Implementing Deep Learning tools and/or techniques for detecting political misinformation on Twitter:

A Literature Review

Abstract

This review explores and defines key concepts, including misinformation, disinformation, and fake news, recognizing the profound impact of misinformation on the integrity of democratic processes. A significant outcome is the identification of Graph Neural Networks (GNNs) as a potent solution in the realm of fake news detection. The study highlights the unique suitability of GNNs in integrating findings from cognitive science and psychology into the detection process, a synthesis that remains largely unexplored.

Introduction

The quality and authenticity of information accessible to the public significantly influence the contemporary democratic landscape. As Lewandowsky et al. (2017) highlight, a well-informed populace is fundamental to the functioning of a democracy. Kuklinski et al. (2000) further assert that citizens must have access to factual information to evaluate public policy and inform their preferences effectively.

However, the information landscape has drastically deteriorated, as evidenced by the World Health Organization's (WHO) 2020 declaration of a global "infodemic" (Van Der Linden, 2022). The proliferation of misinformation, mainly through social media platforms, has emerged as a critical issue. As indicated by Altay et al. (2023), experts advocate for various measures against misinformation, including platform and algorithmic design changes, content moderation, de-platforming of misinformative actors, and crowdsourced detection. The severity of misinformation is such that, according to Lewandowsky et al. (2020), people are willing to support politicians despite acknowledging the falsehood of their statements. This paradoxical scenario underscores the potential role of technology, not only as a contributor to the problem but also as a part of the solution, as Lewandowsky et al. (2017) suggested.

This Literature Review explores the key concepts, definitions, and pertinent background in automatic fake news detection. It highlights promising research developments in Graph Neural Networks (GNNs) (Scarselli et al., 2009) and underscores the importance of incorporating user-specific features in fake news analysis. The review is structured around the research question:

> *"What specific attributes of the social context enhance the efficacy of Graph Neural Network (GNN)-based models in detecting fake news on Twitter?"*

While closely related areas such as Rumour Classification, Truth Classification, Clickbait Detection, and Spammer and Bot Detection are acknowledged, they fall outside the scope of this review. Potential limitations, such as the unavailability of predefined benchmark datasets, are outside the purview of this review.

This paper follows the structure of IMRAD (Introduction, Methods, Results, and Discussion). The *Research Strategy* chapter elucidates the methodology employed in conducting the literature review. Subsequently, the *Review of Literature* chapter presents the field's current state, particularly about the posed research question. Finally, the *Discussion and conclusion* chapter synthesize the findings, discuss their implications, and conclude with insights on why the research question must be addressed.

Research Strategy

In selecting pertinent literature for this review, consideration was given to the prestige of the publishing journal, utilizing metrics like the journal's impact factor where necessary. Additionally, the frequency of citations and the thematic relevance to the current review were critical factors in literature selection.

The literature search spanned from December 1, 2023, to January 6, 2024, and was confined to specific databases: IEEE Xplore Digital Library, ACM Digital Library, arXiv, ScienceDirect, Google Scholar, and Consensus. The inclusion of arXiv was deemed essential, as it is a common platform for disseminating research in computer science. However, it is essential to note that works sourced from preprint servers were approached with heightened scrutiny and caution.

