Review article

Fake news detection: A survey of graph neural network methods

Huyen Trang Phan [a,b], Ngoc Thanh Nguyen [c,*], Dosam Hwang [a,*]

[a] *Department of Computer Engineering, Yeungnam University, Gyeongsan, South Korea*
[b] *Faculty of Information Technology, Nguyen Tat Thanh University, Ho Chi Minh, Vietnam*
[c] *Department of Applied Informatics, Wroclaw University of Science and Technology, Wroclaw, Poland*

ARTICLE INFO

ABSTRACT

The emergence of various social networks has generated vast volumes of data. Efficient methods for capturing, distinguishing, and filtering real and fake news are becoming increasingly important, especially after the outbreak of the COVID-19 pandemic. This study conducts a multiaspect and systematic review of the current state and challenges of graph neural networks (GNNs) for fake news detection systems and outlines a comprehensive approach to implementing fake news detection systems using GNNs. Furthermore, advanced GNN-based techniques for implementing pragmatic fake news detection systems are discussed from multiple perspectives. First, we introduce the background and overview related to fake news, fake news detection, and GNNs. Second, we provide a GNN taxonomy-based fake news detection taxonomy and review and highlight models in categories. Subsequently, we compare critical ideas, advantages, and disadvantages of the methods in categories. Next, we discuss the possible challenges of fake news detection and GNNs. Finally, we present several open issues in this area and discuss potential directions for future research. We believe that this review can be utilized by systems practitioners and newcomers in surmounting current impediments and navigating future situations by deploying a fake news detection system using GNNs.

Contents

* Corresponding authors.
  E-mail addresses: huyentrangtin@ynu.ac.kr, pthtrang@ntt.edu.vn (H.T. Phan), Ngoc-Thanh.Nguyen@pwr.edu.pl (N.T. Nguyen), dshwang@yu.ac.kr (D. Hwang).

1. Introduction

Recently, social networks have contributed to an explosion of information. Social networks have become the main communication channel for people worldwide. However, the veracity of news posted on social networks often cannot be determined. Thus, using social networks is a double-edged sword. Therefore, if the news received from social networks is real, it will be beneficial. Conversely, if this news is fake, it will have many harmful consequences, and the extent of damage when fake news is widely disseminated is incalculable.

Fake news is a source of absolutely inventive information to spread deceptive content or entirely misrepresent actual news articles [1]. Numerous examples of fake news exist. Hunt et al. [2] indicated that during the 2016 US presidential election, the activity of Clinton supporters was affected by the spread of traditional center and left-leaning news from top influencers, whereas the movement of Trump supporters was influenced by the dynamics of top fake news spreaders. Moreover, public opinion manipulation based on the spread of fake news related to the Brexit vote in the United Kingdom was reported. Most recently, the prevalence of fake news has been witnessed during the COVID-19 pandemic. These examples show that the spread of fake news on social networks has a significant effect on many fields. Timely detection and containment of fake news before widespread dissemination is an urgent task. Therefore, many methods have been implemented to detect and prevent the spread of fake news over the past decade, among which the graph neural network (GNN)-based approach is the most recent.

Based on previous studies' findings regarding the benefit of using GNNs for fake news detection, we summarize some main justifications for using GNNs as follows. Existing approaches for fake news detection focus almost exclusively on features related to the content, propagation, and social context separately in their models. GNNs promise to be a potentially unifying framework for combining content, propagation, and social context-based approaches [3]. Fake news spreaders can attack machine learning-based models because these models depend strongly on news text. Making detection models less dependent on the news text is necessary to avoid this issue. GNN-based models can achieve similar or higher performance than modern methods without textual information [4]. GNN-based approaches can provide flexibility in defining the information propagation pattern using parameterized random walks and iterative aggregators [5].

A graph neural network is a novel technique that focuses on using deep learning algorithms over graph structures [6]. Before their application in fake news detection systems, GNNs had been successfully applied in many machine learning and natural language processing-related tasks, such as object detection [7,8], sentiment analysis [9,10], and machine translation [11,12]. The rapid development of numerous GNNs has been achieved by improving convolutional neural networks, recurrent neural networks, and autoencoders through deep learning [13]. The rapid development of GNN-based methods for fake news detection systems on social networks can be attributed to the rapid growth of social networks in terms of the number of users, the amount of news posted, and user interactions. Consequently, social networks naturally become complex graph structures if they are applied independently, which is problematic for previous machine learning-based and deep learning-based fake news detection algorithms. The main reasons for this phenomenon are the dependence of the graph size on the number of nodes and the different numbers of node neighbors. Therefore, some important operations (convolutions) are difficult to calculate in the graph domain. Additionally, the primary assumption of previous machine learning and deep learning-based fake news detection algorithms is that news items are independent. This assumption cannot apply to graph data because nodes can connect to other nodes through various types of relationships, such as citations, interactions, and friendships. GNN-based fake news detection methods have been developed. Although some state-of-the-art results have been achieved (see Table 1), no complete GNN-based fake news detection and prevention system existed when we conducted this study. Faking news on social networks is still a major challenge that needs to be solved (the first justification).

Various survey papers of fake news detection have been published, such as [18–23]. We briefly summarize related work as follows: Vitaly Klyuev et al. [20] presented a survey of different fake news detection methods based on semantics using natural language processing (NLP) and text mining techniques. Additionally, the authors discussed automatic checking and bot detection on social networks. Meanwhile, Oshikawa et al. [21] introduced a survey for fake news detection, focusing only on reviewing NLP-based approaches. Collins et al. [18] presented various variants of fake news and reviewed recent trends in preventing the spread of fake news on social networks. Shu et al. [22] conducted a review on various types of disinformation, factor influences, and approaches that decrease the effects. Khan et al. [19] presented fake news variants, such as misinformation, rumors, clickbait, and disinformation. They provided a more detailed representation of some fake news variant detection methods without limiting NLP-based approaches. They also introduced types of available detection models, such as knowledge-based, fact-checking, and hybrid approaches. Moreover, the authors introduced governmental strategies to prevent fake news and its variants. Mahmud et al. [23] presented a comparative analysis by implementing several commonly used methods of machine learning and GNNs for fake news detection on social media and comparing their performance. No survey papers have attempted to provide a comprehensive and thorough overview of fake news detection using the most current technique, namely, the GNN-based approach (the second justification).

The above two justifications motivated us to conduct this survey. Although some similarities are unavoidable, our survey is different from the aforementioned works in that we focus on description, analysis, and discussion of the models of fake news detection using the most recent GNN-based techniques. We believe that this paper can provide an essential and basic reference for new researchers, newcomers, and systems practitioners in overcoming current barriers and forming future directions when improving the performance of fake news detection systems using GNNs. This paper makes the following four main contributions.

Table 1
A description of the improved performance of the traditional methods compared using GNN-based methods.

Method	Ref	Improved methods	Dataset	Least improved performance
GCAN	[14]	DTC, SVM-TS, mGRU, RFC, tCNN, CRNN, CSI, dEFEND	Twitter15, Twitter16	Accuracy: 18.7%, Accuracy: 19.9%
FANG	[5]	Feature SVM, CSI	Twitter	AUC: 6.07%
SAFER	[15]	HAN, dEFEND, SAFE, CNN, RoBERTa, Maj sharing baseline	FakeNewsNet, FakeHealth	F_1: 5.19%, F_1: 5.00%
Bi-GCN	[16]	DTC, SVM-RBF, SVM-TS, RvNN PPC_RNN+CNN	Weibo, Twitter15, Twitter16	Accuracy: 4.5%, Accuracy: 13.6%, Accuracy: 14.3%
AA-HGNN	[17]	SVM, LIWC, text-CNN, Label propagation, DeepWalk, LINE, GAT, GCN, HAN	PolitiFact BuzzFeed	Accuracy: 2.82% Accuracy:9.34%

- We provide the most comprehensive survey yet of fake news, including similar concepts, characteristics, types of related features, types of approaches, and benchmarks datasets. We redefine similar concepts regarding fake news based on their characteristics. This survey can serve as a practical guide for elucidating, improving, and proposing different fake news detection methods.
- We provide a brief review of existing types of GNN models. We also make necessary comparisons among types of models and summarize the corresponding algorithms.
- We introduce the details of GNN models for fake news detection systems, such as pipelines of models, benchmark datasets, and open source code. These details provide a background and guide experienced developers in proposing different GNNs for fake news prevention applications.
- We introduce and discuss open problems for fake news detection and prevention using GNN models. We provide a thorough analysis of each issue and propose future research directions regarding model depth and scalability trade-offs.

This section justified the problem and highlighted our motivations for conducting this survey. The remaining sections of the paper are ordered as follows. Section 2 introduces the background and provides an overview of fake news, fake news detection, and GNNs. Section 3 presents the survey methodology used to conduct the review. General information on the included papers is analyzed in Section 4. In Section 5, the selected papers are categorized and reviewed in detail. Subsequently, we discuss the comparisons, advantages, and disadvantages of the methods by category in Section 6. Next, the possible challenges of fake news and GNNs are briefly evaluated in Section 7. Finally, we identify several open issues in this area and discuss potential directions for future research in Section 8.

2. Background

2.1. Understanding fake news

What is fake news? News is understood as meta-information and can include the following [24]:

- Source: Publishers of news, such as authors, websites, and social networks.
- Headline: Description of the main topic of the news with a short text to attract readers' attention.
- Body content: Detailed description of the news, including highlights and publisher characteristics.
- Image/Video: Part of the body content that provides a visual illustration to simplify the news content.
- Links: Links to other news sources.

"Fake news" was named word of the year by the Macquarie Dictionary in 2016 [24]. Fake news has received considerable attention from researchers, with differing definitions from various view opinions. In [24], the authors defined fake news as *"a news article that is intentionally and verifiably false"*. Alcott and Gentzkow [2] provided a narrow definition of fake news as *"news articles that are intentionally and verifiably false, and could mislead readers"*. In another definition, the authors considered fake news as *"fabricated information that mimics news media content in form but not in organizational process or intent"* [25]. In [26], the authors considered fake news in various forms, such as false, misleading, or inventive news, including several characteristics and attributes of the disseminated information. In [27], the authors provided a broad definition of fake news as *"false news"* and a narrow definition of fake news as *"intentionally false news published by a news outlet"*. Similar definitions have been employed in previous fake news detection methods [3,4,28,29].

Characteristics of Fake news: Although various definitions exist, most fake news has the following common characteristics.

- Echo chamber effect: Echo chambers [30] can be broadly defined as environments focusing on the opinions of users who have the same political leaning or beliefs about a topic. These opinions are reinforced by repeated interactions with other users with similar tendencies and attitudes. Social credibility [31] and frequency heuristic [31] (i.e., the trend to search for information that conforms to preexisting reviews) may be the reason for the appearance of echo chambers on social networks [24,32–34]. When news does not contain enough information, *Social credibility* can be used to judge its truthfulness. However, many people still perceive it as credible and disseminate it, leading to popular acceptance of such news as credible. A *Frequency heuristic* forms when people frequently hear the news, leading to natural approval of the information, even if it is fake news.
- Intention to deceive [35]: This characteristic is identified based on the hypothesis that *"no one inadvertently produces inaccurate information in the style of news articles, and the fake news genre is created deliberately to deceive"* [25]. Deception is prompted by political/ideological or financial reasons [2,36–38]. However, fake news may also appear and is spread to amuse, to entertain, or, as proposed in [39], *"to provoke"*.
- Malicious account: Currently, news on social networks comes from both real people and unreal people. Although fake news is created and primarily spread by accounts that are not real people, several real people still spread fake news. Accounts created mainly to spread fake news are called malicious accounts [27]. Malicious accounts are divided into three main types: *social bots, trolls, and cyborg users* [24]. *Social bots* are social network accounts controlled by computer algorithms. A social bot is called a malicious account when it is designed primarily to spread harmful information and plays a large role in creating and spreading

fake news [40]. This malicious account can also automatically post news and interact with other social network users. *Trolls* are real people who disrupt online communities to provoke an emotional response from social media users [24]. Trolls aim to manipulate information to change the views of others [40] by kindling negative emotions among social network users. Consequently, users develop strong doubts and distrust them [24]; they will fall into a state of confusion, unable to determine what is real and what is fake. Gradually, users will doubt the truth and begin to believe lies and false information. *Cyborg users* are malicious accounts created by real people; however, they maintain activities by using programs. Therefore, cyborgs are better at spreading false news [24].

- Authenticity: This characteristic aims to identify whether news is factual [27]. *Factual statements* can be proven true or false. Subjective opinions are not considered factual statements. Only objective opinions are considered factual statements. Factual statements can never be incorrect. When a statement is published, it is not a factual statement if it can be disproved [41]. *Nonfactual statements* are statements that we can agree or disagree with. In other words, this news is sometimes wrong, sometimes right or completely wrong. Fake news contains mostly nonfactual statements.
- The information is news: This characteristic [27] reflects whether the information is news.

Based on the characteristics of fake news, we provide a new definition of fake news as follows. **"Fake news"** is news containing nonfactual statements with malicious accounts that can cause the echo chamber effect, with the intention to mislead the public.

Concepts related to Fake news: Various concepts regarding fake news exist. Using the characteristics of fake news, we can redefine these concepts to distinguish them as follows.

- False news [42,43] is news containing nonfactual statements from malicious accounts that can cause the echo chamber effect with undefined intentions.
- Disinformation [44] is news or non-news containing nonfactual statements from malicious accounts that can cause the echo chamber effect, with the intention to mislead the public.
- Cherry-picking [45] is news or non-news containing common factual statements from malicious accounts and can cause the echo chamber effect, with the intention to mislead the public.
- Rumor [46] is news or non-news containing factual or nonfactual statements from malicious accounts and can cause the echo chamber effect with undefined intentions.
- Fake information is news or non-news of nonfactual statements from malicious accounts that can cause the echo chamber effect, with the intention to mislead the public.
- Manipulation [47] is news on markets containing nonfactual statements from malicious accounts that can cause the echo chamber effect, with the intention to mislead the public.
- Deceptive news [2,24,27] is news containing nonfactual statements from malicious accounts that can cause the echo chamber effect, with the intention to mislead the public.
- Satire news [48] is news containing factual or nonfactual statements from malicious accounts that can cause the echo chamber effect, with the intention to entertain the public.
- Misinformation [33] is news or non-news containing nonfactual statements from malicious accounts that can cause the echo chamber effect with undefined intentions.
- Clickbait [49] is news or non-news containing factual or nonfactual statements from malicious accounts that can cause the echo chamber effect, with the intention to mislead the public.

- Fake facts [50] are undefined information (news or non-news) comprising nonfactual statements from malicious accounts that can cause the echo chamber effect, with the intention to mislead the public.
- Propaganda [48] is biased information (news or non-news) comprising undefined statements (factual or nonfactual) regarding mostly political events from malicious accounts and that can cause the echo chamber effect, with the intention to mislead the public.
- Sloppy journalism [19] is unreliable and unverified information (news or non-news) comprising undefined statements shared by journalists that can cause the echo chamber effect, with the intention to mislead the public.

2.2. Fake news detection

2.2.1. What is fake news detection?

Unlike traditional news media, fake news is detected using mainly content-based news features; for social media, social context-based auxiliary features can aid in detecting fake news. Thus, in [24], the authors present a formal definition of fake news detection based on the content-based and context-based features of the news. Given the social interactions ε among n users for news article a, the objective of fake news detection is to predict whether a is an instance of fake news. This objective is defined by a prediction function $F : \varepsilon \to \{0, 1\}$ such that,

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

Herein, Shu and Sliva define prediction function F as a binary classification function because fake news detection comprises distorted information from publishers regarding actual news topics (distortion bias). According to media bias theory [51], a distortion bias is often defined as a binary classification.

Using the above definition of fake news detection, in this paper, we consider fake news detection as a multiclassification task. Given a set of n news $N = \{n_1, n_2, \ldots, n_n\}$ and a set of m labels Ψ, fake news detection identifies a classification function F, such that $F : N \to \Psi$, to map each news $n \in N$ into the true class with the reliable label in Ψ. Corresponding to the concepts related to fake news (see Section 2.1) are concepts related to fake news detection, such as rumor detection and misinformation detection (classification). These concepts are defined similarly to the fake news detection task.

2.2.2. Fake news detection datasets

In this section, we introduce common datasets that have been recently used for fake news detection. These datasets were prepared by combining the English datasets presented in previous papers [19,52,53] and enriched by adding missing datasets. In contrast to other surveys or review papers, we calculated the statistics on 35 datasets, whereas D'Ulizia et al. [52], Sharma et al. [53], and Khan et al. [19] considered only 27, 23, and 10 datasets, respectively. Therefore, we list datasets by domain name, type of concept, type of content, and number of classes. A brief comparison of the fake news datasets is presented in Table 2.

Based on the content presented in Table 2, these datasets can be further detailed as follows:

- ISOT[1]: Both fake news and real news from Reuters; fake news from websites flagged by PolitiFact and Wikipedia.
- Fakeddit: English multimodal fake news dataset including images, comments, and metadata news.

---

1 https://www.uvic.ca/engineering/ece/isot/datasets/fake-news/index.php

Table 2
A comparison among Fake news datasets.

Name of dataset	Domain	Type of concept	Type of content	Number of classes
1- ISOT [54,55]	Politics, society, business, sport, crime, technology, health	Fake news, real news	Text	2
2- Fakeddit [56]	Society, politics	Fake news	Text, image, videos	2, 3, 6
3- LIAR [57]	Politics	Fake news	Text	6
4- FakeNewsNet [58]	Society, politics	Fake news	Text, image	2
5- Stanford Fake News [59]	Society	Fake news, satire	Text, image, videos	2
6- FA-KES [60]	Politics	Fake news	Text	2
7- BREAKING! [61]	Society, politics	Fake news, satire	Text, image	2, 3
8- BuzzFeedNews [24]	Politics	Fake news	Text	4
9- FEVER [62]	Society	Fake news	Text	3
10- FakeCovid [63]	Health, society	Fake news	Text	11
11- CredBank [64]	Society	Rumor	Text	2, 5
12- Memetracker [65]	Society	Fake news, real news	Text	2
13- BuzzFace [66]	Politics, society	Fake news	Text	4
14- FacebookHoax [67]	Science	Fake news	Text	2
15- Higgs-Twitter [68]	Science	Fake news	Text	2
16- Trust and Believe [69]	Politics	Fake news	Text	2
17- Yelp [70]	Technology	Fake news	Text	2
18- PHEME [71]	Society, politics	Rumor	Text	2
19- Fact checking [72]	Politics, society	Fake news	Text	5
20- EMERGENT [73]	Society, technology	Rumor	Text	3
21- Benjamin Political News [74]	Politics	Fake news	Text	3
22- Burfoot Satire News [75]	Politics,economy, technology, society	Satire	Text	2
23- MisInfoText [76]	Society	Fake news	Text	5
24- Ott et al.'s dataset [77]	Tourism	Fake reviews	Text	2
25- FNC-1 [78]	Politics, society, technology	Fake news	Text	4
26- Fake_or_real_news [79]	Politics, society	Fake news	Text	2
27- TSHP-17 [80]	Politics	Fake news	Text	2, 6
28- QProp [81]	Politics	Fake news	Text	2, 4
29- NELA-GT-2018[82]	Politics	Fake news	Text	2, 3, 5
30- TW_info [83]	Politics	Fake news	Text	2
31- FCV-2018[84]	Society	Fake news	Videos, text	2
32- Verification Corpus [85]	Society	Fake news	Videos, text, image	2
33- CNN/Daily Mail [86]	Politics, society, business, sport, crime, technology, health	Fake news	Text	4
34- Tam et al.'s dataset [87]	Politics, technology, science, crime, fraud and scam, fauxtography	Rumor	Text	5
35- FakeHealth [88]	Health	Fake news	Text	2

- LIAR[2]: English dataset with 12,836 short statements regarding politics collected from online streaming and two social networks – Twitter and Facebook – from 2007 to 2016.
- Stanford Fake News: Fake news and satire stories, including hyperbolic support or condemnation of a figure, conspiracy theories, racist themes, and discrediting of reliable sources.
- FA-KES: Labeled fake news regarding the Syrian conflict, such as casualties, activities, places, and event dates.
- BREAKING!: English dataset created using the Stanford Fake News dataset and BS detector dataset[3]. The data, including news regarding the 2016 US presidential election, were collected from web pages.
- BuzzFeedNews[4]: English dataset with 2283 news articles regarding politics collected from Facebook from 2016 to 2017.

