Bipartite Graph Modeling for the Analysis of Fake News Propagation

Carmela Bernardo, Marta Catillo, Antonio Pecchia,
Francesco Vasca, Umberto Villano

Dipartimento di Ingegneria, Università degli Studi del Sannio, Benevento, Italy
{carmela.bernardo,marta.catillo,antonio.pecchia,
francesco.vasca,villano}@unisannio.it

Abstract. The propagation of fake news on online social networks represents one of the main causes of the spread of misinformation in modern society. In this paper, a bipartite graph model is proposed to analyze the relationships between different fake news spreaders regarding different topics represented by keywords derived from the content of the articles. The projections of the weighted graph allow us to identify the spreaders of fakeness that play a central role and the interconnections between topics most affected by fakeness. The effectiveness of the model is demonstrated by applying it to real datasets derived from scraping on the Politifact.com website and considering the weights obtained from fact-checking of the news content.

Keywords: Fake news propagation, graph theory, bipartite graph, social network analysis.

1 Introduction

The rise of social media has revolutionized communication, allowing people to connect, share information, and stay up-to-date. However, a dark side has emerged: a flood of unreliable and often deceptive information, commonly known as **'fake news'**. This surge of online fabrication poses a significant threat to society. One prominent example is the suspected influence of fake news on the 2016 U.S. election [1]. Since then, 'fake news' has become a household term, sparking intense interest from both tech companies and researchers who are eager to dissect its creation, spread, and impact.

Detecting online content that is both false and designed to mislead presents significant technical challenges [14, 16, 24]. These arise from the ease with which social media tools generate and disseminate vast amounts of content, creating a formidable volume for analysis. The diversity of online information further complicates this task, spanning a wide array of topics and adding layers of complexity. Evaluating the accuracy and intent of statements often exceeds the capabilities of automated systems alone, necessitating a collaborative approach blending human expertise with technological solutions. While sources identifying deceptive content are scarce, they serve as foundational elements for fostering cooperative efforts in combating misinformation.

This paper proposes a bipartite graph-based method in order to profile and characterize fake news spreaders. In spite of the plethora of papers, which leverage graphs to make a propagation analysis aimed at the detection of fake news, the body of literature focused on characterizing fake news spreaders and propagation by using social network analysis has received a more limited attention, see [19, 20] and the references therein.

In this paper, we propose to use a bipartite graph in order to explore the link between spreaders and the likelihood of sharing fake news. We model a **weighted bipartite graph**, constructed with data scraped from PolitiFact (`https://www.politifact.com/`, accessed on May 2024). The analysis of the graph sheds light on two pilot dimensions: the **spreader dimension** and the **keyword dimension**. Degree centrality analysis – *unweighted* and *weighted* by the fakeness level of the news – in the bipartite graph allows the identification of central spreaders and keywords in fake news propagation. In such a way, we are able to isolate spreaders that participate across a wide range of topics, which may propagate fake news, and frequently referenced keywords, which could stand for misinformation. The analysis is complemented with the **projection graphs** which allows for the investigation of interaction patterns among nodes of the same category in the fake news propagation. The results show valuable insights in terms of source characterization. For example, social media platform sources exhibit high weighted degrees, thus reflecting extensive fake news publications. Looking at the keywords used together, instead, we can identify some patterns that are often found in fake news. For example, fake news thrives on topics with the keywords *'Ukraine'* or *'Russia'*. By handling massive datasets and uncovering crucial connections, this analysis paves the way for developing custom indicators that can pinpoint and profile individuals and keywords involved in spreading fake news.

The rest of the paper is organized as follows. Sect. 2 presents the background, terminology, and related work in the fake news area. Sect. 3 shows the data collection process. Sect. 4 describes the methodology with the experimental setup. Sect. 5 assesses the models against the collected data and provides results and findings. Sect. 6 concludes the work.

2 Background and Related work

Characterizing and modeling the interaction between users and news can reveal many hidden features, such as user opinions, stances, and membership in specific communities, even if these aspects are not explicitly stated. In addition, analyzing the spread of news across different media sources can provide valuable insights into the bias and trustworthiness of the news content provider. Below we provide some basic notions on the terminology underlying the domain of fake news propagation and highlight the contribution of the article compared to the existing literature on the topic.