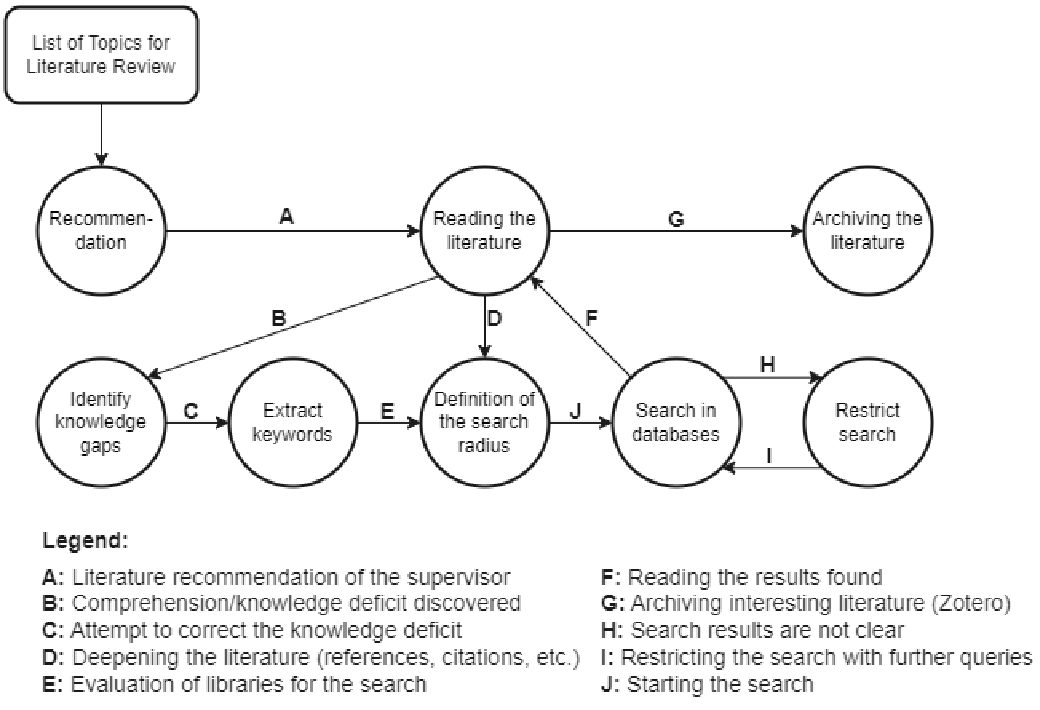
List of Topics for Literature Review
Recommen-dation
Reading the literature
Archiving the literature
Identify knowledge gaps
Extract keywords
Definition of the search radius
Search in databases
Restrict search
A
B
C
D
E
F
G
H
I
J
Legend:
A: Literature recommendation of the supervisor
B: Comprehension/knowledge deficit discovered
C: Attempt to correct the knowledge deficit
D: Deepening the literature (references, citations, etc.)
E: Evaluation of libraries for the search
F: Reading the results found
G: Archiving interesting literature (Zotero)
H: Search results are not clear
I: Restricting the search with further queries
J: Starting the search

Figure 1: Framework used for the literature search based on recommendations from the University of Zurich.

The study employs a structured thematic analysis based on the framework in Figure 1. Detailed documentation of search terms and results is excluded, acknowledging the nonsystematic nature of the review and the non-reproducibility of searches in platforms like Google Scholar.

The initial familiarization with the topic was facilitated through existing literature reviews due to the absence of prior literature mapping.

Review of Literature

The ensuing chapter delineates the outcomes of the conducted literature review.

Definition of misinformation

There is an acknowledged lack of standardised definitions in the scholarly discourse, particularly across disciplines (Altay et al., 2023; Sindermann et al., 2020). Figure 2 presents an Euler diagram depicting the identified terms and their interrelationships.

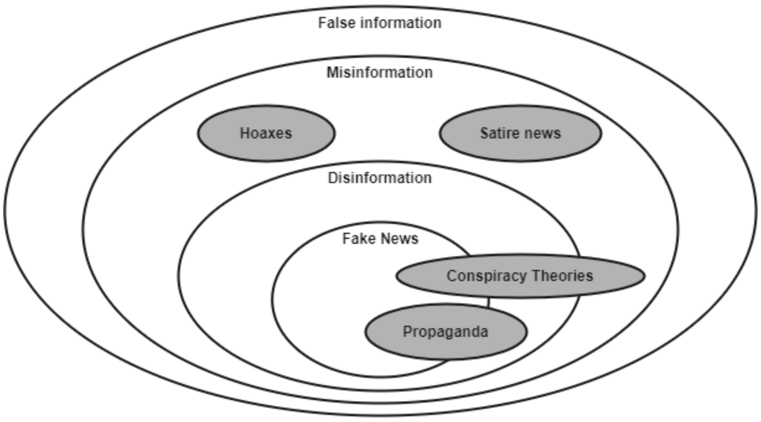

Figure 2: Euler diagram presented herein illustrates the interrelationships among various terms requiring definition.

Lewandowsky et al. (2020) provide a nuanced classification of terms such as false information, misinformation, disinformation, and fake news. Based on their definition, false information is characterized as any data that is factually incorrect, which might not necessarily be disseminated or verifiable. Misinformation, a broader term, includes any false information that is spread, irrespective of an underlying intent to mislead. A more specific category within misinformation is disinformation, defined as intentionally false information propagated explicitly to deceive. In the context of media, Lewandowsky et al. (2020) describe fake news as false information that is often sensational and mimics the format of traditional news media.

