Expert Systems With Applications 153 (2020) 112986

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Review

Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities

Priyanka Meel, Dinesh Kumar Vishwakarma*

Biometric Research Laboratory, Department of Information Technology, Delhi Technological University, New Delhi 110042, India

ARTICLE INFO

Article history:
Received 19 July 2019
Revised 6 September 2019
Accepted 26 September 2019
Available online 4 October 2019

Keywords:
Clickbait
Deep learning
Fraudulent Content
Information Pollution
Machine learning
Opinion Spam
Online Social Networks
Rumour Propagation

ABSTRACT

Internet and social media have become a widespread, large scale and easy to use platform for real-time information dissemination. It has become an open stage for discussion, ideology expression, knowledge dissemination, emotions and sentiment sharing. This platform is gaining tremendous attraction and a huge user base from all sections and age groups of society. The matter of concern is that up to what extent the contents that are circulating among all these platforms every second changing the mindset, perceptions and lives of billions of people are verified, authenticated and up to the standards. This paper puts forward a holistic view of how the information is being weaponized to fulfil the malicious motives and forcefully making a biased user perception about a person, event or firm. Further, a taxonomy is provided for the classification of malicious information content at different stages and prevalent technologies to cope up with this issue form origin, propagation, detection and containment stages. We also put forward a research gap and possible future research directions so that the web information content could be more reliable and safer to use for decision making as well as for knowledge sharing.

1. Introduction

In the era of information overload, restiveness, uncertainty and implausible content all around; information credibility or web credibility refers to the trustworthiness, reliability, fairness and accuracy of the information. Information credibility is the extent up to which a person believes in the content provided on the internet. Every second of time passes by millions of people interacting on social media, creating vast volumes of data, which has many unseen patterns and behavioural trends inside. The data disseminating on the web, social media and discussion forums have become a massive topic of interest for analytics as well as critics as it reflects social behaviour, choices, perceptions and mindset of people. Connectivity on the internet provides people a vivacious and enthusiastic means of entertainment as well as refreshment. A considerable amount of unverified and unauthenticated information travels through these networks, misleading a large population. Thus to increase the trustworthiness of online social networks and mitigate the devastating effects of information pollution; timely detection and containment of false contents circulating on the web are highly required (Nunes & Correia, 2013).

The section of the data on which we are focusing is information pollution i.e. how the contents on the web are being contaminated intentionally or sometimes unintentionally. The false information may be in any format fake review, fake news, satire, hoax, etc. affects the human community in a negative way. Approximately 65% of the US adult population is dependent on social media for daily news (Shao, Ciampaglia, Flammini, & Menczer, 2016). If we grab the information without showing severe concern about its truthfulness, we have to pay in the long run. Social networks information diffusion has strong temporal features: Bursting updates, flooding all platforms with the carnival of information within no time (of course without fact-checking) and finally fast dying feature. Official news media is also losing the trust and confidence; in the rush of securing readership they are releasing eye-catching and sensational headlines with images, the readers do not have the time to read the actual news content; trust the appealing headline and the image. Thus, appealing headlines gives birth to a misunderstood falsified piece of information.

Earlier rumors used to spread at a slow pace, but the advent of internet technologies and popularity of retweeting activities on social networks has fuelled the dissemination of a piece of rumor around the globe at an alarming rate. In 2016, US presidential elections, because of some flows in algorithmic architecture Facebook has become a key distributor of fake news (Zannettou, Sirivianos, Blackburn, & Kourtellis, 2018), which has affected people's choice of the vote and had a tremendous impact on the result

* Corresponding author.
E-mail addresses: priyankameel86@gmail.com (P. Meel), dinesh@dtu.ac.in (D.K. Vishwakarma).

https://doi.org/10.1016/j.eswa.2019.112986
0957-4174/

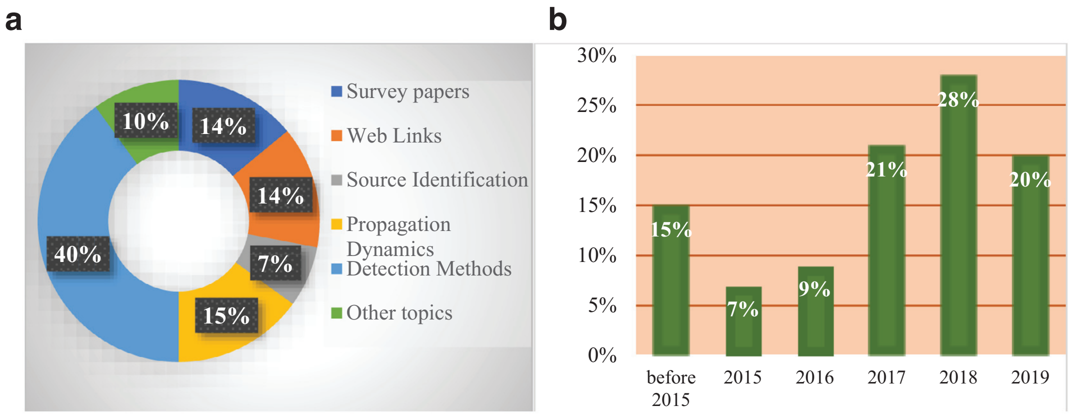

Fig. 1. (a) Topic-wise (b) Year-wise distribution of the refereed literature.

of the election. It is a remarkable example of how fake news accounts had outperformed real news. The main lineage of work done by researchers in web and social media mining is in tweeting behavior analysis, feature extraction, trends and pattern analysis, information diffusion, visualization, anomaly detection, predictive analysis, recommender systems, and situation awareness (Kumar & Shah, 2018; Zhou & Zafarani, 2018; Shelke & Attar, 2019; Zubiaga, Aker, Bontcheva, Liakata, & Procter, 2018). Fake news detection algorithms focus on figure out deep systematic patterns embedded inside the content of news. Another primary feature of detection is transmission behavior that strengthens the diffusion of information, which is of questionable integrity and value.

1.1. Motivation

Social media is a very fast data generating and disseminating platform and every second, millions and million of the users are interacting on web platforms and creating huge volumes of data. But contrary to traditional news sources such as news channels and newspapers, the credibility of contents circulating on social media platforms is questionable due to independence of freedom of expression. Recently, it has been seen that there is a huge increase in the number users (Newman, Fletcher, Kalogeropoulos, Levy, & Nielsen, 2018), who access the social media and web platforms for news and knowledge. Social media contents are governing people's choices of preferences. The term "Fake news" has become widespread after "2016 US presidential elections" where it is assumed that the fraudulent contents circulated during the elections exert considerable effects on the election results. Hence, to outline and analyses the various approaches used to deal with these issues, this work is presented. This work includes the current scenario of information pollution on web in terms of ecosystem, different data sharing and generating platforms, data analytics and fact-checking tools. Our survey methodology focuses on the four different stages of information pollution origin, propagation, detection and intervention.

1.2. Organization of the Paper and Key Contributions

In this work, we have identified various works reported in the field of fake news and rumor detection. Fig. 1 represents the percentage of referred literature according to topic and year of publishing. It is evident from the statistics of Fig. 1(a) that most of the work done is centered on detecting a piece of information as fake or real. Rumour diffusion is a long-studied terminology from mathematical epidemiological models but the term fake news has fuelled drastically in the world's digital communication scenario after 2016 US presidential elections, statistics from Fig. 1(b) also strengthens this fact.

This work provides an overview of the state-of-the-art technologies, models, datasets, and experimental setups for content and behavior analysis of fraudulent information circulating online. This review takes into consideration the broader perspectives of the research conducted by other scholars as on date as well as our analysis of the situation. The flow of information in this survey is structured according to Fig. 2.

A taxonomy of false information, a comprehensive survey of social impact, motivation for spreading false contents, user perception and available state-of-the-art methods of fact-checking is provided in Section 2. Section 3 focuses on the technological aspects of the identification of sources from where the falsified contents are originated.

Different models and diffusion patterns of intended contents for the targeted population is described in detail in Section 4. Section 5 deals with different stylometric and feature-oriented machine learning methods, deep learning and other methods of credibility analysis by which fraudulent contents can be segregated. The same section also details the experimental setups and datasets used by different researchers to address the issue. Countermeasures to aware the social audience who have already been influenced or are about to influence by malicious content are stated in Section 6. Current challenges and potential future scope of research are thoroughly presented in Section 7. Finally, Section 8 details in with the social and methodological findings and Section 9 conclude the work. The main contributions of the work are as under:

- Puts forward a serious concern towards the burning issue of trustworthiness and reliability of web content on social media platforms.
- The fraudulent content of all varieties scattered online is categorized, and the fake information ecosystem is analyzed right from the creation to disposition.
- A piece of detailed information about social media users, commercial analytics tools, popular social media platforms are outlined and discussed.
- The key contributions of earlier state-of-the-arts have been analyzed in terms of their merits and demerits.
- Establishes the significance of fact-checking and credibility analysis in the current scenario of internet-based information broadcasting.
- The current state of online fact-checking tools and APIs for content credibility analysis is presented and deliberated.
- The publicly available datasets are outlined along with experimental settings, highest accuracy, and methods.

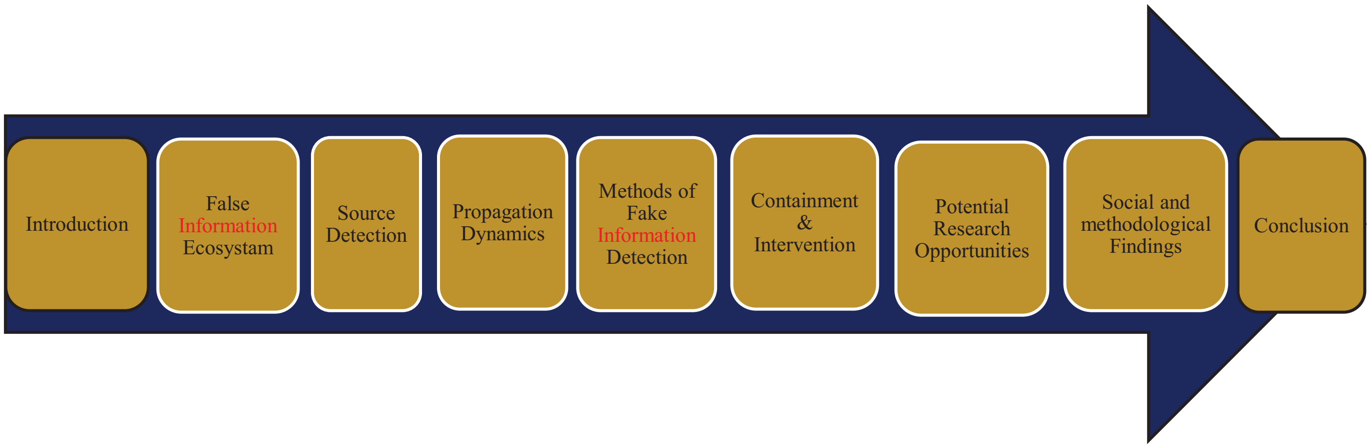

Fig. 2. Organisation of the paper.

- The technological aspects of false content detection from the source viewpoints, propagation, detection, and intervention are analyzed and discussed.
- A thorough analysis of machine learning and deep learning approaches of fake news and rumour detection are presented, which includes their merits and demerits.
- Highlights the contemporary issues in the domain of information pollution that are still unaddressed and needs due attention.
- The state-of-the-art and research gap presented provide insight for deciding the future course of action to combat the knotty question of fraudulent content on web.

2. False information ecosystem

According to the Global digital report 2019 (Newman et al., 2018) out of the world's total population of 7.676 billion, there are 4.388 billion internet users and 3.484 billion social media users. Almost half of the world's total population depends upon the internet for their knowledge. However, how much or up to what extent the circulated facts are verified is still a big question. How much we can rely on the information content that we are browsing every day. False information is created and initiated by a small number of people. People, relations, content and time are four critical dimensions of networked data analysed multi-dimensionally by proposing an iOLAP framework based on polyadic factorization approach (Chi, Zhu, Hino, Gong, & Zhang, 2009). This framework handles all types of networked data such as microblogs, social bookmarking, user comments, and discussion platforms with an arbitrary number of dimensions. Origination, propagation, detection and Intervention are the four main facets of information pollution, which are diagrammatically represented in Fig. 3.

Origination deals with the creation of fake content by a single person, account or multiple accounts. Propagation analyses the reason behind the fast and large-scale spread of fraudulent contents online. The analysis is done by Vosoughi, Deb, and Aral (2018), Horne and Adali (2017) sheds new light on fake news writing style, linguistic features and fraudulent content propagation trends; concludes that falsehood disseminates significantly faster, deeper, farther and more broadly than the truth in all the categories. False news was 70% more likely to be retweet by more number of unique users, as fake stories are more novel, surprising and eye-catching; attracts human attention hence encourages information sharing. Identification of the misinformation and disinformation from the massive volume of social media data using different Artificial Intelligence technologies comes under detection. Finally, intervention methods concentrate to restrict the outspread of false information by spreading the truth.

Fake product review is an emerging field of forgery in online social networks, specifically in the field of e-commerce, as more and more people share their shopping experiences online through reviews (Martens & Maalej, 2019). The customer reviews directly related to the reputation of a product in the E-commerce era. People consider ratings, feedback reviews, and comments by previous buyers to make an opinion on whether to purchase a particular item or not. The algorithms suggested in Elmurngi and Gherbi (2017a), Elmurngi and Gherbi (2017b), Dong et al. (2018) for detecting fake movie reviews are based on sentiment analysis, temporal, statistical features and text classification. Ahmed, Traore, and Saad (2017) use six supervised machine learning classifiers SVM, LSVM, KNN, DT, SGD, LR to detect fake reviews of hotels and fake news articles on the web using text classification. Their experiments achieve a significant accuracy of 90% and 92% respectively. Different content-based, features based, behavior-based and graph-based approaches (Viviani & Pasi, 2017) can be used to detect opinion spams present in different formats of fake reviews, fake comments, social network posting and fake messages. In addition to the mainstream news media; there is also a concept of alternative media (Starbird, 2017) that aims to just present the facts and let readers use their critical thinking to explore reality by means of discussions.

2.1. Categorization of false information

False information which is present in the form of images, blogs, messages, stories, breaking news; generally termed as information pollution has many formats that are not mutually exclusive but at the same time also have some heterogeneity that brings them under a specific category. The categorization of different information pollution formats is represented by means of a Venn diagram in Fig. 4. Table 1 summarizes different categories and impact of fraudulent content on the internet. Although each category has some salient characteristics throughout the paper, we have used the terms interchangeably at many places to provide a complete synergy of information pollution on the digital communication platform.

2.2. Motivation for spreading

Interactions of people on social media give rise to a lot of information content which turns out to be false sometimes intentionally with a predefined motive or unintentionally by mistake.

Fig. 3. Lifecycle of False Information.

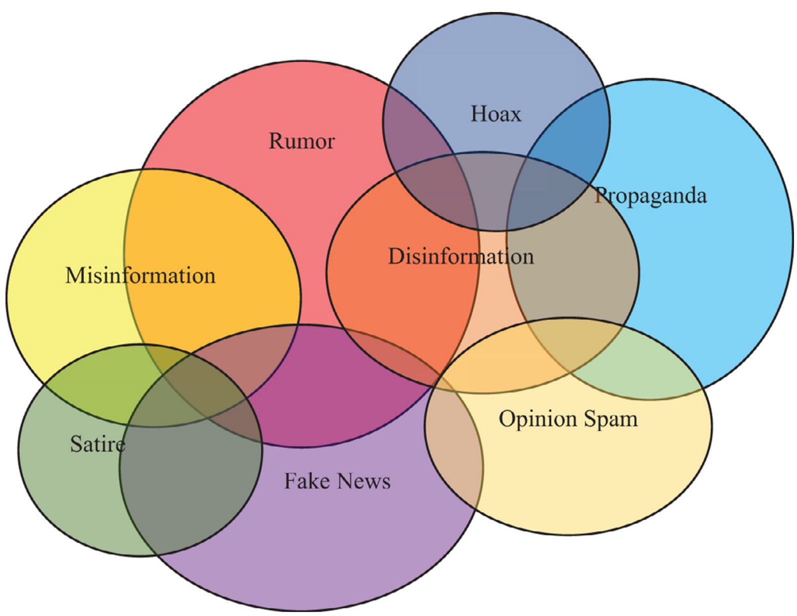

Fig. 4. Venn diagram of false information on social media and web.

The following Table 2 details the key reasons behind the increasing spread of misleading contents on online platforms:

2.3. Social impact

Social networking platforms launched in the past two decades plays a role in social interactions by providing easy to use features to exchange information in many formats. Table 3 summarizes popular social networking platforms along with their customer base and salient features (data source (Global social media ranking, 2019) and Wikipedia). Fig. 5 shows the popularity of statistics on major social platforms (data source (Newman et al., 2018)). Fig. 6(a) and (b) explain some statistics based on age and country about social media users(data source (Newman et al., 2018)). Around the globe, 54% of people express strong concern about "what is real or fake" when thinking about online news. The younger section of the population is under more influence of Internet-based knowledge, and as the age grows according to statistics, this ratio decreases. Table 4 supported by Fig. 7 (a–e) explains some of the prominent havocs created in society in recent years as a consequence of information pollution and classifies them according to the taxonomy provided in Section 2.1.

2.4. User perception

Users perceive the data from social networks based on their intelligence and consciousness about the facts. According to their interests and insight, they can either forward the data assuming as true, discard it assuming as false or becomes neutral to the news (Bronstein, Pennycook, Bear, Rand, & Cannon, 2019). A survey

Table 1
Categorization of false information.

