. These metrics are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{11}$$

$$\text{F1-score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{12}$$

The F1-score is particularly valuable when dealing with imbalanced datasets, as it balances the trade-off between precision and recall.

6 Results and Discussion

In this section, we present the experimental results of the proposed fake news detection model. We first analyze the performance of various model configurations developed in this study, and then compare our best-performing model against existing baseline methods.

6.1 Evaluation of the Proposed Model

In this section, we present a comprehensive evaluation of the PGT model using a Stratified 5-fold cross-validation strategy on the P-HNet dataset. Reporting both individual fold results and overall OOF classification metrics allows for a robust assessment of model stability and limits the risk of overfitting. Additionally, we conduct a systematic ablation study to isolate the contribution of each feature subset.

Performance metrics for each stage of Stratified 5-fold cross-validation, including accuracy and F1-score for the fake news class, are detailed in Table 3. The model demonstrates stable performance across all validation folds, yielding a mean accuracy of $75.86\% \pm 1.55\%$ and a mean F1-score of $66.43\% \pm 2.40\%$ for the fake news class. This low standard deviation across validation folds confirms the robustness of the model in handling the inherent complexities of Persian fake news detection.

To analyze the classification behavior across different news classes, Table 4 reports the detailed classification performance per class using OOF predictions. The model achieves a precision of 0.82 for the Real News class, indicating its high capability in correctly identifying authentic articles. For the Fake News class, a recall of 0.68 highlights the model's ability to identify deceptive content, while a macro-averaged F1-score of 0.74 confirms the balanced performance of the model across classes and its effectiveness in mitigating class imbalance challenges.

Finally, to quantify the contribution of each feature group to the predictive power of the PGT model, we conducted a group-wise ablation study on the test set. Table 5 presents the performance degradation observed when specific feature groups are removed from the complete feature set ($x^{\{\text{Final}\}}$). The results emphasize the critical role of news text representations, propagation network features, and textual emotion features. Removing the LaBSE features leads to the most severe drop in accuracy and F1-score, followed by micro- and macro-level propagation network features and textual emotion features as having the greatest impact.

Table 3: Performance of the PGT model using Stratified 5-fold cross-validation on the P-HNet dataset.

Fold	**Accuracy (%)**	**F1-score (%)**
Fold 1	77.14	67.74
Fold 2	74.29	65.12
Fold 3	74.29	62.81
Fold 4	76.00	68.18
Fold 5	77.59	68.29
Average	**75.86 ± 1.55**	**66.43 ± 2.40**

Based on the empirical results of the ablation experiments in terms of F1-score, the relative significance of the feature subsets is summarized below:

$$x^{\{\text{Stance}\}} < x^{\{\text{MPNN}\}} \approx x^{\{\text{Tem}\}} \approx x^{\{\text{Emo}\}} < x^{\{\text{Other}\}} \ll x^{\{\text{LaBSE}\}}$$

Table 4: Detailed classification performance per class based on Overall OOF predictions.

Class	Precision	Recall	F1-score	Support
Fake News	0.65	0.68	0.66	308
Real News	0.82	0.81	0.81	566
Macro Average	0.74	0.74	0.74	874
Weighted Average	0.76	0.76	0.76	874

Table 5: Ablation study results for the PGT model on the test set of the P-HNet dataset.

Type	Accuracy (%)	F1-score (%)
$x^{\{\text{Final}\}} - x^{\{\text{Tem}\}}$	77.17	69.14
$x^{\{\text{Final}\}} - x^{\{\text{Other}\}}$	79.45	68.09
$x^{\{\text{Final}\}} - x^{\{\text{Stance}\}}$	78.54	70.06
$x^{\{\text{Final}\}} - x^{\{\text{Emo}\}}$	74.89	68.93
$x^{\{\text{Final}\}} - x^{\{\text{LaBSE}\}}$	73.52	60.81
$x^{\{\text{Final}\}} - x^{\{\text{MPNN}\}}$	75.34	69.66
$x^{\{\text{Final}\}} = (\text{PGT})$	**82.19**	**74.51**

The results indicate that although stance network features, temporal features of propagation networks, and micro-level propagation network representations provide improvements to baseline model performance, the primary discriminative power of the model stems from propagation-network-based features, emotional features, and news text representations (LaBSE).