- FakeNewsNet[5]: English dataset with 422 news articles regarding society and politics collected from online streaming and Twitter.
- FEVER: English dataset with 185,445 claims regarding society collected from online streaming.
- FakeCovid: English dataset with 5182 news articles for COVID-19 health and society crawled from 92 fact-checking websites, referring to Poynter and Snopes.
- CredBank[6]: English dataset with 60 million tweets about over 1000 events regarding society collected from Twitter from October 2014 to February 2015.
- Memetracker: English dataset with 90 million documents, 112 million quotes, and 22 million various phrases regarding society collected from 165 million sites.
- BuzzFace: English dataset with 2263 news articles and 1.6 million comments regarding society and politics collected from Facebook from July 2016 to December 2016. This dataset was extended in September 2016.
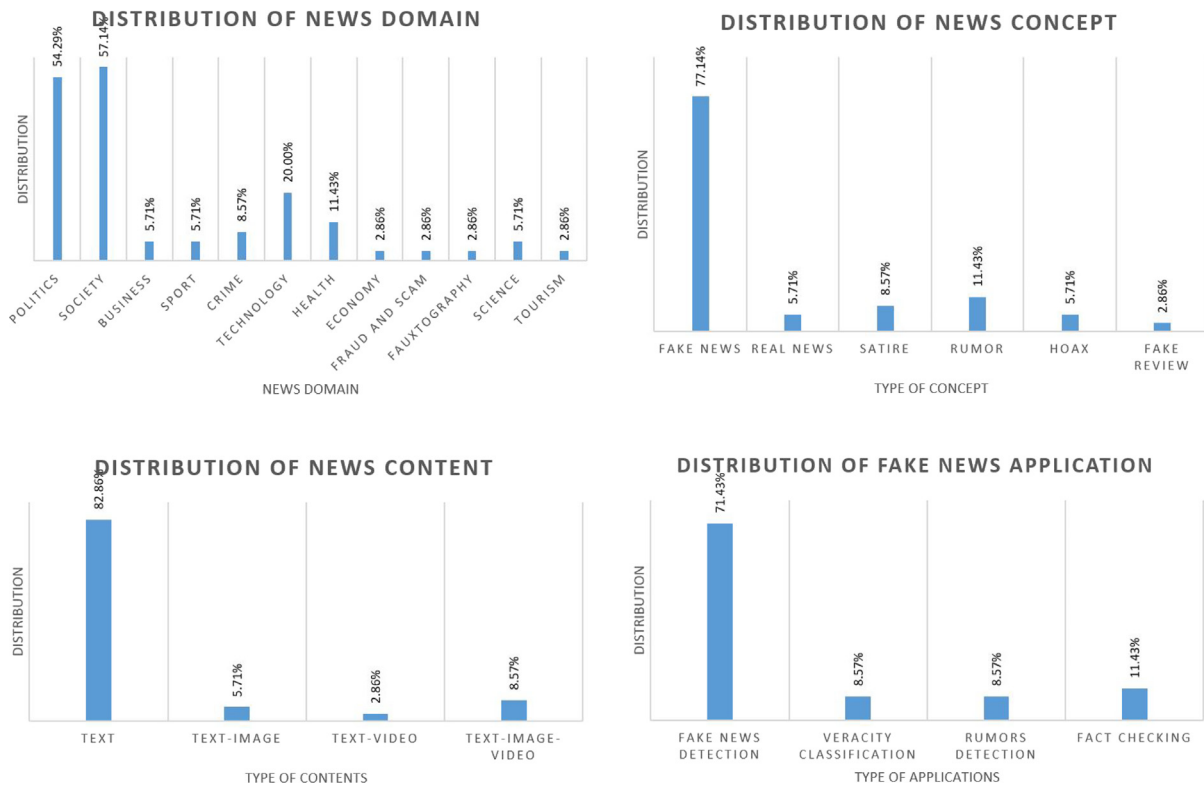
Fig. 1. A comparison among datasets in terms of four criteria.

- FacebookHoax: English dataset with 15,500 hoaxes regarding science collected from Facebook from July 2016 to December 2016. Additionally, this dataset identifies posts with over 2.3 million likes.
- Higgs-Twitter: English dataset with 985,590 tweets posted by 527,496 users regarding the science of the new Higgs boson detection collected from Twitter.
- Trust and Believe: English dataset with information from 50,000 politician users on Twitter. All information was labeled manually or using available learning methods.
- Yelp: English dataset with 18,912 technology fake reviews collected from online streaming.
- PHEME: English and German dataset with 4842 tweets and 330 rumors conversations regarding society and politics collected from Twitter.

Because of the limited number of manuscript pages, we do not describe further datasets. The remaining datasets are presented in the Appendix under Description of Datasets.

Based on the above analysis, we compare the criteria of fake news datasets in Fig. 1, followed by a discussion of observations and the main reason for these observations.

First, regarding the type of news content, 29 of the 35 datasets contained text data (82.86%); three of the 35 datasets comprised text, image, and video data (8.57%), namely, Fakeddit, Stanford Fake News, and Verification Corpus; two of the 35 datasets contained text and image data (5.71%), namely, FakeNewsNet and Breaking; and only one dataset contained text and video data (2.86%). No dataset included separate images or videos because previous fake news detection methods used mainly NLP-based techniques that were highly dependent on text data. Additionally, labeled image or video data are scarce because annotating them is labor intensive and costly.

Second, regarding the news domain, 20 and 19 of the 35 datasets focused on society news (57.14%) and political news

(54.29%), respectively, whereas only one dataset contained economy, fraud/scam, and fauxtography news (2.86%). These findings can be explained by the fact that fake news is more pertinent and widespread in political and societal domains than in other domains [89].

Third, regarding the type of fake news concepts, 27 of the 35 datasets contained the fake news concept (77.14%), followed by rumors (11.43%), satire (8.57%), hoaxes, and real news (5.71%), and finally, fake reviews (2.86%). Therefore, datasets containing the fake news concept are generally used for fake news detection applications because fake news contains false information spread by news outlets for political or financial gains [46].

Finally, regarding the type of applications, the most common application objective of the 35 datasets was fake detection (71.43%), followed by fact-checking (11.43%), veracity classification, and rumor detection (8.57%) because fake news detection applications can be used to solve practical problems. Additionally, fake news detection is the most general application, covering the entire process of classifying false information as true or false. Thus, fake information datasets are the most relevant for collection [52].

2.2.3. Features of fake news detection

The details of extracting and representing useful categories of features from news content and context are summarized in Fig. 2.

Based on the news attributes and discriminative characteristics of fake news, we can extract different features to build fake news detection models. Currently, fake news detection relies mainly on news and context information. In this survey, we categorize factors that can aid fake news detection into seven categories of features: network-, sentiment-, linguistic-, visual-, post-, user-, and latent-based features.

Linguistic-based features: These are used to capture information regarding the attributes of the writing style of the news, such as words, phrases, sentences, and paragraphs. Fake news
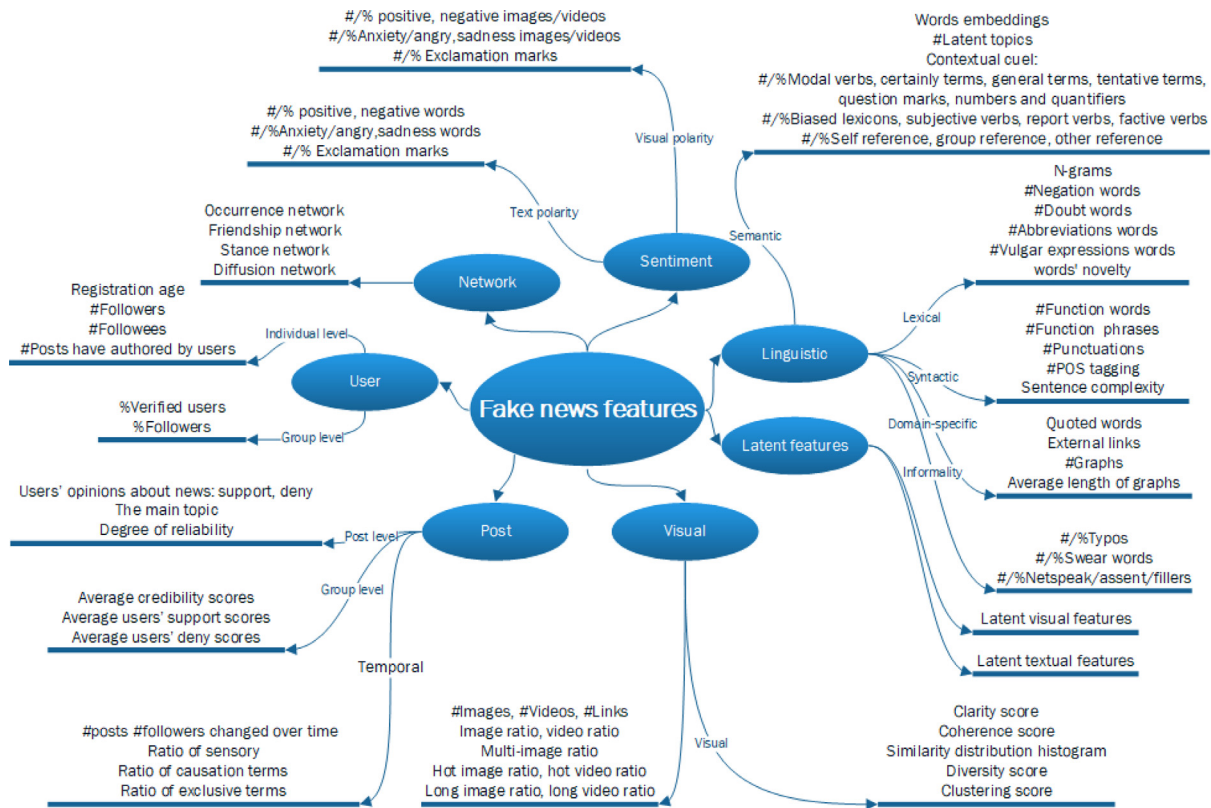
Fig. 2. Categories of features for fake news detection methods.

is created to mislead or entertain the public for financial or political gains. Therefore, based on the intention of fake news, we can easily extract features related to writing styles that often appear only in fake news, such as using provocative words to stimulate the reader's attention and setting sensational headlines. To best capture linguistic-based features, we divide them into five common types: lexical, syntactic, semantic, domain-specific, and informality. Lexical features refer to wording, such as the most salient characters (n-grams) [90,91], frequency of negation words, doubt words, abbreviation words, vulgar words [92], and the novelty of words [93]. Syntactic features capture properties related to the sentence level, such as the number of punctuations [94], number of function words (nouns, verbs, and adjectives) [93], frequency of POS tags [95], and sentence complexity [96,97]. Semantic features capture properties related to latent content, such as the number of latent topics [98] and contextual clues [99]. These features are extracted with state-of-the-art NLP techniques, such as distribution semantics (embedding techniques) and topic modeling (LDA technique) [100]. Domain-specific features capture properties related to domain types in the news, such as quoted words, frequency of graphs, and external links [101]. Informality features capture properties related to writing errors, such as the number of typos, swear words, netspeak, and assent words [27].

Sentiment-based features: This category of features captures properties regarding human emotions or feelings appearing in the news [102,103]. These features are identified and extracted based on the intentions and authenticity characteristics of fake news. They are classified into two groups: visual polarity and text polarity. The critical factors related to visual polarity are the number of positive/negative images/videos, number of anxious/angry/sad images/videos, and number of exclamation marks [27]. These factors capture information similar to the text polarity.

User-based features: This category of features is identified and extracted based on the malicious account characteristics of

fake news, specifically social bots and cyborg users. User-based features are properties related to user accounts that create or spread fake news. These features are classified into two levels, namely, the group level and the individual level [27]. The individual focuses on exploiting fake or real factors regarding each specific user, such as registration age, number of followers, and number of opinions posted by users [102,104]. Meanwhile, the group level focuses on factors regarding the group of users, such as the ratio of users, the ratio of followers, and the ratio of followees [95,105].

Post-based features: This category of features is identified and extracted based on the malicious accounts and news characteristics of fake news. Post-based features are used to capture properties related to users' responses or opinions regarding the news shared. These features are classified into three categories: group, post, and temporal [27]. The post level focuses on exploiting factors regarding each post [28], such as other users' opinions regarding this post (support, deny), main topic, and degree of reliability. The group level focuses on factors regarding all opinions related to this post [106], such as the ratio of supporting opinions, ratio of contradicting opinions, and reliability degree [95,105]. The temporal level notes factors such as the changing number of posts and followers over time and the sensory ratio [105].

Network-based features: Network-based features are employed to extract information regarding the attributes of the media where the news appears and is spread [107]. This category of features is identified and extracted based on the characteristics of fake news, such as the echo chamber, malicious account, and intention. Herein, the extractable features are propagation constructions, diffusion methods, and some factors related to the dissemination of news, for example, density and clustering coefficient. Therefore, many network patterns can form, such as occurrence, stance, friendship, and diffusion [24]. The stance network [106] is a graph with nodes, edges, nodes showing all the

text related to the news, and edges between nodes show similar weights of stances in texts. The co-occurrence network [28] is a graph with nodes showing users and edges indicating user engagement, such as the number of user opinions on the same news. The friendship network [105] is a graph with nodes showing users who have opinions related to the same news and edges showing the followers/followees constructions of these users. The diffusion network [105] is an extended version of the friendship network with nodes that indicate users who have opinions on the same news; the edges show the information diffusion pathways among these users.

Data-driven features: This category of features is identified and extracted based on the data characteristics of fake news, such as the data domain, data concept, data content, and application. The data domain exploits domain-specific and cross-domain knowledge in the news to identify fake news from various domains [108]. The data concept focuses on determining whether concept drift [109] exists in the news. The data content focuses on considering properties related to latent content in the news, such as the number of latent topics [98] and contextual clues [99]. These features are extracted based on state-of-the-art NLP techniques, such as distribution semantics (embedding techniques) and topic modeling (LDA technique) [100].

Visual-based features: Few fake news detection methods have been applied to visual news [24]. This category of features is identified and extracted based on the authenticity, news, and intended characteristics of fake news. Visual-based features are used to capture properties related to news containing images, videos, or links [27,100]. The features in this category are classified into two groups: visual and statistical. The visual level reflects factors regarding each video or image, such as clarity, coherence, similarity distribution, diversity, and clustering score. The statistical level calculates factors regarding all visual content, such as the ratio of images and the ratio of videos.

Latent features: A critical concept that we need to be aware of herein is latent features that are not directly observable, including latent textual features and latent visual features. Latent features are needed to extract and represent latent semantics from the original data more effectively. This category of features is identified and extracted based on the characteristics of fake news, such as the echo chamber, authenticity, and news information. Latent textual features are often extracted by using the news text representation models to create news text vectors. Text representation models can be divided into three groups: contextualized text representations, such as BERT [110], ELMo [111], Non-contextualized text representation, such as Word2Vec [112], FastText [113], GloVe [114], and knowledge graph-based representation, such as Koloski et al. method [115], RotatE [116], QuatE [117], ComplEx [118]. Contextualized text representations are word vectors that can capture richer context and semantic information. Knowledge graph-based representations can enrich various contextual and noncontextual representations by adding human knowledge representations via connections between two entities with their relationship based on knowledge graphs. News text representations can be not only used as inputs for traditional machine learning models [119] but also integrated into deep learning models, such as neural networks [115], recurrent networks [120], and transformers [110,121,122], and GNNs-based models [123–125] for fake news detection. Latent visual features are often extracted from visual news, such as images and videos. Latent visual features are extracted by using neural networks [126] to create a latent visual representation containing an image pixel tensor or matrix.

2.2.4. Fake news detection techniques

Fig. 3 shows an overview of fake news detection techniques. Previous related papers [21,27,42,46,53,79,107] demonstrated that fake news detection techniques are often classified into four categories of approaches: content-based approaches, including knowledge-based and style-based approaches, context-based approaches, propagation-based approaches, multilabel learning-based approaches, and hybrid-based fake news detection approaches. Let Ψ^a be one of the corresponding output classes of the fake news detection task. For example, $\Psi^a \in \{$real, false$\}$ or $\Psi^a \in \{$nonrumor, unverified rumor, false rumor, true rumor$\}$ or $\Psi^a \in \{$true, false$\}$.

Knowledge-based detection: Given news item a with a set of knowledge denoted by triple $K = (S, P, O)$ [127], where $S = \{s_1, s_2, \ldots, s_k\}$ is a set of subjects extracted from news item a, $P = \{p_1, p_2, \ldots, p_k\}$ is a set of predicates extracted from news item a, $O = \{o_1, o_2, \ldots, o_k\}$ is a set of objects extracted from news item a. Thus, $k_i^a = (s_i, p_i, o_i) \in K$, $1 \le i \le n$, is called a knowledge. For example: we have a news as *"John Smith is a famous doctor at a central hospital"*; from this statement, we have $k_i^a = (JohnSmith, Profession, Doctor)$. Assume that we have a set of true knowledge $Kt = (St, Pt, Ot)$, where $kt_l^a = (st_l, pt_l, ot_l) \in Kt$, $1 \le l \le m$. Let G_K be a true knowledge graph including a set of true knowledge, where nodes represent a set of $(St, Ot) \in Kt$ and edges represent a set of $(Pt) \in Kt$, the aim of a knowledge-based fake news detection method is to define a function F to compare $k_i^a = (s_i, p_i, o_i) \in K$ with $kt_l^a = (st_l, pt_l, ot_l) \in Kt$, such that: $F : k_i^a \xrightarrow{G_K} \Psi_i^a$. Function F is used to assign a label $\Psi_i^a \in [0, 1]$ to each triple (s_i, p_i, o_i) by comparing it with all triples (st_l, pt_l, ot_l) on graph G_K, where labels 0 and 1 indicate fake and real, respectively. Function F can be defined as $F(k_i^a, G_K) = Pr($edge p_i is a link from $\hat{s}_i$ to $\hat{o}_i$ on $G_K)$, where Pr is the probability; $\hat{s}_i$ and $\hat{o}_i$ are the matched nodes to s_i and o_i on G_K, respectively. $\hat{s}_i$ and $\hat{o}_i$ are identified as $\hat{s}_i = \text{argmin}_{st_l} |J(s_i, st_l)| < \xi$ and $\hat{o}_i = \text{argmin}_{ot_l} |J(o_i, ot_l)| < \xi$, respectively, where ξ is a certain threshold; $J(s_i, st_l)$ is a function to calculate the distance between s_i and st_i and it is the similar for $J(o_i, ot_l)$. For example, when $|J(s_i, st_l)| = 0$ or $|J(s_i, st_l)| < \xi$, we can regard s_i as the same as st_i. The techniques in this category are proposed based on the authenticity and news characteristics of fake news. The objective of knowledge-based techniques is to employ external sources to fact-check news statements. The fact-checking step aims to identify the truth of a statement corresponding to a specific context [72]. It can be implemented automatically (computational-oriented [128]) or manually (expert-based [101, 129,130], crowd-sourced [67,131]).

Style-based detection: Given a news item a with a set of f_s^a style features, where f_s^a is a set of features regarding the news content. Style-based fake news detection is defined as binary classification to identify whether news item a is fake or real, which means that we have to find a mapping function F such that $F : f_s^a \rightarrow \Psi^a$. The techniques in this category are proposed based on the intention and news characteristics of fake news. The objective of style-based techniques is to capture the distinct writing style of fake news. Fake news employs distinct styles to attract the attention of many people and stand out from ordinary news. The capturing step of the writing styles was built automatically. However, two techniques must be observed as criteria: style representation techniques [132–134] and style classification techniques [28,91,135].

Context-based detection: Given news item a with a set of f_c^a context features, where f_c^a includes news text, news source, news publisher, and news interaction. Context-based fake news detection is defined as the task of binary classification to identify whether news item a is fake or real, which means that we have to find a mapping function F such that $F : f_c^a \rightarrow \Psi^a$. The techniques
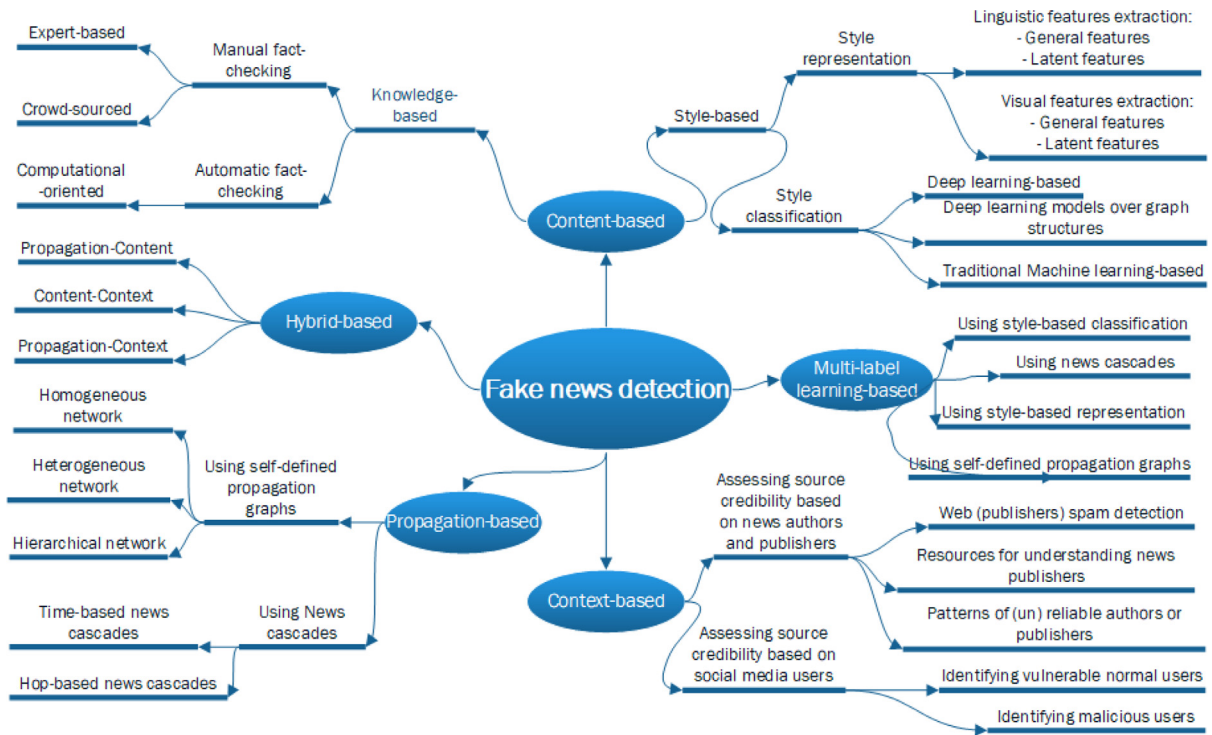
Fig. 3. Categories of fake news detection.

in this category are proposed based on the malicious account and news characteristics of fake news. The objective of source-based techniques is to capture the credibility of sources that appear, publish, and spread the news [27]. Credibility refers to people's emotional response to the quality and believability of news. The techniques in this category are often classified into two approaches: (i) assessing the reliability of sources where the news appeared and is spread based on news authors and publishers [136,137] and (ii) assessing the reliability of sources where the news appeared and is spread based on social media users [105,138,139].