Category	Definition	Impact
Rumor	Unverified piece of information which is not necessarily false; may turn out to be true also	Uncertainty and confusion about facts
Fake News	False information spread under the guise of being authentic news usually spread through news outlets or internet with an intention to gain politically or financially, increase readership, biased public opinion	to damage an agency, entity, or person or gain financial/political profit
Misinformation	Circulating information that becomes false inadvertently as a consequence of an honest mistake, carelessness or cognitive bias	Less harmful but wrong interpretation of facts can lead to big damage
Disinformation	Deliberately deceptive information with a predefined intention	To promote a belief, idea, financial gain or tarnish an opponent's image
Clickbait	The deliberate use of misleading headlines to encourage visitors to click on a particular webpage	To earn advertising revenue, to trigger phishing attacks
Hoax	The false story, especially by means of Joke, prank, humor or malicious deception, used to masquerade the truth	Falsehood is perceived as truth and reality
Satire/parody	Articles that primarily contains humor and irony, no harmful intention but has the potential to fool. The Onion and Satire Wire are sources of satirical news articles.	The motive is fun but sometimes exert adverse effects also
Opinion Spam	Fake or intentionally biased reviews or comments about products and services	untruthful customer opinion
propaganda	Unfairly prejudiced and deceptive information spread in targeted communities according to a predefined strategy to promote a particular viewpoint or political agenda	Political/financial profit
Conspiracy theories	an explanation of an event that invokes a conspiracy by sinister and powerful actors, often political in motivation-based entirely on prejudice or insufficient evidence	Extremely harmful to people and society

Table 2
Motivation behind information pollution.

Motive	Description
Political Intent	to malign the public image of the opponent or promote a person or party
Financial Profit	False-positive information triggers the motivation for large-scale investments and affects stock prices. Fake ratings and reviews of products are intentionally written to increase sales.
Passion for promoting an ideology	A considerable number of people are impassioned about a particular organization, ideology, person or philosophy and they want to spread it by any means.
Fun	For amusement and fun, satirical sites write humorous content that is often mistaken for real news. This is the least severe motive, which does not have many harmful effects because intentions are not usually wrong.
Increase customer base	In the era of Internet-based journalism, online news media is rushing to secure readership and increase customer base. Thus, publishing the stories of questionable integrity and content in the process to lure readers to their websites and platforms
Rush to cover the latest news	In a competition to be the first to cover the story, journalists often publish articles without fact-checking and get millions and millions of views. Truth and veracity become liabilities in the current online journalism with aims to" Publish first, correct if necessary"
Generate advertising revenues	Fake news creators have earned a sizable profit from automated advertising engines like AppNexus, Facebook Ads and Google AdSense (Reilly, 2018) during the 2016 US presidential elections. Earning capital through false advertising news is a significant driving force that an entire cottage industry of practitioners has indulged in this controversial endeavor.
Technological Reasons	Algorithms are structured to endorse things based on popularity, not accuracy (Reilly, 2018), Echo chambers and filter bubbles in search engines (Mohseni, Ragan, & Hu, 2019) are some of the algorithmic flows accounts for biased information circulation. Therefore, they are agnostically promoting the spread of disinformation as fake news is intentionally designed to gain more user attention.
Manipulate public opinion	In a consumer-based economy, public opinion regarding a firm, service, product or people holds significant importance as customers are going to decide the fate of stocks, sales, election results, all types of businesses and many more.

study supported by questionnaires done in 2017 by Ghaisani, Munajat, and Handayani (2017) suggests that users judge the credibility of information available online on certain factors such as a link to other sources, interest in the topic, embedded videos, embedded photos, source of information, writing style, logical explanation, peer comments, similarity with other contents and media, etc. Social media analytics tools are principal source of monitoring, analyzing and managing information floating on social networks in the public domain. They statistically, behaviourally and semantically analyze the data from different aspects to generate reports. Table 5 lists some of the public and commercial social media analytics tools that play a crucial role in providing suggestions and developing mass opinions.

Internet is a major hub for knowledge seekers, but out of the available information which is credible for learners is a question that needs careful attention. A recommendation framework is proposed (Li, Bao, Zheng, & Huang, 2015) for online learning communities by merging user credibility network, domain experts group and user rating matrix, which is based on expertise, influence, longevity and centrality of individuals. This framework provides three categories of recommendations: learning peer recommendations, domain expert's recommendations, and learning resource recommendations. Vox Civitas (Diakopoulos, Naaman, & Kivran-Swaine, 2010) is a social media visual analytics web-based tool developed in 2010 for journalistic inquiry of public sentiments and opinions based on vast message exchange on Twitter. The tool exhibits temporal behaviour by collecting the contents of social media over a specific time window to perform their content analysis based on four factors: relevance, uniqueness, sentiment (positive, negative, controversial and neutral) and keywords (ranked by their TF-IDF scores) to cover the follow-up story angles of certain key events. Whisper (Nan et al., 2012) is a real-time tool that tracks information diffusion process in social media and answers when, where and how an idea is propagated. To trace multiple pathways of community response, information propagation, social-spatial extent and temporal trends, an efficient flux line-drawing method is used.

2.5. Current state of fact checking

Compromised social network accounts can be used for spreading misinformation, tarnish the reputation of opponents or they can cause multi-billion-dollar monetary losses in financial markets. Table 6 lists popular credibility analysis tools that are used to check the authenticity of online content. Credfinder (Alrubaian, Al-Qurishi, Al-Rakhami, Hassan, & Alamri, 2016) is a chrome extension developed and launched in 2016 for assessing real-time credibility of tweeter messages based on content and user-specific features.

Table 3
Facts about social networking platforms.

Name	Logo	Year	No. of active users/month	Salient features
Facebook		2004	2.32 billion	Supports text, images, videos, live videos, and stories; requires valid email with an age of being 13 and older.
Twitter		2006	330 million	Registered users can post, like, and retweet but unregistered users can only read them; multilingual platform, freeware
WhatsApp		2009	1.6 billion	Voice-over IP (VoIP) and messaging service owned by Facebook; supports text, audio, video, images; freeware
Skype		2003	300 million	telecommunications application supports video chat, voice calls, instant messaging, text, audio, video, images and video conferencing
Facebook Messenger		2011	1.3 billion	messaging app and platform; exchange messages, photos, videos, stickers, audio, files, voice and video calling
Snapchat		2011	287 million	Photo and short video sharing platform; messages and pictures are accessible for a short time only after that, they become unavailable to their recipients.
You tube		2005	1.9 billion	video-sharing platform owned by Google; allows users to view, upload, add to playlists, rate, report, share, subscribe to other users and comment on videos
Tumblr		2007	642 million	Supports blogs containing multimedia and short messages.
Instagram		2010	1 billion	video and photo-sharing service owned by Facebook
Viber		2010	260 million	cross-platform, voice over IP and instant messaging service operated by a Japanese company; available in more than 30 languages
Sina Weibo		2009	462 million	biggest social media platforms in China; huge financial success with high revenue, surging stocks, total earnings per quarter and lucrative advertising sales; hybrid mix of Twitter's and Facebook's features.
Pinterest		2010	291 million	photo sharing and visual bookmarking platform; enables users to find new ideas for projects and save them; used as a "catalog of ideas"
LinkedIn		2003	303 million	business and employment-oriented service used for professional networking that operates via websites and mobile apps
Quora		2010	190 million	a place where people can gain knowledge and share by asking and answering questions
WeChat		2011	1.098 billion	Chinese all-in-one communications app for multi-purpose messaging and calling developed by Tencent
QZone		2005	532 million	platform supports photo sharing, music, videos, writing blogs and maintaining diaries; based in China, created by Tencent
Baidu Tieba		2003	300 million	largest Chinese communication platform developed by Chinese search engine company Baidu, available in 3 languages Chinese, Vietnamese, Japanese
QQ		1999	807 million	supports microblogging, shopping, games, movies, music,and voice chat; instant messaging service
TikTok		2016	500 million	known as Douyin in China; supports customizable music videos of up to 60 seconds of length with user-generated special effects and music; Declared world's most downloaded app in the first quarter of 2018.
Reddit		2005	330 million	American social news aggregation, discussion and web content rating website; Links, text posts, images and other contents submitted by registered users can be voted up and down by other members.

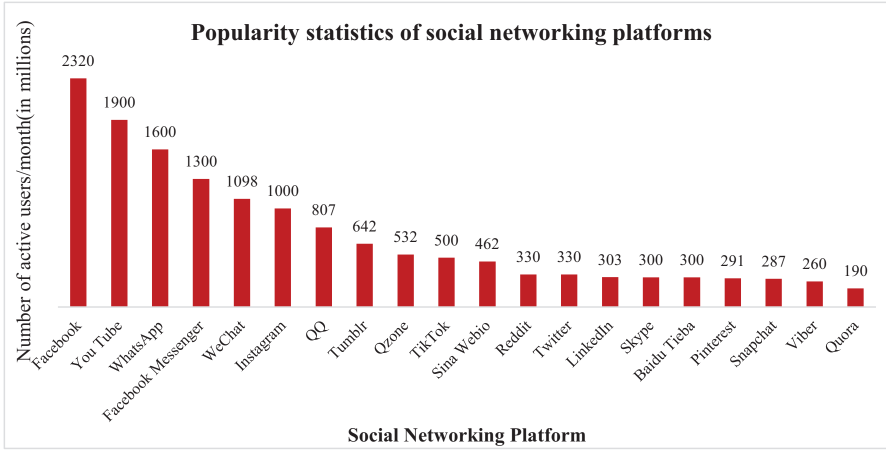

Fig. 5. Number of active users/month of popular social networking platforms (data source (Newman et al., 2018)).

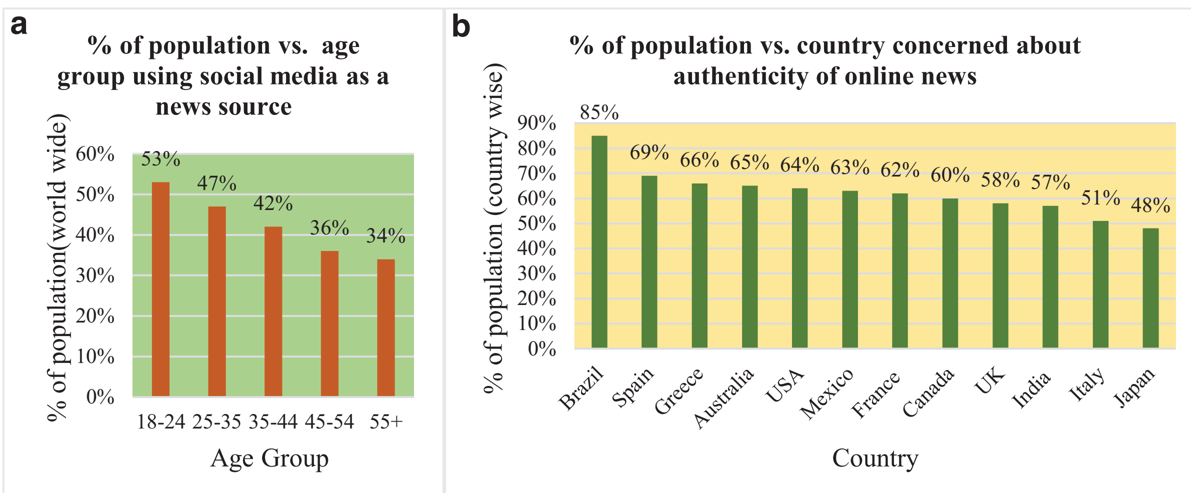

Fig. 6. (a) Social media as a news source according to age group (b) Awareness of people towards news truthfulness (data source (Newman et al., 2018)).

This extension has two major components: A chrome extension (client) that captures the real-time data from the tweeter timeline and a web-based backend (server) that analyses the collected tweets and calculates their credibility. Response time of credfinder is very less and it was extensively tested during 2016 US presidential elections but not as popular as it has no provision to check the images for forgery.

Hoaxy (Shao et al., 2016,2018) is a platform for collection, detection and analysis of fraudulent online content from various viewpoints and its related fact-checking efforts. The collected contents from news websites and social media are fed into a database that is updated on a regular basis and analyzed to extract different hidden patterns. The fact-checking activities initiate on social media almost 10-12 hours after the spread of misinformation. Hoaxy is tested by collecting approximately 1442,295 tweets and articles from 249,659 different users. Because of the limited character length of tweets, URLs of web pages are commonly shared.

COMPA (Egele, Stringhini, Kruegel, & Vigna, 2017) works by building a behavioral profile for every social network account based on message characteristics and stable habits that a user develops over time. Every new message is compared against the already built profile; if it profoundly deviates from the learned behavior, it is triggered as a possible compromise. However, if the attacker is well aware of the capabilities of COMPA the fake message can be designed in such a way that its behavior resembles the actual one, so it can't be detected.The Flux flow (Zhao et al., 2014) is an interactive visual analysis system designed for detecting, exploring and interpreting anomalous conversational threads on Twitter. It incorporates three major components: (a) data preprocessing and storage module (b) backend data analysis module (c) anomaly detection module. Flux flow represents different dimensions of information propagation such as content, topics, temporal dynamics of the spreading, sentiment, relationship and connections among different threads as well as authors.

3. Source detection

Source detection refers to find out a person or location from where the fraudulent information in the social network or web started spreading. Along with other containment methods, identifying the original source of information pollution plays a vital role in reducing online misinformation. In various application domains, origin identification is very important such as Medicine (to find the source of the epidemic), Security (to detect the source of the virus), social network (to identify the origin of the wrong information), financial networks (for finding the reasons of cascade failures), etc. The following Fig. 8 summarizes the steps involved in the source detection process.

A bio-inspired method which solely depends upon the infected time of observers was developed in Liu, Gao, She, and Zhang (2016), proposes a Physarum-inspired Mathematical Model

Fig. 7. Images Spreading Fake news on different social and news media platforms. ((a)-("Chinese salt-buying frenzy," 2011), (b)-("Explosion at the White House," 2013),(c)-(Donald Trump ends school shootings by banning schools | 8Satire," 2019),(d)-("Was Kamal Nath the driver of Rajiv Gandhi?," 2018),(e)-("Was Hillary Clinton photographed with Osama Bin Laden?," 2017)).

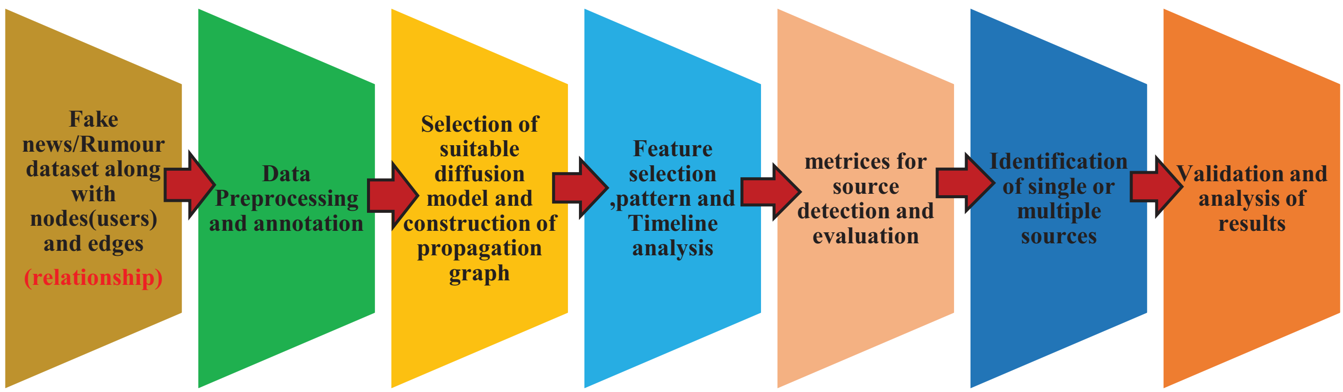

Fig. 8. Steps of source identification of false information.

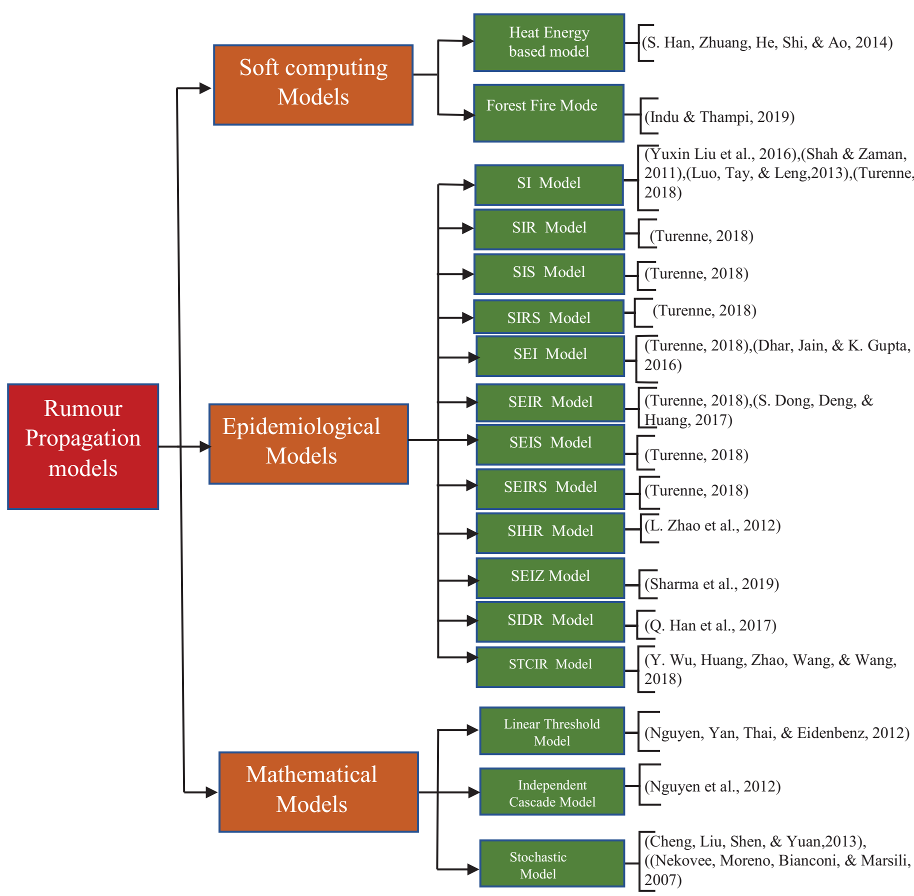

Fig. 9. Classification of different rumor diffusion models. (Han, Zhuang, He, Shi, and Ao (2014); Luo, Tay, and Leng (2013); Turenne (2018); Dhar, Jain, and Gupta (2016); Dong, Deng, and Huang (2017); Sharma et al. (2019)).