6.2 Comparison with Previous Methods

As the baseline methods were originally developed for non-Persian datasets, reproducing all their original features on Persian data faced certain constraints. Nevertheless, to ensure a rigorous and fair comparison, each method was implemented using the closest feasible configuration. For the DSS model, the onset of news propagation was defined as the timestamp of the first tweet associated with a given news item. Regarding feature extraction, G12 was omitted because the propagation networks in our dataset are limited to a maximum depth of two, making G12 redundant with G11. Similarly, among the stance network features, C16 was modified; since computing this feature requires counting all cliques, which leads to exponentially complex calculations due to the high density of Persian networks, it was replaced with counting only maximal cliques. All stance network features were thus extracted, except for the exclusions and adaptations detailed in Section 4.3.

Regarding the HPFN baseline (Shu et al. 2020), the original architecture is built upon a total of 32 features categorized into three primary groups: Structural (14 features), Temporal (13 features), and Linguistic (5 features), extracted from both micro-level and macro-level propagation networks. The breakdown of these features and their screening process for the present study are as follows:

1. **Structural Features (S_1 to S_{14}):** Comprising 9 macro-level (S_1 to S_9) and 5 micro-level (S_{10} to S_{14}) features, 12 were successfully implemented, while 2 features (S_8 and S_9, related to bot behavior detection) were excluded due to dataset limitations.
2. **Temporal Features (T_1 to T_{13}):** Consisting of 8 macro-level (T_1 to T_8) and 5 micro-level (T_9 to T_{13}) features. Due to the limited depth of propagation networks in our dataset, 3 features (T_6, T_{12}, and T_{13}) were redundant with T_1, T_9, and T_{11}, respectively, and were removed accordingly.
3. **Linguistic Features (L_1 to L_5):** All 5 features are computed directly at the micro-level. To adapt these features for the Persian language, positive and negative sentiment scores were derived using a pre-trained Persian sentiment analysis model based on ParsBERT.[6] Feature L_3 was removed due to redundancy with L_4 caused by the propagation depth constraint, leaving the remaining 4 features (L_1, L_2, L_4, L_5) in the final baseline.

As illustrated in Table 6, the proposed PGT model outperforms all baseline approaches on the test set, achieving the highest accuracy of 82.19% and an F1-score of 74.51%. Among the baselines, while DSS and HPFN leverage propagation information, their performance is substantially constrained by missing features that could not be extracted from Persian data, as well as distinct propagation patterns in Persian social media. This marked superiority demonstrates that effectively integrating news text representations, propagation network features, and emotional features provides significantly greater discriminative power for fake news detection compared to existing baseline models.

Table 6: Performance comparison between the proposed PGT model and baseline methods on the P-HNet dataset.

Model	Accuracy (%)	F1-score (%)
DSS (Optimized)	56.62	45.71
DSS (Default Parameters)	54.79	44.07
HPFN	66.21	42.19
PGT (Proposed)	**82.19**	**74.51**

7 Conclusion and Future Work

This study explored the feasibility of fake news detection by leveraging diverse feature sets derived from propagation and stance networks. We introduced PGT, a hybrid model designed for Persian fake news detection that integrates textual emotions, news content representations, micro-level propagation network representations, structural features from both micro- and macro-level propagation networks, and stance network features to effectively discriminate between authentic and fabricated news.