2.1 Fundamentals on fake news

The term fake news has become the *'de facto'* expression for identifying false information in mainstream media, particularly concerning online content, which gained significant prominence during and after the 2016 U.S. Presidential Campaign. However, research on fake news generally uses a more restrictive definition. According to [1], a fake news is *'a news article that is intentionally and verifiably false'*. This definition focuses on two crucial elements: **intent** and **verifiability**. Fake news, therefore, refers to news articles deliberately crafted to deceive or misinform readers, which can be proven false through other sources. To clearly understand the spread of fake news, it is essential to discuss its key components [24]. These components can be divided into four main categories: *creator/spreader*, *target victim*, *content*, and *social context*.

Creator/Spreader. The creators of online fake news can be either humans or non-humans. Human creators include both well-intentioned individuals who unknowingly publish false information and malicious users who intentionally create fake news.

Target Victims. Victims are the primary targets of online fake news. They can be users on social media or other online news platforms. Depending on the purpose of the fake news, targets can include students, voters, parents, senior citizens, and others.

News Content. News content encompasses the main body of the news, which includes both physical elements (e.g., title, body text, multimedia) and non-physical elements (e.g., purpose, sentiment, topics).

Social Context. Social context refers to how news is distributed across the Internet. This includes user network analysis (examining how online users engage with the news) and broadcast pattern analysis (studying the temporal patterns of dissemination).

2.2 Graph-based and tree-based approaches

Automated analysis of fake news, containing intentionally distorted facts, is a major focus for the scientific community [4, 6, 7]. Drawing inspiration from the expressive *graph-based* or *tree-based* structures, a substantial body of research incorporates methods based on trees propagation and graph structures to encode the news dissemination patterns and user responses [11, 22]. In [23], the authors propose two graph neural network-based algorithms – GLO-PGNN and ENS-PGNN – designed for rumor detection in social networks. Both algorithms operate in two stages: first, they learn representations for each node in the propagation graph; second, they classify rumors using the representations obtained in the initial stage. Experimental results on the public Tweet dataset show that these approaches significantly outperform several recent state-of-the-art baselines in both rumor detection and early detection tasks. In [13], instead, it is addressed the problem of fake news detection based on news content using knowledge graphs. The authors leverage knowledge graph embeddings to compute semantic similarities, which helps manage incomplete and imprecise knowledge graphs. They use a basic knowledge graph embedding model, namely TransE, to evaluate the effectiveness of these methods in content-based fake news detection. Again, the authors in [3] introduce KG-MFEND, a novel framework utilizing knowledge graphs for multi-domain fake news detection. This model enhances performance by improving BERT and incorporating external knowledge to mitigate domain differences at the word level. The authors construct a new knowledge graph that includes multi-domain knowledge and injects entity triples to create a sentence tree, enriching the news background knowledge. Results indicate that KG-MFEND shows strong generalization capabilities across single, mixed, and multiple domains, outperforming current state-of-the-art methods in multi-domain fake news detection. In [12], it is described a graph-based approach for identifying organized astroturfing campaigns behind fake news articles. The method combines established techniques from graph mining and representation learning, transforming posts into informative diffusion trees. These diffusion trees serve as input for training a graph attention network (GAT) classifier. Kumar et al. [8] introduce the BCTree-LSTM model, which transforms the propagation tree into a binarized constituency tree structure to enhance rumor and stance detection. In contrast, Ma et al. [10] leverage tree kernels to measure the similarity between different propagation tree structures for identifying rumors on Twitter. Additionally, the authors in [9] propose the recurrent neural networks (RvNN) model, which captures discriminative features by utilizing the non-sequential propagation structure of rumors. This approach involves representing news propagation through a propagation tree and then learning its propagation process using RvNN.

Our contribution. The widespread and harmful nature of fake news on the Internet has become a significant concern, driving the development of automatic fake news detection systems. This is essentially why most graph-based and tree-based research focuses on fake news detection. However, our approach is different. Rather than considering the fake news detection facets, we leverage graphs in order to develop a **conceptual model** for the analysis of propagation patterns characterized by two distinct dimensions: *(i)* the origin of the fake news (the **'creator/spreader' dimension**); *(ii)* the topic of the fake news (the **'news content' dimension**). As a result, our perspective is strongly focused on the concepts of *characterization* and *profiling*.

3 Data Collection

The majority of the studies in the literature have been validated using real online social networks. The key components of these networks, such as the number of nodes and their interrelationships, are represented in several datasets available on the web and used as benchmark [15]. There are also some popular datasets of fake news. Some of them are: **FakeNewsNet** [18], **LIAR** [21], **PHEME** [25], and **Buzzface** [17].