Table 4
Few examples of false information.

Fake Information/News	Classification	Truth and Impact	Reference
Radiation leakage in Japan could pollute seawater and sea salt, so additionally iodized salt could help to protect people from nuclear radiation	Rumour Fig. 7(a)	Caused salt-buying frenzy in China; shopkeepers charged 10 times higher than average prices; Beijing supermarkets run out of salt	Chinese salt-buying frenzy (2011)
Breaking: Two Explosions in the White House and Barack Obama is injured	Fake news/Disinformation Fig. 7(b)	The news was announced from the hacked Twitter account of the Associated Press; before the news was clarified costs 10 billion USD losses	Explosion at the White House (2013)
shootouts and kidnapping by drug gangs happening near schools in Veracruz	Rumour	Rumour triggers severe chaos in the city resulting in 26 car crashes; spread through Facebook and Twitter, as people left their cars in the middle of a street and rushed to pick up their children from school	Mexico 'Twitter terrorism' charges cause uproar (2015)
Six hundred murders take place in Chicago during the second weekend of August 2018.	Disinformation/Fake news	The actual number of murders was one; Created fear and anxiety in society	Did 600 Murders Take Place in Chicago Last Weekend? (2018)
Donald Trump ends school shootings by banning schools	Satire Fig. 7(c)	An article published by a satire website spread as a breaking news	Donald Trump ends school shootings by banning schools | 8Satire (2019)
Newly appointed Madhya Pradesh Chief Minister Kamal Nath was former Prime Minister Rajiv Gandhi's driver	Misinformation Fig. 7(d)	Kamal Nath had shared an image on Rajiv Gandhi's birth anniversary, from his official Twitter handle in which he is driving the car, and Rajiv Gandhi is sitting by his side	Was Kamal Nath the driver of Rajiv Gandhi? (2018)
North Korea Opening its doors to Christians	Rumour	A bogus story published in a notorious fake news web site claimed without evidence. The Magazine aims at spreading the good news to devout Christian readers	North Korea Opening its doors to Christians (2018)
Don't have Paracetamol tablets, it contains the 'Machupo' virus!	Hoax	The Machupo virus, which spreads through direct contact with infected rodents, is only known to be found in South America; no cases have been reported in India so far.	Do paracetamol tablets contain 'machupo' virus? (2019)
"Recall these fantastic, mind-boggling photographs of how Bin Laden was hosted in the White House," Russia's Foreign Ministry spokeswoman Maria Zakharova has commented on the photograph showing Osama Bin Laden was hosted in the White House.	Propaganda Fig. 7(e)	Osama Bin Laden's photograph has been superimposed on a photo of Mrs. Clinton meeting musician Shubhashish Mukherjee at an event in 2004. This fake image is shared on social media in Russia.	Was Hillary Clinton photographed with Osama Bin Laden? (2017)

Table 5
Public and commercial social media analytics tools.

Analytics tool	Category/function	Data source	Salient features	Reference
Crowdbooster	Analytics, Marketing, Management	Twitter, Facebook, LinkedIn	Trip adviser/shopping/online city cabs/can Schedule Unlimited Tweets and Posts, Follower Evaluation	Crowdbooster (2019)
Vox Civitas	Analytics, Automatic content analysis	Twitter	Journalistic Inquiry to study public opinions after an event	Diakopoulos et al. (2010)
Whisper	Visualization, Tracing information diffusion process	Social Networks (Twitter etc.)	Visualize social-spatial extent, temporal trends, and community response to a topic	Nan et al. (2012)
Talkwalker	Analytics, marketing	Social networks, blogs, news websites	Analyze real-time conversations across social network blogs, news websites, and forums in 187 languages. It provides a wide range of data statistics related to mentions, sentiment, distribution of conversations, etc.	Talkwalker (2009)
Google analytics	Web analytics service	all social networks	tracks and reports website traffic, users activities such as session duration, pages per session, bounce rate, etc, have a real-time insight of visitors currently on the website;	Google Analytics (2005)
Hootsuite	social media management, listening, publishing, and analytics	Twitter, Facebook, Instagram, LinkedIn, Google + and YouTube.	Improve the effectiveness of ads and broadens the reach of posts; Customize reports in multiple matrices and formats; Track brand mentions better by integrating with specialized tools like Brandwatch and Talkwalker.	Hootsuite (2008)
Snalytics	Optimize story-based content	Snapchat and Instagram	Create and manage stories with feature-rich publishing; provides different matrices of story popularity and reading	Snaplytics (2015)

of misinformation source detection under the constraint of limited observers and SI model of the diffusion process. The model gives higher locating accuracy and less error rate when compared to experimental results of four benchmark networks with traditional Gaussian and GaussianSI model. Shelke and Attar (2019) provides a state of the art survey of different source detection methodologies in case of single and multiple misinformation source along with different available datasets and experimental setups. A two-stage source localization algorithm for probabilistic weighted graph (Louni & Subbalakshmi, 2018) is designed which models the heterogeneity of social relationships by using probabilistically varying weights for the edges. In the first stage of the algorithm,

Table 6
List of Fact-checking platforms.

Name	Salient features	Reference
TwitterTrails	An interactive online tool for investigating the propagation characteristics, refutation of stories shared on Twitter, origin, and trustworthiness	Finn, Metaxas, and Mustafaraj (2014)
TweetCred	A real-time web-based system with a rating between '1 and 7' to assesses the credibility of each tweet in the twitter timeline.	Gupta, Kumaraguru, Castillo, and Meier (2014)
Hoaxy	A platform for collection, detection and analysis of online misinformation and its related fact-checking efforts.	Shao et al. (2016)
Emergent	Web-based automatic real-time rumor tracker; tracks social media mentions of URLs associated rumors.	Emergent (2019)
CredFinder	Analyses user and content features to find out the credibility of tweets. Works in real-time as an extension of the Chrome Browser.	Alrubaian et al. (2016)
RumorLens	A tool to aid journalists in segregating posts that spread a specific rumor on Twitter, by traversing the size and distribution of the audience.	Resnick, Carton, Park, Shen, and Zeffer (2014)
COMPA	System to detect compromised social network accounts. Message characteristics and behavioral user profiles are used for misinformation detection.	Egele et al. (2017)
FluxFlow	Interactive visual analysis system to detect, explore and interpret anomalous conversational threads in twitter	Zhao et al. (2014)
REVEAL	Verification of social media content mainly concentrating on image authenticity from a journalistic and enterprise outlook.	REVEAL (2014)
InVID	The platform supports authentication, fraud detection, reliability and accuracy checking of newsworthy video content and files spread via social media	InVID (2017)
ClaimBuster	Allows users to perform live fact-checking with the help of finding out factual claims	Hassan, Arslan, Li, and Tremayne (2017)
TruthOrFiction	Covers Politics, religion, nature, aviation, food, medical, etc., Email rumors are classified in truth and Fiction	Truth or Fiction - Fact Check (2019)
Snopes	Covers all domains of the news; label videos and News articles in 12 categories, True; Mostly true; Mixture; Mostly false; False; Unproven; Outdated; Miscaptioned; Correct attribution; Misattributed; Scam; Legend	Snopes.com (1994)
FactCheck	Intends to reduce the level of confusion and deception in U.S. politics. Analyses TV ads, debates, speeches, interviews and news and labels them as True; No evidence; False	FactCheck.org (2003)
PolitiFact	Covers American politics; After fact-checking labels articles as True, Mostly True, Half True, Mostly False, False and Pants on fire	Fact-checking U.S. politics-PolitiFact (2007)
Fake News Tracker	Predicting fake news from data collected automatically from social context and news, also provides effective visualization facilities using NLP and deep neural networks	Shu, Mahudeswaran, and Liu (2019)

the most likely candidate cluster to contain the source of the rumor is identified. In the second stage, the source is estimated from the set of nodes inside the most likely candidate cluster. To minimize the estimation error of source and analyze the rumor centrality maximum likelihood estimator (Shah & Zaman, 2011) is used that examines the asymptomatic behavior of infected nodes in detail for regular trees, general trees and general graphs. Along with the infection source the infection region i.e. a subset of nodes infected by each source in a network is identified considering SI propagation model with homogeneous spreading rates based on approximations of the infection sequence count. Choi et al. (2017) and Choi, Moon, Shin, and Yi (2016) identify rumor source using different approaches such as batch queries, interactive queries, Maximum-A-Posteriori-Estimator(MAPE). Zhu and Ying (2016) gtries to identify source using a path-based approach and Zhang, Zhang, Lv, and Yin (2016) estimates spreading source in network based on observer nodes.

4. Propagation dynamics

The majority of the research in the propagation dynamics of misinformation is done in line with epidemic models, which categorizes the people in different classes then derives equations to perform steady-state analysis. People who never heard the rumor, Ignorant are similar to **Susceptible (S)**, those who are spreading rumors, Spreaders are similar to **Infective (I)** and people who heard rumor but do not spread it, Stiflers similar to **Removed (R)**. The dynamics of rumor spreading on homogeneous network LiveJournal are studied in Zhao et al. (2011) with consideration of forgetting rate, spreading rate, stifling rate and average degree using SIR (Susceptible-Infected-Removed) epidemiological model. The same group of researchers further extended their work by adding a new category of people **Hibernators (H)**, coming from the spreaders due to forgetting mechanism and later becoming spreaders again due to remembering mechanism in SHIR(Susceptible-Infected-Hibernator-Removed) model (Zhao et al., 2012).SIDR(Spreader-Ignorant-Doubter-Stifler) model is proposed in Han, Miao, and Fan (2017).

Mean-field equations and steady-state analysis are done to study SHIR rumor diffusion model in social networks. Another model based on users forget and remember mechanism is presented by Gu, Li, and Cai (2008) in which an individual's state keeps on switching between active(with the message) and inactive(without message). Fig. 9 classifies prominent methods of rumor spreading available in literature in three major categories: Soft computing, epidemiological and mathematical approaches.

A nature-inspired approach based on forest fire model is proposed by Indu and Thampi (2019) to figure-out the diffusion path of rumors and find out the most influential users in rumor diffusion. The model evaluates the probability of each node to be affected by misinformation and finally identify all the rumor affected nodes to estimate the complete range of rumor spread. The study concluded that only a few users have tweeted the rumour and 90% of the messages are retweets. Mendoza, Poblete, and Castillo (2010) analyzed the propagation dynamics, follower-followees relationship, number of tweets per user, the vocabulary of tweets, retweet behavior for conformed truths and rumors supported by a case study of 2010 earthquake in Chile. The research concluded that false stories are questioned much more than confirmed truths.

A rumour propagation model for emergency situations based on the interactions of seven stakeholders of population ignorant(I), Wise(W), spreader (S), unbeliever (U), indifferent (IN), Opponent(O) and reasonable immune(RI) is proposed using an active immune mechanism (Chen, Song, & Zhou, 2017). Experiments show that network properties profoundly affect the diffusion process. Rumour propagation analysis on online social site BlogCatalog is done by formalizing a dataset of an undirected graph G(V,E) contains 10,312 nodes and 3 33 983 edges using stochastic epidemic model (Cheng, Liu, Shen, & Yuan, 2013). The complex structure of social networks can be modeled using different graphical formats such as Assortative correlated scale-free networks, Uncorrelated scale-free networks, Homogeneous networks, Inhomogeneous networks and Random Graphs. Analysis of rumor diffusion in complex structures is done by using the stochastic model (Nekovee, Moreno, Bianconi, & Marsili, 2007), which are further analyzed by analytical and numerical solutions of mean-field equations.

A content-based probabilistic model (Mondal, Pramanik, Bhattacharya, Boral, & Ghosh, 2018) utilized four properties of rumor propagation temporal, structural, linguistic and social tie for identification of unverified tweets in the aftermath of a disaster in an early stage. The salient feature of the approach is a tweet that has at least one rumor propagation feature is being extracted, and its probability of being a rumor is analyzed. Another key finding of the method is that rumours contain high sentiments, generally dominated by words related to social ties and actions like hearsay.

5. Methods of false information detection

The important part of information pollution is to detect fake news and rumors. Different artificial intelligence algorithms along with cognitive psychology and mathematical models are used to identify false contents. The following section provides a detailed explanation of datasets, experimental settings, methods of training, validation, testing used in various machine learning and deep learning technologies.

5.1. Datasets and experimental settings

Different formats of datasets are used for content and behavioral analysis such as text tweets, images, headlines, news articles, product and services reviews, URLs, readers' comments, suggestions, discussions on particular events, etc. Most of the researchers have used Twitter API for collection and analysis of rumors and fake news as a data source. Zubiaga et al. (2018) describe a detailed method of accessing, collecting and storing data using Twitter, Sina Weibo and Facebook's API along with their limitations. FakeNewsNet (Shu, Mahudeswaran, Wang, Lee, & Liu, 2018) is a data repository that contains two comprehensive datasets PolitiFact and GossipCop to facilitate research in the field of fake news analysis. These datasets collect multi-dimensional information from news content, social context and spatiotemporal data from diverse news domains. Details of some of the widely used datasets and experimental setups are given in Table 7. Comparative analysis of the features and usefulness of FakeNewsNet from other publicly available datasets LIAR (Wang, 2017), BS Detector, CREDBANK (Mitra & Gilbert, 2015), BuzzFace (Santia & Williams, 2018) and FacebookHoax (Tacchini, Ballarin, Vedova, Moret, & Alfaro, 2017) is presented in Shu et al. (2018). Al-Qurishi et al. (2015) describe in detail different criteria and methods of selecting the best possible open-source tool for data gathering from social media and web.

5.2. Handcrafted feature extraction

Machine learning is a prominent technology in designing models for detecting false content. The effectiveness of these algorithms mainly depends on pattern analysis and feature extraction. Table 8 summarises key handcrafted features used in earlier state-of-the-arts to design machine-learning models.

5.3. Network structure

Network structures are innovative methods of credibility assessment of a target article (Chen et al., 2019; Zhou, Zafarani, Shu, & Liu, 2019). A model is being constructed in Dynamic Relational Networks (Ishida & Kuraya, 2018) by using related news articles that are mutually evaluating each other's credibility based on the facts of who, what, where, when, why and how. Each article unit contains one article node and many fact nodes. Nodes of one article unit are mutually evaluated by consistency among their fact nodes with another available article. For fairness of evaluation, each user can build his network by using a bottom-up approach. Structure of small world peer-to-peer social networks (Wang, Moreno, & Sun, 2006) and large web-based social networks spanning large geographical areas (Csányi & Szendroi, 2004) are analyzed through various modeling techniques to deduce some important characteristics of propagation and area related properties. In the case of small-world network, the connectivity between users is scale-free in the form of undirected, directed and weighted graphs. Fig 10. represents some of the network structures being constructed for credibility assessment.

To model network structures and user connectivity of online social networks, scalable synthetic graph generators are used. They provide a wide variety of generative graph models that can be used by researchers to generate graphs based on different extracted features such as propagation, temporal, connectivity,follower-followee, etc. Some of the tools and their characteristics are summarized in Table 9.

5.4. Machine learned classifiers

A novel approach of multistage credibility analysis is proposed in Alrubaian, Al-qurishi, Al-rakhami, Rahman, and Alamri (2015) with five stages: Feature extraction, Relative importance assignment, naïve Bayes classifier, Opinion mining and finally overall assessment which classifies the tweets as credible or non-credible. In Elmurngi and Gherbi (2017a, b) the authors have done a comparative study of supervised machine learning algorithms using sentiment analysis and text classification on movie review. Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbour (KNN-IBK), KStar (K*) and Decision Tree (DT-J48) on two different movie reviews dataset V1.0 and V2.0 are tested (Elmurngi & Gherbi, 2017b). The results draw a strong conclusion that SVM outperforms all other supervised machine-learning algorithms by giving a prediction accuracy of 76% and 81.35% on the two datasets, respectively. Fig. 11 explains generalized machine learning architecture used by various algorithms listed in Fig. 12 and Table 10 for fake news detection.

Elmurngi and Gherbi (2017a) compare NB, K-NN (with $K = 1$, $K = 3$, $K = 5$, $K = 7$), SVM and DT algorithms on movie review datasets with stop words and without using stop words. Stop words are the list of alphabets that convey no meaning, frequently used in a language and retains a high probability of confusing the classifier. Thus, it needs to filter out stop words before training a classifier to reduce memory requirements and better classification accuracy. In addition to the User, Message, content, Topic and sentiment features Castillo, Mendoza, and Poblete (2011) also utilized propagation characteristics for the construction of a decision tree in DT-J48 algorithm. Decision tree algorithm gives 89% prediction accuracy more than SVM, Naïve Bayes and Decision rule on the same dataset of 747 most trending news topics each of which contains almost 10,00 tweets. Amazon Mechanical Turk is used for "Human Intelligence Task" (HIT) of manually categorizing the news cases into "almost certainly true", "likely to be false", "almost certainly false" and "can't decide". Supervised machine learning classifiers are then trained and tested the accuracy with comparative analysis based on precision, recall and F-scores. A significant conclusion is that newsworthy topics tend to have deep propagation trees and trustworthy news is propagated through authors that have already written several messages with many re-posts.