Propagation-based detection: Given news item a with a set of f_p^a propagation patterns features for news. Propagation-based fake news detection is defined as binary classification to identify whether news item a is fake or real, which means that we have to develop a mapping function F such that $F : f_p^a \rightarrow \Psi^a$. The techniques in this category are proposed based on the echo chamber effect and news characteristics of fake news. The objective of propagation-based techniques is to capture and extract information regarding the spread of fake news. That is, the methods in this category aim to detect fake news based on how people share it. These techniques are often grouped into two small categories: (i) using news cascades [140,141] and (ii) using self-defined propagation graphs [142–145].

Multilabel learning-based detection: Let $\chi \in R^d$ be the d-dimension input feature matrix; hence, news item $a = [a_1, \ldots, a_d] \in \chi$; and let $\Gamma = \{real, fake\}^l$ be the label matrix, such that $\Psi = [\Psi_1, \ldots, \Psi_l] \in \Gamma$, where l is the number of class labels. Given a training set $\{(a, \Psi)\}$, the task of multilabel learning detection is to learn a function $F : \chi \rightarrow \Gamma$ to predict $\hat{\Psi} = F(a)$. Multilabel learning-based detection is a learning method where each news item in the training set is associated with a set of labels. The techniques in this category are proposed based on the echo chamber effect and news characteristics of fake news. The objective of multilabel learning-based techniques is to capture and extract information regarding the news content and the news latent text. The techniques in this category are often

classified into four approaches: (i) using style-based representation [17,115,146,147]; (ii) using style-based classification [15, 29,148–151]; (iii) using news cascades [140,152]; and (iv) using self-defined propagation graphs [4,16,125,153].

Hybrid-based detection: This method is a state-of-the-art approach for fake news detection that simultaneously combines two previous approaches, such as content-context [154,155], propagation-content [147,156], and context-propagation [4,14]. These hybrid methods are currently of interest because they can capture more meaningful information related to fake news. Thus, they can improve the performance of fake news detection models.

A critical issue that needs to be discussed is fake news early detection. Early detection of fake news provides an early alert of fake news by extracting only the limited social context with a suitable time delay compared with the appearance of the original news item. Knowledge-based methods are slightly unsuitable for fake news early detection because these methods depend strongly on knowledge graphs; meanwhile, newly disseminated news often generates new information and contains knowledge that has not appeared in knowledge graphs. Style-based methods can be used for fake news early detection because they depend mainly on the news content that allows us to detect fake news immediately after news appears and has not been spread. However, style-based fake news early detection methods are only suitable for a brief period because they rely heavily on the writing style, which creators and spreaders can change. Propagation-based methods are unsuitable for fake news early detection because news that is not yet been disseminated often contains very little information about its spread. To the best of our knowledge, context-based methods are most suitable for fake news early detection because they depend mainly on the news surroundings, such as news sources, news publishers, and news interactions. This feature allows us to detect fake news immediately after news appears and has not been spread by using website spam detection [157], distrust link pruning [158], and user behavior analysis [159] methods. In general, early detection of fake news is only suitable for a brief period because human

Table 3
Descriptions of notations.

Notations	Descriptions
$\lvert . \rvert$	The length of a set
G	A graph
V	The set of nodes in a graph
v	A node in a graph
E	The set of edges in a graph
e_{ij}	An edge between two nodes v_i, v_j in a graph
A	The graph adjacency matrix
D	The degree matrix of A. $D_{ii} = \sum_{j=1}^{n} A_{ij}$
n	The number of nodes
m	The number of edges
r	The set of relations of edges
d	The dimension of node feature vector
c	The dimension of edge feature vector
$x_{v_i, v_j}^{e} \in R^c$	The feature vector of edge e_{ij}
$x_v^n \in R^d$	The feature vector of node v
$X^e \in R^{m \times c}$	The edge feature matrix of a graph
$X \in R^{n \times d}$	The node feature matrix of a graph
$X^{(t)} \in R^{n \times d}$	The node feature matrix at the time step t

intelligence is limitless. When an early detection method of fake news is applied, it will not be long until humans create an effective way to combat it. This issue is still a major challenge for the fake news detection field.

2.3. Understanding graph neural networks

In this section, we provide the background and definition of a GNN. The techniques, challenges, and types of GNNs are discussed in the following section. Before presenting the content of this section, we introduce the notations used in this paper in Table 3.

2.3.1. What is a graph?

Before we discuss deep learning models on graph structures, we provide a more formal description of a graph structure. Formally, a simple graph is presented as $G = \{V, E\}$, where $V = \{v_1, v_2, \ldots, v_n\}$ is the set of nodes, and $E = \{e_{11}, e_{12}, \ldots, e_{nn}\}$ is the set of edges where $e_{ij} = (v_i, v_j) \in E$, $1 \le i, j \le n$. In which, v_i and v_j are two adjacent nodes. The adjacency matrix A is a $n \times n$ matrix with

$$A_{ij} = \begin{cases} 1, & if \, e_{ij} \in E, \\ 0, & if \, e_{ij} \notin E. \end{cases} \tag{2}$$

We can create improved graphs with more information from simple graphs, such as attributed graphs [6], multi-relational graphs [160].

Attributed graphs are the extended version of simple graphs. They are obtained by adding the node attributes X or the edge attributes X^e, where $X \in R^{n \times d}$ is a node feature matrix with $x_v^n \in R^d$ indicating the feature vector of a node v; $X^e \in R^{m \times c}$ is an edge feature matrix with $x_{v_i, v_j}^e \in R^c$ indicating the feature vector of an edge e_{ij}.

Spatial–Temporal graphs are special cases of attributed graphs, where the node attributes automatically change over time. Therefore, let $X^{(t)}$ be a feature matrix of the node representations at tth time step, a spatial–temporal graph is defined as $G^{(t)} = \{V, E, X^{(t)}\}$, where $X^{(t)} \in R^{n \times d}$.

Multi-relational graphs are another extension version of simple graphs that include edges with different types of relations τ. In these cases, we have $e_{ij} = (v_i, v_j) \in E \rightarrow e_{ij} = (v_i, \tau, v_j) \in E$. Each edge has one relation adjacency matrix A_τ. The entire graph can be created an adjacency tensor $\mathcal{A} \in R^{n \times r \times n}$. The multi-relational graphs can be divided into two subtypes: heterogeneous and multiplex graphs.

Heterogeneous graphs: Here, nodes can be divided into different types. That means $V = V_1 \cup V_2 \cup \ldots \cup V_k$, where for $i \neq j$, $V_i \cap V_j = \varnothing$. Meanwhile, edges must generally satisfy the conditions following the node types. Then, we have $e_{ij} = (v_i, \tau, v_j) \in E \rightarrow e_{ij} = (v_i, \tau_h, v_j) \in E$, where $v_i \in V_t$, $v_j \in V_k$ and $t \neq k$.

Multiplex graphs: Here, graphs are divided into a set of k layers, where each node belongs to one layer, and each layer has a unique relation called the intralayer edge type. Another edge type is the interlayer edge type. The interlayer connects the same node across the layers. That means $G = \{G_i, i \in \{1, 2, \ldots, k\}\}$, $G_i = \{V_i, E_i\}$, with $V_i = \{v_1, v_2, \ldots, v_n\}$, $E_i = E_i^{intra} \cup E_i^{inter}$, $E_i^{intra} = \{e_{lj} = (v_l, v_j), v_l, v_j \in V_i\}$, $E_i^{inter} = \{e_{lj} = (v_l, v_j), v_l \in V_i, v_j \in V_h, 1 \le h \le k, h \neq i\}$.

2.3.2. What are graph neural networks?

GNNs are created using deep learning models over graph structure data, which means deep learning models deal with Euclidean space data; in contrast, GNNs [6,161–163] deal with non-Euclidean domains. Assume that we have a graph $G = \{V, E\}$ with adjacency matrix A and node feature matrix (or edge feature matrix) X (or X^e). Given A and X as inputs, the main objective of a GNN is to find the output, i.e., node embeddings and node classification, after the k-th layer is: $H^{(k)} = F(A, H^{(k-1)}; \theta^{(k)})$, where F is a propagation function; θ is the parameter of function F, and when $k = 1$, then $H^{(0)} = X$. The propagation function has a number of forms. Let $\sigma(\cdot)$ be a non-linear activation function, e.g., ReLU; $W^{(k)}$ is the weight matrix for layer k; $\hat{A}$ is the normalized adjacency matrix and calculated as $\hat{A} = D^{-0.5} A^\top D^{-0.5}$ with D, is the diagonal degree matrix of $A^\top$, that is calculated as $D_{ii} = \sum_j A_{ij}^\top$; $A^\top = A + I$ with I is the identity matrix. A simple form of the propagation function is often used: $F(A, H^{(k)}) = \sigma(AH^{(k-1)}W^{(k)})$. In addition, the propagation function can be improved to be suitable for GNN tasks as follows:

For the node classification task, function F often takes the following form [164]:

$$F(A, H^{(k)}) = \sigma(\hat{A} H^{(k-1)} W^{(k)}) \tag{3}$$

For the node embeddings task, function F often takes the following form [165]:

$$F(A, H^{(k)}) = \sigma((Q\phi(H_e^{(k-1)} M_e) Q^\top \odot \hat{A}) H^{(k-1)} W^{(k)}) \tag{4}$$

where Q is a transformer representing whether edge e is connected to the given node and $Q^\top = T + I$; M_e the learnable matrix for the edges; ϕ is the diagonalization operator; $\odot$ is the element-wise product; $H_e^{(k-1)}$ is the hidden feature matrix of edges in the $(k-1)$-th layer, where $H_e^0 = X^e$ (X^e is the edge feature matrix). The $Q\phi(H_e^{(k-1)} M_e) Q^\top$ is to normalize the feature matrix of edges. The $Q\phi(H_e^{(k-1)} M_e) Q^\top \odot \hat{A}$ is to fuse the adjacency matrix by adding the information from edges.

More choices of the propagation function in GNNs are detail presented in Refs. [13,165]. Early neural networks were applied to acyclic graphs by Sperduti et al. [166] in 1997. In 2005, Gori et al. [167] introduced the notion of GNNs, which were further detailed by Scarselli et al. [168] in 2009 and by Gallicchio et al. [169] in 2010. According to Wu et al. [6], GNNs can be divided into four main taxonomies: conventional GNNs, graph convolutional networks, graph autoencoders, and spatial–temporal graph neural networks. In the next subsections, we introduce the categories of GNNs.

Conventional graph neural networks (GNNs*) which are an extension of recurrent neural networks (RNNs), were first introduced by Scarselli et al. [168] by considering an information diffusion mechanism, where states of nodes are updated and information is exchanged until a stable equilibrium is obtained [167,168]. In these GNNs, the function F is also defined as Eq. (3). However, the feature matrix of the k-th layer $H^{(k)}$ is updated using different equation as follows:

$$H_{v_j}^{(k)} = \sum_{v_i \in N(v_j)} F(x_{v_j}^n, x_{(v_i, v_j)}^e, x_{v_i}^n, H_{v_i}^{(k-1)}) \tag{5}$$

where $N(v_j)$ is the set of neighbor nodes of node v_j, F is a parametric function, $H_{v_j}^{(k)}$ is the feature vector of node v_j for the k-th layer, and $H_{v_j}^{(0)}$ is a random vector.

Graph convolutional networks (GCNs) were first introduced by Kipf and Welling [164]. They are capable of representing graphs and show outstanding performance in various tasks. In these GNNs, after the graph is constructed, the function F is also defined as Eq. (3). However, the recursive propagation step of a GCN at the k-th convolution layer is given by:

$$H^{(1)} = \sigma(\hat{A}H^{(0)}W^{(1)} + b^{(1)}) \tag{6}$$

Hence,

$$H^{(2)} = \sigma(\hat{A}H^{(1)}W^{(2)} + b^{(2)}) \tag{7}$$

That means:

$$H^{(k)} = \sigma(\hat{A}H^{(k-1)}W^{(k)} + b^{(k)}) \tag{8}$$

where $H^{(0)} = X$. $\sigma(\cdot)$ is an activation function. $W^{(k)} \in R^{m \times d}$, $k = \{1, 2, 3, \ldots\}$ is a transition matrix created for the k-th layer. $b^{(1)}$ and $b^{(2)}$ are the biases of two layers.

Graph autoencoders (GAEs) are deep neural architectures with two components: (i) the encoder, which converts nodes on the graph into a vector space of latent features, and (ii) the decoder, which decodes the information on the graph from the latent feature vectors. The first version of GAEs was introduced by Kipf and Welling [170,171]. In these GNNs, the form of function F is redefined as the following Equation:

$$F(\tilde{A}, H^{(k)}) = \sigma(\tilde{A}H^{(k-1)}W^{(k)}) \tag{9}$$

where $\tilde{A} = \varphi(ZZ^\top)$ is the reconstructed adjacency matrix and φ is the activation function of the decoder composition. Z is the output of the encoder composition. In these GAEs, the GCNs are used in the encoder step to create the embedding matrix; therefore, Z is calculated based on Eq. (3). Thus, $Z = F(\hat{A}, H^{(k)})$ with $F(\cdot)$ corresponds to the case of GCNs. $Z^\top$ is the transpose matrix of Z.

Spatial–temporal graph neural networks (STGNNs) in various real-world tasks are dynamic as both graph structures and graph inputs. To represent these types of data, a spatial–temporal graph is constructed as introduced in Section 2.3.1. However, to capture the dynamicity of these graphs, STGNNs have been proposed for modeling the inputs containing nodes with dynamic and interdependency. STGNNs can be divided into two approaches: RNN-based and CNN-based methods.

For the RNN-based approach, to capture the spatial–temporal relation, the hidden states of STGNNs are passed to a recurrent unit based on graph convolutions [172–174]. The propagation function form of STGNNs is also shown in Eq. (3). However, the value of the k-th layer is calculated as follows:

$$H^{(t)} = \sigma(WX^{n(t)} + UH^{(t-1)} + b) \tag{10}$$

where $X^{n(t)}$ is the node feature matrix at time step t. After using graph convolutions, Eq. (10) is recalculated as follows:

$$H^{(t)} = \sigma(GCN(X^{n(t)}, \hat{A}; W) + GCN(H^{(t-1)}, \hat{A}; U) + b) \tag{11}$$

where GCN is one of GCNs model. $U \in R^{n \times n}$ is the eigenvector matrix ranked by eigenvalues with $U^\top U = I$.

For the CNN-based approach, RNN-based approaches recursively handle spatial–temporal graphs. Thus, they must iterate the propagation process and therefore they have limitations regarding the propagation time and gradient explosion or vanishing problems [175–177]. CNN-based approaches can solve these problems by exploiting parallel computing to achieve stable gradients and low memory.

Attention-based graph neural networks (AGNNs) [178] remove all intermediate fully connected layers and replace the propagation layers with an attention mechanism that maintains the structure of the graph [179]. The attention mechanism allows learning a dynamic and adaptive local summary of the neighborhoods to obtain more accurate predictions [180]. The propagation function form of the AGNN is shown in Eq. (3). However, the AGNN includes graph attention layers. In each layer, a shared, learnable linear transformation $M \in R^{t_h \times d_h}$, where h is the number of the t-th hidden layer, d_h is the dimensional of the t-th hidden layer, is used for the input features of every node as follows:

$$H^{(t)} = \sigma(M^{(t)}H^{(t-1)}) \tag{12}$$

where the row-vector of node v_i defined as follows:

$$H^{(t)v_i} = \sum_{v_j \in N(v_i) \cup \{i\}} M_{ij}^{(t-1)} H_j^{(t-1)} \tag{13}$$

where

$$M_{ij}^{(t-1)} = \varphi([\beta^{(t-1)} cos(H_i^{(t-1)}, H_j^{(t-1)})]_{v_j \in N(v_i) \cup \{i\}}) \tag{14}$$

where $\beta^{(t-1)} \in R$ is an attention-guided parameter of propagation layers. Note that the value of β of propagation layers is changed over hidden states. $\varphi(\cdot)$ is the activation function of propagation layer.

3. Survey methodology

In this study, we conducted a systematic review of fake news detection articles using GNN methods, including three primary steps: "literature search," "selection of eligible papers," and "analyzing and discussing" [181]. The research methodology is illustrated in Fig. 4:

The literature search is used to select peer-reviewed and English-language scientific papers containing the following keywords: "GNN" OR "graph neural network" OR "GCN" OR "graph convolutional network" OR "GAE" OR "graph autoencoder" OR "AGNN" OR "attention-based graph neural network" combined with "fake news" OR "false news" OR "rumour" OR "rumor" OR "hoax" OR "clickbait" OR "satire" OR "misinformation" combined with "detection". These keywords were extracted from Google Scholar, Scopus, and DBLP from January 2019 to the end of Q2 2021.

The selection of eligible papers is used to exclude the non-explicit papers on fake news detection using GNNs. To select the explicit papers, we specify a set of exclusion/inclusion criteria. The inclusion criteria were as follows: written in English, published after 2019, peer-reviewed, and retrieved full-text. The exclusion criteria were as follows: papers of reviews, surveys, and comparisons or only presented mathematical models.

Analysis and discussion papers are used to compare the surveyed literature and capture the main challenges and interesting open issues that aim to provide various unique future orientations for fake news detection.

By the above strategy, a final total of 27 papers (5 papers in 2019, 16 papers in 2020, and 6 papers for the first 6 months of 2021) are selected for a comprehensive comparison and analysis. These selected papers are classified into four groups based on GNN taxonomies (see Section 2.3.2), including conventional GNN-based, GCN-based, AGNN-based, and GAE-based methods. In the next step, eligible papers are analyzed via the criteria of the method's name, critical idea, loss function, advantage, and disadvantage.
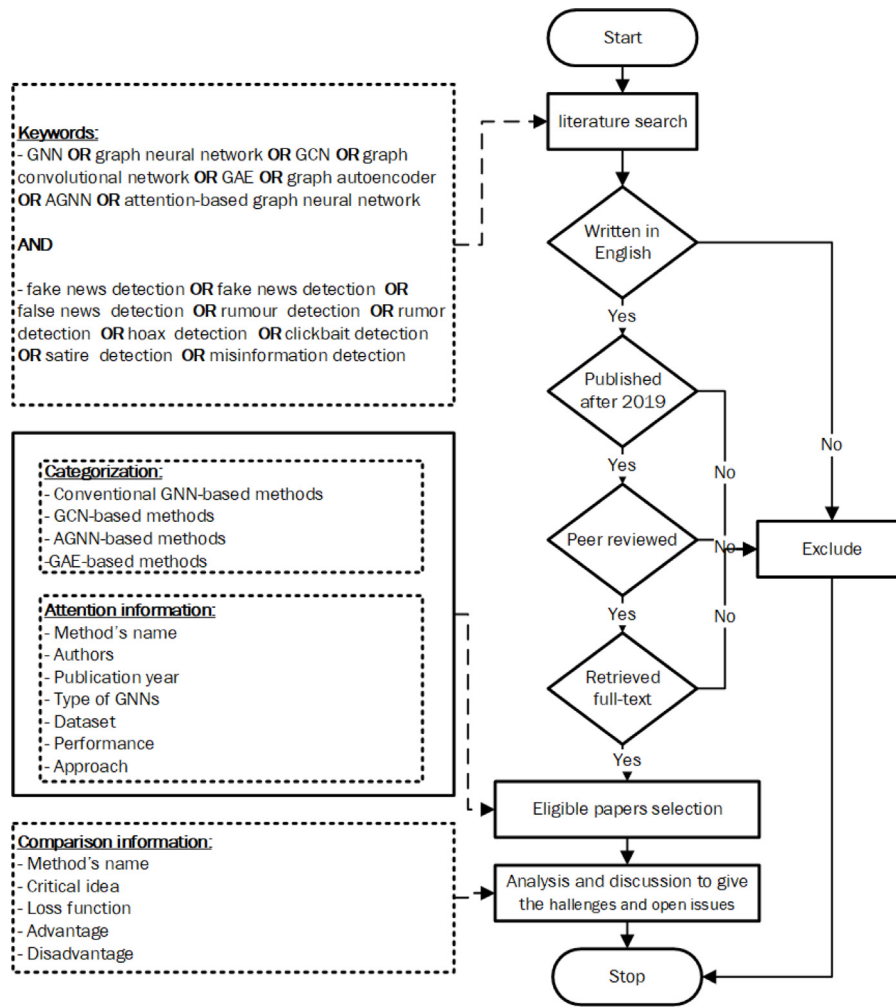
Keywords:
- GNN **OR** graph neural network **OR** GCN **OR** graph convolutional network **OR** GAE **OR** graph autoencoder **OR** AGNN **OR** attention-based graph neural network

AND

- fake news detection **OR** fake news detection **OR** false news detection **OR** rumour detection **OR** rumor detection **OR** hoax detection **OR** clickbait detection **OR** satire detection **OR** misinformation detection

Categorization:
- Conventional GNN-based methods
- GCN-based methods
- AGNN-based methods
- GAE-based methods

Attention information:
- Method's name
- Authors
- Publication year
- Type of GNNs
- Dataset
- Performance
- Approach

Comparison information:
- Method's name
- Critical idea
- Loss function
- Advantage
- Disadvantage

Fig. 4. Flow diagram of research methodology.