When creating a dataset for fake news research, identifying whether the news is fake often relies on the evaluations of established **fact-checking sites**. These sites offer an original truth rating for each piece of news based on thorough investigations conducted by experts. Given that real-time news typically consist of a blend of information, a simple binary classification may not adequately capture the complexity of the issue. Fact-checking websites enable the verification of the authenticity of various types of online news by informing users whether the information is true, false, or somewhere in between. Some popular fact-checking sites are **PolitiFact** (`https://www.politifact.com/`), **Snopes** (`https://www.snopes.com/`), and **Suggest** also called **GossipCop** (`https://www.suggest.com/`).

3.1 Scraping fact-checking websites

For the experimentation described below, we collected data directly from the PolitiFact fact-checking site, taking advantage of the **web scraping technique**. Web scraping is a valuable process that automates extracting information from the World Wide Web. Han et al. [5] show that this method could be proven to be cost-efficient as opposed to manually filtering out large sums of data. We opted for such a solution in order to have the most up-to-date information available about topics of public interest. To scrape data, we used the following strategy:

I Selection of the website to be scraped (i.e., PolitiFact in our study);

II Identification of the news items to be scraped. Specifically, this includes selecting a statement to classify, source (creator/spreader), date, and the classification of the statement. The classification categories – **fakeness levels** – are: *True*, *Mostly-true*, *Half-true*, *Mostly-false*, *False*, and *Pants on fire*;

III At this point, the scraping starts. In order to do this, we first need to become familiar with the structure of the articles. We leverage the browser developer tools to pinpoint the `HTML` tags surrounding the data we aim to extract. Specifically, we identify the tags encompassing the information for extraction. Within the `<a>` tag there is the text (e.g., *'Donald Trump'*) we want to scrape and store into our dataset for spreader. It is worth pointing out that the target or classification of the statement is within the image (i.e., PolitiFact state indicator in the web page). More specifically, it is within the `<alt>` attribute of the image. This will tell us if the statement is *True*, *Mostly-true*, *Half-true*, *Mostly-false*, *False*, and *Pants on fire*.

By means of the `BeautifulSoup` Python library, we iterate the previous process to scrape all the needed information.

3.2 Creating the dataset

In order to prepare data for bipartite graph processing, we have implemented a Python script. It scrapes PolitiFact based on two user-provided inputs: *(i)* **Number of pages** (N) (this sets the number of pages of results to return) and *(ii)* **News spreader** (this can be a person, a social media platform, an organization, or a committee). The script then outputs the N most recent fact checks on PolitiFact related to statements from the chosen news spreader. It is worth pointing out that the script can also search for specific terms. By entering a keyword as input it returns the N most recent relevant pages scraped from PolitiFact. The results are stored in comma-separated values (`csv`) files specially crafted to apply the bipartite graph approach (more on this later). Each record consists of categorical features that provide context data. These are: *Author* (the 'fact-checker', that is, the expert who verified the news and labeled it), *Statement* (the precise text of the statement being fact-checked), *Spreader* (the source of the news (speaker), which can be a person, a social media post, an organization), *Date* (date of statement revision in Month-Day-Year format), *Target* (the label assigned by the reviewer), and *Keyword* (keyword by which the statement was searched). In order to collect data for our experimentation, we run the script by fixing 6 keywords (*'Ukraine'*, *'Russia'*, *'Zelensky'*, *'Putin'*, *'Trump'*, and *'Biden'*) and $N = 25$ pages. Given that each PolitiFact page contains at most 5 statements, the generated dataset encompasses 652 statements. It is worth pointing out that the scraping process also captures records with additional labels – *No Flip*, *Half Flip*, and *Full Flop* – beyond those previously mentioned. We decided to remove data with these labels because we only consider the labels related to news fakeness that we mentioned earlier.

4 Method

The analysis of fake news propagation proposed in this research study is conducted using a bipartite graph, which represents the relationships between speakers and keywords associated with non-true news.

4.1 Bipartite graphs

A graph is defined as a pair $\mathcal{G} = \{\mathcal{N}, \mathcal{E}\}$, where $\mathcal{N}$ is the set of nodes (or vertices) and $\mathcal{E}$ is the set of edges (or arcs) between the nodes. A graph is said to be bipartite if its nodes can be partitioned into two disjoint sets such that no edges exist between nodes in the same set [2]. Formally, a bipartite graph is defined as a triple $\mathcal{G} = \{\mathcal{U}, \mathcal{V}, \mathcal{E}\}$, where $\mathcal{U}$ and $\mathcal{V}$ denote the two disjoint sets of nodes and $\mathcal{E} \subseteq \mathcal{U} \times \mathcal{V}$ is the set of edges. The sets $\mathcal{U}$ and $\mathcal{V}$ can represent different categories, classes, or types of entities within the graph.