Table 7
Datasets used for fake information analysis.

Reference	Dataset details	size	Fake news/Rumour	Real news/non-rumour	Information used for detection	Experimental setting and accuracy
Yang et al. (2018)	Dataset contains multiple information such as the title, text, image, author, website	20,015 news articles	11,941	8074	Title, text and image	80%data is used for training,10% for validation and 10%for testing
Singhania et al. (2017)	20,372 fake articles from 19 fake news websites listed in Politifact.20,937 Genuine news articles from 9 websites listed in Forbes	41,307 total news articles	20,372	20,937	Headline and text	20% dataset is used for training,10% validation and 70% testing; accuracy 99.77%
Elmurngi and Gherbi (2017b)	movie review datasets	total 2000 movie reviews	1009 (v2.0)	991 (v2.0)	Text reviews	50% positive and 50% negative reviews; Accuracy 81.35%
Elmurngi and Gherbi (2017b)	Movie reviews	Total 1400 movie reviews	732	668	Text reviews	50% positive and 50% negative reviews; Accuracy 81.35%
Elmurngi and Gherbi (2017a)	Movie text review dataset	Total 2000 reviews	1009	991	Text reviews	50% positive and 50% negative reviews are there; Accuracy 81.75%
Shao et al. (2016)	tweets containing fake news and fact-checking	1442,295	1287,769	154,526	Text tweets and URLs	(80-90) % of the data are original tweets and retweets while (10-20) % are quotes and replies
Castillo et al. (2011)	By using "Twitter Monitor" over a period of 2 months total 2524 cases are detected, out of which 747 were labelled as news	747 cases/news (almost 10,000) tweets per news	302 (likely to be false or almost certainly false)	306(almost certainly true)	Text tweets with content and propagation analysis related to each news	18.6% (139) news cases were labeled as ambiguous; Accuracy 89%
Kwon et al. (2013)	Total 102 topics are retained after pre-processing each contains at least 60 tweets	102 events with at least 60 tweets/event	47	55	Content analysis of text tweets	Each event is annotated first by human experts as rumour or non-rumour; Accuracy 92%
Mendoza et al. (2010)	Total 4727,524 preliminary tweets by 716,344 different users	7 rumours and 7 confirmed news topics	7 false news events; average 24,029 tweets/event	7 confirmed truths; average 16,871 tweets/event	Checked by human experts for annotation	19.8% of the tweets are replied to original tweets
Hamidian and Diab (2015)	Annotated Twitter dataset of 9000 tweets	5 established rumours are taken	Tweets related to each rumour are classified as Not rumour, endorse rumour, Denies rumour, Question rumour, Neutral and Undetermined tweets		Content analysis of Twitter text	80% data is used for training,10% development, and 10% testing
Ma et al. (2018b)	Two public twitter datasets contain 1381 and 1181 propagation trees respectively	Total 2562 rumour propagation trees	Tweets are classified in four classed: non-rumour, false rumour, true rumour, verified rumour		Text tweets in the form of a tree structure	Experiments conducted on two versions for each tree bottom-up and top-down, by flipping the edges' direction, achieve 73% highest accuracy
Ma et al. (2016)	Twitter dataset of 992 events,491,229 users,1101,985 total posts with average 1,111posts/event	992 events	498 rumours	494 non-rumours/real facts	Hidden representations of contextual information in text tweets over time	10% events are used for model tuning, rest 90% are used for training and testing in a ratio of 3:1; Accuracy 91
Ma et al. (2016)	Sina Weibo dataset of 4,664events,2746,818users,3805,656 total posts with average 816 posts/event	4664 events	2313 rumours	2351 non-rumours/real facts	Hidden patterns of contextual information in text messages over time	10% events are used for model tuning, rest 90% are used for training and testing in a ratio of 3:1; accuracy 91%
Zhang et al. (2018)	Dataset contains 14,055 news articles related to 152 subjects created by 3634 users	14,055 total text articles	Six different credibility labels are assigned to each article (True, Mostly True, Half True, Half False, Mostly false, Totally False)		Tweets and text articles are analyzed for news, creator, and subject	Detection is done by constructing a deep diffusive network model
Thota et al. (2018)	49,973 unique pair of news headlines and article pair of 1684 events.	49,973	Stance between the headline and article is defined as 'agree', 'disagree', 'discuss' or 'unrelated'		Headline, article text and cosine similarity	67% data is used for training and 33% for testing with 3-fold cross-validation; accuracy 94.31%

(*continued on next page*)

Table 7 (*continued*)

Ajao, Bhowmik, and Zargari (2018)	5800 tweets centered at 5 rumour stories	5800 tweets	Each tweet is classified as rumour and non-rumour		Message text and image	10-fold cross-validation and zero paddings is done; accuracy 82%
Perez-Rosas et al. (2017)	480 news excerpts approx. 5 sentences/news	480 news excerpts	240	240	Text of 5 sentences/ news	5-fold cross validation; accuracy 78
Perez-Rosas et al. (2017)	200 news articles focusing celebrities 17 sentences/article	200 news collected from web	100	100	Text of 17 sentences/article	5-fold cross validation; accuracy 78%
Ahmed et al. (2017)	1600 total reviews of 20 most popular hotels in Chicago	1600 hotel reviews	800	800	Text reviews	80% dataset is used for training, 20% for testing with 5-fold cross-validation; Accuracy 90%
Ahmed et al. (2017)	25, 200 political news articles	25,200 articles	12 600	12 600	Text features from news articles	80% dataset is used for training, 20% for testing with 5-fold cross-validation; Accuracy 92%
Jin, Cao, Zhang, Zhou, and Tian (2017)	Real world multimedia dataset from Sina Weibo contains 50,287 tweets and 25,953images	Total 146 events,49,713 tweets, 25 513 images from 42, 310 distinct users	73 events,23 456 tweets,10 231 images and 21 136 users	73 events, 26,257 tweets,15 287 images and 22 584 users	Text and Image visual & statistical features	83.6% highest accuracy,4-fold cross validation for training/validation of each model
Sivasangari, Pandian, Santhya (2018)	5912 total text-tweets of 4 events	5912 text-tweets	2021 rumours	3891 non rumours	Text and user features	Highest accuracy of 90.02%
Elmurngi and Gherbi (2018)	3 different movie review datasets are used	1400 reviews, 2000 reviews, 10,662 reviews	700, 1000, 5331	700, 1000, 5331	Text and sentiment features	Weka tool is used, highest accuracy 81.35%
Vosoughi et al. (2017)	Total 9,38,806 tweets Collected from 3 major events, snopes.com and factcheck.org	Total 209 rumours	113 false rumour	96 true rumour	Text, propagation and sentiment features	Correctly predicts the veracity of 75% rumours. Two HMMs are trained one for true and one for false rumours.
Del Vicario, Quattrociocchi, Scala, and Zollo (2019)	Italian official newspapers Facebook pages for real news and Italian websites Facebook pages for fake news	Total of 75 Facebook pages	17 pages of fake news	58 pages of official news	Content, user, structural, semantic and sentiment features	60% data is used for training and 40% for testing; 77% accuracy for early detection and 91% for fake news detection after spreading.

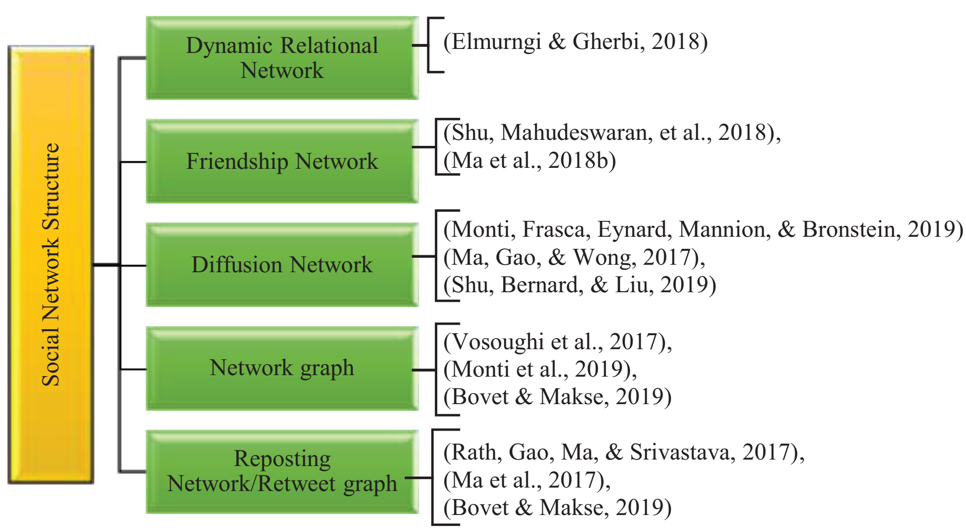

Fig. 10. Different Network structures used in credibility assessment methods. (Ma, Gao, and Wong (2017); Shu, Bernard, and Liu (2019); Bovet and Makse (2019)).

In 2013, three machine learning classifiers Decision tree, random forest, and SVM are used to classify a topic/event as rumor or non-rumor with precision and recall in the range 87% to 92% using rumor propagation as the prominent characteristic (Kwon, Cha, Jung, Chen, & Wang, 2013). To obtain these 11 temporal features, 15 structural features and 65 linguistic features are extracted using LIWC tool to categorize 102 events with at least 60 tweets per event. The temporal feature proposes a new periodic time series model named PES (Periodic external shocks) that considers daily and external shock cycles. Structural properties related to the propagation process are extracted by using the Friendship network and Diffusion Network. Perez-Rosas, Kleinberg, Lefevre, and Mihalcea (2017) focus on linguistic differences based on 2131 features between fake news and legitimate news. They obtained 78% accuracy in detecting fake news on two different datasets using linear SVM classifier.

A credibility analysis system (Alrubaian, Al-Qurishi, Hassan, & Alamri, 2018) based on four components is designed. These com-

Table 8
List of features used for false information detection.

S. No.	Feature	Reference
1	**Text/Content Specific Features**	Chen, Lin, and Yuan (2017), Alrubaian et al. (2018), Elmurngi and Gherbi (2017a; b), Castillo et al. (2011); Diakopoulos et al. (2010), Kwon et al. (2013), Hamidian and Diab (2015), Ma et al. (2018b), Zhang et al. (2018), Perez-Rosas et al. (2017), Ahmed et al. (2017), Yang et al. (2018), Zhang et al. (2015), Elmurngi and Gherbi (2018), Varol et al. (2017), Vosoughi et al. (2017), Del Vicario et al. (2019), Lukasik et al. (2016), Sahana, Pias, Shastri, and Mandloi (2015), Ma, Gao, Wei, Lu, and Wong (2015)
	C1: Number of characters	Castillo et al. (2011), Perez-Rosas et al. (2017), Zhang et al. (2015), Del Vicario et al. (2019)
	C2: Number of words	Del Vicario et al. (2019), Yang et al. (2018), Castillo et al. (2011), Ahmed et al. (2017), Varol et al. (2017), Sahana et al. (2015)
	C3: Number of sentences	Yang et al. (2018), Del Vicario et al. (2019)
	C4: Number of words/Sentence	Yang et al. (2018)
	C5: Question mark, exclamation and capital letters, punctuation	Yang et al. (2018), Castillo et al. (2011), Perez-Rosas et al. (2017), Del Vicario et al. (2019), Sahana et al. (2015), Ma et al. (2015)
	C6: Negations (no, not)	Yang et al. (2018), Del Vicario et al. (2019), Sahana et al. (2015)
	C7: Exclusive words (but, without, however)	Yang et al. (2018)
	C8: First person pronouns (I, we, my)	Ma et al. (2015), Yang et al. (2018), Castillo et al. (2011), Perez-Rosas et al. (2017), Jin et al. (2017)
	C9: Second person Pronouns (you, your)	Yang et al. (2018), Castillo et al. (2011), Perez-Rosas et al. (2017), Jin et al. (2017)
	C10: third person pronouns (he, she)	Yang et al. (2018), Castillo et al. (2011), Perez-Rosas et al. (2017), Jin et al. (2017)
	C11: Sentiment Analysis of text (positive, negative, any other)	Elmurngi and Gherbi (2017b), Elmurngi and Gherbi (2017a),Castillo et al. (2011), Diakopoulos et al. (2010), Kwon et al. (2013), Hamidian and Diab (2015), Perez-Rosas et al. (2017), Ahmed et al. (2017), Alrubaian et al. (2018), Jin et al. (2017), Elmurngi and Gherbi (2018), Vaghela and Patel (2018), Varol et al. (2017), Vosoughi et al. (2017), Del Vicario et al. (2019), Yang et al. (2018), Sahana et al. (2015), Ma et al. (2015)
	C12: Unigram, Bigram, Ngrams	Hamidian and Diab (2015), Perez-Rosas et al. (2017), Ahmed et al. (2017)
	C13: frequently used words	Zhang et al. (2018), Perez-Rosas et al. (2017), Ahmed et al. (2017)
	C14: Number of adjectives, noun, verbs, adverbs	Ahmed et al. (2017)
2	**Image Specific features**	Yang et al. (2018), Jin et al. (2017)
	I1: No. of faces	Yang et al. (2018)
	I2: Resolution of image	Yang et al. (2018)
	I3: Image visual features	Jin et al. (2017)
	I4: Image statistical features	Jin et al. (2017)
3	**User/Account Specific Features**	Alrubaian et al. (2015,2016), Castillo et al. (2011), Aphiwongsophon and Chongstitvatana (2018), Indu and Thampi (2019), Alrubaian et al. (2018), (V et al., 2018), Varol et al. (2017), Vosoughi et al. (2017), Del Vicario et al. (2019), Shu, Mahudeswaran et al. (2018), Shu, Wang, Le, Lee, and Liu (2018), Sahana et al. (2015), Ma et al. (2015), Wu, Yang, and Zhu (2015)
	U1: Is user/account verified	Alrubaian et al. (2015), Castillo et al. (2011), Aphiwongsophon and Chongstitvatana (2018), Indu and Thampi (2019), Jin et al. (2017), Shu, Wang, Le et al. (2018), Sahana et al. (2015), Wu et al. (2015)
	U2: Gender	Alrubaian et al. (2015,2018), Shu, Wang, Le et al. (2018), Wu et al. (2015)
	U3: User Image	Alrubaian et al. (2015), Sahana et al. (2015), Ma et al. (2015)
	U4: User Name	Alrubaian et al. (2015,2016), Aphiwongsophon and Chongstitvatana (2018), Sivasangari, Pandian, Santhya (2018)
	U5: No. of Followers	Alrubaian et al. (2015, 2016), Castillo et al. (2011), Aphiwongsophon and Chongstitvatana (2018), Indu and Thampi (2019), Alrubaian et al. (2018), Jin et al. (2017), Varol et al. (2017), Shu, Mahudeswaran et al. (2018), Shu, Wang, Le et al. (2018), Sahana et al. (2015), Wu et al. (2015)
	U6: No. of Following	Alrubaian et al. (2015,2016), Castillo et al. (2011), Aphiwongsophon and Chongstitvatana (2018), Indu and Thampi (2019), Alrubaian et al. (2018), Jin et al. (2017), Varol et al. (2017), Shu, Mahudeswaran et al. (2018), Shu, Wang, Le et al. (2018), Sahana et al. (2015)
	U7: User Location	Alrubaian et al. (2015), Aphiwongsophon and Chongstitvatana (2018), Shu, Mahudeswaran et al. (2018), Wu et al. (2015)
	U8: No. of tweets/messages	Alrubaian et al. (2015), Castillo et al. (2011), Indu and Thampi (2019), Shu, Mahudeswaran et al. (2018), Wu et al. (2015)
	U9: User/Account Created Date/Time	Alrubaian et al. (2015), Castillo et al. (2011), Aphiwongsophon and Chongstitvatana (2018), Indu and Thampi (2019), Wu et al. (2015)
	U10: Account Status	Alrubaian et al. (2015), Aphiwongsophon and Chongstitvatana (2018), Sahana et al. (2015)
	U11: User Orientation	Alrubaian et al. (2015), Alrubaian et al. (2018)
	U12: Number of Friends	Alrubaian et al. (2016), Castillo et al. (2011), Aphiwongsophon and Chongstitvatana (2018), Varol et al. (2017), Sahana et al. (2015), Ma et al. (2015), Wu et al. (2015)
	U13: Account completion (profile information is complete)	Indu and Thampi (2019), Wu et al. (2015)
4	**Message Specific features**	Alrubaian et al. (2015), Alrubaian et al. (2016), Castillo et al. (2011), Hamidian and Diab (2015), Aphiwongsophon and Chongstitvatana (2018), Indu and Thampi (2019), Alrubaian et al. (2018), Jin et al. (2017), Zhang et al. (2015), Varol et al. (2017), Del Vicario et al. (2019), Shu, Mahudeswaran et al. (2018), Sahana et al. (2015), Ma et al. (2015), Wu et al. (2015)
	M1: Message with URL	Alrubaian et al. (2015), Castillo et al. (2011), Alrubaian et al. (2018), Jin et al. (2017), Zhang et al. (2015), Varol et al. (2017), Sahana et al. (2015), Ma et al. (2015), Wu et al. (2015)

(*continued on next page*)