4. Quantitative analysis of eligible papers

Previous fake news detection approaches have mainly used machine learning [74,92,182–184] and deep learning [95,120, 185–189] for classifying news as fake or real, rumor or not a rumor, and spam or not spam. Various surveys and review papers regarding fake news detection using machine learning and deep learning have been published. In this paper, we discuss in detail the most current GNN-based fake news detection approaches. Using the research methodology in Section 3, a final total of 27 papers published after 2019 using GNNs for fake news detection were selected for a more detailed review in the following subsections. Table 4 presents comparisons among previous studies in terms of model name, referral code (Table 5), authors, year of publication, type of GNN, datasets, performance, and approach-based fake news detection.

Using the relationships among the information in Table 4, we compare quantitatively surveyed methods in terms of four distribution criteria of GNN-based fake news detection approaches, as shown in Fig. 5.

The number of surveyed papers from 2019 to 2021 (the end of Q2) regarding fake news detection using GNNs shows that this problem is attracting increasing attention from system practitioners (increasing 40.74% from 2019 to 2020). Although in 2021, only 22.22% of articles on fake news detection focused on using GNNs, Q2 has not yet ended, and we believe that the last two quarters of the year will produce more articles in this field, considering the

outbreak of fake news related to COVID-19 and the challenges of this problem.

With regard to the type of news concepts employed (types of objectives), 14 of the 27 surveyed papers are related to fake news detection (51.85%), followed by rumors and spam detection (29.63%, 7.41%), whereas other types of detection constitute only 3.7%. A likely reason for these results is the creation and spread of fake news correspond to active economic and political interests. That is, if fake news is not detected and prevented in a timely manner, people will suffer many deleterious effects. Additionally, as analyzed above, an equally important reason is that datasets used for fake news detection are now richer and more fully labeled than other datasets (see Section 2.2.2).

With regard to GNN-based techniques, the authors predominantly (74.07%) used GCNs for fake news detection models, followed by GNN-based methods (14.81%), GAE, and AGNN (3.7%). This choice is attributable to the suitability of GCNs for graph representations in addition to having achieved state-of-the-art performance in a wide range of tasks and applications [13].

Finally, one-third of the propagation-based and content-based approaches (33.33%) were published, followed by hybrid-based (22.22%) and context-based (11.11%) approaches. This result is attributable to propagation-based and context-based approaches using mainly news information on network structures, users, and linguistics simultaneously. This information is most consequential for fake news detection.

Table 4
Comparison of surveyed methods using GNNs for fake news detection.

Method's name	Authors	PY and TG	Dataset	Performance	Approach-based
1-Monti et al.	[3]	2019, GCN	Tweets	ROC AUC:92.7%	Propagation
2- !GAS	[123]	2019, GCN	Spam dataset	F_1: 82.17%	Context
3-MGCN	[190]	2019, GCN	Liar	Acc: 49.2%	Content
4-!Chang Li et al.	[191]	2019, GCN	10,385 news articles	Acc: 67.03–88.89%	Content
5-Benamira et al.[2]	[192]	2019, AGNN	Horne et al. [74]	Acc: 70.45–84.25%	Content
		2019, GCN		Acc: 72.04–84.94%	
6-Marion et al.[1]	[153]	2020, GCN	FakeNewsNet	Acc: 73.3%	Propagation
7-Yi Han et al.	[4]	2020, GNN*	FakeNewsNet:		Propagation, context
			Politifact	Acc: 79.2–80.3%	
			GossipCop	Acc: 82.5–83.3%	
8-FakeNews	[193]	2020, GNN*	Covid-19 tweets	ROC: 95%	Content
9-GCAN[3]	[14]	2020, GCN	Twitter 15 [140],	Acc: 87.67%	Context, propagation
			Twitter 16 [140]	Acc: 90.84%	
10-Nguyen et al.	[155]	2020, GCN	Tweets	Task 1: MCC: 36.1–41.9%	Content
				Task 2: MCC: −8.1–1.51%	
11-Pehlivan et al.[4]	[194]	2020, GCN	Covid-19 Tweets	GCN: T-MCC: 2%	Content
				DGCNN: T-MCC: 2.3%	
				M-FCN: T-MCC: 3.5%	
12-*Bi-GCN	[16]	2020, GCN	Weibo [99],	Acc: 96.8%	Propagation
			Twitter 15 [140],	Acc: 88.6%	
			Twitter 16[140]	Acc: 88.0%	
13-VGCN-ItalianBERT[5]	[195]	2020, GCN	1600 images with metadata	F_1: 84.37%	Content
14-*GCNSI	[196]	2020, GCN	Karate [197], Dolphin [198],	Improve the	Propagation
			Power grid [199], Jazz [200],	best method	
			Ego-Facebook	by about 15%	
15-SAFER[6]	[15]	2020, GNN*	FakeNewsNet,	F_1 above 92.97%	Context
			FakeHealth	F_1 above 58.34%	
16-!GCNwithMRF	[201]	2020, GCN	Twitter [202,203]	Acc: 79.2–83.9%	Propagation
17-*Lin et al.[7]	[124]	2020, GAE	Weibo [99],	Acc: 93.4–94.4%	Propagation
			Twitter 15[140]	Acc: 84–85.6%	
			Twitter 16[140]	Acc: 85.2–88.1%	
18-*Malhotra et al.	[147]	2020, GCN	Twitter 15 [99]	Acc: 86.6%	Propagation, content
			Twitter 16 [140]	Acc: 86.5%	
19-!FauxWard	[149]	2020, GCN	Comments on Twitter	Acc: 71.09%	Content
			Comments on Reddit	Acc: 75.36%	
20- KZWANG[8]	[156]	2020, GCN	Weibo [99],	Acc: 95.0%	Propagation
			Twitter 15[140],	Acc: 91.1%	
			Twitter 16[140]	Acc: 90.7%	
21- FANG[9]	[5]	2020, GNN*	FakeNewsNet, PHEME	AUC: 75.18%	Context, content
22-*GraphSAGE	[125]	2021, GCN	Twitter [140]	Acc: 69.0–77.0%	Propagation
			PHEME	Acc: 82.6–84.2%	
23- Bert-GCN Bert-VGCN	[150]	2021, GCN	Covid-19 and 5G tweets	MCC: 33.12–47.95%	Content
				MCC: 39.10–49.75%	
24-*Lotfi	[204]	2021, GCN	PHEME	F_1: 80% (rumor)	Content, propagation
				F_1: 79% (non-rumor)	
25-*SAGNN	[151]	2021, GCN	Twitter 15[140],	Acc: 79.2–85.7%	Content
			Twitter 16[140]	Acc: 72.6–86.9%	
26-AA-HGNN	[17]	2021, AGNN	Fact-checking,	Acc: 61.55%	Content, context
			BuzzFeedNews	Acc: 73.51%	
27-*EGCN	[154]	2021, GCN	PHEME	Acc: 63.8–84.1%	Propagation

PY and TG: Publication year and Type of GNNs. MCC: Matthews correlation coefficient. T-MCC: MCC for test dataset. M-FCN: MALSTM-FCN model.

Table 5
Code source.

Refer	Code source
1	https://github.com/MarionMeyers/fake_news_detection_propagation
2	https://github.com/bdvllrs/misinformation-detection-tensor-embeddings
3	https://github.com/l852888/GCAN
4	https://github.com/titu1994/MLSTM-FCN
5	https://github.com/dbmdz/berts#italian-bert
6	https://github.com/shaanchandra/SAFER
7	https://github.com/lhbrichard/rumor-detection
8	https://github.com/shanmon110/rumordetection
9	https://github.com/nguyenvanhoang7398/FANG

5. Literature survey

In this section, we survey papers using graph neural networks for fake news detection. Based on GNN taxonomies (see Section 2.3.2), we categorized GNN-based fake news detection methods into conventional GNN-based, GCN-based, AGNN-based, and GAE-based methods, as shown in Table 6.

- Conventional GNN-based methods (GNN*) are pioneering GNN-based fake news detection methods. These methods
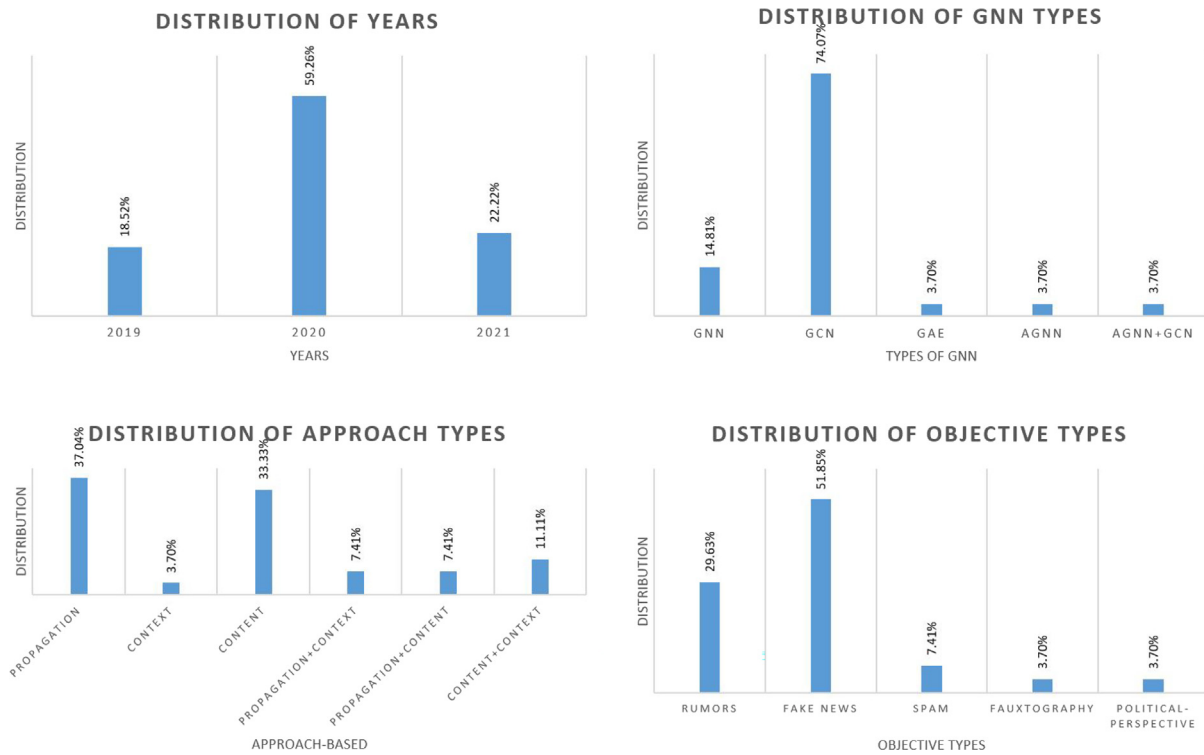
Fig. 5. A comparison of four distribution criteria of GNN-based fake news detection approaches.

Table 6
GNN-based detection methods categorization.

Category	Publication
Conventional GNN	[4,5,15,192,193]
GCN	[3,14,16,123,125,147,149–151,153–156,194,201,204]
AGNN	[17,192]
GAE	[124]

apply a similar set of recurrent parameters to all nodes in a graph to create node representations with better and higher levels.

- GCN-based methods (GCN) often use the convolutional operation to create node representations of a graph. Unlike the conventional GNN-based approach, GCN-based methods allow integrating multiple convolutional layers to improve the quality of node representations.
- AGNN-based methods are constructed mainly by feeding the attention mechanism into graphs. Thus, AGNNs are used to effectively capture and aggregate significant neighbors to represent nodes in the graph.
- GAE-based methods are unsupervised learning approaches to encode nodes on a graph into a latent vector and decode the encoded information to reconstruct the graph data to create node representations by integrating latent information.

Most approaches proposed in the surveyed papers for detecting false information are used to solve a classification problem task that involves associating labels such as rumor or nonrumor and true or false with a particular piece of text. In using GNNs for fake news detection, researchers have employed mainly conventional GNNs and GCNs to achieve state-of-the-art results. On the other hand, some researchers have applied other approaches, such as GAE and AGNN, to predict their conforming labels.

5.1. Detection approach based on GNNs*

GNNs* represent the first version of GNNs and improved the performance of fake news detection of machine learning and deep learning methods that use non-Euclidean data.

Han et al. [4] exploited the capability of GNNs* using non-Euclidean data to detect the difference between news propagation methods on social networks. They then classified the news into two labels of fake and real news by training two instances of GNNs*. In the first case, GNNs* were trained on complete data. The second case involved training GNNs* using partial data. In the second case, unlike conventional GNNs, two techniques – gradient episodic memory and elastic weight consolidation – were used to build GNNs* with continual learning aimed at the early detection of fake propagation patterns. This method can obtain superior performance without considering any text information compared with state-of-the-art models. In particular, time and cost are saved as the dataset grows when training the new data because the entire dataset is not retrained. However, one major limitation is that the strong forgetting occurrence is not solved by extracting more features, including the "universal" features. Hamid et al. [193] introduced a method to detect malicious users who spread misinformation by analyzing tweets related to conspiracy theories between COVID-19 and 5G networks. This method includes two substrategies: (i) content-based fake news detection and (ii) context-based fake news detection. The second strategy is implemented based on GNNs* to train the GNNs* representation and to classify 5G networking comments into three categories: nonconspiracy, conspiracy, and other conspiracies. The obtained performance in terms of average ROC is quite good (0.95%) because it captures mostly information related to the textual and structural aspects of news. However, neither textual nor structural information was used simultaneously. Nguyen et al. [5] proposed a model named FANG for fake news detection based on the news context by considering the following steps: (i) extracting features regarding the news, such as the source,

users, and their interactions, and posting a timeline; (ii) constructing two subhomogeneous graphs, namely, news source and user; and (iii) using an unsupervised model over two subgraphs separately to model neighbor relations. Moreover, the authors used pretrained detection networks to detect news content as extension information. FANG can capture a news context with higher fidelity than recent graphical and nongraphical models. In particular, FANG still achieves robustness even with limited training data. However, features such as users and their interactions are extracted before being fed into the FANG. Therefore, some errors regarding textual encoding and emotion detection can occur, and they are provided to FANG. Another limitation is the rapid obsolescence of contextual datasets because we cannot retrieve hyperlinks and other traces at the query time, as they might no longer be available.

Unlike other contemporary work on GNN*, the fake news detection task is introduced above. Chandra et al. [15] presented a method called SAFER with three distinct features. (i) They constructed a GNN* model with the same heterogeneous input graph for two types of edges and nodes. (ii) They determined context features by exploiting the impact of online social communities without using user profiles. (iii) They only used the network information of online users to evaluate these communities' roles, but their results were still better than those of previous approaches. The authors proposed a relational GNN* and a hyperbolic GNN* to model user and community relations. The relational GNN obtained better results than conventional GNNs. However, the results of the hyperbolic GNN* were comparable only to the other GNN*. Therefore, modeling users/communities for truly hierarchical social network datasets is a challenge that needs to be addressed in the future.

5.2. Detection approach based on GCNs

The GCN-based approach is a category of methods that are used mostly for fake news detection and rely on GNNs. GCNs are an extension of GNNs that derive the graph structure and integrate node information from neighborhoods based on a convolutional function. GCNs can represent graphs and achieve state-of-the-art performance on various tasks, including fake news detection.

Lu et al. [14] presented a novel method for fake news detection of tweets called GCAN, which includes five main steps as follows: (i) extract quantified features related to users; (ii) convert words in news tweets into vectors; (iii) represent aware propagation methods of tweets among users; (iv) capture the correlation between tweet context and user interactions and between tweet context and user propagation; and (v) classify tweets as fake or real news by combining all learned representations. GCAN exhibits outstanding performance with reasonable explainability. The main contribution of this study is the integration of dual coattention mechanisms with GCNs. The first mechanism simultaneously captures the relations between the tweet context and user interactions. The second mechanism simultaneously captures the relations between the tweet context and user propagation. This method can be considered an enriching version of GCNs. The form of GCAN was improved from GCNs as follows: $H^s = tanh(W_s S + (W_g G)F^\top)$, $H^g = tanh(W_g G + (W_s S)F)$, where S represents the embeddings of the relations between the tweet context and user interactions; G represents the embeddings of the relations between tweet context and user propagation; W_s and W_g represent matrices of learnable parameters; F and $F^\top$ are the transformation matrix and its transpose, respectively.

Based on the inherent aggregation mechanism of the GNN, Zhang et al. [151] proposed a simplified GNN method called SAGNN to calculate the degree of interaction between Twitter

users and other users for rumor detection. Unlike the convolution layer of the conventional GCN, the SAGNN does not contain the weight matrix W. Moreover, the identification of the adjacency matrix was different from that of conventional GCNs. Thus, the two layers in the SAGNN are defined as $H^{(1)} = \sigma(H^{(0)}E)$, where $H^{(1)}$ is called the embedding layer and E is the word embedding matrix. Hence, $H^{(2)} = \sigma(\tilde{A}H^{(1)})$, where $H^{(2)}$ is called the aggregation layer; and $\tilde{A} = I + uB + vC$, where u, v are learnable parameters of SAGNN. Matrix B is calculated as follows: if v_i is the parent of v_j then $B_{ij} = 1$ otherwise 0; whereas matrix C is defined as if v_i is a child of v_j then $C_{ij} = 1$ otherwise 0.

Ke et al. [156] constructed a heterogeneous graph, namely, KZWANG, for rumor detection by capturing the local and global relationships on Weibo between sources, reposts, and users. This method comprises three main steps as follows: (i) word embeddings convert text content of news into vectors using a multihead attention mechanism,

$$T = MultiHead(Q, K, V) = Concat(head_1, \ldots, head_h)W^o \qquad (15)$$

where $head_i = attention(QW_i^Q, KW_i^K, VW_i^V)$ with $Q \in R^{n_q \times d}$, $K \in R^{n_k \times d}$, $V \in R^{n_v \times d}$ are sentences of query, key, and value; n_q, n_k, n_v are the number of words in each sentence; and $attention(Q, K, V) = Softmax(\frac{QK^\top}{\sqrt{d_k}})V$; (ii) propagation and interaction representations are learned via GCNs; and (iii) graph construction builds a model of potential interactions among users: $P = H^{(k)} = \sigma(\hat{A}H^{(k-1)}W^{(k-1)})$. The difference between this model and conventional GCNs is that KZWANG is a combination of the news text representation using a multihead attention mechanism and propagation representation using GCNs. Thus, the outputs of the GCN layer and the multihead attention layer are the inputs of rumor classification: $R = Softmax(TP + b)$, where T is the text representation matrix. P is the propagation representation matrix. R is the output of the whole model.