Bipartite graphs can be weighted or unweighted. In an unweighted bipartite graph, edges denote the presence of connections between nodes from the two different partitions. Conversely, in a weighted bipartite graph, each edge is assigned a weight that represents the strength or importance of the relationship. The connections are described by a $|\mathcal{U}| \times |\mathcal{V}|$ matrix B, known as biadjacency matrix, with $|\mathcal{U}|$ and $|\mathcal{V}|$ denoting the size of the two disjoint sets of nodes. For an unweighted bipartite graph, each entry B_{ij} is binary: $B_{ij} = 1$ if there exists an edge between node $i \in \mathcal{U}$ and node $j \in \mathcal{V}$, and $B_{ij} = 0$ otherwise. For a weighted bipartite graph, B_{ij} represents the weight of the edge connecting nodes $i \in \mathcal{U}$ and $j \in \mathcal{V}$, with $B_{ij} = 0$ if no edge exists between them.

In the analysis of bipartite graphs, the centrality of a node can be determined by examining its degree, which is quantified by either the number of edges connected to that node or the total weight of these edges. In particular, in an unweighted bipartite graph, the degree of node i is the number of edges incident to i. Nodes in an unweighted bipartite graph can achieve a maximum degree equal to the number of nodes in the opposite set. In a weighted bipartite graph, the degree of a node can be distinguished into two types: weighted and unweighted. The unweighted degree of node i indicates the number of edges connected to i, similar to the standard degree in unweighted bipartite graphs. The weighted degree of node i is given by the sum of the weights of the edges connected to i.

In the context of bipartite graphs, projection is a method used to derive two different graphs, each composed exclusively of nodes from one of the two disjoint sets, thus allowing the analysis of the relationships among a particular set of nodes. The projection on $\mathcal{U}$ is a graph that includes only the nodes in $\mathcal{U}$, where two nodes are connected if in the bipartite graph they share at least one common neighboring node from $\mathcal{V}$. Similarly, the projection on $\mathcal{V}$ is a graph containing the nodes of $\mathcal{V}$ connected if in the bipartite graph they share at least one common neighbor from $\mathcal{U}$. The projection graphs are described by adjacency matrices $A_{\mathcal{U}}$ and $A_{\mathcal{V}}$, which are derived from biadjacency matrix B as $A_{\mathcal{U}} = BB^\top$ and $A_{\mathcal{V}} = B^\top B$. The single entry of these matrices represents the measure of association between two nodes of the same partition, determined by their shared neighbors in the original bipartite graph.

4.2 Fake news analysis with bipartite graphs

The method proposed in this research study for the analysis of fake news propagation introduces the use of a weighted bipartite graph, constructed with data collected from PolitiFact. In this graph, the two disjoint sets of nodes represent the spreaders and the keywords, with an edge connecting a spreader and a keyword if the spreader has published at least one non-true news related to that keyword. The weight of the edge depends on the number of non-true news articles/posts published by that spreader about that keyword, weighted by the fakeness level of each news. Table 1 shows the correspondence between the fakeness level and the numerical value that is used to compute the total weight of an edge. In particular, the weight of the edge connecting spreader s with keyword k is given by

$$B_{sk} = \sum_{i=0}^{5} \frac{i}{5} N_{sk}^{i}, \tag{1}$$

where $i/5$ denotes the numerical value assigned to fakeness level i as per Table 1, N_{sk}^{i} is the number of news published by spreader s about keyword k with fakeness level i, and i ranging from 0 to 5 corresponds to the different fakeness levels from *True* to *Pants on Fire*. By taking into account both the fakeness level and the number of news articles/posts, this approach provides a representative measure of the fakeness of all news from a spreader on a given topic.

Fakeness level	**Numerical value**
True	0
Mostly-true	0.2
Half-true	0.4
Mostly-false	0.6
False	0.8
Pants on fire	1

Table 1: Correspondence between the fakeness level of news and the numerical value.