Table 8 (*continued*)

	M2: Message with hashtag #	Alrubaian et al. (2015), Alrubaian et al. (2016), Castillo et al. (2011), Hamidian and Diab (2015), Aphiwongsophon and Chongstitvatana (2018), Indu and Thampi (2019), Alrubaian et al. (2018), Jin et al. (2017), Sahana et al. (2015)
	M3: Message with mention @	Alrubaian et al. (2015), Alrubaian et al. (2016), Castillo et al. (2011), Aphiwongsophon and Chongstitvatana (2018), Alrubaian et al. (2018)
	M4: Message Source	Alrubaian et al. (2015), Zhang et al. (2015)
	M5: Message Created date	Alrubaian et al. (2015)
	M6: Number of replies	Alrubaian et al. (2015), Alrubaian et al. (2018), Sivasangari, Pandian, Santhya (2018), Del Vicario et al. (2019), Shu, Mahudeswaran et al. (2018)
	M7: Number of mentioned @	Alrubaian et al. (2015), Alrubaian et al. (2016), Castillo et al. (2011), Aphiwongsophon and Chongstitvatana (2018), Varol et al. (2017), Ma et al. (2015)
	M8: Number of hashtags #	Alrubaian et al. (2015), Alrubaian et al. (2016), Castillo et al. (2011), Aphiwongsophon and Chongstitvatana (2018), Varol et al. (2017), Ma et al. (2015)
	M9: Time of Posting message	Alrubaian et al. (2016), Shu, Mahudeswaran et al. (2018)
	M10: Is Retweeted	Castillo et al. (2011), Hamidian and Diab (2015), Alrubaian et al. (2018)
	M11: Contains emoticon (smile, Frown)	Castillo et al. (2011), Hamidian and Diab (2015)
	M12: Retweet count for a particular message	Sivasangari, Pandian, Santhya (2018), Shu, Mahudeswaran et al. (2018)
	M13: Total No. of Likes	Indu and Thampi (2019), Jin et al. (2017), Sivasangari, Pandian, Santhya (2018), Del Vicario et al. (2019), Shu, Mahudeswaran et al. (2018)
5	**Propagation Features**	Castillo et al. (2011), Varol et al. (2017), Vosoughi et al. (2017), Jin et al. (2017), Del Vicario et al. (2019)
	P1: Degree of root in propagation tree	Castillo et al. (2011)
	P2: Total No. of tweets in largest subtree of root	Castillo et al. (2011)
	P3: Maximum and Average degree of a node (except root)	Castillo et al. (2011),Jin et al. (2017)
	P4: Depth of the propagation tree	Castillo et al. (2011)
	P5: Max size of a level in the propagation tree	Castillo et al. (2011)
	P6: Size of max subtree	Jin et al. (2017)
6	**Temporal Features**	Kwon et al. (2013), Aphiwongsophon and Chongstitvatana (2018), Sivasangari, Pandian, Santhya (2018), Varol et al. (2017), Vosoughi et al. (2017), Shu, Mahudeswaran et al. (2018), Shu, Wang, Le et al. (2018), Lukasik et al. (2016), Ma et al. (2015), Buntain and Golbeck (2017), Poddar, Hsu, Lee, and Subramaniyam (2018)
7	**Structural Features**	Kwon et al. (2013), Ma et al. (2018b), Del Vicario et al. (2019), Buntain and Golbeck (2017)
8	**Linguistic Features**	Kwon et al. (2013), Perez-Rosas et al. (2017), Varol et al. (2017), Vosoughi et al. (2017)

Table 9
Scalable synthetic social network graph generators.

Synthetic graph generator	Salient features	Ref.
Darwini	Can be used efficiently to study propagation and detection of false contents by means of generating different social connections in the form of a graph for which darwini can produce local clustering coefficient, degree distributions, node page rank, eigenvalues and many other matrices.	Edunov, Logothetis, Wang, Ching, and Kabiljo (2016)
DataSynth	Scalable synthetic graph generator with customizable schemas and properties. Introduces novel features of representing the correlation between the structure of a graph and properties.	Prat-Pérez et al. (2017)
BTER	Capture clustering coefficient and degree distribution, useful in reproducing graphs with massive community structure network.	Kolda, Pinar, Plantenga, and Seshadhri (2013)
Myriad	A toolkit for expressive data generator programs can generate nodes and edges data for visualizing and experimenting online social network connections. The naive feature is that can be executed in a massively parallel manner.	Alexandrov, Tzoumas, and Markl (2012)
R-MAT	“recursive matrix “a simple, parsimonious graph model that can quickly generate realistically weighted, directed and bipartite graphs. Diverse real social network and web connectivity graphs can be well approximated by an R-MAT model with appropriate choice of parameters.	Chakrabarti, Zhan, and Faloutsos (2004)
LFR	Graph generator used to evaluate community detection algorithms. Capable of clustering large graphs that exceed main memory using external memory.	Hamann, Meyer, Penschuck, Tran, and Wagner (2018)
gMark	schema-driven, domain-independent, highly configurable and extensible graph instance, and query workload generator. Practical usability has increased many folds with its customizable schemas for varied application domains.	Bagan et al. (2016)
Apache Spark framework	Basic properties of power-law distribution of the number of user communities, the dense intersections of social networks, and others are used to generate a graph similar in structure to existing social networks. A very small amount of resources and faster execution speed in comparison with other similar generators.	Belov and Vovchok (2017)
Attributes Synthetic Generator (ASG)	Consider feature similarity and label homophily among individuals when forming links in the network. To tune the social network parameters exactly to the generated network particle, swarm optimization is used. shared similarity among individuals to form the links in the network. Statistics taken from real OSNs are used to form the nodes attributes. Time efficient and require only limited parameter optimization.	Kiran and Gita (2014)
Multi-Link Generator (MLG)	Follows the preferential attachment model for handling multiple networks that contain different link types. The model starts with few nodes and as networks grow, more nodes and links are added to the model. MLG is scalable and efficient in time and parameter optimization.	Kiran and Gita (2014)

ponents include (a) Reputation component, measures user reputation and expertise (b) Feature ranking algorithm, weighting each feature according to its relative importance (c) User experience component, rank users based on their relevance on a given topic and (d) Credibility classifier engine, distinguishes between credible and non-credible contents. These four components work together in an algorithmic way to authenticate tweets and users. The major loophole of the proposed method (Alrubaian et al., 2018) is that they have not incorporated any temporal, location and image features which could be crucial in many circumstances. Images have

Table 10
Features used by various machine-learning classifiers.

Features	ML classifier									
	SVM	KNN	NB	K*	DT	DR	RF	SGD	LoG	LR
User features	Shu, Mahudeswaran et al. (2018), Castillo et al. (2011), Jin et al. (2017), Del Vicario et al. (2019), Aphiwongsophon and Chongstitvatana (2018)	Del Vicario et al. (2019), Varol et al. (2017)	Castillo et al. (2011), Sivasangari, Pandian, Santhya (2018), Alrubaian et al. (2018,2015), Aphiwongsophon and Chongstitvatana (2018)	Jin et al. (2017)	Castillo et al. (2011), Del Vicario et al. (2019), Alrubaian et al. (2018), Sahana et al. (2015)	Castillo et al. (2011)	Jin et al. (2017), Alrubaian et al. (2018), Lorek, Suehiro-Wicinski, Jankowski-Lorek, and Gupta (2015)	–	Del Vicario et al. (2019), Jin et al. (2017)	–
Message features	Shu, Mahudeswaran et al. (2018), Castillo et al. (2011), Jin et al. (2017), Del Vicario et al. (2019), Aphiwongsophon and Chongstitvatana (2018)	Del Vicario et al. (2019), Varol et al. (2017)	Castillo et al. (2011), Sivasangari, Pandian, Santhya (2018), Alrubaian et al. (2018), Alrubaian et al. (2015), Aphiwongsophon and Chongstitvatana (2018)	–	Castillo et al. (2011), Del Vicario et al. (2019), Alrubaian et al. (2018), Sahana et al. (2015)	Castillo et al. (2011)	Alrubaian et al. (2018)	–	Del Vicario et al. (2019), Zhang et al. (2015)	–
Sentiment analysis	Elmurngi and Gherbi (2017a,b), Jin et al. (2017), Castillo et al. (2011), Elmurngi and Gherbi (2018), Del Vicario et al. (2019)	Elmurngi and Gherbi (2017a,b,2018), Del Vicario et al. (2019), Varol et al. (2017)	Elmurngi and Gherbi (2017a,b), Castillo et al. (2011), Elmurngi and Gherbi (2018), Alrubaian et al. (2018)	Elmurngi and Gherbi (2017b,2018)	Elmurngi and Gherbi (2017a,b), Castillo et al. (2011), Del Vicario et al. (2019), Alrubaian et al. (2018)	Castillo et al. (2011)	Kwon et al. (2013), Alrubaian et al. (2018)	–	Del Vicario et al. (2019)	–
Text features/ Classification	Elmurngi and Gherbi (2017a,b), Ahmed et al. (2017), Jin et al. (2017), Del Vicario et al. (2019), Ma et al. (2015)	Elmurngi and Gherbi (2017a,b), Ahmed et al. (2017), Del Vicario et al. (2019), Varol et al. (2017)	Elmurngi and Gherbi (2017a,b), Sivasangari, Pandian, Santhya (2018), Elmurngi and Gherbi (2018), Alrubaian et al. (2018)	Elmurngi and Gherbi (2017b), Jin et al. (2017)	Elmurngi and Gherbi (2017a,b), Ahmed et al. (2017), Del Vicario et al. (2019)	–	Jin et al. (2017), Alrubaian et al. (2018), Fairbanks et al. (2018)	Ahmed et al. (2017)	Jin et al. (2017), Del Vicario et al. (2019), Zhang et al. (2015), Fairbanks et al. (2018)	Ahmed et al. (2017)

(*continued on next page*)

Table 10 (*continued*)

Topic features	Castillo et al. (2011)	–	Castillo et al. (2011), Alrubaian et al. (2018)	–	Castillo et al. (2011), Alrubaian et al. (2018)	Castillo et al. (2011)	Alrubaian et al. (2018)	–		–
Propagation features	Castillo et al. (2011), Jin et al. (2017), Del Vicario et al. (2019)	Del Vicario et al. (2019), Varol et al. (2017)	Castillo et al. (2011)	Jin et al. (2017)	Castillo et al. (2011), Del Vicario et al. (2019)	Castillo et al. (2011)	Kwon et al. (2013), Jin et al. (2017)	–	Jin et al. (2017), Del Vicario et al. (2019)	–
Structural features	Del Vicario et al. (2019)	Del Vicario et al. (2019)	–	–	Hamidian and Diab (2015), Del Vicario et al. (2019)	–	Kwon et al. (2013), Fairbanks et al. (2018)	–	Fairbanks et al. (2018), Del Vicario et al. (2019)	–
Linguistic features	Perez-Rosas et al. (2017)	Varol et al. (2017)	–	–	–	–	Kwon et al. (2013)	–	–	–
Temporal features	Ma et al. (2015), Aphiwongsophon and Chongstitvatana (2018)	Varol et al. (2017)	Shu, Mahudeswaran et al. (2018), Sivasangari, Pandian, Santhya (2018), Aphiwongsophon and Chongstitvatana (2018)	–	–	–	Kwon et al. (2013)	–	–	–

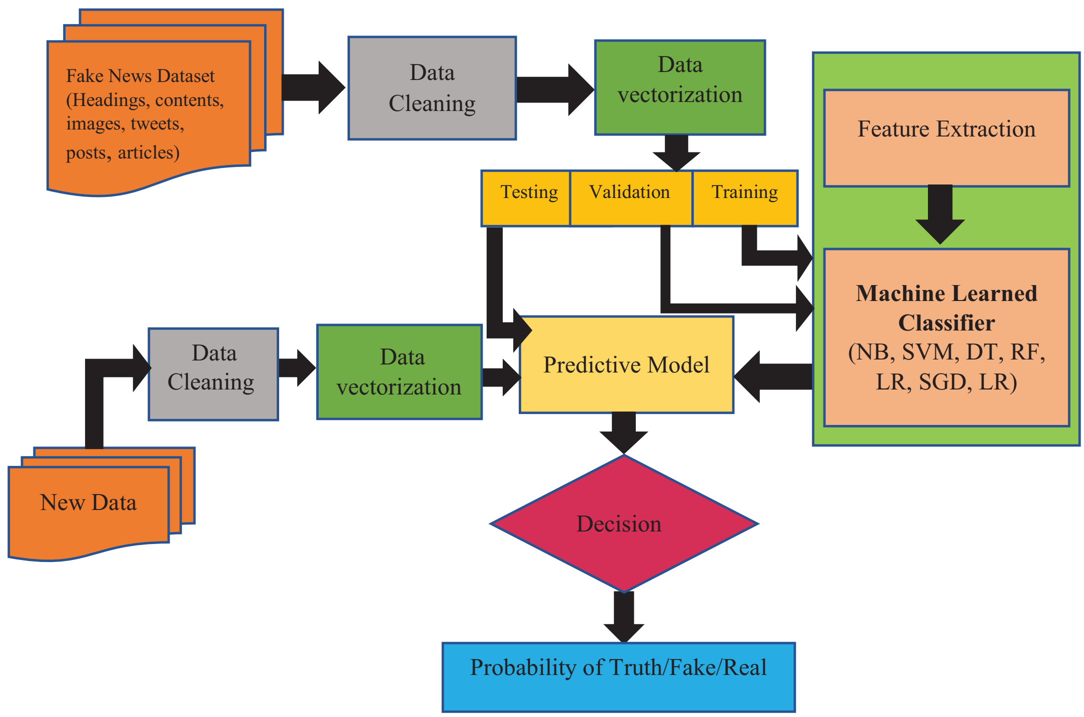

Fig. 11. Machine learning architecture of fake information detection.

a significant influence on microblog news propagation, as it is very efficient to describe an event with attached photographs. A novel attempt (Jin, Cao, Guo, Zhang, & Luo, 2017) explores five different visual and seven statistical features along with existed user, text and propagation features to train the SVM, Logistic Regression, K star and Random forest model with four-fold cross-validation obtains 83.6% highest accuracy. The proposed efficiency can be further enhanced by extracting the latent image features using kernel operations.

A salient example of the spread of fake news that has devastating adverse effects is health rumors (Viviani & Pasi, 2017; Zhang, Zhang, & Li, 2015). The internet health rumors including larger headlines tend to be fake. The information that contains precise numerical or textual data such as the name of a person or place is more likely to be true. The web hyperlinks to established trusted sources increased credibility many folds. These are some of the critical observations drawn after analyzing a dataset of 453 health rumors using a logistic regression supervised machine learning model, out of which 113 are true, and 340 are false (Zhang et al., 2015). Text classification along with sentiment analysis (Elmurngi & Gherbi, 2018) proved to be an effective method of fake review detection on three different movie review datasets. One of the largest number of features 487 (Varol, Ferrara, Menczer, & Flammini, 2017) related to content, user, sentiment, message, temporal, linguistic, propagation is used for early detection of promoted social media campaigns with supervised KNN-DTW (K-nearest neighbor with dynamic time wrapping) having 75% accuracy score for early detection and above 95% after trending. The text content and structural based two separate approaches using logistic regression and random forest with TF-IDF and doc2vec embeddings (Fairbanks, Fitch, Knauf, & Briscoe, 2018) are analysed for bias detection as well as credibility assessment. Experimental results show that the content-based model outperforms for bias detection as it reveals patterns and peculiarities in the underlying text of dataset. The structural model detects fake news perfectly because of the presence of adversarial writing process as fake news authors deliberately opt a different writing style to convince readers.

5.5. Deep learning

The major disadvantage of machine-learning-based models is that they are dependent on hand-crafted features that require exhaustive human efforts as well as meticulous, detailed and biased. Thus, recent technologies are shifting the trend towards deep learning-based models. Deep models extract hidden features and representations in text, images, sentiments, structure and variation in the context of tweets over time to detect fraudulent information diffused online.

Fig. 13 explains generalized deep learning architecture used in various algorithms classified in Fig. 14 and Table 11 for false content detection. A model named Text Image Convolutional Neural Network (TI-CNN) was proposed by Yang et al. (2018) focuses on explicit features and latent features extracted from text and images using CNN. Then the text and image branches are concatenated and finally, the sigmoid activation layer is used to provide the final label to the news. The explicit text features used in this research have some interesting findings based on the results of TI-CNN model on their dataset. According to the findings, fake news has a smaller number of words and sentences than real news. Much of the fake news is spread as tweets and hyperlink on social media with no headings. To become more appealing and draw the reader's attention falsified contents have more no. of exclamation marks, capital letters, and question marks and exhibits a negative sentiment. Images supporting the fraudulent news stories are most of the time irrelevant; contain less no. of human faces, exhibit vague information and very low resolution.

Zhang, Cui, Fu, and Gouza (2018)devised a novel deep diffusive network model of fake news detection based on hybrid feature extraction. Recurrent Neural Network learns explicit features based on most frequently used words extracted by Bag-of-words and latent features incorporate a GRU in the hidden layer. The proposed method is compared with several baseline machine learning and deep learning-based models where it outperforms on many criteria. Still, the technique has a lot of future scope of improvement by adding many temporal, structural and pragmatic features.

Table 11
Input data format used by deep learning classifiers.