[6] https://huggingface.co/HooshvareLab/bert-fa-base-uncased-sentiment-snappfood

The experimental results demonstrated that PGT outperforms baseline approaches. These findings suggest that combining propagation network features with news content attributes offers a superior solution compared to relying on singular information sources. Furthermore, the ablation study confirmed that news text representations, news propagation network features, and textual emotion features, in conjunction with other extracted features, are the most critical factors influencing the overall performance of the model.

Future research could expand the dataset by incorporating a larger volume of news articles, users, and thematic categories to further assess the model's robustness. Moreover, categorizing news by topic and examining feature efficacy within specific domains would be beneficial. User-level information, such as follower counts, could also be integrated to reflect user credibility within the dissemination process. Finally, incorporating visual representations of propagation networks alongside news text and micro-level network embeddings may provide complementary signals to further enhance detection performance.

Another promising direction is to extend the proposed PGT framework toward a Mixture-of-Experts (MoE) architecture for fake news detection. Inspired by the multi-expert strategy employed in Matina (Hosseinbeigi et al. 2025), future research could develop a set of specialized fake news detection models, each trained to assess news credibility from a distinct perspective. For example, separate experts could focus on semantic inconsistencies in news content, source credibility, user behavioral patterns, information propagation structures, or topic-specific indicators. A gating mechanism could then dynamically select or weight these experts according to the characteristics of each news item, integrating their complementary assessments into a final credibility prediction. This extension could facilitate more adaptive and comprehensive fake news detection while allowing each expert to specialize in a particular dimension of misinformation.

Author contributions All authors have contributed significantly to the development and completion of this research. Sara Bourbour Hosseinbeigi led the conceptual development of the study, designed the methodological framework, and supervised the overall research execution. Mohammad Javad Fatemi was actively involved in data acquisition, performed the analysis, and contributed to interpreting the results. Both authors participated in writing, reviewing, and approving the final version of the manuscript.

Funding The authors received no financial funding for this research project.

Data availability The P-HNet dataset analyzed during the current study is publicly available on Hugging Face at https://huggingface.co/datasets/MohammadJavadFatemi/P-HNet.

Declarations

Competing interests The authors declare that they have no competing interests.

Ethics approval This study does not involve any human participants or animal subjects.

Consent to participate Not applicable.

Consent to publish Not applicable.

References

Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. Journal of economic perspectives **31**(2), 211–236 (2017) https://doi.org/10.1257/jep.31.2.211

Akensert: Message-passing neural network (MPNN) for molecular property prediction. keras. https://keras.io/examples/graph/mpnn-molecular-graphs. Accessed 14 July 2026 (2021)

Albahar, M.: A hybrid model for fake news detection: Leveraging news content and user comments in fake news. IET Information Security **15**(2), 169–177 (2021) https://doi.org/10.1049/ise2.12021

Alghamdi, J., Luo, S., Lin, Y.: A comprehensive survey on machine learning approaches for fake news detection. Multimedia Tools and Applications **83**(17), 51009–51067 (2024) https://doi.org/10.1007/s11042-023-17470-8

Alonso, M.A., Vilares, D., Gómez-Rodríguez, C., Vilares, J.: Sentiment analysis for fake news detection. Electronics **10**(11), 1348 (2021) https://doi.org/10.3390/electronics10111348

Bazmi, P., Asadpour, M., Shakery, A.: Multi-view co-attention network for fake news detection by modeling topic-specific user and news source credibility. Information Processing & Management **60**(1), 103146 (2023) https://doi.org/10.1016/j.ipm.2022.103146

Behpouyan: Behpouyan-Sentiment. Hugging Face. https://huggingface.co/Behpouyan/Behpouyan-Sentiment. Accessed 14 July 2026 (2024)

Bourbour Hosseinbeygi, S., SeifKashani, M., Momtazi, S., Fatemi, M.: Fake-news detection by 3 modalities: text + user history + propagation network. In: Proceedings of the 3rd Regional Symposium on Innovation in Science and Technology (ReSIST 2026), Kufa, Iraq (2026). January 2–3, 2026. https://civilica.com/doc/2738958