Lotfi et al. [204] introduced a model that includes two GCNs: (i) a GCN of tweets, such as source and reply as $T = H^{(k)} = \sigma(\hat{A}_T H^{(k-1)}W^{(k-1)})$; (ii) GCN of users, such as interaction among users as $Re = H^{(k)} = \sigma(\hat{A}_{Re}H^{(k-1)}W^{(k-1)})$ where A_T is the adjacency matrix of the GCN of tweets and determined as if ($tweet$ i replies to $tweet$ j) or ($i = j$) then $A_T^{ij} = 1$ otherwise 0. Meanwhile, A_{Re} is the adjacency matrix of the GCN of users and defined as follows: if ($user$ i sent m $tweets$ to $user$ i in $conversation$) or ($i = j$) then $A_{Re}^{ij} = 1$ otherwise 0. And $H^{(0)} = X$ is determined as: if ($there$ is $high$ $frequency$ $words$ j in $tweet$ i) or (the $propagation$ $time$ is $interval$ $between$ the $reply$ $tweet$ i and the $source$ $tweet$) then $X^{ij} = 1$ otherwise 0 . Unlike other models, the authors constructed two independent GCNs and then concatenated them into one fully connected layer for fake news detection as $Softmax((T \oplus Re)W + b)$, where $\oplus$ is the concatenation function.

Vu et al. [125] presented a novel method called GraphSAGE for rumor detection based on propagation detection. In contrast to other propagation-based approaches, this method proposes a graph propagation embedding method based on a GCN to convert the news propagation procedure and their features into vector space by aggregating the node feature vectors and feature vectors of their local neighbors into a combination vector. Thus, the difference between the GraphSAGE model and the traditional GCN models concerns the aggregator functions, which are divided into the following aggregators: (i) Convolutional aggregator:

$$h_{v_j}^k = \sigma(W^k \frac{h_{v_j}^{k-1} + \sum_i h_{v_i}^{k-1}}{|N(v_j)| + 1}), \forall v_i \in N(v_j) \qquad (16)$$

(ii) LSTM aggregator:

$$\triangle_k^{lstm} = LSTM(\{h_{v_i}^{k-1}, \forall v_i \in N(v_j)\}) \qquad (17)$$

(iii) Pooling aggregator:

$$\triangle_k^{pool} = max(\sigma^{pool}(W^{pool}h_{v_i}^{k-1} + b^{pool})), \forall v_i \in N(v_j) \tag{18}$$

(iv) Linear aggregator:

$$\triangle_k^{linear} = \frac{\sum_i w_i h_{v_i}^{k-1}}{\sum_i w_i}, \forall v_i \in N(v_j) \tag{19}$$

where w_i is the weight vector of neighbor h_{v_i}. GraphSAGE efficiently integrates features such as content, social, temporal, and propagation structures. These features can aggregate significant information to train an algorithm for determining whether news are rumors or not. However, this method requires data on the entire propagation procedure of the posted news. In some cases, if the posted news does not obtain response opinions when it is spread, the accuracy of the GraphSAGE model can be reduced.

Bian et al. [16] proposed a Bi-GCN model with two propagation operations, top-down (TD-GCN) and bottom-up (BU-GCN), to detect two essential characteristics of rumors, dispersion, and propagation. Bi-GCN was constructed as follows: (i) High-level node representations as $H_{TD}^{(k)} = \sigma(\hat{A}_{TD}H^{(k-1)}W_{TD}^{(k)})$ and $H_{BU}^{(k)} = \sigma(\hat{A}_{BU}H^{(k-1)}W_{BU}^{(k)})$. (ii) Root feature enhancement as follows $\tilde{H}_{TD}^{(k)} = concat(H_{TD}^{(k)}, (H_{TD}^{(k-1)})^{root})$ and $\tilde{H}_{BU}^{(k)} = concat(H_{BU}^{(k)}, (H_{BU}^{(k-1)})^{root})$ where $concat$ is a concatenate function; $root$ indicates the root node; (iii) Node representations are fed into the pooling aggregator as $S_{TD} = mean(\tilde{H}_{TD}^{(2)})$ and $S_{BU} = mean(\tilde{H}_{BU}^{(2)})$, and then concatenated them into one fully connected layer for fake news detection as $\hat{y} = Softmax(concat(S_{TD}, S_{BU}))$. This model can capture both the propagation of rumor patterns using the TD-GCN and the dispersion of rumor structures using the BU-GCN. Additionally, hidden information of the news is extracted through layers of GCN to increase the influence of rumor roots. However, TD-GCN and BU-GCN were still constructed independently.

Bai et al. [154] constructed a graph called an SR graph, where the node feature matrix X is determined by word vectors; and the adjacent matrix A is defined as follows: $if\ tweet\ i\ replies\ to\ tweet\ j,\ then\ A_{ij} = 1\ otherwise\ 0$. Using the SR-graph, the authors proposed an EGCN model for rumor detection as $PG = H^{(k)} = \sigma(\hat{A}H^{(k-1)}W^{(k)})$ with a node proportion allocation mechanism as $PT = TextCNN(A, X)$ where $TextCNN$ indicates a conventional CNN model. Let n and m be the number of nodes in the current SR-graph and the max SR-graph, respectively, we have the feature output of the EGCN by $Y = PG \times \frac{n}{m} + PT(1 - \frac{n}{m})$. And the output of the EGCN is determined by $\hat{y} = Softmax(FC(Y))$. This model focuses on exploiting the impact of news content on the propagation procedure of rumors. However, the EGCN requires data on the entire conversation regarding the posted news. In some cases, if the posted news does not obtain response opinions when it is spread, its accuracy can be reduced.

Multidepth GCNs introduced by Hu et al. [190] combine the similarity of news to distinguish them as fake or real via degrees of differences. This method can solve the significant challenge of fake news detection, which is automatic fake news detection for short news items with limited information, for example, headlines. Instead of stacking the GCN layer to merge information over a long distance, the authors computed the different distance proximity matrices to describe the relationship between nodes and explicitly protect the multigranularity information, thus improving the node representation process with the diversity information. Therefore, it performed k-step proximity to create different depth proximity matrices before feeding to the GCN. For the step k-th proximity, the output is defined as $z_k = \hat{A}^k ReLU(\hat{A}^k XW_k^{(0)})W_k^{(1)}, k = 1, 2, 3, \ldots$, where node feature matrix X contains word embeddings and representation of credit history. $\hat{A}^k$ is the k-th proximity matrix as $\hat{A}^k = \underbrace{\hat{A} \times \hat{A} \times \cdots \times \hat{A}}_{k}$,

where $A_{ij} = 1\ if\ i, j\ have\ the\ same\ job-title, otherwise\ A_{ij} = 0$. Then, the multi-depth information is aggregated to create final representation $P_j = \sum_{i=1}^{m} \alpha_i z_i$ using an attention mechanism as $\alpha_i = \frac{exp(u_i)}{\sum_{l=1}^{m} exp(u_l)}$, where $u_i = tanh(W_i z_i + b_i)$.

Nguyen et al. [155] introduced two methods, textual-based and graph-based, for false news detection regarding COVID-19 and 5G conspiracy. For the first method, the authors detected fake news using a combination of a pretrained BERT model and a multilayer perceptron model to capture both the textual features and metadata of tweets. For the second method, the author used a GCN with nine extracted features at each node, such as page rank, hub, and authority, for content-based fake news detection. After implementing the two methods, the authors proved that the performance of the first approach is better than that of the second. Thus, metadata play a significant role in fake news detection. Therefore, improving the efficiency of fake news detection by extracting metadata features for GCN should be considered in the future. Regarding COVID-19 and 5G conspiracies, Pehlivan [194] introduced structure-based fake news detection to evaluate the performance of existing models. Unlike other methods, the author used only the temporal features of networks without considering textual features. Two state-of-the-art models were selected to evaluate the GCN and DGCN [205]. Additionally, the authors used their temporal features to test the multivariate long short-term memory fully convolutional network method [206]. Node feature matrices of the GCN and DGCN are created based on the following values: degree centrality, closeness centrality, betweenness centrality, load centrality, harmonic centrality, #cliques, clustering coefficient, square clustering coefficient, and average neighbor degree. The node feature matrix of the multivariate long short-term memory fully convolutional network is created with the average clustering coefficient, #graph cliques, #connected components, local efficiency, #isolates, and normalized time distance to the source tweet. Li et al. [191] introduced a propagation-based method for determining political perspective by focusing on the social contextual information of news. In this study, the GCN is used with an adjacency matrix to capture and represent the social context via feature extractions: sharing actions, following actions regarding political news, and a node feature matrix are used to capture the hidden content of news via word embeddings. Meyers et al. [153] showed the significant role of propagation features in fake news detection models. The authors first constructed a propagation graph to present important information and then used a random forest classifier to train the graph and create node embeddings. Finally, the GCN model was used to predict the authenticity of tweets. Unlike other propagation graphs, the authors constructed the following graph: Let $G = \{V, E\}$ denote the propagation graph, where V is a set of nodes including tweet nodes and retweet nodes and E is a set of edges connected between a tweet node and its retweet node with a time weight. Thus, this propagation graph includes a set of subgraphs, where each subgraph includes a tweet node and its retweet nodes, and its depth never exceeds 1.

A social spammer detection model [201] was built with a combination of the GCN and Markov random field (MRF) models. First, the authors used convolutions on directed graphs to explicitly consider various neighbors. They then presented three influences of neighbors on a user's label (follow, follower, reciprocal) using a pairwise MRF. Significantly, the MRF is formulated as an RNN for multistep inference. Finally, MRF layers were stacked on top of the GCN layers and trained via an end-to-end process of the entire model. Unlike conventional GCNs, this model uses an improved forward propagation rule

$$Q = H^{(l+1)} = \sigma(D_i^{-1}A_iH^{(l)}W_i^{(l)} + D_o^{-1}A_oH^{(l)}W_o^{(l)}$$

$$+ \hat{D}_b^{-1/2} \hat{A}_b \hat{D}_b^{-1/2} H^{(l)} W_b^{(l)}) \tag{20}$$

where A_i, A_o, A_b are types of neighbors; $\hat{A}_b = A_b + I$; D_i, D_o, and $\hat{D}_b$ are degree matrices of A_i, A_o, and A_b, respectively; The node feature matrix $X = H^{(0)}$ is created based on BoW features. Then, the authors initialized the posterior probabilities of the MRF layer with the GCN output as

$$R = Softmax(logH^{(k)} - A_i Q \begin{bmatrix} -w & w' \\ -w & -w \end{bmatrix}$$
$$- A_o Q \begin{bmatrix} -w & -w \\ w' & -w \end{bmatrix} - A_b Q \begin{bmatrix} -w & w' \\ w' & -w \end{bmatrix}) \tag{21}$$

where $w, w' \geq 0$ are two learnable parameters to measure homophily and heterophily strength of MRF model. This method demonstrated the superiority of the combination of GCN and MRF layers. A multistep MRF layer is essential to convergence. However, the node feature matrix was created simply with the bag-of-words method. This limitation can be improved using state-of-the-art embedding models in the future.

A novel GCN framework, called FauxWard [149], is proposed for fauxtography detection by exploiting news characteristics, such as linguistic, semantic, and structural attributes. The authors modeled fauxtography detection as a classification problem and used GCNs to solve this problem. FauxWard is similar to traditional GCN models; however, unlike these models, it adds a cluster-based pooling layer between graph convolutional layers to learn the node representation more efficiently. The cluster-based pooling layer first assigns neighbor nodes into clusters based on the node vectors of the previous graph convolution layers and then learns a cluster representation as the input of the back graph convolution layer. It performs graph convolution by $\tilde{A}^{(k)} = C^{(k-1)\top} \tilde{A}^{(k-1)} C^{(k-1)}$, where $\tilde{A}^{(k)}$ is the updated adjacency matrix; $C^{(k)}$ is the clustering matrix obtained after the k-th graph convolution layer, such that $H^{(k)} = C^{(k-1)\top} \sigma(\tilde{A}^{(k-1)} H^{(k-1)} W^{(k-1)})$, where $H^{(0)} = X$ be a node feature matrix. Unlike conventional GCNs, this X is created by concatenating text content, such as linguistic, sentiment, endorsement, and image content, such as metadata.

Malhotra et al. [147] introduced a method of combining RoBERTa and BiLSTM ($TD = Bi(RoTa(tweet))$, where $RoTa$ indicates a RoBERTa model [207] and Bi indicates a BiLSTM model) and GCN methods ($GD = H^{(k)} = \sigma(\hat{A} H^{(k-1)} W^{(k)})$, where $H^{(0)} = X$ is a node feature matrix by concatenating eleven features, such as friend count, follower count, followee count, etc.) for rumor detection as $\hat{y} = Softmax(concat(TD, GD))$. This model is based on rumor characteristics, such as propagation and content. It exploits features regarding the structure, linguistics, and graphics of tweet news.

Vlad et al. [195] produced a multimodal multitask learning method based on two main components: meme identification and hate speech detection. The first combines GCN and an Italian BERT for text representation, whereas the second is an image representation method, which varies among different image-based structures. The image component employed VGG-16 with five CNN stacks [208] to represent images. The text component used two mechanisms to represent text, namely, Italian BERT attention and convolution. This model is multimodal because it considers features related to the text and image content simultaneously. Meanwhile, Monti et al. [3] introduced a geometric deep learning-based fake news detection method by constructing heterogeneous graph data to integrate information related to the news, such as user profile and interaction, network structure, propagation patterns, and content. Given a URL u with a set of tweets mentioned u, the authors constructed a graph $G_u = \{V, E\}$. V is a set of nodes corresponding to tweets and

their posters. E is a set of edges expressing one of four relations between two nodes: follow, followed, spreading, and spread. This graph has node feature matrix X and adjacency matrix A. X is created by characterizing user features, such as profiles, network structure, and tweet content. However, matrix A is defined as follows: if (node v_j spreads tweet of node v_i) or (node v_i spreads tweet of node v_j) or (node v_i follows node v_j) or (node v_j follows node v_i) then $A_{ij} = 1$ otherwise 0; Given matrices X and A, similar to traditional GCNs, the authors utilized a four-layer GCN: two convolutional layers for node representation and two fully connected layers to predict the news as fake or real. However, unlike some previous GCNs, in this proposal, one attention mechanism in the filters [178] and the mean pooling are used to decrease the feature vectors' dimension for each convolutional layer. SELU [209] is employed as a nonlinearity activation function for the entire network.

Li et al. [123] presented a GCN-based antispam method for large-scale advertisements named GAS. Unlike previous GCNs, in the GAS model, a combination graph is constructed by integrating the nodes and edges of the heterogeneous graph and a homogeneous graph to capture the local and global comment contexts. The GAS is defined in the following steps: (i) Graphs construction: The authors constructed two types of graphs named Xianyu graph and comment graph. The first graph was denoted by $G = \{U, I, E\}$ where U, I are sets of nodes representing users and their items, respectively, and E is a set of edges representing comments. An adjacency of this graph is created as follows: if user i makes comment e to item j, then $A_{ij}^X = 1$ other− wise 0. The second graph is constructed by connecting nodes expressing comments that have the similar meaning. That means if comment i has similar meaning with j, then $A_{ij}^C = 1$ otherwise 0. (ii) GCN on Xianyu graph: Let $h_e^{(l)}, h_{U(e)}^{(l)}$ and $h_{I(e)}^{(l)}$ be the l-th layer node embeddings of edge, user, and item, respectively, $z_e = h_e^{(l)} = \sigma(W_E^{(l)} \cdot concat(h_e^{(l-1)}, h_{U(e)}^{(l-1)}, h_{I(e)}^{(l-1)}))$ where $h_e^{(0)} = TN(w_0, w_1, \ldots, w_n)$ and $U(e), I(e)$ are user node and item node of edge e. Let $h_{N(u)}^{(l)}, h_{N(i)}^{(l)}$ are neighbor embeddings of node u, i. TN stands by TextCNN model [210]. w_k is the word vector of word k in tweet. Hence,

$$h_{N(u)}^{(l)} = \sigma(W_U^{(l)} \cdot att(h_u^{(l-1)}, concat(h_e^{(l-1)}, h_i^{(l-1)}))) \tag{22}$$

where $\forall e = (u, i) \in E(u)$ and

$$h_{N(i)}^{(l)} = \sigma(W_I^{(l)} \cdot att(h_i^{(l-1)}, concat(h_i^{(l-1)}, h_e^{(l-1)}))), \tag{23}$$

where $\forall e = (u, i) \in E(i)$, and $E(u)$ is the edge connected to u; att a stands of attention mechanism. From that, we have: $z_u = h_u^{(l)} = concat(W_U^{(l)} \cdot h_u^{(l)}, h_{N(u)}^{(l)})$ and $z_i = h_i^{(l)} = concat(W_I^{(l)} \cdot h_i^{(l)}, h_{N(i)}^{(l)})$. (iii) GCN on the comment graph: in this step, authors used the GCN model proposed in [211] to represent nodes on the comment graph into node embeddings as $p_e = GCN(X^C, A^C)$, where X^C is node feature matrix. (iv) GAS classifier: The output of GAS model is defined as $y = classifier(concat(z_i, z_u, z_e, p_e))$.

5.3. Detection approach based on AGNNs

Ren et al. [17] introduced a novel approach, called AA-HGNN, to model user and community relations as a heterogeneous information network (HIN) for content-based and context-based fake news detection. The primary technique used in AA-HGNN involves improving the node representation process by learning the heterogeneous information network. In this study, the AGNNs use two levels of an attention mechanism: the node learns the same neighbors' weights and then represents them by aggregating the neighbors' weights corresponding to each type-specific neighbor and a schema to learn the nodes' information, thus obtaining the optimal weight of the type-specific neighbor representations. Assume that we have a news HIN and a

news HIN schema, denoted by $G = \{V, E\}$ and $S_G = \{V_T, E_T\}$. Let $V = \{C \cup N \cup S\}$ with C (creators), N (news), S (subjects); and $E = \{E_{c,n} \cup E_{n,s}\}$. Let $V_T = \{\theta_n, \theta_c, \theta_s\}$ and $E_T = \{write, belongsto\}$ denotes types of nodes and types of links. Node-level attention is defined as $h'_{n_i} = M^{\theta_n} \cdot h_{n_i}$, where $n_i \in N$, h_{n_i} is the feature vector of node n_i. M^{θ_n} is the transformation matrix for type θ_n. Let $T \in \{C \cup N \cup S\}$, $t_j \in T$ belongs to type-neighbor θ_t and $t_j \in neighbor_{n_i}$. Let $e_{ij}^{\theta_t} = att(h'_{n_i}, h'_{t_j}; \theta_t)$ is the importance degree of node t_j for n_i, where att be a node-level attention mechanism with the attention weight coefficient as $\alpha_{ij}^{\theta_t} = Softmax_j(e_{ij}^{\theta_t})$ Then, the schema node is calculated by aggregating from the neighbor's features as $T_{n_i} = \sigma(\sum_{t_j \in neighbor_{n_i}} \alpha_{ij}^{\theta_t} \cdot h'_{t_j})$. Let $\omega_i^{\theta_t} = schema(WT_{n_i}, WN_{n_i})$ is the importance degree of schema node T_{n_i}, where $schema$ be a schema-level attention mechanism, N_{n_i} is a schema node corresponding to neighbors of node n_i. And the final fusion coefficient is calculated as $\beta_i^{\theta_t} = Softmax_t(\omega_i^{\theta_t})$. From that, we have a node representation as $r_{n_i} = \sum_{\theta_t \in V_T} \beta_i^{\theta_t} \cdot T_{n_i}$. AA-HGNN can still achieve excellent performance without using much-labeled data because it benefits from adversarial active learning. It can also be used for other actual tasks relating to heterogeneous graphs because of its high generalizability.

Benamira et al. [192] proposed content-based fake news detection methods for binary text classification tasks. The objective was a GNN-based semisupervised method to solve the problem of labeled data limitations. This method comprises the following steps: news embedding; news representation based on k nearest-neighbor graph inference; and news classification based on GNNs, such as AGNN [179] and GCN [164], which are conventional GNNs without improvements or updates.

5.4. Detection approach based on GAEs

Using the autoencoder special graph data, Kipf [170] used GAE to encode graphs to represent latent structure information in graphs. GAEs are used in various fields, such as recommendation systems [212] and link prediction [213], with reasonable performance. Recently, researchers have begun to apply GAEs for fake news detection. The previous studies for fake news detection models based on GAEs are summarized in Table 10.

Lin et al. [124] proposed a model to capture textual, propagation, and structural information from news for rumor detection. The model includes three parts: an encoder, a decoder, and a detector. The encoder uses a GCN to represent news text to learn information, such as text content and propagation. The decoder uses the representations of the encoder to learn the overall news structure. The detector also uses the representations of the encoder to predict whether events are rumors. The decoder and detector are simultaneously implemented. These parts are generally defined as follows: (i) Encoder component: Two layers of the GCN are used to enhance the learning ability:

$$H^{(1)} = GCN(X, A) = \hat{A}\sigma(\hat{A}XW^{(0)}W^{(1)}) \tag{24}$$

and

$$H^{(2)} = GCN(H^{(1)}, A) = \hat{A}\sigma(\hat{A}H^{(1)}W^{(1)}W^{(2)}) \tag{25}$$

where σ is ReLU function. X represents word vectors that are created by determining the TF-IDF values, and the adjacent matrix A is defined as follows: *if node v_i responds to node v_j, then $A_{ij} = 1$ otherwise 0*. Then, the GCN is used to learn a Gaussian distribution for variational GAE as $z = \mu + \epsilon\sigma$, where $\mu = GCN(H^{(1)}, A)$ and $log\sigma = GCN(H^{(1)}, A)$ (μ, σ, and ϵ are the mean, standard deviation, and standard sample of the Gaussian distribution, respectively). (ii) Decoder component: In this step, an inner product

($\odot$) is used to reconstruct the adjacency matrix as $\tilde{A} = \odot(ZZ^\top)$ where Z is the matrix of distributions z. (iii) Detector component: This step aims to represent the latent information and classify the news. It is defined as $S = MP(Z)$ where MP stands for the mean-pooling operator. Finally, the output layer of this model is defined as $\hat{y} = Softmax(SW + b)$ where W is the parameter matrix of the fully connected layer.