Degree centrality analysis in the bipartite graph allows the identification of central spreaders and keywords in fake news propagation. In particular, the unweighted degree of a spreader provides a measure of the diversity of topics it covers, indicating the number of keywords it is associated with. This metric highlights spreaders that participate across a wide range of topics, thus resulting in possible propagators of fake news. On the other hand, the weighted degree of a spreader represents its influence across different keywords based on the cumulative impact of its news articles/posts, weighted by their assigned fakeness levels. This metric measures the perceived influence of the spreader in propagating fake news across different topics. As regards the keywords, the unweighted degree measures their popularity or attractiveness based on the number of spreaders that have published related news. Therefore, a high unweighted degree indicates that multiple spreaders frequently reference the keyword. The weighted degree of a keyword, instead, reflects its susceptibility to disseminating fake news. Keywords with a high weighted degree not only capture interest but also contribute to the spread of misinformation.

The projection graphs from the bipartite graph allow for the analysis of interaction patterns among nodes of the same category in the fake news propagation. In the projection on the spreaders, two nodes are connected if they have published non-true news articles/posts about at least one common keyword, thus representing how spreaders are connected based on shared topics. The weight of an edge between two spreaders reflects their propensity to publish on the same topics with high fakeness levels, and a high weight indicates a stronger correlation in their spread of misinformation. In the projection on the keywords, two nodes are connected if they have been addressed by at least one common spreader. The weight of an edge between two keywords measures the intensity of their association or similarity. A high weight denotes that the corresponding keywords are frequently discussed by the same spreaders in non-true articles/posts, highlighting prevalent topics in the spread of fake news.

5 Results

In order to evaluate the effectiveness of the bipartite graph modeling for the analysis of fake news propagation, we used the dataset including articles and posts related to the keywords *'Biden'*, *'Putin'*, *'Russia'*, *'Trump'*, *'Ukraine'*, and *'Zelensky'*, as described in Section 3. Following the method proposed in the previous section, a bipartite graph is constructed as shown in Fig. 1, where the red nodes represent the spreaders and the light blue nodes are the keywords. For the sake of readability, the labels of the spreaders are omitted, while those of the keywords are reported in the figure. The edge thickness is proportional to the edge weight, denoting the intensity of association between spreaders and keywords. The graph consists of 96 nodes (90 spreaders and 6 keywords) and 187 edges with an average weight of 2.68. A visual analysis of the bipartite graph shows that a small fraction of spreaders exhibit a high frequency of publications associated with high fakeness levels across all the selected keywords. However, a more detailed analysis of these aspects can be obtained by examining the degree centrality and projection graphs.

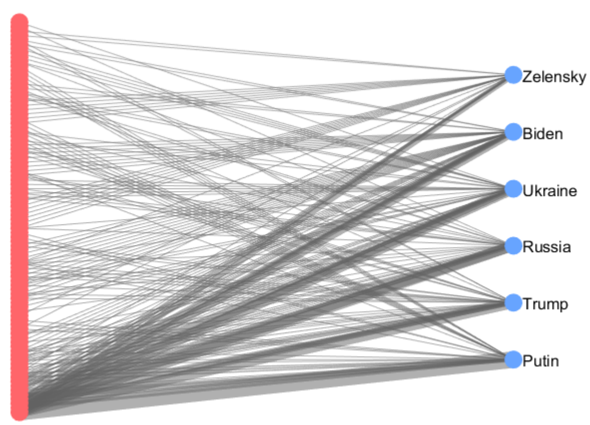

Fig. 1: Bipartite graph modeling fake news propagation, where the red and light blue nodes represent the spreaders and the keywords, respectively.

Figure 2 reports the unweighted and weighted degrees of the 15 spreaders having the highest respective degree, where the spreaders are sorted according to the degree values. The average unweighted degree of all spreaders is 2.08, indicating that, on average, they have discussed 2 of the keywords considered. The average weighted degree is 5.57, which reflects the overall intensity and frequency of fake news dissemination among the spreaders. The unweighted degree plot shows that 5 spreaders have a maximum value of 6, corresponding to the number of keywords considered for this analysis. However, this does not measure their propensity to publish fake news but only the variety of keywords used in their articles/posts. The weighted degree, which accounts for the number and fakeness level of published news, can provide more insights into how spreaders influence the dissemination of fake news. Social media platforms (*'Facebook posts'* and *'Instagram posts'*) exhibit high weighted degrees, thus reflecting extensive fake news publications. This is expected given their large user base and tendency to share inaccurate information. Moreover, 13 of the 15 spreaders appear in both barplots (unweighted and weighted degrees), indicating that these spreaders frequently cover the topics considered and are significant in fake news propagation. *'Robert F. Kennedy Jr.'* is included in the weighted degree barplot but not in the unweighted one since he has frequently published fake news articles/posts related only to the keywords *'Biden'* and *'Trump'*, thus resulting in an unweighted degree of 2 and large weighted degree. Therefore, this degree analysis allows us to preliminary classify spreaders based on their propensity to publish fake news articles/posts.