Contrastingly, Shu et al. (2017) adopt a narrower perspective on fake news, defining it as a news article that is both intentionally and verifiably false. The focus on verifiability is paramount, especially in the realm of automatic fake news detection, as it underscores the importance of maintaining accuracy, reliability, objectivity, and ethical integrity in the research field.

Interdisciplinary nature

Misinformation constitutes a focal point of research across a range of disciplines, including cognitive science, sociology, media and communication studies, psychology, and computer science, as highlighted by Altay et al. (2023). The interdisciplinary nature of this research is critical, especially in efforts to mitigate the proliferation of fake news and to confront the fundamental issues it uncovers, a stance actively advocated by (Lazer et al., 2018). In addressing these challenges, Lewandowsky et al. (2017) emphasize the necessity of integrating technological solutions with psychological insights. This approach, which they term "technocognition," advocates for a synergistic, interdisciplinary methodology to effectively counter the spread and impact of misinformation.

Fake news detection

Shu et al. (2017) formally define the detection of fake news using a prediction function, denoted as F, which is expressed as $F\colon \varepsilon \rightarrow \{0,1\}$. In this definition ε represents a set of tuples, $\varepsilon = \{e_{it}\}$, that encapsulate the dynamics of how news disseminates over time across a network of n users, denoted as $U = \{u_1, u_2, \dots, u_n\}$, and their associated posts $P = \{p_1, p_2, \dots, p_n\}$ relating to a specific news article a on social media platforms. Each tuple in this set, specified as $e_{it} = \{u_i, p_i, t\}$, corresponds to an instance where a user u_i shares the news article a through post p_i at a specific time t. The prediction function F is tasked with determining the veracity of the news article a, assigning a value of 1 if the article is deemed fake news and 0 otherwise. The notation can be summarized as:

$$F(a) = \begin{cases} 1, & if\ a\ is\ a\ fake\ news \\ 0, & otherwise. \end{cases}$$

For learning, diverse features can be strategically extracted from social media to enhance model efficacy. Figure 3, informed by the work of Shu et al. (2017), presents a mind map illustrating a range of features previously employed in this domain. The incorporation of several feature types within a model constitutes multi-aspect learning.

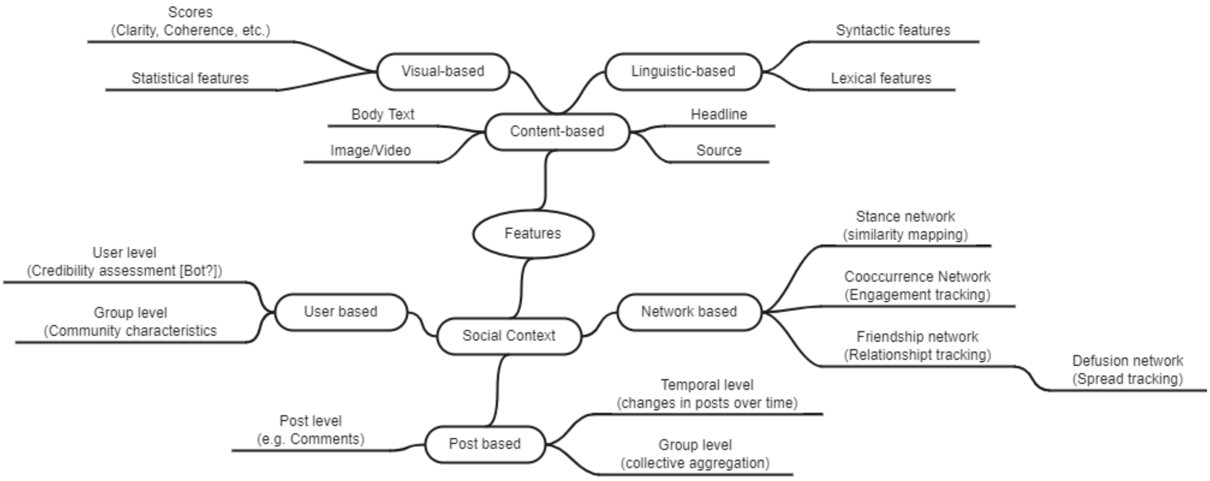

Figure 3: Collection of features for fake news detection based on Shu et al. (2017)

The categorization of fake news detection models, as shown in Figure 4, is closely related to, but not entirely dependent on, the features they utilize.