6. Discussion

6.1. Discussion on GNNs*-based methods

Previous studies for fake news detection models based on GNNs* are compared in Table 7.

We presented the main steps, advantages, and disadvantages of GNN*-based methods for fake news detection. Some of our assessments are as follows: Regarding the extracted features, [4] used only user-based features; [5] used features based on networks, users, and linguistics; and [193] used linguistic-based features (textual analysis). Meanwhile, [15] used features related to networks and linguistics. Regarding graph structure, [4,5,193] constructed a homogeneous graph. However, unlike [4,193] only one graph was constructed, and [5] created two subgraphs to represent news sources and news users. Meanwhile, [15] built a heterogeneous graph with two types of nodes and edges. However, although the graph structure of [15] is better than that of the other three models, [192] provides the best performance. This result may be because [5] can better extract meaningful features in fake news detection. Therefore, to develop new GNN*-based models in the future, more attention should be given to extracting excellent features and building good standard data instead of focusing on improving the graph structure.

6.2. Discussion on GCNs-based methods

Previous studies for fake news detection models based on GCNs are compared in Tables 8 and 9.

We presented the main steps, advantages, and disadvantages of GCN-based methods for fake news detection. In our assessments, methods such as [3,14,16,123,191], and [196], show the best efficiency, where two methods are used for fake news detection, two for rumor detection, and two for spam classification. Regarding the two papers in the first category, [3] was the first to apply GCNs for fake news detection. This method focuses on extracting user-based, network-based, and linguistic-based features to build propagation-based heterogeneous GCNs. The authors determined that this proposal can obtain a more promising result than content-based methods. Conversely, [14] is an enriched GCN with a dual coattention mechanism. This method uses user-based and linguistic-based features to construct homogeneous GCNs with a dual coattention mechanism. In our assessment, although [14] used dual coattention mechanisms, the efficiency was still lower than that in [3]. Noticeably, this result is attributable mainly to more features being extracted by [3] than by [14]. Additionally, the graph structure used in [3] was evaluated as better than the structure used in [14]. Moving forward, we hope to improve the performance of fake news detection methods by building dual coattention heterogeneous GCNs using user-based, network-based, and linguistic-based features simultaneously. For the two papers in the second category, both methods were built to detect rumors by propagation-based GCNs. The difference is that [16] constructed bidirectional GCNs to capture the rumor dispersion structure and rumor propagation patterns simultaneously. Meanwhile, [196] created unidirectional GCNs based on the information of multiorder neighbors to capture rumor sources.

Table 7

Comparison of GNNs* methods.

Method [Ref]	Critical idea	Loss function	Advantage	Disadvantage
Benamira et al. [192]	– Focus on analyzing the news content using semi-supervised learning – Binary classification model	Cross-entropy loss	– Can obtain high efficacy with limited labeled data	– Have not been evaluated with big data and multi-labeled data
Yi Han et al. [4]	– Propagation-based fake news detection method – Focus on continual learning and incremental training techniques – Use two techniques: EWC and GEM	Elastic Weight Consolidation loss	– Can improve the performance of conventional methods without using any text information – Can handle unseen data and new data	– Ignore the selection of features or the finding of "universal" features
FakeNews [193]	– Focus on two tasks aiming to analysis and detect fake news via news textual and news structure	NA	– Can use to multi-classification tasks – Binary classification task obtains significantly higher performance than the ternary ones	– Tasks implement separately corresponding to information textual and structure
SAFER [15]	– Contextual fake news detection method – Focus on combining information: content nature, user behaviors, and users social network	NA	– Improve the performance of the traditional GNNs – Can add more layers to identify more efficacy neighborhood	– Sensitive to bottleneck and over-smoothing problems
FANG [5]	– Contextual fake news detection method – Focus on representation quality by capturing sharing patterns and social structure	The total of unsupervised proximity loss, self-super-vised stance loss, and supervised fake news loss	– Improve representation quality – Can use to a limited training dataset – Can capture temporal patterns of fake news	– Stance detection and textual encode have not been jointly optimized – Sharing contents and hyperlinks become obsolete quickly

In our view, [16] can outperform [196] because rumor detection, rumor propagation, and dispersion are more critical than rumor sources. For the two papers in the last category, [123,191] also proposed similar methods for spam detection using social context-based GCNs. The different points are that [123] built a model integrating heterogeneous and homogeneous graphs to capture both local and global news contexts. In contrast, [191] constructed only one heterogeneous graph to capture the general news context. In our opinion, the model presented in [123] is more comprehensible, can be reimplemented, and yields slightly better results than the method in [191]. The reason for this result is that building each type of graph is suitable for the capture and integration of each type of context, which can capture the news context more comprehensively than constructing one graph for all contexts. Thus, when building fake news detection models based on GNNs, different graphs should be constructed to capture each specific type of information and then perform the fusion step. This approach promises to provide better performance than building one type of graph to capture all types of information. We maximally limit the construction of a general graph and then divide it into specific types because the breakdown of the graph can easily result in the loss of information on the relationship among edges.

6.3. Discussion on AGNNs- and GAEs-based methods

Previous studies for fake news detection models based on AGNNs and GAEs are compared in Table 10.

We presented the main steps, advantages, and disadvantages of the two methods in the AGNN and AGE categories for fake news detection. Evidently, [17] presented a more detailed fake news detection method than [192]. Additionally, the method in [17] was proposed after that in [192]; thus, it is better than [192]. For example, [192] constructed a homogeneous graph, whereas [17] created a heterogeneous graph. The heterogeneous graph was evaluated as superior to the homogeneous graph because it can capture more meaningful information. Therefore, it obtains better results than [192]. Meanwhile, the Lin et al. [124] method uses a conventional GCN variant to encode the latent representation of graphs. This method can capture the entire structural information efficiency. It can thus enrich traditional GCNs by adding two more components, namely, the decoder and detector. However, this study focused only on user-based and linguistic-based features, ignoring network-based features; therefore, the desired effect is not expected.

7. Challenges

7.1. Fake news detection challenges

Based on recent publications in the field of fake news detection, we summarized and classified challenges into five categories, where each category of challenge corresponds to one category of fake news detection. The details of each type of challenge are shown in Fig. 6. The following presents significant challenges that can become future directions in fake news detection.

Deepfake [214] is a hyperrealistic, digitally controlled video that shows people saying or doing things that never truly happened or composite documents generated based on artificial intelligence techniques. Given the sophistication of these counterfeiting techniques, determining the veracity of the public appearances or influencer claims is challenging owing to fabricated descriptions. Therefore, Deepfake currently poses a significant challenge to fake news detection.

The hacking of influencers' accounts to spread fake news or disinformation about a speech by celebrities themselves is also a unique phenomenon in fake news detection. However, this information will be quickly removed when the actual owner of these accounts discovers and corrects them. However, at the time of its spread, this information causes extremely harmful effects. Instantly detecting whether the posts of influencers are fake has thus become an important challenge.

News may be fake at one point in time and real at another. That is, the news is real or fake, depending on the time it is said and spread. Therefore, real-time fake news detection has not yet been thoroughly addressed.

Table 8
Comparison of GCN methods.

Method [Ref]	Critical idea	Loss function	Advantage	Disadvantage
Monti et al. [3]	– Focus on analyzing the news content, users, social structure and propagation using geometric deep learning – Binary classification task	Hinge loss	– Enable to integrate heterogeneous data – Obtain very high performance with big real data	– Only implement with the binary classification task
!GAS [123]	– Focus on capturing the global and local contexts of the news – Integrate graphs of homogeneous and heterogeneous – Binary classification task	Regression loss	– Can solve spam problems like adversarial actions and scalability – Obtain high performance with large-scale data – Can apply to online news	– Only implement with the binary classification task
Marion et al. [153]	– Focus on capturing the propagation features of the news using geometric deep learning – Binary classification task	NA	– Can apply to the non-URL-limited news	– Use a very broad definition – Apply to a single data source – Not high generalizability
GCAN [14]	– Focus on the relation of original tweet and retweet and the co-influence of the user interaction and original tweet – Use the dual co-attention mechanism – Binary classification task	Cross-entropy loss	– Can early detection – Can detect a tweet story as fake using only short-text tweet without needing user comments and network structure – Explainable of fake reasons	– Apply to a single data source – Not high generalization
Pehlivan et al. [194]	– Focus on the temporal features of the network structure without considering any textual features – Binary classification task	Cross-entropy loss	– Can apply to metadata	– Not promising performance – The data is split not reasonable for training, testing, validation
*Bi-GCN [16]	– Focus on analyzing features related to dispersion and propagation of the news – Construct a top-down graph to learn rumor spread and a bottom-up graph to capture rumor dispersion – Multi classification task	Cross-entropy loss	– Have an early detection mechanism – Can detect rumors in real-time – Obtain much higher performance than state-of-the-art methods	– Not high generalization
*GCNSI [196]	– Focus on identifying multiple sources of rumor without any knowledge related to news propagation – Improve the previous GCN models by modifying the enhanced node representations and loss function – Multi classification task	Sigmiod cross-entropy loss	– First model based on multiple sources of the rumor – Improve the performance of the state-of-the-art methods by about 15%	– Have to retrain the model if the graph structure is changed – Take quite much time to train and obtain suitable parameters
!GCNwithMRF [201]	– The first semi-supervised model focus on continuously integrating both methods of feature-based and propagation-based – Use the deep learning model with a refined MRF layer on directed graphs to enable the end-to-end training – Multi classification task	Cross-entropy loss	– Obtain superior effectiveness – Can ensure convergence	– Use simple BoW for features representation

Constructing benchmark datasets and determining the standard feature sets corresponding to each approach for fake news detection remain challenges.

Kai Shu et al. [215] constructed the first fake news detection methods by effectively extracting content, context, and propagation features simultaneously through four embedding components: news content, news users, user-news interactions, and publisher news relations. Then, these four embeddings were fed into a semisupervised classification method to learn a classification function for unlabeled news. In addition, this method can be used for fake news early detection. Ruchansky et al. [28] constructed a more accurate fake news prediction model by extracting the behavior of users, news, and the group behavior of fake news propagators. Then, three features were fed into the architecture, including three modules as follows: (i) use a recurrent neural network to capture the temporal activity of a user on given news via news and propagator behaviors; (ii) learn the news source via user behavior; and (iii) integrate the previous two modules for fake news detection with high accuracy. From this survey of literature, we see that the most effective approaches combine features regarding content, context, and propagation. Although these combination methods may have high complexity

regarding the algorithms used, the many extracted features, and high feature dimensions, they can simultaneously capture various aspects of fake news. Therefore, the most efficacious and least costly extraction of content, propagation patterns, and users' stance simultaneously is not only a promising solution but also a significant challenge for fake news detection.

7.2. Challenges related to graph neural networks

Based on studying the related literature, this section summarizes some challenges of GNN-based methods and then identifies possible future directions.

Most conventional GNNs utilize undirected graphs and edge weights as binary values (1 and 0) [216] unsuitable for many actual tasks. For example, in graph clustering, a graph partition is sought that satisfies two conditions: (i) the difference between the weights of edges among unlike groups is as low as possible; (ii) the difference in the weights of edges among similar groups is as high as possible. Here, if the weight of the edges is a binary value, the given problem cannot be solved using this graph. Therefore, future studies can construct graphs with the weights of edges as the actual values representing the relationship among the nodes as much as possible.

Table 9
Comparison of GCN methods (continued).

Method [Ref]	Critical idea	Loss function	Advantage	Disadvantage
*Malhotra et al. [147]	– Focus on combining features related to text and users – Use the geometric deep learning with RoBERTa-based embedding – Multi classification task	Cross-entropy loss	– Enable for more efficiently features extraction	– Evaluated by a limited dataset – Overfitting test error
!FauxWard [149]	– Focus on features related to the linguistic and semantic of the user comments and the user network structure – Use the geometric deep learning on a user comment network – Binary classification model	Cross-entropy loss	– Obtain a significant performance within a short time window	– Not directly analyze the content of the news containing an image-centric
KZWANG [156]	– Focus on depth integrating of contextual information and propagation structure – Use multi-head attention mechanism to create contextual representation without extracting any features – Multi classification model	Cross-entropy loss	– Have an early detection mechanism – Can create a better semantic-integrated representation – Improve performance significantly	– Random split for validation data and manual split for training, testing data
*GraphSAGE [125]	– Focus on determining patterns propagation-based characteristics and information related to the content, social network structure, and delay time – Use a graph embedding technique to integrate information of graph structure and node features – Multi classification model	Cross-entropy loss	– High generalization for unseen data – Reduce the detection error of state-of-the-art methods down to 10% – Efficiently integrate features related to the whole propagated post	– Can reduce the performance if not use full information of post (original and response) in the spread process
Bert-GCN Bert-VGCN [150]	– Focus on using features related to the content of news text – Improve the other GCN-based models using BERT-based embeddings – Multi classification model	NA	– Can create better word representations – Can improve the performance of the conventional GCN method significantly	– Not high generalization – No suitable augmentation data to improve features extraction and avoid overfitting
*Lotfi [204]	– Focus on information of text content, spread time, social network structure – Construct weighted graphs based on users interaction in conversations – Binary classification model	Cross-entropy loss	– Obtain high efficacy in early detection – Can improve the performance of the state-of-the-art methods significantly	– Strong depend on the full information of both original tweet and response tweets of conversations
*SAGNN [151]	– Focus on capturing the information of users interactions – Improve the conventional GCN models by adding one or more aggregation layer – Multi classification model	Cross-entropy loss	– Optimal capture of user's interactions – Capture better the different features between rumors and non-rumors	– Not high-performance generalization due to only comparing with one baseline method
*EGCN [154]	– Focus on fully extracting features related to text content and structure – Construct weighted graphs of source-replies relation for conversations – Binary classification model	NA	– Can obtain comparable performance or better than machine learning methods – Can use the information of the global and local structure simultaneously	– Not high generalization

Table 10
Comparison of AGNN and GAE methods.

Method [Ref]	Critical idea	Loss function	Advantage	Disadvantage
Benamira et al. [192]	– Focus on analyzing the news content using semi-supervised learning – Binary classification model	Cross-entropy loss	– Can obtain good efficacy with limited labeled data	– Have not been evaluated with big data and multi-labeled data
AA-HGNN [17]	– The first model using adversarial active learning for fake news detection – Improve the conventional GCN models by using a new hierarchical attention mechanism for node representation – Multi classification model	Cross-entropy loss	– Support early detection stage – Still obtain high performance with limited training data – Can extract information as text and structure simultaneously	– Not compare the efficacy with the context-based methods
Lin et al. [124]	– Focus on integrating the information related to text, propagation, and network structure – Include three parts: encoder, decoder, and detector – Multi classification model	The sum of Cross-entropy loss and KL divergence loss	– The first GAE-based rumor detection method – Can create better and high-level node representations – Obtain better efficacy than other the latest methods	– Low performance for the non-rumor class – Not high generalization for the performance

For NLP tasks, GNNs have not represented node features by capturing the context of a paragraph or an entire sentence. Alternatively, these methods have also overlooked the semantic relationships among phrases in the sentences. For example, for sentiment classification tasks, we have the sentence *"The smell of this milk tea is not very fragrant."* This sentence includes a fuzzy sentiment phrase, namely *"not very fragrant"*. Some approaches classify this sentence as expressing a positive sentiment because they only focus on *"fragrant"*, ignoring the role of both *"not"* and *"very"*, whereas other models determine the expression as
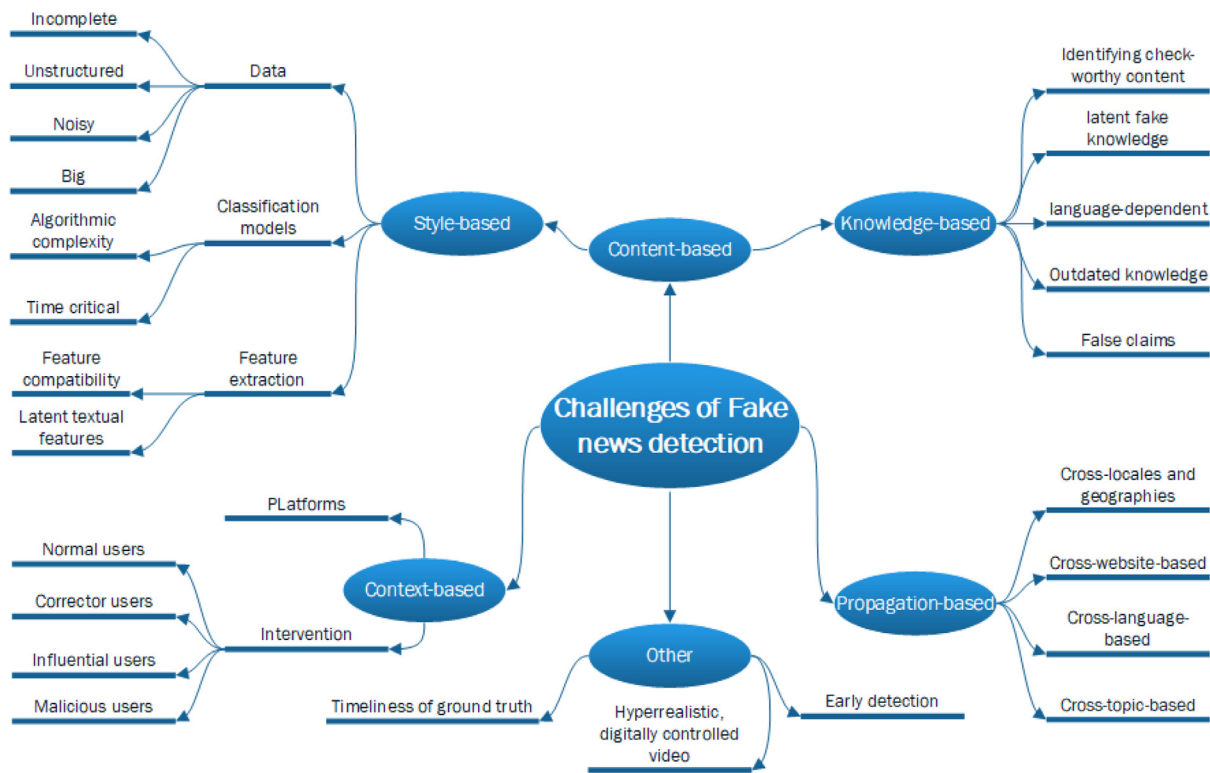
Fig. 6. List of challenges of fake news detection.

a negative sentiment because they ignore the impact of *"very"*. Therefore, future directions for improving GNN-based models should focus on determining node features based on sentence embeddings or significant phrase embeddings.

Capturing context, content, semantic relations, and sentiment knowledge simultaneously in sentences is essential for GNN-based NLP tasks. Meanwhile, only a few studies have incorporated some of these features by flexible GNNs to improve the efficiency of NLP tasks, including fake news detection. For instance, in [217], the authors extracted common sense knowledge and syntax via GNNs, whereas in [218], the authors constructed a single text-based GNN by representing document-word relations and word co-occurrence. To the best of our knowledge, no GNN has simultaneously considered all the content, contexts, common sense knowledge, and semantic relations. This task remains an exciting challenge for NLP task-based GNNs.

So far, GCNs have been limited to a few layers (two or three) owing to the vanishing gradient, which limits their real-world applications. For example, GCNs in [217,219,220] stopped at two layers because of the vanishing gradient error. Therefore, constructing deep fuzzy GCNs of syntactic, knowledge, and context by using the deep learning algorithm over the combination graph of the fuzzy syntactic graph, the fuzzy knowledge graph and the fuzzy context graph can solve the aforementioned limitations of previous methods for aspect-level sentiment analysis.

8. Conclusion and open issues

GNN-based fake news detection is relatively new. Thus, the number of published studies is limited. Although we did not implement methods presented in the 27 studies on the same datasets and did not evaluate their efficiency on the same comparison criteria, the 27 papers surveyed here show that this method initially obtained excellent results. Additionally, many challenges need to be addressed to achieve more comprehensive results, which we discussed at the end of the corresponding

sections. Nonetheless, given the 27 surveyed papers, promising results are expected in the future. By addressing these challenges, we hope to improve the effectiveness of GNN-based fake news detection. The following paragraphs analyze some challenges for GNN-based fake news detection and discuss future directions.