Bourbour, S., SeifKashani, M., Aali, F.: Claim verification in persian: A hybrid pipeline with llms and rationale-quality classification. In: 2026 International Interdisciplinary Conference on Artificial Intelligence: Engineering, Health, Finance and Humanities (IICAI 2026), pp. 1–6. IEEE, ??? (2026). https://doi.org/10.1109/IICAI70155.2026.11620856

Choudhary, A., Arora, A.: Linguistic feature based learning model for fake news detection and classification. Expert Systems with Applications **169**, 114171 (2021) https://doi.org/10.1016/j.eswa.2020.114171

Davoudi, M., Moosavi, M.R., Sadreddini, M.H.: Dss: A hybrid deep model for fake news detection using propagation tree and stance network. Expert Systems with Applications **198**, 116635 (2022) https://doi.org/10.1016/j.eswa.2022.116635

Dair, K., Sadirmekova, Z.: Fake news detection on twitter using machine learning. In: 2026 IEEE 6th International Conference on Smart Information Systems and Technologies (SIST), pp. 1–6 (2026). https://doi.org/10.1109/SIST61674.2026.11596077 . IEEE

Ebiwonjumi, O., Ahakonye, L.A.C., Kim, D.-S.: Examining software engineering practices in the pre-ai and post-ai era. In: 2026 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp. 651–656 (2026). https://doi.org/10.1109/ICAIIC68212.2026.11454249 . IEEE

Eghbalzadeh, H., Hosseini, B., Khadivi, S., Khodabakhsh, A.: Persica: A persian corpus for multi-purpose text mining and natural language processing. In: 6th International Symposium on Telecommunications (IST), pp. 1207–1214 (2012). https://doi.org/10.1109/ISTEL.2012.6483172 . IEEE

Feng, F., Yang, Y., Cer, D., Arivazhagan, N., Wang, W.: Language-agnostic bert sentence embedding. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (volume 1: Long Papers), pp. 878–891 (2022). https://doi.org/10.18653/v1/2022.acl-long.62

Gao, T., Yang, J., Peng, W., Jiang, L., Sun, Y., Li, F.: A content-based method for sybil detection in online social networks via deep learning. IEEE Access **8**, 38753–38766 (2020) https://doi.org/10.1109/ACCESS.2020.2975877

Hosseini, M., Jafari, M.: A novel content-based approach for fake news detection using transformer model: A case study of covid-19 dataset. In: Proceedings of International Conference on Web Research (ICWR), pp. 45–52 (2024). https://doi.org/10.1109/ICWR61162.2024.10533339 . IEEE

Hosseinbeigi, S.B., SeifKashani, M., Seraj, J., Taherinezhad, F., Nafisi, A., Nadi, F., Barati, I., Hasani, H., Amiri, M., Masoudi, M.: Matina: A culturally-aligned Persian language model using multiple LoRA experts. In: Findings of the Association for Computational Linguistics: ACL 2025, pp. 20874–20889. Association for Computational Linguistics, Vienna, Austria (2025). https://doi.org/10.18653/v1/2025.findings-acl.1074 . https://aclanthology.org/2025.findings-acl.1074/

Hosseinbeigi, S.B., Taherinezhad, F., Faili, H., Baghbani, H., Nadi, F., Amiri, M.: Matina: A large-scale 73B token Persian text corpus. In: Proceedings of

the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pp. 9143–9157. Association for Computational Linguistics, Albuquerque, New Mexico (2025). https://doi.org/10.18653/v1/2025.naacl-long.462 . https://aclanthology.org/2025.naacl-long.462/

Hu, L., Yang, T., Zhang, L., Zhong, W., Tang, D., Shi, C., Duan, N., Zhou, M.: Compare to the knowledge: Graph neural fake news detection with external knowledge. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (volume 1: Long Papers), pp. 754–763 (2021). https://doi.org/10.18653/v1/2021.acl-long.62