Input data	Deep learning							
	CNN	Deep NN	Recursive NN	Recurrent NN	LSTM	GRU	MLP	Dense NN
Text and image	Yang et al. (2018), Ajao et al. (2018)	–	–	Ajao et al. (2018), Jin et al. (2017)	Ajao et al. (2018), Jin et al. (2017), Khattar, Goud, Gupta, and Varma (2019)	–	–	–
Text and headline	–	Singhania et al. (2017)	Davis and Proctor (2017)	Peng (2018), Esmaeilzadeh et al. (2019), Borges, Martins, and Calado (2019)	Davis and Proctor (2017), Esmaeilzadeh et al. (2019), Sadiq, Wagner, Shyu, and Feaster (2019), Conforti et al. (2018)	Davis and Proctor (2017)	Davis and Proctor (2017)	–
Text tweets/messages/news	Poddar et al. (2018), Ma, Gao, and Wong (2018a), Chen, Liu, and Kao (2017)	Ma et al. (2018a)	Ma et al. (2018b)	Ma et al. (2016), Poddar et al. (2018), Rath, Gao, Ma, and Srivastava (2017), Wu and Liu (2018)	Ma et al. (2016), Wu and Liu (2018)	Ma et al. (2016), Rath et al. (2017)	–	–
Text tweet, news article, creator, subject, metadata	Wang (2017), Monti, Frasca, Eynard, Mannion, and Bronstein (2019), Roy et al. (2018), Qian et al. (2018), García Lozano, Lilja, Tjörnhammar, and Karasalo (2017), Karimi, Roy, Saba-Sadiya, and Tang (2018)	–	–	Zhang et al. (2018), Girgis, Amer, and Gadallah (2018), Chuan et al. (2019)	Wang (2017), Roy et al. (2018), Girgis et al. (2018), Zhang, Lipani, Liang, and Yilmaz (2019), Karimi et al. (2018)	Zhang et al. (2018), Girgis et al. (2018), Chuan et al. (2019)	Roy et al. (2018), Zhang et al. (2019)	–
Headline, article, cosine similarity b/w headline and article	–	–	–	–	–	–	–	Thota et al. (2018)

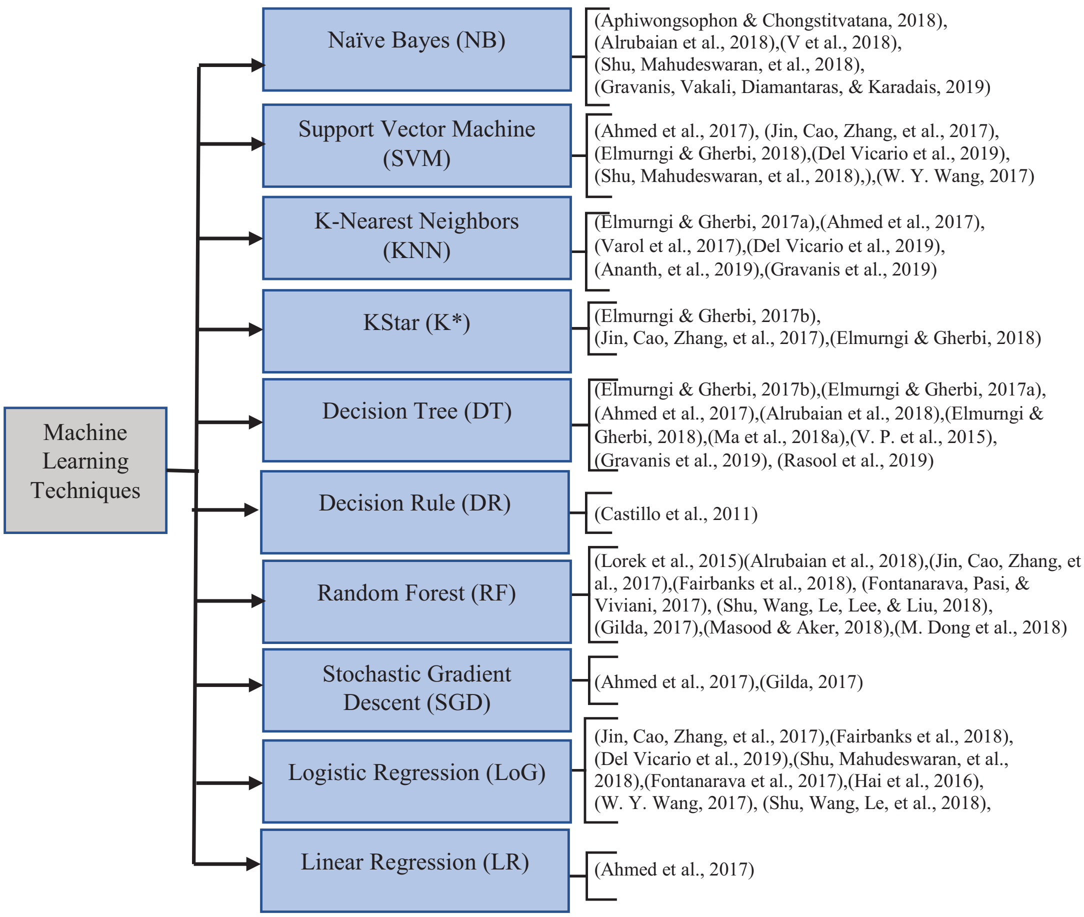

Fig. 12. Classification of different machine learning techniques. (The references cited in this figure are Gravanis, Vakali, Diamantaras, and Karadais (2019); Ananth, Radha, Prema, and Niranjan (2019); Rasool, Butt, Shaukat, and Akram (2019); Fontanarava, Pasi, and Viviani (2017); Gilda (2017); Masood and Aker (2018); Hai et al. (2016)).

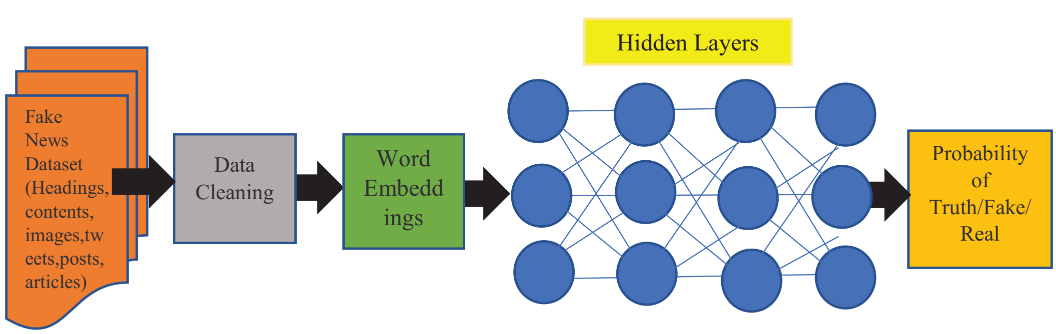

Fig. 13. Deep learning Architecture of fake information detection.

In 2017, an automated detector of fake news was proposed by Singhania, Fernandez, and Rao (2017) with 96.24% accuracy using three bottom-up(word, sentence, headline) levels of hierarchical attention network(3HAN) where attention weights are given to different parts of an article based on their importance. The pre-training of the 3HAN model outperforms with an accuracy of 99.77% as compared to normal 3HAN. The main advantage of attention mechanism is that along with increased accuracy, it provides a complete visualization of the internal classification process at different layers whereas non-attention-based models work like a black box. Ma et al. (2016) proposed models based on Recurrent neural networks in 2016 and Recursive neural networks in 2018 (Ma, Gao, & Wong, 2018b). Recurrent NN uses hidden contextual representations and their variations over time to train the tanh-RNN basic model which is further enhanced by using LSTM-1, GRU-1 and GRU-2 along with embeddings to improve the accuracy up to 91 %. Also, another method of detecting rumors is proposed based on top-down and bottom-up tree-structured Recursive neural networks which deeply integrates the structural and textual propertied of tweets for detecting rumors at early stages. TD-RvNN and BU-RvNN do not incorporate user, message and sentiment-oriented features is a major drawback of the proposed method which needs to be addressed.

Thota, Tilak, Ahluwalia, and Lohia (2018) have done significant work towards stance detection between the headline and text article using dense neural networks, classifying the stance in four categories ‘agree’, ’disagree’, discuss’ and ‘unrelated’. Three types of embedding TF-IDF, Bag of words and Word2vec with cosine similarity between the headline and text are used with dense neural networks, giving accuracies of 94.31%, 89.23% and 75.67% respectively. Roy, Basak, Ekbal, and Bhattacharyya (2018) uses an ensemble framework of CNN and LSTM along with Multi-layer perceptron

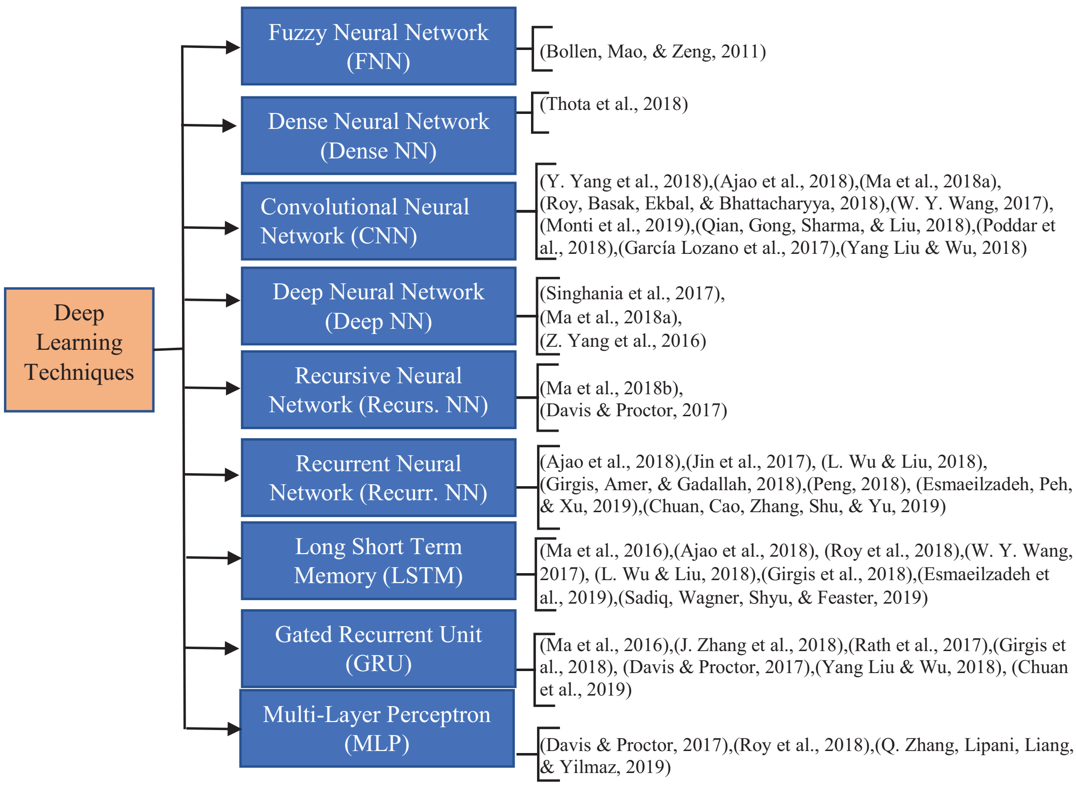

Fig. 14. Classification of different deep learning techniques. (The references cited in this figure are Bollen, Mao, and Zeng (2011); Yang et al. (2016); Liu and Wu (2018)).

model (MLP). Peng (2018) proposed RNN with Bimodal Distribution Removal (BDR) algorithm using tanh and Word2Vec. A novel two-level CNN with collective user intelligence (Qian, Gong, Sharma, & Liu, 2018), abstractive text summarization (Esmaeilzadeh, Peh, & Xu, 2019) using neural models and framework of Emotion-based fake news detection(EFN) (Chuan, Cao, Zhang, Shu, & Yu, 2019) with RNN and GRU are proposed to achieve better model accuracy. Cross-level stance detection (Conforti, Pilehvar, & Collier, 2018) in news articles with LSTM for credibility analysis takes into account the stance of supporting articles along with the main news article. CSI (Capture, Score and Integrate) (Ruchansky, Seo, & Liu, 2017) is a hybrid deep model for deception detection that integrates multiple new technologies.

5.6. Other methods

Cognitive psychology is a method of analyzing human perceptions. The cognitive process examines four main ingredients coherency of the message, credibility of the source, consistency of message, general acceptability of message using collaborative filtering property of social networks to detect misinformation, disinformation, and propaganda (Kumar & Geethakumari, 2014). The proposed genetic framework measures the credibility of the source of information as well as the quality of new ideas on twitter dataset with 90% accuracy. A system Rumour Gauge (Vosoughi, Mohsenvand, & Roy, 2017) is designed to accurately predict the veracity of real-world rumors on Twitter before verification by trusted channels using Hidden Markov Model. However, the system is incapable of differentiating between malicious and accidental misinformation. The stance is the overall position held by a person towards an idea, object or belief. Review of different methods of rumor identification using stance classification in four categories of supporting, denying, Querying and Commenting is presented in Zubiaga et al. (2018). The work is done in various areas of knowledge-based, style-based, propagation-based, user-based and credibility based fake news detection including manual as well as automatic fact-checking in homogeneous, heterogeneous and hierarchical networks are summarized by Zhou and Zafarani (2018). Kumar and Shah (2018) focused on three types of false information opinions based on fake reviews, Fact-based hoaxes and intent-based disinformation.

O'Brien, Simek, and Waugh (2019) proposed an iterative Graph-based method of credibility classification. Tri-relationship among publisher, news piece and user (Shu, Wang, & Liu, 2019) explores the role of social context for trustworthiness analysis. Shu, Wang, and Liu (2018) try to improve fake news detection accuracy by exploring different characteristics of social media user profiles based on experienced and naïve users. Hawkes process (Lukasik et al., 2016) is a probabilistic framework of fake news detection. Investigative journalism and wisdom of crowd (Liu, Nourbakhsh, Li, Fang, & Shah, 2015), unsupervised Bayesian network (Shu et al., 2019), filter out misleading and false websites (Aldwairi & Alwahedi, 2018) are some of the other prominent methods of content analysis. Veracity analysis of fake news by scrapping and authenticating the web search is proposed in Vishwakarma, Varshney, and Yadav (2019).

6. Containment and intervention

Twitter data is extensively used to analyze the rumor spread during and after the Great Japan Earthquake of March 11, 2011 (Miyabe, Nadamoto, & Aramaki, 2014), performing a comparative study of disaster and normal situation tweets and spreading patterns. The work concluded with establishing the fact that rumor tweets spread easily, but rumor disaffirmation tweets do not spread more than a few nodes in the network. Anti-rumour news and campaigns are used to alleviate the spreading of rumor. Software developers and technology firms have

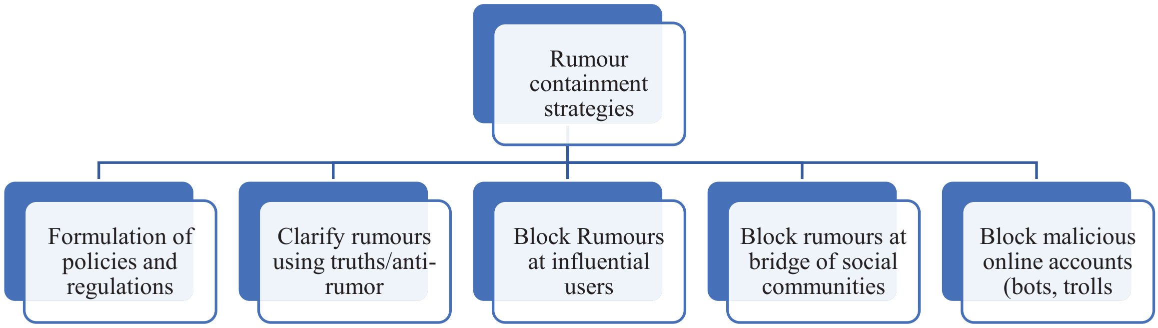

Fig. 15. Classification of rumor containment strategies.

begun developing human-driven mechanisms as well as tools to identify and quarantine fake news. Mainstream news organizations also constitute teams of fact-checkers and investigating units. Fig. 15 classifies some of the prominent technologies used to intervene in the spread of malicious content online.

Significant efforts for the mitigation of fraudulent content are done in Nguyen, Yan, Thai, and Eidenbenz (2012) by identifying a set of highly influential nodes, which are decontaminated first and in turn diffuse the confirmed news in their complete friend network. In a small size, social network GVS (Greedy viral stopper) algorithm is used to find out the set of most influential nodes. If the network structure is very vast, then the community-based heuristic algorithm is used. The highest disadvantage of this method is that is has assumed that facts and misinformation spread with the same rate in a network which proves out to be false in many research studies.

The authors in Starbird, Maddock, Orand, Achterman, and Mason (2014) found that the crowd has the potential to self-correct. Corrections to the misinformation emerge in the social networks themselves but are muted and not widely propagated. In order to mitigate the rumors in Vehicular Social Networks (Wu, Huang, Zhao, Wang, & Wang, 2018), a specially authorized node is introduced in each network which has the responsibility of spreading anti-rumor messages to spread correct information. Blocking rumors at highly influential users and at the community bridges are two main strategies of proactive measures along with the remedial method of spreading truths to mitigate information pollution. A mathematical model (Wen et al., 2014) based on the categorization of the population in susceptible, defended, recovered(active, immunized), infected(contagious, misled) is introduced to investigate the methods of rumor containment with parameters of degree, betweenness, core, overlapped and separated. By predicting, the possible future path of rumor propagation can try to block it at influential users and bridge of social communities.