Benchmark data: Recently, some researchers have argued that when training a system, data affect system performance more than algorithms do [221]. However, in our assessment, we had no graph benchmark data for fake news detection in the graph learning community. Graph-based fake news detection benchmarks may present an opportunity and direction for future research.

Compatible hardware: With the rapid growth of Deepfake, graphs to represent these data will become more complex. However, the more scalable GNNs are, the higher the price and complexity of the algorithms is. Scientists often use graph clustering or graph sampling to solve this problem, ignoring the information loss of the graph using these techniques. Therefore, in the future, graph scalability may be solved by developing dedicated hardware that fits the graph structure. For example, GPUs were a considerable leap forward in lowering the price and increasing the speed of deep learning algorithms.

Fake news early detection: Early detection of fake news involves detecting fake news at an early stage before it is widely disseminated so that people can intervene early, prevent it early, and limit its harm. Early detection of fake news must be accomplished as soon as possible because the more widespread fake news is, the more likely it is that the authentication effect will take hold, meaning that people will be likely to believe the information. Currently, for fake news early detection, people often focus on analyzing the news content and the news context, which leads to three challenges. First, new news often appears to bring new knowledge, which has not been stored in the existing trust knowledge graph and cannot be updated immediately at the time the news appears. Second, fake news tends to be written with the same content but with different deceptive writing

styles and to appear simultaneously in many various fields. Finally, limited information related to news content, news context, news propagation, and latent information can adversely affect the performance of GNN-based detection methods.

Dynamic GNNs: Most graphs used in the current GNN-based fake news detection methods have a static structure that is difficult to update in real time. In contrast, news authenticity can change continuously over time. Therefore, it is necessary to construct dynamic graphs that are spatiotemporally capable of changing with real-time information.

Heterogeneous GNNs: The majority of current GNN-based fake news detection models construct homogeneous graphs. However, it is difficult to represent all the news texts, images, and videos simultaneously on these graphs. The use of heterogeneous graphs that contain different types of edges and nodes is thus a future research direction. New GNNs are suitable for heterogeneous graphs, which are required in the fake news detection field.

Multiplex GNNs: As analyzed in Section 7.2, most GNN-based fake news detection approaches have focused on independently using propagation, content, or context features for classification. Very few methods have used a combination of two of the three features. No approach uses a hybrid of propagation, content, and context simultaneously in one model. Therefore, this issue is also a current challenge in fake news detection. In the future, research should build GNN models by constructing multiplex graphs to represent news propagation, content, and context in the same structure.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article

Appendix. Description of datasets

- Fact-checking: The English dataset with 221 statements regarding society and politics was collected from online streaming.
- EMERGENT: The English dataset with 300 claims and 2595 associated article headlines regarding society and technology were collected from online streaming and Twitter.
- Benjamin Political News: The English dataset with 225 stories regarding politics was collected from online streaming from 2014 to 2015.
- Burfoot Satire News[7]: The English dataset with 4233 news articles regarding economy, politics, society, and technology was collected from online streaming.
- MisInfoText[8]: The English dataset with 1692 news articles regarding society was collected from online streaming.
- Ott et al.'s dataset: The English dataset with 800 reviews regarding tourism was collected from TripAdvisor social media.
- FNC-1: The English dataset with 49,972 articles regarding politics and society were collected from online streaming.
- Fake_or_real_news: The English dataset with 6337 articles regarding politics and society was collected from online streaming.

- TSHP-17: The English dataset with 33,063 articles regarding politics was collected from online streaming.
- QProp[9]: The English dataset with 51,294 articles regarding politics was collected from online streaming.
- NELA-GT-2018[10]: The English dataset with 713,000 articles regarding politics was collected from online streaming from February 2018 to November 2018.
- TW_info: The English dataset with 3472 articles regarding politics was collected from Twitter from January 2015 to April 2019.
- FCV-2018: The dataset, including 8 languages with 380 videos and 77,258 tweets regarding society, was collected from three social networks, namely YouTube, Facebook, and Twitter from April 2017 to July 2017.
- Verification Corpus: The dataset including 4 languages with 15,629 posts regarding 17 society events (hoaxes) was collected from Twitter from 2012 to 2015.
- CNN/Daily Mail: The English dataset with 287,000 articles regarding politics, society, crime, sport, business, technology, and health was collected from Twitter from April 2007 to April 2015.
- Tam et al.'s dataset: The English dataset with 1022 rumors and 4 million tweets regarding politics, science, technology, crime, fauxtography, and fraud/scam was collected from Twitter from May 2017 to November 2017.
- FakeHealth[11]: The English dataset with 500,000 tweets, 29,000 replies, 14,000 retweets, and 27,000 user profiles with timelines and friend lists regarding health were collected from Twitter.

References

[1] A. Bovet, H.A. Makse, Influence of fake news in Twitter during the 2016 US presidential election, Nature Commun. 10 (1) (2019) 1–14.
[2] H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election, J. Econ. Perspect. 31 (2) (2017) 211–236.
[3] F. Monti, F. Frasca, D. Eynard, D. Mannion, M.M. Bronstein, Fake news detection on social media using geometric deep learning, 2019, arXiv preprint arXiv:1902.06673.
[4] Y. Han, S. Karunasekera, C. Leckie, Graph neural networks with continual learning for fake news detection from social media, 2020, arXiv preprint arXiv:2007.03316.
[5] V.-H. Nguyen, K. Sugiyama, P. Nakov, M.-Y. Kan, Fang: Leveraging social context for fake news detection using graph representation, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 1165–1174.
[6] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, S.Y. Philip, A comprehensive survey on graph neural networks, IEEE Trans. Neural Netw. Learn. Syst. 32 (1) (2020) 4–24.
[7] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788.
[8] W. Shi, R. Rajkumar, Point-gnn: Graph neural network for 3d object detection in a point cloud, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1711–1719.
[9] G. Chen, Y. Tian, Y. Song, Joint aspect extraction and sentiment analysis with directional graph convolutional networks, in: Proceedings of the 28th International Conference on Computational Linguistics, 2020, pp. 272–279.
[10] C. Zhang, Q. Li, D. Song, Aspect-based sentiment classification with aspect-specific graph convolutional networks, 2019, arXiv preprint arXiv:1909.03477.
[11] J. Bastings, I. Titov, W. Aziz, D. Marcheggiani, K. Sima'an, Graph convolutional encoders for syntax-aware neural machine translation, 2017, arXiv preprint arXiv:1704.04675.
[12] D. Marcheggiani, J. Bastings, I. Titov, Exploiting semantics in neural machine translation with graph convolutional networks, 2018, arXiv preprint arXiv:1804.08313.

---

7 http://www.csse.unimelb.edu.au/ research/lt/resources/satire/
8 https://github.com/ sfu-discourse-lab/MisInfoText

9 http://proppy.qcri.org/about.html
10 https://doi.org/10.7910/DVN/ULHLCB
11 https://tinyurl.com/y36h42zu

[13] S. Zhang, H. Tong, J. Xu, R. Maciejewski, Graph convolutional networks: A comprehensive review, Comput. Soc. Networks 6 (1) (2019) 1–23.

[14] Y.-J. Lu, C.-T. Li, GCAN: Graph-aware co-attention networks for explainable fake news detection on social media, 2020, arXiv preprint arXiv: 2004.11648.

[15] S. Chandra, P. Mishra, H. Yannakoudakis, M. Nimishakavi, M. Saeidi, E. Shutova, Graph-based modeling of online communities for fake news detection, 2020, arXiv preprint arXiv:2008.06274.

[16] T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, J. Huang, Rumor detection on social media with bi-directional graph convolutional networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, (no. 01) 2020, pp. 549–556.

[17] Y. Ren, B. Wang, J. Zhang, Y. Chang, Adversarial active learning based heterogeneous graph neural network for fake news detection, in: 2020 IEEE International Conference on Data Mining, ICDM, IEEE, 2020, pp. 452–461.

[18] B. Collins, D.T. Hoang, N.T. Nguyen, D. Hwang, Trends in combating fake news on social media–A survey, J. Inf. Telecomedian. 5 (2) (2021) 247–266.

[19] T. Khan, A. Michalas, A. Akhunzada, Fake news outbreak 2021: Can we stop the viral spread? J. Netw. Comput. Appl. (2021) 103112.

[20] V. Klyuev, Fake news filtering: Semantic approaches, in: 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO, IEEE, 2018, pp. 9–15.

[21] R. Oshikawa, J. Qian, W.Y. Wang, A survey on natural language processing for fake news detection, 2018, arXiv preprint arXiv:1811.00770.

[22] K. Shu, A. Bhattacharjee, F. Alatawi, T.H. Nazer, K. Ding, M. Karami, H. Liu, Combating disinformation in a social media age, Wiley Interdiscipl. Rev.: Data Min. Knowl. Discov. 10 (6) (2020) e1385.

[23] F.B. Mahmud, M.M.S. Rayhan, M.H. Shuvo, I. Sadia, M.K. Morol, A comparative analysis of graph neural networks and commonly used machine learning algorithms on fake news detection, in: 2022 7th International Conference on Data Science and Machine Learning Applications, CDMA, IEEE, 2022, pp. 97–102.

[24] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: A data mining perspective, ACM SIGKDD Explor. Newsl. 19 (1) (2017) 22–36.

[25] D.M. Lazer, M.A. Baum, Y. Benkler, A.J. Berinsky, K.M. Greenhill, F. Menczer, M.J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, et al., The science of fake news, Science 359 (6380) (2018) 1094–1096.

[26] T. Quandt, L. Frischlich, S. Boberg, T. Schatto-Eckrodt, Fake news, in: The International Encyclopedia of Journalism Studies, John Wiley & Sons, Inc. Hoboken, NJ, USA, 2019, pp. 1–6.

[27] X. Zhou, R. Zafarani, Fake news: A survey of research, detection methods, and opportunities, 2, 2018, arXiv preprint arXiv:1812.00315.

[28] N. Ruchansky, S. Seo, Y. Liu, Csi: A hybrid deep model for fake news detection, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017, pp. 797–806.

[29] K. Shu, L. Cui, S. Wang, D. Lee, H. Liu, Defend: Explainable fake news detection, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 395–405.

[30] M. Cinelli, G.D.F. Morales, A. Galeazzi, W. Quattrociocchi, M. Starnini, The echo chamber effect on social media, Proc. Natl. Acad. Sci. 118 (9) (2021).

[31] C. Paul, M. Matthews, The Russian "firehose of falsehood" propaganda model, Rand Corporation 2 (7) (2016) 1–10.

[32] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H.E. Stanley, W. Quattrociocchi, Echo chambers in the age of misinformation, 2015, arXiv preprint arXiv:1509.00189.

[33] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H.E. Stanley, W. Quattrociocchi, The spreading of misinformation online, Proc. Natl. Acad. Sci. 113 (3) (2016) 554–559.

[34] M. Del Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, W. Quattrociocchi, Echo chambers: Emotional contagion and group polarization on Facebook, Sci. Rep. 6 (1) (2016) 1–12.

[35] J.L. Egelhofer, S. Lecheler, Fake news as a two-dimensional phenomenon: A framework and research agenda, Ann. Int. Commun. Assoc. 43 (2) (2019) 97–116.

[36] V. Bakir, A. McStay, Fake news and the economy of emotions: Problems, causes, solutions, Digit. Journalism 6 (2) (2018) 154–175.

[37] B. Franklin, B. McNair, Fake News: Falsehood, Fabrication and Fantasy in Journalism, Routledge, 2017.

[38] E.C. Tandoc Jr., Z.W. Lim, R. Ling, Defining "fake news" a typology of scholarly definitions, Digit. Journalism 6 (2) (2018) 137–153.

[39] C. Wardle, Fake news. It's complicated, First Draft 16 (2017) 1–11.

[40] K. Stahl, Fake news detection in social media, Calif. State Univ. Stanislaus 6 (2018) 4–15.

[41] N.R. Hanson, A note on statements of fact, Analysis 13 (1) (1952) 24.

[42] F. Pierri, S. Ceri, False news on social media: A data-driven survey, ACM Sigmod Record 48 (2) (2019) 18–27.

[43] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, Science 359 (6380) (2018) 1146–1151.

[44] N. Kshetri, J. Voas, The economics of "fake news", IT Prof. 19 (6) (2017) 8–12.

[45] E.J. Fox, S.J. Hoch, Cherry-picking, J. Mark. 69 (1) (2005) 46–62.

[46] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, R. Procter, Detection and resolution of rumours in social media: A survey, ACM Comput. Surv. 51 (2) (2018) 1–36.

[47] F. Allen, D. Gale, Stock-price manipulation, Rev. Financ. Stud. 5 (3) (1992) 503–529.

[48] V.L. Rubin, Y. Chen, N.K. Conroy, Deception detection for news: Three types of fakes, Proc. Assoc. Inf. Sci. Technol. 52 (1) (2015) 1–4.

[49] Y. Chen, N.J. Conroy, V.L. Rubin, Misleading online content: Recognizing clickbait as" false news", in: Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, 2015, pp. 15–19.

[50] B. Hofmann, Fake facts and alternative truths in medical research, BMC Med. Ethics 19 (1) (2018) 1–5.

[51] M. Gentzkow, J.M. Shapiro, D.F. Stone, Media bias in the marketplace: Theory, in: Handbook of Media Economics, vol. 1, Elsevier, 2015, pp. 623–645.

[52] A. D'Ulizia, M.C. Caschera, F. Ferri, P. Grifoni, Fake news detection: A survey of evaluation datasets, PeerJ Comput. Sci. 7 (2021) e518.

[53] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, Y. Liu, Combating fake news: A survey on identification and mitigation techniques, ACM Trans. Intell. Syst. Technol. 10 (3) (2019) 1–42.

[54] H. Ahmed, I. Traore, S. Saad, Detection of online fake news using n-gram analysis and machine learning techniques, in: International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, Springer, 2017, pp. 127–138.

[55] H. Ahmed, I. Traore, S. Saad, Detecting opinion spams and fake news using text classification, Secur. Privacy 1 (1) (2018) e9.

[56] K. Nakamura, S. Levy, W.Y. Wang, R/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection, 2019, arXiv preprint arXiv:1911.03854.

[57] W.Y. Wang, " Liar, liar pants on fire": A new benchmark dataset for fake news detection, 2017, arXiv preprint arXiv:1705.00648.

[58] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media, Big Data 8 (3) (2020) 171–188.

[59] J. Golbeck, M. Mauriello, B. Auxier, K.H. Bhanushali, C. Bonk, M.A. Bouzaghrane, C. Buntain, P. Chanduka, P. Cheakalos, J.B. Everett, et al., Fake news vs satire: A dataset and analysis, in: Proceedings of the 10th ACM Conference on Web Science, 2018, pp. 17–21.

[60] F.K.A. Salem, R. Al Feel, S. Elbassuoni, M. Jaber, M. Farah, Fa-kes: A fake news dataset around the Syrian war, in: Proceedings of the International AAAI Conference on Web and Social Media, Vol. 13, 2019, pp. 573–582.

[61] A. Pathak, R.K. Srihari, BREAKING! Presenting fake news corpus for automated fact checking, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, 2019, pp. 357–362.

[62] J. Thorne, A. Vlachos, C. Christodoulopoulos, A. Mittal, Fever: A large-scale dataset for fact extraction and verification, 2018, arXiv preprint arXiv:1803.05355.

[63] G.K. Shahi, D. Nandini, FakeCovid–A multilingual cross-domain fact check news dataset for COVID-19, 2020, arXiv preprint arXiv:2006.11343.

[64] T. Mitra, E. Gilbert, Credbank: A large-scale social media corpus with associated credibility annotations, in: Ninth International AAAI Conference on Web and Social Media, 2015.

[65] J. Leskovec, L. Backstrom, J. Kleinberg, Meme-tracking and the dynamics of the news cycle, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009, pp. 497–506.

[66] G.C. Santia, J.R. Williams, Buzzface: A news veracity dataset with Facebook user commentary and egos, in: Twelfth International AAAI Conference on Web and Social Media, 2018.

[67] E. Tacchini, G. Ballarin, M.L. Della Vedova, S. Moret, L. de Alfaro, Some like it hoax: Automated fake news detection in social networks, 2017, arXiv preprint arXiv:1704.07506.

[68] M. De Domenico, A. Lima, P. Mougel, M. Musolesi, The anatomy of a scientific rumor, Sci. Rep. 3 (2013) 2980.

[69] T. Khan, A. Michalas, Trust and believe-should we? Evaluating the trustworthiness of Twitter users, in: 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom, IEEE, 2020, pp. 1791–1800.

[70] R. Barbado, O. Araque, C.A. Iglesias, A framework for fake review detection in online consumer electronics retailers, Inf. Process. Manage. 56 (4) (2019) 1234–1244.

[71] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, R. Procter, Detection and resolution of rumours in social media: A survey, 2017, Preprint, arXiv 1704.

[72] A. Vlachos, S. Riedel, Fact checking: Task definition and dataset construction, in: Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science, 2014, pp. 18–22.

[73] W. Ferreira, A. Vlachos, Emergent: A novel data-set for stance classification, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016, pp. 1163–1168.

[74] B. Horne, S. Adali, This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 11, (no. 1) 2017.

[75] C. Burfoot, T. Baldwin, Automatic satire detection: Are you having a laugh? in: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, 2009, pp. 161–164.

[76] F. Torabi Asr, M. Taboada, Big data and quality data for fake news and misinformation detection, Big Data Soc. 6 (1) (2019) 2053951719843310.

[77] M. Ott, Y. Choi, C. Cardie, J.T. Hancock, Finding deceptive opinion spam by any stretch of the imagination, 2011, arXiv preprint arXiv:1107.4557.

[78] B. Riedel, I. Augenstein, G.P. Spithourakis, S. Riedel, A simple but tough-to-beat baseline for the fake news challenge stance detection task, 2017, arXiv preprint arXiv:1707.03264.

[79] P.S. Dutta, M. Das, S. Biswas, M. Bora, S.S. Saikia, Fake news prediction: A survey, Int. J. Sci. Eng. Sci. 3 (3) (2019) 1–3.

[80] H. Rashkin, E. Choi, J.Y. Jang, S. Volkova, Y. Choi, Truth of varying shades: Analyzing language in fake news and political fact-checking, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 2931–2937.

[81] A. Barrón-Cedeno, I. Jaradat, G. Da San Martino, P. Nakov, Proppy: Organizing the news based on their propagandistic content, Inf. Process. Manage. 56 (5) (2019) 1849–1864.

[82] J. Nørregaard, B.D. Horne, S. Adalı, Nela-gt-2018: A large multi-labelled news dataset for the study of misinformation in news articles, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 13, 2019, pp. 630–638.

[83] Y. Jang, C.-H. Park, Y.-S. Seo, Fake news analysis modeling using quote retweet, Electronics 8 (12) (2019) 1377.

[84] O. Papadopoulou, M. Zampoglou, S. Papadopoulos, I. Kompatsiaris, A corpus of debunked and verified user-generated videos, Online Inf. Rev. (2019).

[85] C. Boididou, S. Papadopoulos, M. Zampoglou, L. Apostolidis, O. Papadopoulou, Y. Kompatsiaris, Detection and visualization of misleading content on Twitter, Int. J. Multimedia Inf. Retr. 7 (1) (2018) 71–86.

[86] H. Jwa, D. Oh, K. Park, J.M. Kang, H. Lim, Exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (bert), Appl. Sci. 9 (19) (2019) 4062.

[87] N.T. Tam, M. Weidlich, B. Zheng, H. Yin, N.Q.V. Hung, B. Stantic, From anomaly detection to rumour detection using data streams of social platforms, Proc. VLDB Endow. 12 (9) (2019) 1016–1029.

[88] E. Dai, Y. Sun, S. Wang, Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 14, 2020, pp. 853–862.

[89] J.A. Hill, S. Agewall, A. Baranchuk, G.W. Booz, J.S. Borer, P.G. Camici, P.-S. Chen, A.F. Dominiczak, Ç. Erol, C.L. Grines, et al., Medical misinformation: Vet the message! 2019.

[90] M. Potthast, S. Köpsel, B. Stein, M. Hagen, Clickbait detection, in: European Conference on Information Retrieval, Springer, 2016, pp. 810–817.

[91] V.L. Rubin, N. Conroy, Y. Chen, S. Cornwell, Fake news or truth? Using satirical cues to detect potentially misleading news, in: Proceedings of the Second Workshop on Computational Approaches To Deception Detection, 2016, pp. 7–17.

[92] S. Vosoughi, M.N. Mohsenvand, D. Roy, Rumor gauge: Predicting the veracity of rumors on Twitter, ACM Trans. Knowl. Discov. Data (TKDD) 11 (4) (2017) 1–36.

[93] Y. Qin, D. Wurzer, V. Lavrenko, C. Tang, Spotting rumors via novelty detection, 2016, arXiv preprint arXiv:1611.06322.

[94] V. Qazvinian, E. Rosengren, D. Radev, Q. Mei, Rumor has it: Identifying misinformation in microblogs, in: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 2011, pp. 1589–1599.