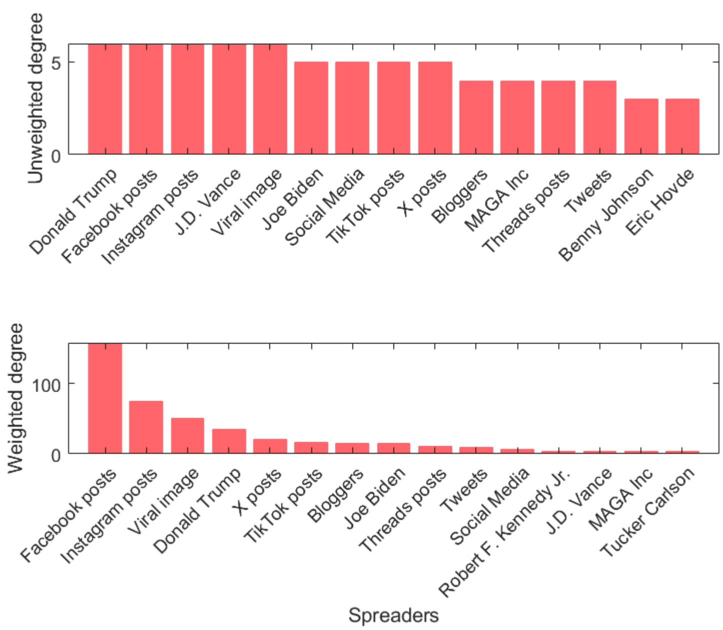

Fig. 2: Unweighted and weighted degrees of the 15 spreaders with the highest respective degrees.

In order to investigate the role of the spreaders with respect to the others, we analyze the projection on the spreaders shown in Fig. 3, where the edge thickness again corresponds to the edge weight. The projection graph exhibits a form of clustering among the spreaders, thus indicating how spreaders are grouped according to the keywords they have discussed. Moreover, it is evident that 5 nodes have a central role in connecting different groups within the graph. These nodes correspond to the spreaders with the highest unweighted degree (*'Donald Trump'*, *'Facebook posts'*, *'Instagram posts'*, *'J.D. Vance'*, and *'Viral Imagine'*) since they share at least one keyword with every other spreader. Therefore, for the definition of projection, each of these central nodes has 89 incident edges, i.e., the maximum possible given the 90 spreaders. This suggests that heterogeneity in terms of discussed keywords leads to more connections in the projection graph on the spreaders, reflecting a tendency to propagate more fake news.

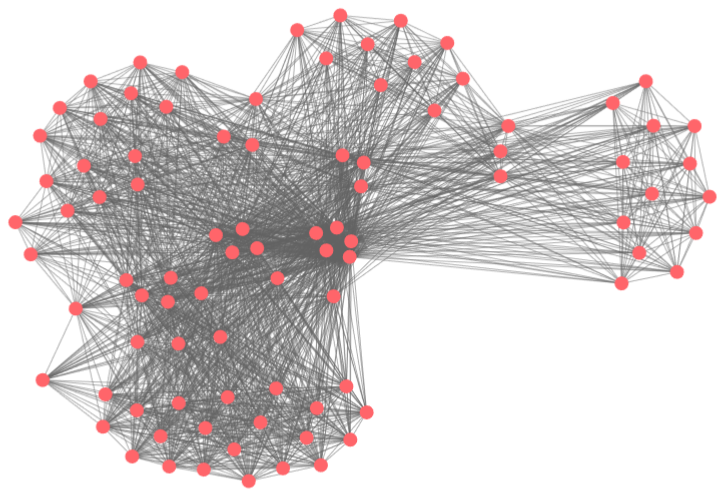

Fig. 3: Projection graph on the spreaders.