Formulation of policies and regulations for contents posted on social media and legal laws for wrongdoers will motivate the users to think rationally before resharing or posting. Social bots, which are social media accounts operated by computer algorithms can give a wrong impression pertaining to the popularity of information and endorsed by many people that enable the echo chamber effect for the propagation of fake news. Apart from social bots, cyborg users and trolls are also malicious accounts that amplify the spread of fake news must be blocked (Shu, Mahudeswaran et al., 2018). Community signals, user's flags and expert opinions (Tschiatschek, Singla, Rodriguez, Merchant, & Krause, 2018) leverage the detection as well as minimize the spread of fraudulent information by stopping the propagation paths. "Fake news game" (Roozenbeek & van der Linden, 2019) is an educational game that provides key containment strategies to inoculate the public against the risk of fake news.

7. Potential research opportunities

A lot of work has been done in the past years to make online content more reliable and trustful and some of the key areas remain unaddressed. The following section highlights the current research gap and potential future direction of work. Quick and real-time detection of the source is useful to control the spread of false information and reduce the adverse impact on society. Real-time collected datasets, automatic detection of rumors and finding its original source is a challenging issue.

- **Cross-platform detection**: As people have accounts on various social networking websites and sometimes, they spread the rumor across their different social networks, in such cases source detection becomes somewhat difficult. Along with this propagation of false information from one web community to another i.e. cross-platform spread and detection has become a significant challenge for tracking in front of the researchers.
- **Real-time learning**: Deployment of a web-based application for fact-checking which can learn in real-time from new manually fact-checked articles and provides real-time detection of fraudulent information.
- **Unsupervised models**: Current work is mainly done by using supervised learning approaches. Unsupervised models need to be developed due to massive unlabelled data from social media.
- **Datasets**: The establishment of convincing gold standard datasets in this field is highly required as most of the research is being done on customized datasets. Because of the lack of publicly available large-scale datasets a benchmark comparison between different algorithms cannot be done.
- **Multilingual platform**: Most of the work focuses on linguistic features in English language text. Other popular and regional languages (multilingual platform for fake news detection) are not considered yet.
- **Complex and dynamic network structure**: The veracity classification task becomes a prediction task if we are doing it before its resolution and requires a huge amount of supporting evidence. The issue further complicates because of the complex and dynamic network structure of social platforms.
- **Early detection**: Detecting fake news at the early stage is a highly challenging task before it becomes widespread so that timely actions can be taken for its mitigation and intervention. After fake news has become widespread and gained users' trust, it's almost impossible to change people's perception.
- **Cross-domain analysis**: Most of the existing approach focuses only on one way of deception detection either in the form of content, propagation, style, etc. Cross-

domain analysis, including multiple aspects such as topic-website-language-images-URL, helps in identifying unique non-varying characteristics, provides early accurate detection of fraudulent content.

- **Deep learning**: Deep learning technologies can address all formats of information text, image, speech and video. Deep architecture is customizable to a new class of problem and it bypasses feature engineering, which is the most time consuming but necessary part of a machine-learning framework. However, the disadvantage of deep learning technologies is that they require a considerable amount of time for model training with a relatively massive amount of data and do not provide interpretations of what the model has actually learned, so inside the model it is almost a black box type of processing.
- **Multimedia false information detection**: fabricated and manipulated audio, images and videos need developing data analytics, computer vision and signal processing techniques. To discover signature characteristics of manipulated and fabricated multimedia machine learning and deep learning algorithms are highly required.
- **Bridging echo chambers**: Social media is prone to form echo chambers when a user's existing beliefs, views are reinforced and he is not aware of the opposite beliefs. Therefore, further research is required to bridge the conflicting echo chambers in order to effectively exchange the opposing beliefs to readers so that polarization can be reduced. It also helps in truth discovery by making users think judiciously and rationally on multiple dimensions.

8. Social and methodological findings

Information pollution is a real-time practical issue that is being faced by each one of us every day on sharing or consuming a piece of information from WhatsApp, Twitter, Facebook, Instagram, Quora or other social networking and online platforms. The domain of this study is highly pertinent to pragmatic human life. This survey study focused on theories from mathematics, natural language processing, artificial intelligence, machine learning, deep learning, data sciences and human psychology to solve the current burning issue of truthfulness and credibility analysis of web content. Most of the work done to date in literature establishes reasonable theories by drawing synergy between the practical problem domain and available methods. Few positive attempts to avail the methods for everyday use have also been done in terms of APIs and extension to web browsers for fact-checking. Despite all these attempts still there are several functionalities to be inbuilt for real-time fact-checking by incorporating all possible scenarios of contaminating the online content. In this study we tried to summarize the practical social issue that can be methodologically solved using computer science and data analytics theories.

9. Conclusion

Information pollution, fake news, rumours, misinformation, disinformation has become a by-product of the digital communication ecosystem, which proves to be very dangerous. This review work presents the impact analysis, characterization, compare and comprehensively evaluate the current scenario of methods, technologies, tools to quarantine the malice of information pollution. This paper tries to provide a holistic view of information pollution ecosystem in terms of taxonomy of fraudulent contents, lifecycle of a complete ecosystem, different social digital communication platforms, primary driving forces behind disinformation spread and different credibility analysis platforms. Then provides a completely technical standpoint to the issue right from creation to disposition focusing on source identification, propagation dynamics, detection methods and containment strategies. Approximately 40% of the studied research concentrated on detection of false content using machine learning and deep learning implicit as well as explicit feature engineering and pattern analysis techniques. Finally, open issues and challenges are also highlighted to further explore potential research opportunities.

This work may be helpful to the new researchers to understand the different components of digital online communication from a social and technical perspective. Multilingual cross-platform fake news spreading, complex and dynamic network structure, huge volumes of unlabelled real-time data and early detection of rumors are some challenging issues that are still unaddressed and need further research. Improving the reliability and future of online information ecosystem is a joint responsibility of the social community, digital policymakers, administration, technical and research scholars.

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.

Credit authorship contribution statement

Priyanka Meel: Software, Validation, Investigation, Visualization, Writing - original draft, Data curation. **Dinesh Kumar Vishwakarma:** Conceptualization, Methodology, Formal analysis, Writing - review & editing, Supervision, Project administration, Funding acquisition, Resources.

References

Ahmed, H., Traore, I., & Saad, S. (2017). Detecting opinion spams and fake news using text classification. *Security and Privacy, 1*(1) e9. doi:10.1002/spy2.9.

Ajao, O., Bhowmik, D., & Zargari, S. (2018). Fake news identification on Twitter with hybrid CNN and RNN models. In *Proceedings of the 9th international conference on social media and society* (pp. 226–230). ACM. (July) Copenhagen, Denmark. doi:10.1145/3217804.3217917.

Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. *Procedia Computer Science, 141*, 215–222. doi:10.1016/j.procs.2018.10.171.

Alexandrov, A., Tzoumas, K., & Markl, V. (2012). Myriad: Scalable and expressive data generation. *Proceedings of the VLDB Endowment, 5*(12), 1890–1893. doi:10.14778/2367502.2367530.

Al-Qurishi, M., Al-Rakhami, M., Alrubaian, M., Alarifi, A., Rahman, S. M. M., & Alamri, A. (2015). Selecting the best open source tools for collecting and visualzing social media content. In *2015 2nd world symposium on web applications and networking, WSWAN* (pp. 1–6). Sousse, Tunisia. doi:10.1109/WSWAN.2015.7210346.

Alrubaian, M., Al-Qurishi, M., Al-Rakhami, M., Hassan, M. M., & Alamri, A. (2016). CredFinder: A real-time tweets credibility assessing system. In *Proceedings of the 2016 IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM 2016* (pp. 1406–1409). San Francisco, USA. doi:10.1109/ASONAM.2016.7752431.

Alrubaian, M., Al-qurishi, M., Al-rakhami, M., Rahman, S. M. M., & Alamri, A. (2015). A multistage credibility analysis model for microblogs. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 1434–1440). Paris, France. doi:10.1145/2808797.2810065.

Alrubaian, M., Al-Qurishi, M., Hassan, M. M., & Alamri, A. (2018). A credibility analysis system for assessing information on Twitter. *IEEE Transactions on Dependable and Secure Computing, 15*(4), 661–674. doi:10.1109/TDSC.2016.2602338.

Ananth, S., Radha, K., Prema, S., & Niranjan, K. (2019). Fake news detection using convolution neural network in deep learning. *International Journal of Innovative Research in Computer and Communication Engineering, 7*(1), 49–63. doi:10.15680/IJIRCCE.2019.

Aphiwongsophon, S., & Chongstitvatana, P. (2018). Detecting fake news with machine learning method. In *IEEE 15th international conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON)* (pp. 528–531). Chiang Rai, Thailand. doi:10.1109/ECTICon.2018.8620051.

Bagan, G., Bonifati, A., Ciucanu, R., Fletcher, G. H. L., Lemay, A., & Advokaat, N. (2016). gMark: Schema-driven generation of graphs and queries. *IEEE Transactions on Knowledge and Data Engineering, 29*(4), 856–869. doi:10.1109/TKDE.2016.2633993.

Belov, Y. A., & Vovchok, S. I. (2017). Generation of a social network graph by using apache spark. *Automatic Control and Computer Sciences, 51*(7), 678–681. doi:10.3103/s0146411617070264.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8. doi:10.1016/j.jocs.2010.12.007.

Borges, L., Martins, B., & Calado, P. (2019). Combining similarity features and deep representation learning for stance detection in the context of checking fake news. *ACM Journal of Data and Information Quality (JDIQ), 11*(3), 1–24.

Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications, 10*(1), 1–14. doi:10.1038/s41467-018-07761-2.

Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *Journal of Applied Research in Memory and Cognition, 8*(1), 108–117. doi:10.1016/j.jarmac.2018.09.005.

Buntain, C., & Golbeck, J. (2017). Automatically identifying fake news in popular Twitter threads. In *2nd IEEE international conference on smart cloud, SmartCloud 2017* (pp. 208–215). New York, USA. doi:10.1109/SmartCloud.2017.40.

Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on Twitter. In *20th international conference on World Wide Web. ACM, 2011* (pp. 675–684). Hyderabad, India.

Chakrabarti, D., Zhan, Y., & Faloutsos, C. (2004). R-MAT: A recursive model for graph mining. In *Proceedings of the 2004 SIAM international conference on data mining* (pp. 442–446). Society for Industrial and Applied Mathematics. Lake Buena Vista, Florida, USA. doi:10.1137/1.9781611972740.43.

Chen, J., Song, Q., & Zhou, Z. (2017). Agent-based simulation of rumor propagation on social network based on active immune mechanism. *Journal of Systems Science and Information, 5*(6), 571–584. doi:10.21078/jssi-2017-571-14.

Chen, S., Lin, L., & Yuan, X. (2017). Social media visual analytics. *Computer Graphics Forum, 36*(3), 563–587. doi:10.1111/cgf.13211.

Chen, X., Zhou, F., Zhang, K., Trajcevski, G., Zhong, T., Zhang, F., & Technologies, I. (2019). Information diffusion prediction via recurrent cascades convolution. In *IEEE 35th international conference on data engineering(ICDE)* (pp. 770–781). Macau SAR, China. doi:10.1109/ICDE.2019.00074.

Chen, Y.-C., Liu, Z.-Y., & Kao, H.-Y. (2017). IKM at SemEval-2017 Task 8: Convolutional Neural Networks for stance detection and rumor verification. In *11th international workshop on semantic evaluations (SemEval-2017)* (pp. 465–469). Vancouver, Canada. doi:10.18653/v1/s17-2081.

Cheng, J. J., Liu, Y., Shen, B., & Yuan, W. G. (2013). An epidemic model of rumor diffusion in online social networks. *The European Physical Journal B, 86*(1). doi:10.1140/epjb/e2012-30483-5.

Chi, Y., Zhu, S., Hino, K., Gong, Y., & Zhang, Y. (2009). iOLAP: A framework for analyzing the internet, social networks, and other networked data. *IEEE Transactions on Multimedia, 11*(3), 372–382. doi:10.1109/TMM.2009.2012912.

Chinese salt-buying frenzy. (2011).

Choi, J., Moon, S., Shin, J., & Yi, Y. (2016). Estimating the rumor source with anti-rumor in social networks. In *International conference on network protocols, ICNP* (pp. 1–6). Singapore. doi:10.1109/ICNP.2016.7785329.

Choi, J., Moon, S., Woo, J., Son, K., Shin, J., & Yi, Y. (2017). Rumor source detection under querying with untruthful answers. *IEEE conference on computer communications (INFOCOM)* Atlanta, USA. doi:10.1109/INFOCOM.2017.8057194.

Chuan, G., Cao, J., Zhang, X., Shu, K., & Yu, M. (2019). Exploiting emotions for fake news detection on social media. arXiv:1903.01728.

Conforti, C., Pilehvar, M. T., & Collier, N. (2018). Towards automatic fake news detection : Cross-level stance detection in news articles. In *The first workshop on fact extraction and VERification (FEVER)* (pp. 40–49). Brussels, Belgium.

Crowdbooster. (2019).

Csányi, G., & Szendroi, B. (2004). Structure of a large social network. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 69*(3) 036131(1-5). doi:10.1103/PhysRevE.69.036131.

Davis, R., & Proctor, C. (2017). *Fake news, real consequences : Recruiting neural networks for the fight against fake news*. 1–8.

Del Vicario, M., Quattrociocchi, W., Scala, A., & Zollo, F. (2019). Polarization and fake news: Early warning of potential misinformation targets. *ACM Transactions on the Web, 13*(2) 10(1-22).

Dhar, J., Jain, A., & Gupta, V. K. (2016). A mathematical model of news propagation on online social network and a control strategy for rumor spreading. *Social Network Analysis and Mining, 6*(1), 1–9. doi:10.1007/s13278-016-0366-5.

Diakopoulos, N., Naaman, M., & Kivran-Swaine, F. (2010). Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *2010 IEEE Symposium on Visual Analytics Science and Technology* (pp. 115–122). Salt lake City, Utah, USA. doi:10.1109/VAST.2010.5652922.

Did 600 Murders Take Place in Chicago Last Weekend? (2018).

Do paracetamol tablets contain 'machupo' virus? (2019).

Donald Trump ends school shootings by banning schools | 8Satire. (2019).

Dong, M., Yao, L., Wang, X., Benatallah, B., Huang, C., & Ning, X. (2018). Opinion fraud detection via neural autoencoder decision forest. *Pattern Recognition Letters*, 1–9. doi:10.1016/j.patrec.2018.07.013.

Dong, S., Deng, Y. Bin, & Huang, Y. C. (2017). SEIR model of rumor spreading in online social network with varying total population size. *Communications in Theoretical Physics, 68*(4), 545–552. doi:10.1088/0253-6102/68/4/545.

Edunov, S., Logothetis, D., Wang, C., Ching, A., & Kabiljo, M. (2016). Darwini: Generating realistic large-scale social graphs. arXiv:1610.00664.

Egele, M., Stringhini, G., Kruegel, C., & Vigna, G. (2017). Towards detecting compromised accounts on social networks. *IEEE Transactions on Dependable and Secure Computing, 14*(4), 447–460.

Elmurngi, E., & Gherbi, A. (2017a). An empirical study on detecting fake reviews using machine learning techniques. In *IEEE seventh international conference on innovative computing technology (INTECH)* (pp. 107–114). Luton. UK.

Elmurngi, E., & Gherbi, A. (2017b). Detecting fake reviews through sentiment analysis using machine learning techniques. In *Sixth international conference on data analytics* (pp. 65–72). Barcelona, Spain.

Elmurngi, E., & Gherbi, A. (2018). Fake reviews detection on movie reviews through sentiment analysis using supervised learning techniques. *International Journal on Advances in Systems and Measurements, 11*(1 & 2), 196–207.

Emergent. (2019).

Esmaeilzadeh, S., Peh, G. X., & Xu, A. (2019). Neural abstractive text summarization and fake news detection. arXiv:1904.00788.

Explosion at the White House. (2013).

FactCheck.org. (2003).

Fact-checking U.S. politics- PolitiFact. (2007).

Fairbanks, J., Fitch, N., Knauf, N., & Briscoe, E. (2018). Credibility assessment in the news : Do we need to read ? *MIS2*.

Finn, S., Metaxas, P. T., & Mustafaraj, E. (2014). Investigating rumor propagation with TwitterTrails. arXiv:1411.3550.

Fontanarava, J., Pasi, G., & Viviani, M. (2017). An ensemble method for the credibility assessment of user-generated content. In *International conference on web intelligence* (pp. 863–868). ACM. Leipzig, Germany. doi:10.1145/3106426.3106464.

García Lozano, M., Lilja, H., Tjörnhammar, E., & Karasalo, M. (2017). Mama Edha at SemEval-2017 Task 8: Stance classification with CNN and rules. In *11th international workshop on semantic evaluations (SemEval-2017)* (pp. 481–485). Vancouver, Canada. doi:10.18653/v1/s17-2084.

Ghaisani, A. P., Munajat, Q., & Handayani, P. W. (2017). Information credibility factors on information sharing activites in social media. In *2017 IEEE second international conference on informatics and computing (ICIC)* (pp. 1–5). Papua, Indonesia.

Gilda, S. (2017). Evaluating machine learning algorithms for fake news detection. In *IEEE student conference on research and development: inspiring technology for humanity, SCOReD* (pp. 110–115). Putrajaya, Malaysia. doi:10.1109/SCORED.2017.8305411.

Girgis, S., Amer, E., & Gadallah, M. (2018). Deep learning algorithms for detecting fake news in online text. In *IEEE 13th international conference on computer engineering and systems, (ICCES)* (pp. 93–97). Cairo, Egypt. doi:10.1109/ICCES.2018.8639198.

Global social media ranking. (2019).

Google Analytics. (2005).

Gravanis, G., Vakali, A., Diamantaras, K., & Karadais, P. (2019). Behind the cues: A benchmarking study for fake news detection. *Expert Systems with Applications, 128*, 201–213. doi:10.1016/j.eswa.2019.03.036.