Khoshraftar, S., An, A.: A survey on graph representation learning methods. ACM Transactions on Intelligent Systems and Technology **15**(1), 1–55 (2024) https://doi.org/10.1145/3633518

Kudugunta, S., Ferrara, E.: Deep neural networks for bot detection. Information Sciences **467**, 312–322 (2018) https://doi.org/10.1016/j.ins.2018.08.019

Khattar, D., Goud, J.S., Gupta, M., Varma, V.: Mvae: Multimodal variational autoencoder for fake news detection. In: The World Wide Web Conference, pp. 2915–2921 (2019). https://doi.org/10.1145/3308558.3313552

Kuntur, S., Wróblewska, A., Ganzha, M., Paprzycki, M., Sachdeva, S.: Fake news detection: It's all in the data! Applied Sciences **16**(3), 1585 (2026) https://doi.org/10.3390/app16031585

Lv, J., Gao, Y., Li, L., Shi, L., Li, S.: Multi-modal fake news detection: A comprehensive survey on deep learning technology, advances, and challenges. Journal of King Saud University Computer and Information Sciences **37**(9), 306 (2025) https://doi.org/10.1007/s44443-025-00317-7

Monti, F., Frasca, F., Eynard, D., Mannion, D., Bronstein, M.M.: Fake news detection on social media using geometric deep learning. arXiv preprint arXiv:1902.06673 (2019)

Mayank, M., Sharma, S., Sharma, R.: Deap-faked: Knowledge graph based approach for fake news detection. In: 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 47–51 (2022). https://doi.org/10.1109/ASONAM55673.2022.10068653 . IEEE

Meel, P., Vishwakarma, D.K.: Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. Expert Systems with Applications **153**, 112986 (2020) https://doi.org/10.1016/j.eswa.2019.112986

Nasir, J.A., Khan, O.S., Varlamis, I.: Fake news detection: A hybrid cnn-rnn based deep learning approach. International journal of information management data insights **1**(1), 100007 (2021) https://doi.org/10.1016/j.jjimei.2020.100007

Oshikawa, R., Qian, J., Wang, W.Y.: A survey on natural language processing for fake news detection. In: Proceedings of the Twelfth Language Resources and Evaluation Conference, pp. 6086–6093 (2020)

Park, M., Chai, S.: Constructing a user-centered fake news detection model by using classification algorithms in machine learning techniques. IEEE Access **11**, 71517–71527 (2023) https://doi.org/10.1109/ACCESS.2023.3294613

Phan, H.T., Nguyen, N.T., Hwang, D.: Fake news detection: A survey of graph neural network methods. Applied Soft Computing **139**, 110235 (2023) https://doi.org/10.1016/j.asoc.2023.110235

Pan, J.Z., Pavlova, S., Li, C., Li, N., Li, Y., Liu, J.: Content based fake news detection using knowledge graphs. In: International Semantic Web Conference, pp. 669–683 (2018). https://doi.org/10.1007/978-3-030-00671-6_39 . Springer

Qi, P., Cao, J., Yang, T., Guo, J., Li, J.: Exploiting multi-domain visual information for fake news detection. In: 2019 IEEE International Conference on Data Mining (ICDM), pp. 518–527 (2019). https://doi.org/10.1109/ICDM.2019.00062 . IEEE

Ramalingam, D., Chinnaiah, V.: Fake profile detection techniques in large-scale online social networks: A comprehensive review. Computers & Electrical Engineering **65**, 165–177 (2018) https://doi.org/10.1016/j.compeleceng.2017.05.020

Rostami, R.R., Karbasi, S.: Detecting fake accounts on twitter social network using multi-objective hybrid feature selection approach. Webology **17**(1) (2020) https://doi.org/10.14704/WEB/V17I1/a204

Ruchansky, N., Seo, S., Liu, Y.: Csi: A hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 797–806 (2017). https://doi.org/10.1145/3132847.3132877