A similar analysis can also be performed for the keywords. Figure 4 shows the degrees of the keywords, sorted according to their values. The average unweighted and weighted degrees are 31.17 and 83.57, respectively, thus indicating that the selected keywords are frequently discussed and highly susceptible to fakeness. Although *'Biden'* is the most mentioned keyword, as evident by its highest unweighted degree, topics related to the Ukraine-Russia war seem more affected by fake news. Indeed, the keywords *'Putin'*, *'Russia'*, and *'Ukraine'* exhibit the highest weighted degrees. The keyword *'Zelensky'* has the lowest weighted degree, not due to a lower susceptibility to misinformation, but because it is the least mentioned among the keywords considered, as shown by its unweighted degree.

The projection on the keywords illustrated in Fig. 5 results in a complete graph, i.e., each node is directly connected to every other node through an edge. This is because the keywords selected for the analysis are interrelated, and each pair of keywords has likely been discussed by at least one spreader. The large edge weights further indicate a strong correlation between the keywords, especially for those related to the Ukraine-Russia war. By combining the heterogeneity of

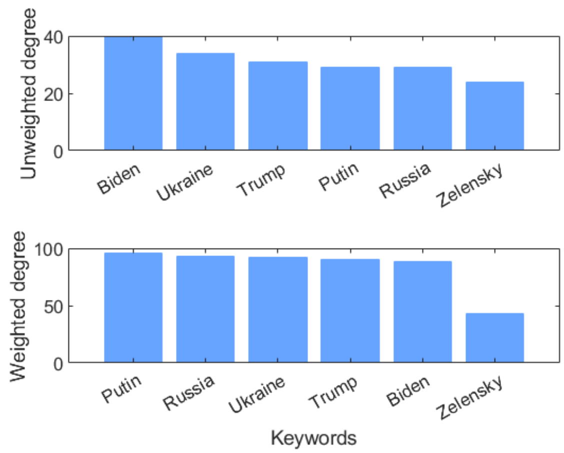

Fig. 4: Unweighted and weighted degrees of the keywords.

the spreaders and the similarity of the keywords, it is possible to determine the primary spreaders of fake news about one keyword, which may likely be the same for related keywords.

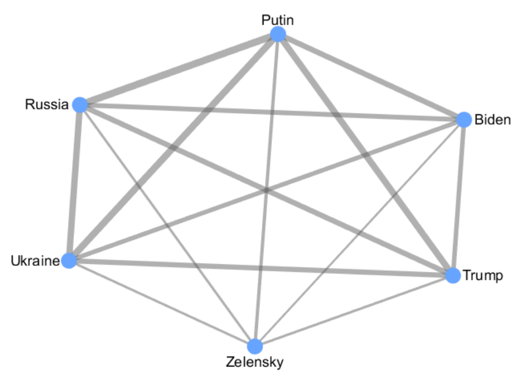

Fig. 5: Projection graph on the keywords.

6 Conclusion

In this study, we investigated fake news propagation modeled as a bipartite graph, which allows the identification of central spreaders and keywords involved in the spread of misinformation. In particular, we introduced the concepts of bipartite graph, degree centrality, and projection and demonstrated their practical relevance in the context under analysis. Moreover, using web scraping techniques, we collected a dataset from Politifact consisting of articles and posts related to widely-discussed topics, such as the Ukraine-Russia war. This dataset was used to validate our proposed approach based on bipartite graphs, demonstrating its effectiveness in analyzing fake news propagation.

Future directions could involve applying a similar analysis on datasets extracted by choosing different, possibly unrelated keywords. Moreover, a network analysis of the projection graphs introducing measures such as betweenness and closeness centralities might provide further insights into the roles of spreaders and keywords in fake news propagation. Clustering analysis based on similar interaction patterns or community detection could also be conducted to identify groups or communities among spreaders and keywords.

Acknowledgments. This work has been partially supported by project SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU.
The authors wish to thank Francesco Pio Briuolo for his help with the initial application of the method to the dataset.