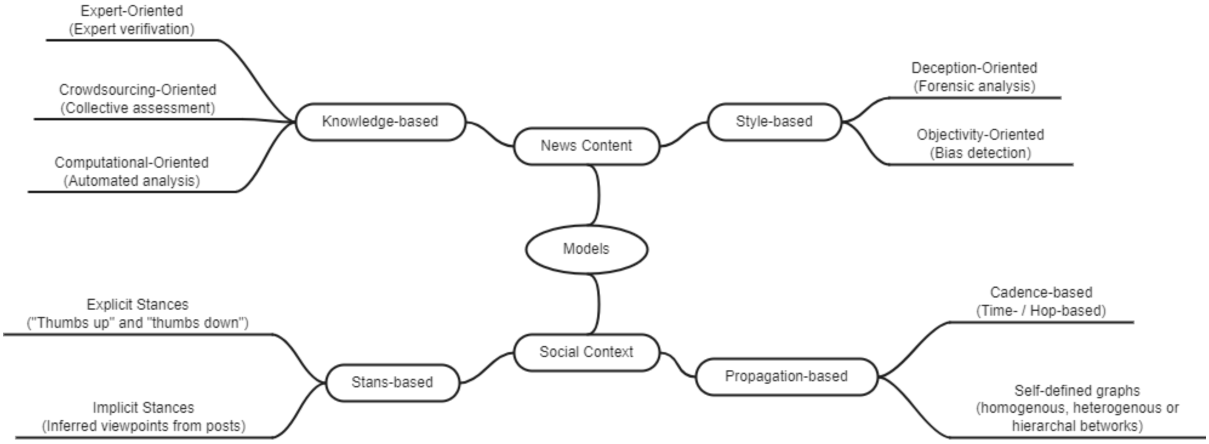

Figure 4: Mindmap illustrating various model types for fake news detection (Shu et al., 2017).

In their systematic literature review, Phan et al. (2023) offer a comprehensive synthesis of existing research in the field, including an in-depth examination of feature detection methodologies and model frameworks.

The rise of graph neural networks (GNN)

A graph (Gross et al., 2014), denoted as $G = (V, E)$, is composed of two distinct sets: $V = \{v_1, v_2, \dots, v_n\}$, representing vertices (nodes), and E, a subset of the set of unordered pairs of these vertices, symbolized as $E \subseteq \{\{u, v\} \mid u, v \in V\}$, representing edges. The vertices V are the fundamental units of the graph, while the edges E are the connections or links between pairs of vertices. Each edge in this graph is associated with either one (in the case of a loop) or two distinct vertices, referred to as its endpoints.

In their seminal work, Scarselli et al. (2009) introduced the concept of Graph Neural Networks (GNNs), pioneering a method for parameter estimation within graph structures and their constituent nodes. These networks are designed to learn a function, $\tau(G, n)$ which predicts specific attributes or features of a given node n in the vertex set V, or alternatively, $\tau(G)$ to infer global properties of the entire graph G. To enhance learning accuracy, additional data is typically integrated into the graph as a feature matrix $\hat{X} = [\hat{t}_0, \hat{t}_1, \dots, \hat{t}_{\mathrm{N}}]$. This matrix is composed of individual node feature vectors, where N equals the number of vertices, denoted as |V|. Consequently, the graph G is represented more comprehensively as $G = (V, E, \hat{X})$, encapsulating vertices V, edges E, and the feature matrix $\hat{X}$.

The evolution of Graph Transformer Networks (GTNs) represents a significant advancement in graph-based data analysis, as evidenced by the foundational contributions of Kipf and Welling (2017), Hamilton et al. (2018), and Yun et al. (2019), among others. These GTNs distinguish themselves by their innate capability to directly utilise the relational information embedded within graph data. This stands in contrast to conventional neural networks, which necessitate extensive pre-processing to adapt to the structural intricacies of graph data.

In a notable application of GTNs, Soga et al. (2024) demonstrated the efficacy of these networks in encoding stance data derived from interactions between posts and users.

These results illustrate the potential of GTNs in extracting insights from intricate network-based data (stance- and propagation-based), marking a significant stride in the domain of graph-based learning algorithms.