[95] A. Zubiaga, M. Liakata, R. Procter, Learning reporting dynamics during breaking news for rumour detection in social media, 2016, arXiv preprint arXiv:1610.07363.

[96] E.J. Briscoe, D.S. Appling, H. Hayes, Cues to deception in social media communications, in: 2014 47th Hawaii International Conference on System Sciences, IEEE, 2014, pp. 1435–1443.

[97] A.Y. Chua, S. Banerjee, Linguistic predictors of rumor veracity on the internet, in: Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1, 2016, pp. 387–391.

[98] J. Ito, J. Song, H. Toda, Y. Koike, S. Oyama, Assessment of tweet credibility with LDA features, in: Proceedings of the 24th International Conference on World Wide Web, 2015, pp. 953–958.

[99] J. Ma, W. Gao, P. Mitra, S. Kwon, B.J. Jansen, K.-F. Wong, M. Cha, Detecting rumors from microblogs with recurrent neural networks, 2016.

[100] Z. Jin, J. Cao, Y. Zhang, J. Zhou, Q. Tian, Novel visual and statistical image features for microblogs news verification, IEEE Trans. Multimed. 19 (3) (2016) 598–608.

[101] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, B. Stein, A stylometric inquiry into hyperpartisan and fake news, 2017, arXiv preprint arXiv:1702.05638.

[102] C. Castillo, M. Mendoza, B. Poblete, Information credibility on twitter, in: Proceedings of the 20th International Conference on World Wide Web, 2011, pp. 675–684.

[103] S. Hamidian, M.T. Diab, Rumor detection and classification for Twitter data, 2019, arXiv preprint arXiv:1912.08926.

[104] X. Hu, J. Tang, H. Liu, Online social spammer detection, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 28, (no. 1) 2014.

[105] S. Kwon, M. Cha, K. Jung, W. Chen, Y. Wang, Prominent features of rumor propagation in online social media, in: 2013 IEEE 13th International Conference on Data Mining, IEEE, 2013, pp. 1103–1108.

[106] Z. Jin, J. Cao, Y. Zhang, J. Luo, News verification by exploiting conflicting social viewpoints in microblogs, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30, (no. 1) 2016.

[107] A. Bondielli, F. Marcelloni, A survey on fake news and rumour detection techniques, Inform. Sci. 497 (2019) 38–55.

[108] A. Silva, L. Luo, S. Karunasekera, C. Leckie, Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, (no. 1) 2021, pp. 557–565.

[109] T.R. Hoens, R. Polikar, N.V. Chawla, Learning from streaming data with concept drift and imbalance: An overview, Progress Artif. Intell. 1 (1) (2012) 89–101.

[110] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018, arXiv preprint arXiv:1810.04805.

[111] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, 2018, CoRR abs/1802.05365, arXiv preprint arXiv:1802.05365, 1802.

[112] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, A. Joulin, Advances in pre-training distributed word representations, 2017, arXiv preprint arXiv:1712.09405.

[113] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, T. Mikolov, Learning word vectors for 157 languages, 2018, arXiv preprint arXiv:1802.06893.

[114] J. Pennington, R. Socher, C.D. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2014, pp. 1532–1543.

[115] B. Koloski, T.S. Perdih, M. Robnik-Šikonja, S. Pollak, B. Škrlj, Knowledge graph informed fake news classification via heterogeneous representation ensembles, Neurocomputing (2022).

[116] Z. Sun, Z.-H. Deng, J.-Y. Nie, J. Tang, Rotate: Knowledge graph embedding by relational rotation in complex space, 2019, arXiv preprint arXiv:1902.10197.

[117] S. Zhang, Y. Tay, L. Yao, Q. Liu, Quaternion knowledge graph embeddings, Adv. Neural Inf. Process. Syst. 32 (2019).

[118] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, G. Bouchard, Complex embeddings for simple link prediction, in: International Conference on Machine Learning, PMLR, 2016, pp. 2071–2080.

[119] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, Automatic detection of fake news, 2017, arXiv preprint arXiv:1708.07104.

[120] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, 2014, arXiv preprint arXiv:1406.1078.

[121] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[122] A. Glazkova, M. Glazkov, T. Trifonov, g2tmn at Constraint @Aaai2021: Exploiting CT-BERT and Ensembling Learning for COVID-19 Fake News Detection, Springer, 2021, pp. 116–127.

[123] A. Li, Z. Qin, R. Liu, Y. Yang, D. Li, Spam review detection with graph convolutional networks, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 2703–2711.

[124] H. Lin, X. Zhang, X. Fu, A graph convolutional encoder and decoder model for rumor detection, in: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA, IEEE, 2020, pp. 300–306.

[125] D.T. Vu, J.J. Jung, Rumor detection by propagation embedding based on graph convolutional network, Int. J. Comput. Intell. Syst. 14 (1) (2021) 1053–1065.

[126] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.

[127] M. Nickel, K. Murphy, V. Tresp, E. Gabrilovich, A review of relational machine learning for knowledge graphs, Proc. IEEE 104 (1) (2015) 11–33.

[128] A.L. Ginsca, A. Popescu, M. Lupu, Credibility in information retrieval, Found. Trends Inf. Retr. 9 (5) (2015) 355–475.

[129] O. Etzioni, M. Banko, S. Soderland, D.S. Weld, Open information extraction from the web, Commun. ACM 51 (12) (2008) 68–74.

[130] A. Magdy, N. Wanas, Web-based statistical fact checking of textual documents, in: Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents, 2010, pp. 103–110.

[131] L. De Alfaro, V. Polychronopoulos, M. Shavlovsky, Reliable aggregation of Boolean crowdsourced tasks, in: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, vol. 3, (no. 1) 2015.

[132] Y. Chen, N.K. Conroy, V.L. Rubin, News in an online world: The need for an "automatic crap detector", Proc. Assoc. Inf. Sci. Technol. 52 (1) (2015) 1–4.

[133] V.L. Rubin, N.J. Conroy, Y. Chen, Towards news verification: Deception detection methods for news discourse, in: Hawaii International Conference on System Sciences, 2015, pp. 5–8.

[134] Z. Wei, J. Chen, W. Gao, B. Li, L. Zhou, Y. He, K.-F. Wong, An empirical study on uncertainty identification in social media context, in: Social Media Content Analysis: Natural Language Processing and beyond, World Scientific, 2018, pp. 79–88.

[135] B. Shi, T. Weninger, Fact checking in heterogeneous information networks, in: Proceedings of the 25th International Conference Companion on World Wide Web, 2016, pp. 101–102.

[136] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, Fakenewsnet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media, 2018, arXiv preprint arXiv:1809.01286.

[137] N. Sitaula, C.K. Mohan, J. Grygiel, X. Zhou, R. Zafarani, Credibility-based fake news detection, in: Disinformation, Misinformation, and Fake News in Social Media, Springer, 2020, pp. 163–182.

[138] D. Mocanu, L. Rossi, Q. Zhang, M. Karsai, W. Quattrociocchi, Collective attention in the age of (mis) information, Comput. Hum. Behav. 51 (2015) 1198–1204.

[139] M. Tambuscio, G. Ruffo, A. Flammini, F. Menczer, Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks, in: Proceedings of the 24th International Conference on World Wide Web, 2015, pp. 977–982.

[140] J. Ma, W. Gao, K.-F. Wong, Detect Rumors in Microblog Posts Using Propagation Structure Via Kernel Learning, Association for Computational Linguistics, 2017.

[141] L. Wu, H. Liu, Tracing fake-news footprints: Characterizing social media messages by how they propagate, in: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018, pp. 637–645.

[142] Y. Liu, Y.-F.B. Wu, Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[143] K. Shu, D. Mahudeswaran, S. Wang, H. Liu, Hierarchical propagation networks for fake news detection: Investigation and exploitation, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 14, 2020, pp. 626–637.

[144] K. Wu, S. Yang, K.Q. Zhu, False rumors detection on sina weibo by propagation structures, in: 2015 IEEE 31st International Conference on Data Engineering, IEEE, 2015, pp. 651–662.

[145] X. Zhou, R. Zafarani, Network-based fake news detection: A pattern-driven approach, ACM SIGKDD Explor. Newsl. 21 (2) (2019) 48–60.

[146] S. Alonso-Bartolome, I. Segura-Bedmar, Multimodal fake news detection, 2021, arXiv preprint arXiv:2112.04831.

[147] B. Malhotra, D.K. Vishwakarma, Classification of propagation path and tweets for rumor detection using graphical convolutional networks and transformer based encodings, in: 2020 IEEE Sixth International Conference on Multimedia Big Data, BigMM, IEEE, 2020, pp. 183–190.

[148] X. Zhou, J. Wu, R. Zafarani, SAFE: Similarity-aware multi-modal fake news detection, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2020, pp. 354–367.

[149] L. Shang, Y. Zhang, D. Zhang, D. Wang, Fauxward: A graph neural network approach to fauxtography detection using social media comments, Soc. Netw. Anal. Min. 10 (1) (2020) 1–16.

[150] A. Paraschiv, G.-E. Zaharia, D.-C. Cercel, M. Dascalu, Graph convolutional networks applied to fakenews: Corona virus and 5G conspiracy, Univ. Politeh. Bucharest Sci. Bull. Ser. C-Electr. Eng. Comput. Sci. (2021) 71–82.

[151] L. Zhang, J. Li, B. Zhou, Y. Jia, Rumor detection based on SAGNN: Simplified aggregation graph neural networks, Mach. Learn. Knowl. Extr. 3 (1) (2021) 84–94.

[152] J. Ma, W. Gao, K.-F. Wong, Rumor Detection on Twitter with Tree-Structured Recursive Neural Networks, Association for Computational Linguistics, 2018.

[153] M. Meyers, G. Weiss, G. Spanakis, Fake news detection on Twitter using propagation structures, in: Multidisciplinary International Symposium on Disinformation in Open Online Media, Springer, 2020, pp. 138–158.

[154] N. Bai, F. Meng, X. Rui, Z. Wang, Rumour detection based on graph convolutional neural net, IEEE Access 9 (2021) 21686–21693.

[155] N. Tuan, P. Minh, FakeNews detection using pre-trained language models and graph convolutional networks, in: Multimedia Evaluation Benchmark Workshop 2020, MediaEval 2020, 2020.

[156] Z. Ke, Z. Li, C. Zhou, J. Sheng, W. Silamu, Q. Guo, Rumor detection on social media via fused semantic information and a propagation heterogeneous graph, Symmetry 12 (11) (2020) 1806.

[157] B.D. Horne, J. Nørregaard, S. Adalı, Different spirals of sameness: A study of content sharing in mainstream and alternative media, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 13, 2019, pp. 257–266.

[158] G. Bachi, M. Coscia, A. Monreale, F. Giannotti, Classifying trust/distrust relationships in online social networks, in: 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing, IEEE, 2012, pp. 552–557.

[159] G. Wang, X. Zhang, S. Tang, H. Zheng, B.Y. Zhao, Unsupervised clickstream clustering for user behavior analysis, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016, pp. 225–236.

[160] W.L. Hamilton, Graph representation learning, Synthesis Lect. Artif. Intell. Mach. Learn. 14 (3) (2020) 1–159.

[161] J. Bruna, W. Zaremba, A. Szlam, Y. LeCun, Spectral networks and locally connected networks on graphs, 2013, arXiv preprint arXiv:1312.6203.

[162] M. Niepert, M. Ahmed, K. Kutzkov, Learning convolutional neural networks for graphs, in: International Conference on Machine Learning, PMLR, 2016, pp. 2014–2023.

[163] R. Ying, J. You, C. Morris, X. Ren, W.L. Hamilton, J. Leskovec, Hierarchical graph representation learning with differentiable pooling, 2018, arXiv preprint arXiv:1806.08804.

[164] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2016, arXiv preprint arXiv:1609.02907.

[165] X. Jiang, R. Zhu, S. Li, P. Ji, Co-embedding of nodes and edges with graph neural networks, IEEE Trans. Pattern Anal. Mach. Intell. (2020).

[166] A. Sperduti, A. Starita, Supervised neural networks for the classification of structures, IEEE Trans. Neural Netw. 8 (3) (1997) 714–735.

[167] M. Gori, G. Monfardini, F. Scarselli, A new model for learning in graph domains, in: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005, Vol. 2, IEEE, 2005, pp. 729–734.

[168] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, IEEE Trans. Neural Netw. 20 (1) (2008) 61–80.

[169] C. Gallicchio, A. Micheli, Graph echo state networks, in: The 2010 International Joint Conference on Neural Networks, IJCNN, IEEE, 2010, pp. 1–8.

[170] T.N. Kipf, M. Welling, Variational graph auto-encoders, 2016, arXiv preprint arXiv:1611.07308.

[171] Y. Wang, B. Xu, M. Kwak, X. Zeng, A simple training strategy for graph autoencoder, in: Proceedings of the 2020 12th International Conference on Machine Learning and Computing, 2020, pp. 341–345.

[172] Y. Li, R. Yu, C. Shahabi, Y. Liu, Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, 2017, arXiv preprint arXiv:1707.01926.

[173] Y. Seo, M. Defferrard, P. Vandergheynst, X. Bresson, Structured sequence modeling with graph convolutional recurrent networks, in: International Conference on Neural Information Processing, Springer, 2018, pp. 362–373.

[174] J. Zhang, X. Shi, J. Xie, H. Ma, I. King, D.-Y. Yeung, Gaan: Gated attention networks for learning on large and spatiotemporal graphs, 2018, arXiv preprint arXiv:1803.07294.

[175] Z. Wu, S. Pan, G. Long, J. Jiang, C. Zhang, Graph wavenet for deep spatial-temporal graph modeling, 2019, arXiv preprint arXiv:1906.00121.

[176] S. Yan, Y. Xiong, D. Lin, Spatial temporal graph convolutional networks for skeleton-based action recognition, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[177] B. Yu, H. Yin, Z. Zhu, Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting, 2017, arXiv preprint arXiv:1709.04875.

[178] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, 2017, arXiv preprint arXiv:1710.10903.

[179] K.K. Thekumparampil, C. Wang, S. Oh, L.-J. Li, Attention-based graph neural network for semi-supervised learning, 2018, arXiv preprint arXiv:1803.03735.

[180] J.B. Lee, R. Rossi, X. Kong, Graph classification using structural attention, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 1666–1674.

[181] A. Liberati, D.G. Altman, J. Tetzlaff, C. Mulrow, P.C. Gøtzsche, J.P. Ioannidis, M. Clarke, P.J. Devereaux, J. Kleijnen, D. Moher, The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration, J. Clin. Epidemiol. 62 (10) (2009) e1–e34.

[182] S. Afroz, M. Brennan, R. Greenstadt, Detecting hoaxes, frauds, and deception in writing style online, in: 2012 IEEE Symposium on Security and Privacy, IEEE, 2012, pp. 461–475.

[183] J. Lafferty, A. McCallum, F.C. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, 2001.

[184] S. Vosoughi, Automatic Detection and Verification of Rumors on Twitter (Ph.D. thesis), Massachusetts Institute of Technology, 2015.

[185] O. Ajao, D. Bhowmik, S. Zargari, Fake news identification on twitter with hybrid CNN and RNN models, in: Proceedings of the 9th International Conference on Social Media and Society, 2018, pp. 226–230.

[186] A. Jacovi, O.S. Shalom, Y. Goldberg, Understanding convolutional neural networks for text classification, 2018, arXiv preprint arXiv:1809.08037.

[187] S. Volkova, K. Shaffer, J.Y. Jang, N. Hodas, Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2017, pp. 647–653.

[188] F. Yu, Q. Liu, S. Wu, L. Wang, T. Tan, et al., A convolutional approach for misinformation identification, in: IJCAI, 2017, pp. 3901–3907.

[189] A. Zubiaga, M. Liakata, R. Procter, G. Wong Sak Hoi, P. Tolmie, Analysing how people orient to and spread rumours in social media by looking at conversational threads, PLoS One 11 (3) (2016) e0150989.

[190] G. Hu, Y. Ding, S. Qi, X. Wang, Q. Liao, Multi-depth graph convolutional networks for fake news detection, in: CCF International Conference on Natural Language Processing and Chinese Computing, Springer, 2019, pp. 698–710.

[191] C. Li, D. Goldwasser, Encoding social information with graph convolutional networks forpolitical perspective detection in news media, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 2594–2604.

[192] A. Benamira, B. Devillers, E. Lesot, A.K. Ray, M. Saadi, F.D. Malliaros, Semi-supervised learning and graph neural networks for fake news detection, in: 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM, IEEE, 2019, pp. 568–569.

[193] A. Hamid, N. Shiekh, N. Said, K. Ahmad, A. Gul, L. Hassan, A. Al-Fuqaha, Fake news detection in social media using graph neural networks and nlp techniques: A COVID-19 use-case, 2020, arXiv preprint arXiv:2012.07517.

[194] Z. Pehlivan, On the pursuit of fake news: From graph convolutional networks to time series, in: Multimedia Evaluation Benchmark Workshop 2020, MediaEval 2020, 2020.

[195] G.-A. Vlad, G.-E. Zaharia, D.-C. Cercel, M. Dascalu, UPB@ DANKMEMES: Italian memes analysis-employing visual models and graph convolutional networks for meme identification and hate speech detection (short paper), in: EVALITA, 2020.

[196] M. Dong, B. Zheng, N. Quoc Viet Hung, H. Su, G. Li, Multiple rumor source detection with graph convolutional networks, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 569–578.

[197] W.W. Zachary, An information flow model for conflict and fission in small groups, J. Anthropol. Res. 33 (4) (1977) 452–473.

[198] D. Lusseau, K. Schneider, O.J. Boisseau, P. Haase, E. Slooten, S.M. Dawson, The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations, Behav. Ecol. Sociobiol. 54 (4) (2003) 396–405.

[199] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442.

[200] P.M. Gleiser, L. Danon, Community structure in jazz, Adv. Complex Syst. 6 (04) (2003) 565–573.

[201] Y. Wu, D. Lian, Y. Xu, L. Wu, E. Chen, Graph convolutional networks with markov random field reasoning for social spammer detection, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, (no. 01) 2020, pp. 1054–1061.

[202] K. Lee, B.D. Eoff, J. Caverlee, Seven months with the devils: A long-term study of content polluters on twitter, in: Fifth International AAAI Conference on Weblogs and Social Media, 2011.

[203] C. Yang, R. Harkreader, J. Zhang, S. Shin, G. Gu, Analyzing spammers' social networks for fun and profit: A case study of cyber criminal ecosystem on Twitter, in: Proceedings of the 21st International Conference on World Wide Web, 2012, pp. 71–80.

[204] S. Lotfi, M. Mirzarezaee, M. Hosseinzadeh, V. Seydi, Detection of rumor conversations in Twitter using graph convolutional networks, Appl. Intell. 51 (7) (2021) 4774–4787.

[205] C. Data61, Stellargraph machine learning library, 2018, Publication Title: GitHub Repository. GitHub.

[206] N.J. Vickers, Animal communication: When i'm calling you, will you answer too? Curr. Biol. 27 (14) (2017) R713–R715.

[207] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, 2019, arXiv preprint arXiv:1907.11692.

[208] T. Pires, E. Schlinger, D. Garrette, How multilingual is multilingual BERT? 2019, arXiv preprint arXiv:1906.01502.

[209] G. Klambauer, T. Unterthiner, A. Mayr, S. Hochreiter, Self-normalizing neural networks, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 972–981.

[210] Y. Zhang, B. Wallace, A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification, 2015, arXiv preprint arXiv:1510.03820.

[211] W.L. Hamilton, R. Ying, J. Leskovec, Inductive representation learning on large graphs, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 1025–1035.

[212] Q. Xu, F. Shen, L. Liu, H.T. Shen, Graphcar: Content-aware multimedia recommendation with graph autoencoder, in: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018, pp. 981–984.

[213] R.v.d. Berg, T.N. Kipf, M. Welling, Graph convolutional matrix completion, 2017, arXiv preprint arXiv:1706.02263.

[214] F.J. García-Ull, DeepFakes: The next challenge in fake news detection, Anàlisi: Quaderns de Comunicació I Cultura (64) (2021) 0103–120.

[215] K. Shu, S. Wang, H. Liu, Beyond news contents: The role of social context for fake news detection, in: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, 2019, pp. 312–320.

[216] U. Von Luxburg, A tutorial on spectral clustering, Stat. Comput. 17 (4) (2007) 395–416.

[217] J. Zhou, J.X. Huang, Q.V. Hu, L. He, SK-GCN: Modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification, Knowl.-Based Syst. 205 (2020) 106292.

[218] L. Yao, C. Mao, Y. Luo, Graph convolutional networks for text classification, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, (no. 01) 2019, pp. 7370–7377.

[219] K. Bijari, H. Zare, E. Kebriaei, H. Veisi, Leveraging deep graph-based text representation for sentiment polarity applications, Expert Syst. Appl. 144 (2020) 113090.

[220] M. Zhang, T. Qian, Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2020, pp. 3540–3549.

[221] A. Wissner-Gross, Datasets over algorithms, Edge. Com. Retr. 8 (2016).