References

1. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. Journal of Economic Perspectives **31**(2), 211–236 (2017)
2. Asratian, A.S., Denley, T.M., Häggkvist, R.: Bipartite graphs and their applications, vol. 131. Cambridge University Press (1998)
3. Fu, L., Peng, H., Liu, S.: KG-MFEND: an efficient knowledge graph-based model for multi-domain fake news detection. The Journal of Supercomputing **79**, 1–28 (2023)
4. Hakak, S., Khan, W.Z., Bhattacharya, S., Reddy, G.T., Choo, K.R.: Propagation of fake news on social media: Challenges and opportunities. In: Computational Data and Social Networks. pp. 345–353. Springer International Publishing (2020)
5. Han, S., Anderson, C.K.: Web scraping for hospitality research: Overview, opportunities, and implications. Cornell Hospitality Quarterly **62**(1), 89–104 (2021)
6. Kaur, S., Kumar, P., Kumaraguru, P.: Automating fake news detection system using multi-level voting model. Soft Computing **24**, 9049–9069 (2020)
7. Kozik, R., Pawlicka, A., Pawlicki, M., Choraś, M., Mazurczyk, W., Cabaj, K.: A meta-analysis of state-of-the-art automated fake news detection methods. IEEE Transactions on Computational Social Systems pp. 1–11 (2023)
8. Kumar, S., Carley, K.: Tree LSTMs with convolution units to predict stance and rumor veracity in social media conversations. In: Korhonen, A., Traum, D., Màrquez, L. (eds.) Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 5047–5058. Association for Computational Linguistics (2019)
9. Ma, J., Gao, W., Joty, S., Wong, K.: An attention-based rumor detection model with tree-structured recursive neural networks. ACM Transactions on Intelligent Systems and Technology **11**(4) (2020)
10. Ma, J., Gao, W., Wong, K.: Detect rumors in microblog posts using propagation structure via kernel learning. In: Barzilay, R., Kan, M. (eds.) Proc. of the 55th Annual Meeting of the Association for Computational Linguistics. pp. 708–717. Association for Computational Linguistics (2017)

11. 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)
12. Michail, D., Kanakaris, N., Varlamis, I.: Detection of fake news campaigns using graph convolutional networks. International Journal of Information Management Data Insights **2**(2), 100104 (2022)
13. Pan, J.Z., Pavlova, S., Li, C., Li, N., Li, Y., Liu, J.: Content based fake news detection using knowledge graphs. In: Vrandečić, D., Bontcheva, K., Suárez-Figueroa, M.C., Presutti, V., Celino, I., Sabou, M., Kaffee, L., Simperl, E. (eds.) The Semantic Web – ISWC 2018. pp. 669–683. Springer International Publishing, Cham (2018)
14. Phan, H.T., Nguyen, N.T., Hwang, D.: Fake news detection: A survey of graph neural network methods. Applied Soft Computing **139**, 110235 (2023)
15. Raponi, S., Khalifa, Z., Oligeri, G., Di Pietro, R.: Fake news propagation: A review of epidemic models, datasets, and insights. ACM Transactions on the Web **16**(3) (2022)
16. Ruffo, G., Semeraro, A., Giachanou, A., Rosso, P.: Studying fake news spreading, polarisation dynamics, and manipulation by bots: A tale of networks and language. Computer Science Review **47**, 100531 (2023)
17. Santia, G., Williams, J.: Buzzface: A news veracity dataset with Facebook user commentary and egos. Proc. of the International AAAI Conference on Web and Social Media **12**(1), 531–540 (2018)
18. 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)
19. Shu, K., Mahudeswaran, D., Wang, S., Liu, H.: Hierarchical propagation networks for fake news detection: investigation and exploitation. In: Proc. of the International Conference on Web and Social Media. pp. 1–10. The Association for the Advancement of Artificial Intelligence (2019)
20. Sivasankari, S., Vadivu, G.: Tracing the fake news propagation path using social network analysis. Soft Computing **26**, 12883–12891 (2022)
21. Wang, W.Y.: "liar, liar pants on fire": A new benchmark dataset for fake news detection. In: Barzilay, R., Kan, M. (eds.) Proc. of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp. 422–426. Association for Computational Linguistics (2017)
22. Wu, L., Liu, H.: Tracing fake-news footprints: Characterizing social media messages by how they propagate. In: Proc. of the 11th International Conference on Web Search and Data Mining. p. 637–645. ACM (2018)
23. Wu, Z., Pi, D., Chen, J., Xie, M., Cao, J.: Rumor detection based on propagation graph neural network with attention mechanism. Expert Systems with Applications **158**, 113595 (2020)
24. Zhang, X., Ghorbani, A.A.: An overview of online fake news: Characterization, detection, and discussion. Information Processing Management **57**(2), 102025 (2020)
25. Zubiaga, A., Liakata, M., Procter, R., Wong Sak Hoi, G., Tolmie, P.: Analysing how people orient to and spread rumours in social media by looking at conversational threads. PLOS ONE **11**(3), 1–29 (2016)