Role of the social media user

Building upon the promising results of stance-based approaches as demonstrated by Soga et al. (2024), the field of cognitive sciences contributes many factors, many of which have undergone experimental validation, that play a significant role in influencing the acceptance and propagation of fake news (Altay et al., 2023; Arin et al., 2023; Ecker et al., 2022; Kuklinski et al., 2000; Lewandowsky et al., 2017; Shu et al., 2017; Sindermann et al., 2020; Van Der Linden, 2022).

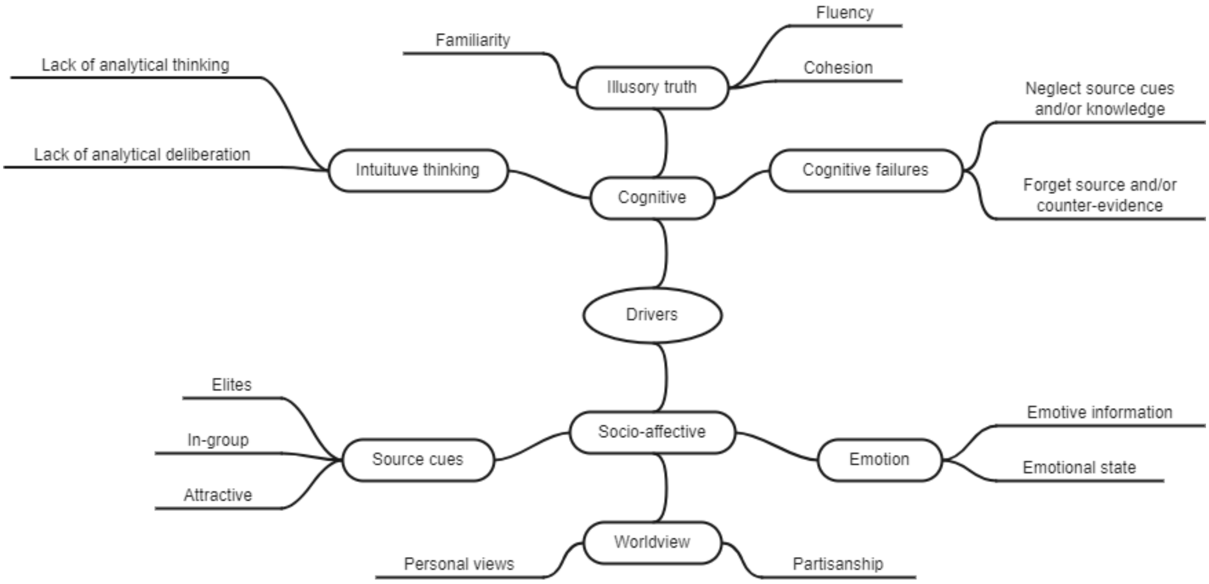

Figure 5: Cognitive and socio-affective determinants that contribute to susceptibility to fake news (Ecker et al., 2022).

Sindermann et al. (2020) and Altay et al. (2023) agree that cognitive biases and trust factors significantly influence susceptibility to fake news. Individuals often misjudge the accuracy of news based on their political attitudes, overrating congruent news and underrating incongruent news. Additionally, they identify specific factors that increase susceptibility to fake news, including biases, partisanship, identity, and trust in media and political systems. Analytical thinking, however, is found to reduce this susceptibility. Arin et al. (2023) contribute to this discourse by highlighting that accidental sharing of fake news is more common than deliberate sharing, with a notable decrease in accidental sharing among older demographics. They also find that older, male, higher-income, and politically left-leaning individuals are more proficient in detecting fake news.

Gaps in the literature

Despite identifying several gaps in the existing literature (Altay et al., 2023; Phan et al., 2023), this study has uncovered a significant gap in using Graph Neural Networks (GNN) for models that include cognitive or socio-affective factors, with Soga et al. (2024) being a rare exception but limited to stance-based features. Additionally, user-centric research in this field is limited, with notable work by Mu and Aletras (2020) focusing on predicting user behaviour related to misinformation spread based on linguistic features.