Segura-Bedmar, I., Alonso-Bartolome, S.: Multimodal fake news detection. Information **13**(6), 284 (2022) https://doi.org/10.3390/info13060284

Samadi, M., Momtazi, S.: Fake news detection: deep semantic representation with enhanced feature engineering. International Journal of Data Science and Analytics **20**(2), 325–336 (2025) https://doi.org/10.1007/s41060-023-00387-8

Samadi, M., Mousavian, M., Momtazi, S.: Persian fake news detection: Neural representation and classification at word and text levels. Transactions on Asian and Low-Resource Language Information Processing **21**(1), 1–11 (2021) https://doi.org/10.1145/3472620

Sarlakifar, F., Mortazavi, M.M., Shamsfard, M.: Emorecbigru: Emotion recognition in persian tweets with a transformer-based model, enhanced by bidirectional gru. International Journal of Information & Communication Technology Research (2251-6107) **16**(3) (2024) https://doi.org/10.61186/itrc.16.3.35

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big data **8**(3), 171–188 (2020) https://doi.org/10.1089/big.2020.0062

Shu, K., Mahudeswaran, D., Wang, S., Liu, H.: Hierarchical propagation networks for fake news detection: Investigation and exploitation. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 14, pp. 626–637 (2020). https://doi.org/10.1145/3373464.3373473

Soga, K., Yoshida, S., Muneyasu, M.: Exploiting stance similarity and graph neural networks for fake news detection. Pattern Recognition Letters **177**, 26–32 (2024) https://doi.org/10.1016/j.patrec.2023.11.019

Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G.S., On, B.-W.: Fake news stance detection using deep learning architecture (cnn-lstm). IEEE Access **8**, 156695–156706 (2020) https://doi.org/10.1109/ACCESS.2020.3019735

Uppada, S.K., Manasa, K., Vidhathri, B., Harini, R., Sivaselvan, B.: Novel approaches to fake news and fake account detection in osns: user social engagement and visual content centric model. Social Network Analysis and Mining **12**(1), 52 (2022) https://doi.org/10.1007/s13278-022-00878-9

Xu, R., Shang, J., Guan, M., Wang, J., Cui, H., Min, G.: Mese: Mining emotional and semantic evolution from user comments for fake news detection. IEEE Transactions on Knowledge and Data Engineering (2026) https://doi.org/10.1109/TKDE.2026.3690971

Yuan, C., Ma, Q., Zhou, W., Han, J., Hu, S.: Early detection of fake news by utilizing the credibility of news, publishers, and users based on weakly supervised learning. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 5444–5454 (2020). https://doi.org/10.18653/v1/2020.coling-main.475

Yu, P., Xia, Z., Fei, J., Lu, Y.: A survey on deepfake video detection. Iet Biometrics **10**(6), 607–624 (2021) https://doi.org/10.1049/bme2.12031

Zhang, X., Ghorbani, A.A.: An overview of online fake news: Characterization, detection, and discussion. Information Processing & Management **57**(2), 102025 (2020) https://doi.org/10.1016/j.ipm.2019.03.004

Zheng, H.: Research evolution and hotspots analysis of disinformation induced by generative artificial intelligence (2020–2025). Computers and Artificial Intelligence

3(3), 258–270 (2026) https://doi.org/10.70267/cai.26v3n3.258270

Zhou, X., Jain, A., Phoha, V., Zafarani, R.: Fake news early detection: A theory-driven model. Digital Threats: Research and Practice, 1 (2), 1–25 (2020). https://doi.org/10.1145/3377478

Zhou, X., Zafarani, R.: Network-based fake news detection: A pattern-driven approach. ACM SIGKDD explorations newsletter **21**(2), 48–60 (2019) https://doi.org/10.1145/3373464.3373473

Zhou, X., Zafarani, R.: A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR) **53**(5), 1–40 (2020) https://doi.org/10.1145/3395046