Discussion and conclusion

This literature review has delved into the intricate relationship between cognitive science, psychology, and the phenomenon of fake news. The studies from these fields illuminate a crucial interplay between cognitive and socio-affective factors influencing individuals' likelihood of believing and sharing fake news. This relationship underscores a vital insight: these same cognitive and socio-affective factors could be leveraged by users to recognize fake news. Given this connection, it becomes evident that factors beyond the standard parameters, extending from the individual to the group level, should be integrated into graph-based fake news detection models. Despite the promising potential several alternative applications and contributions emerge:

1. Predicting Susceptibility to Fake News: The identified features could be instrumental in predicting when an individual is susceptible to certain already recognized fake news.
2. Focus on Influential Individuals: The research could support in better classifying and predicting the role of individuals with a large reach. By prioritizing these individuals, more targeted and manual fake news detection efforts could be deployed.
3. Interdisciplinary Contributions: Answering the research question could be invaluable to other fields, even if the results would not contribute to a better fake news detection performance.
4. Including more user-based features could yield benefits in detecting other forms of misinformation than fake news.

In conclusion, while the primary goal of integrating cognitive and socio-affective factors into fake news detection models remains pivotal, the broader implications of this research extend beyond this objective. Whether in predicting individual susceptibility, identifying key influencers, or contributing to interdisciplinary knowledge, the insights garnered from this review offer a multifaceted perspective on addressing the challenges of fake news in the digital age.

References

Altay, S., Berriche, M., Heuer, H., Farkas, J., Rathje, S., 2023. A survey of expert views on misinformation: Definitions, determinants, solutions, and future of the field. HKS Misinfo Review. https://doi.org/10.37016/mr-2020-119

Arin, K.P., Mazrekaj, D., Thum, M., 2023. Ability of detecting and willingness to share fake news. Sci Rep 13, 7298. https://doi.org/10.1038/s41598-023-34402-6

Ecker, U.K.H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L.K., Brashier, N., Kendeou, P., Vraga, E.K., Amazeen, M.A., 2022. The psychological drivers of misinformation belief and its resistance to correction. Nat Rev Psychol 1, 13–29. https://doi.org/10.1038/s44159-021-00006-y

Gross, J.L., Yellen, J., Zhang, P. (Eds.), 2014. Handbook of graph theory, Second edition. ed, Discrete mathematics and its applications. CRC Press, Taylor & Francis Group, Boca Raton.

Hamilton, W.L., Ying, R., Leskovec, J., 2018. Inductive Representation Learning on Large Graphs.

Kipf, T.N., Welling, M., 2017. Semi-Supervised Classification with Graph Convolutional Networks.

Kuklinski, J.H., Quirk, P.J., Jerit, J., Schwieder, D., Rich, R.F., 2000. Misinformation and the Currency of Democratic Citizenship. The Journal of Politics 62, 790–816. https://doi.org/10.1111/0022-3816.00033

Lazer, D.M.J., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S.A., Sunstein, C.R., Thorson, E.A., Watts, D.J., Zittrain, J.L., 2018. The science of fake news. Science 359, 1094–1096. https://doi.org/10.1126/science.aao2998

Lewandowsky, S., Cook, J., Lombardi, D., 2020. Debunking Handbook 2020. https://doi.org/10.17910/B7.1182

Lewandowsky, S., Ecker, U.K.H., Cook, J., 2017. Beyond misinformation: Understanding and coping with the "post-truth" era. Journal of Applied Research in Memory and Cognition 6, 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

Mu, Y., Aletras, N., 2020. Identifying Twitter users who repost unreliable news sources with linguistic information. PeerJ Computer Science 6, e325. https://doi.org/10.7717/peerj-cs.325

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

Scarselli, F., Gori, M., Ah Chung Tsoi, Hagenbuchner, M., Monfardini, G., 2009. The Graph Neural Network Model. IEEE Trans. Neural Netw. 20, 61–80. https://doi.org/10.1109/TNN.2008.2005605

Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H., 2017. Fake News Detection on Social Media: A Data Mining Perspective 19.

Sindermann, C., Cooper, A., Montag, C., 2020. A short review on susceptibility to falling for fake political news. Current Opinion in Psychology 36, 44–48. https://doi.org/10.1016/j.copsyc.2020.03.014

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

Van Der Linden, S., 2022. Misinformation: susceptibility, spread, and interventions to immunize the public. Nat Med 28, 460–467. https://doi.org/10.1038/s41591-022-01713-6

Yun, S., Jeong, M., Kim, R., Kang, J., Kim, H.J., 2019. Graph Transformer Networks.