Gu, J., Li, W., & Cai, X. (2008). The effect of the forget-remember mechanism on spreading. *The European Physical Journal B, 62*(2), 247–255. doi:10.1140/epjb/e2008-00139-4.

Gupta, A., Kumaraguru, P., Castillo, C., & Meier, P. (2014). TweetCred: Real-time credibility assessment of content on Twitter. In *International conference on social informatics* (pp. 228–243). Cham: Springer. Barcelona, Spain.

Hai, Z., Zhao, P., Cheng, P., Yang, P., Li, X.-L., & Li, G. (2016). Deceptive review spam detection via exploiting task relatedness and unlabeled data. In *Conference on empirical methods in natural language processing* (pp. 1817–1826). Austin, Texas, USA. doi:10.18653/v1/d16-1187.

Hamann, M., Meyer, U., Penschuck, M., Tran, H., & Wagner, D. (2018). I/O-efficient generation of massive graphs following the LFR benchmark. *Journal of Experimental Algorithmics, 23*(2), 1–33. doi:10.1145/3230743.

Hamidian, S., & Diab, M. (2015). Rumor detection and classification for Twitter data. In *SOTICS 2015: The fifth international conference on social media technologies, communication, and informatics* (pp. 71–77). Barcelona, Spain.

Han, Q., Miao, F., & Fan, W. (2017). Rumor spreading and monitoring deployment in online social networks. In *17th IEEE international conference on communication technology rumor* (pp. 1347–1351). Chengdu, China.

Han, S., Zhuang, F., He, Q., Shi, Z., & Ao, X. (2014). Energy model for rumor propagation on social networks. *Physica A: Statistical Mechanics and Its Applications, 394*, 99–109. doi:10.1016/j.physa.2013.10.003.

Hassan, N., Arslan, F., Li, C., & Tremayne, M. (2017). Toward automated fact-checking: Detecting check-worthy factual claims by ClaimBuster. In *23rd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1803–1812). Halifax, Nova Scotia, Canada. doi:10.1145/3097983.3098131.

Hootsuite. (2008).

Horne, B. D., & Adali, S. (2017). 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 *Eleventh international AAAI conference on web and social media* (pp. 759–766). Montreal, Quebec, Canada.

Indu, V., & Thampi, S. M. (2019). A nature - inspired approach based on Forest Fire model for modeling rumor propagation in social networks. *Journal of Network and Computer Applications, 125*, 28–41. doi:10.1016/j.jnca.2018.10.003.

InVID. (2017).

Ishida, Y., & Kuraya, S. (2018). Fake news and its credibility evaluation by dynamic relational networks: A bottom up approach. *Procedia Computer Science, 126*, 2228–2237. doi:10.1016/j.procS.2018.07.226.

Jin, Z., Cao, J., Guo, H., Zhang, Y., & Luo, J. (2017). Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In *25th ACM international conference on multimedia* (pp. 795–816). Mountain View, California, USA. doi:10.1145/3123266.3123454.

Jin, Z., Cao, J., Zhang, Y., Zhou, J., & Tian, Q. (2017). Novel visual and statistical image features for microblogs news verification. *IEEE Transactions on Multimedia, 19*(3), 598–608. doi:10.1109/TMM.2016.2617078.

Karimi, H., Roy, P. C., Saba-Sadiya, S., & Tang, J. (2018). Multi-source multi-class fake news detection. In *27th international conference on computational linguistics* (pp. 1546–1557). Santa Fe, New-Mexico, USA.

Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). MVAE: Multimodal variational autoencoder for fake news detection. In *The world wide web conference* (pp. 2915–2921). ACM. San Francisco, USA. doi:10.1145/3308558.3313552.

Kiran, L., & Gita, S. (2014). *Synthetic generators to simulate social networks (No. SAND2014-19132C)*. Albuquerque, NMUnited States: Sandia National Lab.(SNL-NM).

Kolda, T. G., Pinar, A., Plantenga, T., & Seshadhri, C. (2013). A scalable generative graph model with community structure. *SIAM Journal on Scientific Computing, 36*(5), C424–C452. doi:10.1137/130914218.

Kumar, K. K., & Geethakumari, G. (2014). Detecting misinformation in online social networks using cognitive psychology. *Human-Centric Computing and Information Sciences, 4*(1), 1–22. doi:10.1186/s13673-014-0014-x.

Kumar, S., & Shah, N. (2018). False information on web and social media: A survey. arXiv:1804.08559.

Kwon, S., Cha, M., Jung, K., Chen, W., & Wang, Y. (2013). Prominent features of rumor propagation in online social media. In *IEEE international conference on data mining, ICDM* (pp. 1103–1108). Dallas, Texas, USA. doi:10.1109/ICDM.2013.61.

Li, Y., Bao, H., Zheng, Y., & Huang, Z. (2015). Social analytics framework to boost recommendation in online learning communities. In *IEEE 15th international conference on advanced learning technologies: Advanced technologies f I(CALT)2015* (pp. 405–406). Hualien, Taiwan. doi:10.1109/ICALT.2015.100.

Liu, X., Nourbakhsh, A., Li, Q., Fang, R., & Shah, S. (2015). Real-time rumor debunking on Twitter. In *24th ACM International on conference on information and knowledge management* (pp. 1867–1870). Melbourne, Australia. doi:10.1145/2806416.2806651.

Liu, Yang, & Wu, Y. B. (2018). 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* (pp. 354–361). Orleans, Louisiana, USA.

Liu, Yuxin, Gao, C., She, X., & Zhang, Z. (2016). A bio-inspired method for locating the diffusion source with limited observers. In *2016 IEEE Congress on Evolutionary Computation, CEC 2016* (pp. 508–514). Vancouver, Canada. doi:10.1109/CEC.2016.7743836.

Lorek, K., Suehiro-Wicinski, J., Jankowski-Lorek, M., & Gupta, A. (2015). Automated credibility assessment on Twitter. *Computer Science, 16*(2), 157–168. doi:10.7494/csci.2015.16.2.157.

Louni, A., & Subbalakshmi, K. P. (2018). Who spread that rumor: Finding the source of information in large online social networks with probabilistically varying internode relationship strengths. *IEEE Transactions on Computational Social Systems, 5*(2), 335–343. doi:10.1109/TCSS.2018.2801310.

Lukasik, M., Srijith, P. K., Vu, D., Bontcheva, K., Zubiaga, A., & Cohn, T. (2016). Hawkes processes for continuous time sequence classification: An application to rumour stance classification in Twitter. In *54th annual meeting of the association for computational linguistics* (pp. 393–398). Berlin, Germany. doi:10.18653/v1/p16-2064.

Luo, W., Tay, W. P., & Leng, M. (2013). Identifying infection sources and regions in large networks. *IEEE Transactions on Signal Processing, 61*(11), 2850–2865. doi:10.1109/TSP.2013.2256902.

Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K.-F., & Cha, M. (2016). Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the twenty-fifth international joint conference on artificial intelligence(IJCAI-16)* (pp. 826–833). New York, USA.

Ma, J., Gao, W., Wei, Z., Lu, Y., & Wong, K.-F. (2015). Detect rumors using time series of social context information on microblogging websites. In *24th ACM international on conference on information and knowledge management - CIKM '15* (pp. 1751–1754). Melbourne, Australia. doi:10.1145/2806416.2806607.

Ma, J., Gao, W., & Wong, K.-F. (2017). Detect rumors in microblog posts using propagation structure via kernel learning. In *55th annual meeting of the association for computational linguistics* (pp. 708–717). Vancouver, Canada. doi:10.18653/v1/p17-1066.

Ma, J., Gao, W., & Wong, K.-F. (2018a). Detect rumor and stance jointly by neural multi-task learning. In *Companion proceedings of the the web conference 2018* (pp. 585–593). International World Wide Web Conferences Steering Committee. Lyon, France. doi:10.1145/3184558.3188729.

Ma, J., Gao, W., & Wong, K.-F. (2018b). Rumor detection on Twitter with tree-structured recursive neural networks. In *Proceedings of the 56th annual meeting of the association for computational linguistics (Volume 1: Long Papers)* (pp. 1980–1989). Melbourne, Australia.

Martens, D., & Maalej, W. (2019). Towards understanding and detecting fake reviews in app stores. *Empirical Software Engineering*, 1–40. doi:10.1007/s10664-019-09706-9.

Masood, R., & Aker, A. (2018). The fake news challenge: Stance detection using traditional machine learning approaches. In *10th international joint conference on knowledge discovery, knowledge engineering and knowledge management (KMIS)* (pp. 128–135). Seville, Spain. doi:10.5220/0006898801280135.

Mendoza, M., Poblete, B., & Castillo, C. (2010). Twitter under crisis: Can we trust what we RT? In *First workshop on social media analytics. ACM, 2010* (pp. 71–79). Washington D.C., USA.

Mexico "Twitter terrorism" charges cause uproar. (2015).

Mitra, T., & Gilbert, E. (2015). CREDBANK : A large-scale social media corpus with associated credibility annotations. In *Ninth international AAAI conference on web and social media* (pp. 258–267). Oxford, England.

Miyabe, M., Nadamoto, A., & Aramaki, E. (2014). How do rumors spread during a crisis?Analysis of rumor expansion and disaffirmation on Twitter after 3.11 in Japan. *International Journal of Web Information Systems, 10*(4), 394–412. doi:10.1108/ijwis-04-2014-0015.

Mohseni, S., Ragan, E. D., & Hu, X. (2019). Open issues in combating fake news: Interpretability as an opportunity. arXiv:1904.03016.

Mondal, T., Pramanik, P., Bhattacharya, I., Boral, N., & Ghosh, S. (2018). Analysis and early detection of rumors in a post disaster scenario. *Information Systems Frontiers, 20*(5), 961–979. doi:10.1007/s10796-018-9837-8.

Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. M. (2019). Fake news detection on social media using geometric deep learning. 1–15 arXiv:1902.06673.

Nan, Cao, Yu-Ru, Lin, Xiaohua, Sun, Lazer, D., Shixia, L., & Huamin, Qu (2012). Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics, 18*(12), 2649–2658. doi:10.1109/tvcg.2012.291.

Nekovee, M., Moreno, Y., Bianconi, G., & Marsili, M. (2007). Theory of rumour spreading in complex social networks. *Physica A: Statistical Mechanics and Its Applications, 374*(1), 457–470. doi:10.1016/j.physa.2006.07.017.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2018). *Reuter Institute Digital News Report 2018*. doi:10.2139/ssrn.2619576.

Nguyen, N. P., Yan, G., Thai, M. T., & Eidenbenz, S. (2012). Containment of misinformation spread in online social networks. In *4th annual ACM web science conference. ACM, 2012.* (pp. 213–222). Evanston, Illinois, USA. doi:10.1145/2380718.2380746.

North Korea Opening its doors to Christians. (2018).

Nunes, M., & Correia, J. (2013). Improving trust using online credibility sources and social network quality in P2P marketplaces. In *IEEE 2013 8th Iberian conference on information systems and technologies (CISTI)* (pp. 1–4). AISTI. Lisbon, Portugal.

O'Brien, K., Simek, O., & Waugh, F. (2019). Collective classification for social media credibility estimation. In *52nd Hawaii international conference on system sciences* (pp. 2235–2243). Hawaii, USA. doi:10.24251/hicss.2019.271.

Peng, L. (2018). Fake news detection : Sequence models. *Annual bio-inspired conference* Manchester, UK.

Perez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2017). Automatic detection of fake news. (August) arXiv:1708.07104.

Poddar, L., Hsu, W., Lee, M. L., & Subramaniyam, S. (2018). Predicting stances in twitter conversations for detecting veracity of rumors: A neural approach. In IEEE 30th international conference on tools with artificial intelligence, ICTAI (pp. 65–72). Volos, Greece. doi:10.1109/ICTAI.2018.00021.

Prat-Pérez, A., Guisado-Gámez, J., Salas, X. F., Koupy, P., Depner, S., & Bartolini, D. B. (2017). Towards a property graph generator for benchmarking. In *Proceedings of the fifth international workshop on graph data-management experiences & systems* (pp. 1–6). ACM. Chicago, USA. doi:10.1145/3078447.3078453.

Qian, F., Gong, C., Sharma, K., & Liu, Y. (2018). Neural user response generator : Fake news detection with collective user intelligence. In *Twenty-seventh international joint conference on artificial intelligence (IJCAI-18) Neural* (pp. 3834–3840). Stockholm, Sweden.

Rasool, T., Butt, W. H., Shaukat, A., & Akram, M. U. (2019). Multi-label fake news detection using multi-layered supervised learning. In *2019 11th international conference on computer and automation engineering - ICCAE* (pp. 73–77). Bangkok, Thailand. doi:10.1145/3313991.3314008.

Rath, B., Gao, W., Ma, J., & Srivastava, J. (2017). From retweet to believability: Utilizing trust to identify rumor spreaders on Twitter. In *IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 179–186). Sydney, Australia. doi:10.1145/3110025.3110121.

Reilly, I. (2018). F for fake: Propaganda! Hoaxing! Hacking! Partisanship! and Activism! in the fake news ecology. *The Journal of American Culture, 41*(2), 139–152. doi:10.1111/jacc.12834.

Resnick, P., Carton, S., Park, S., Shen, Y., & Zeffer, N. (2014). RumorLens : A system for analyzing the impact of rumors and corrections in social media. In *Computational journalism conference* (pp. 10121–10701). New York, USA.

REVEAL. (2014).

Roozenbeek, J., & van der Linden, S. (2019). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research, 22*(5), 570–580. doi:10.1080/13669877.2018.1443491.

Roy, A., Basak, K., Ekbal, A., & Bhattacharyya, P. (2018). A deep ensemble framework for fake news detection and classification. (2018) arXiv:1811.04670.

Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. In *2017 ACM on conference on information and knowledge management* (pp. 797–806). Singapore. doi:10.1145/3132847.3132877.

Sadiq, S., Wagner, N., Shyu, M. L., & Feaster, D. (2019). High dimensional latent space variational autoencoders for fake news detection. In *IEEE international conference on multimedia information processing and retrieval, MIPR 2019* (pp. 437–442). San Jose, California, USA. doi:10.1109/MIPR.2019.00088.

Sahana, V. P., Pias, A. R., Shastri, R., & Mandloi, S. (2015). Automatic detection of rumoured tweets and finding its origin. In *2015 International conference on computing and network communications, CoCoNet* (pp. 607–612). Trivandrum, Kerala, India. doi:10.1109/CoCoNet.2015.7411251.

Santia, G. C., & Williams, J. R. (2018). BuzzFace : A news veracity dataset with facebook user commentary and egos. In *Proceedings of the twelfth international AAAI conference on web and social media (ICWSM)* (pp. 531–540). Palo Alto, California, USA.

Shah, D., & Zaman, T. (2011). Rumors in a network: Who's the culprit? In *IEEE transactions on information theory: 57* (pp. 5163–5181). doi:10.1109/TIT.2011.2158885.

Shao, C., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2016). Hoaxy: A platform for tracking online misinformation. In *25th international conference companion on world wide web* (pp. 745–750). International World Wide Web Conferences Steering Committee. Montreal, Quebec, Canada. doi:10.1145/2872518.2890098.

Shao, C., Hui, P. M., Wang, L., Jiang, X., Flammini, A., Menczer, F., & Ciampaglia, G. L. (2018). Anatomy of an online misinformation network. *PLoS ONE, 13*(4), 1–23. doi:10.1371/journal.pone.0196087.

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news : A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology, 10*(3), 1–42. doi:10.1145/3305260.

Shelke, S., & Attar, V. (2019). Source detection of rumor in social network – A review. *Online Social Networks and Media, 9*, 30–42. doi:10.1016/J.OSNEM.2018.12.001.

Shu, K., Bernard, H. R., & Liu, H. (2019). Studying fake news via network analysis: Detection and mitigation. In *Emerging research challenges and opportunities in computational social network analysis and mining* (pp. 43–65). Springer, Cham. doi:10.1007/978-3-319-94105-9.

Shu, K., Mahudeswaran, D., & Liu, H. (2019). FakeNewsTracker: A tool for fake news collection, detection, and visualization. *Computational and Mathematical Organization Theory, 25*(1), 60–71. doi:10.1007/s10588-018-09280-3.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). FakeNewsNet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media. arXiv:1809.01286.

Shu, K., Wang, S., Le, T., Lee, D., & Liu, H. (2018). Deep headline generation for clickbait detection. In IEEE international conference on data mining, ICDM (pp. 467–476). Singapore. doi:10.1109/ICDM.2018.00062.

Shu, K., Wang, S., & Liu, H. (2018). Understanding user profiles on social media for fake news detection. In IEEE 1st conference on multimedia information processing and retrieval, MIPR (pp. 430–435). Florida, USA. doi:10.1109/MIPR.2018.00092.

Shu, K., Wang, S., & Liu, H. (2019). Beyond news contents: The role of social context for fake news detection. In *Twelfth ACM international conference on web search and data mining* (pp. 312–320). Melbourne, Australia. doi:10.1145/3289600.3290994.

Shu, K., Wang, S., Liu, H., Yang, S., Gu, R., & Wu, F. (2019). Unsupervised fake news detection on social media: A generative approach. *33rd AAAI conference on artificial intelligence*. Hawaii, USA.

Singhania, S., Fernandez, N., & Rao, S. (2017). 3HAN : A deep neural network for fake news detection. In International conference on neural information processing. Springer, Cham, 2017*., (November)* (pp. 572–581). Guangzhou, China. doi:10.1007/978-3-319-70096-0.

Snaplytics. (2015).

Snopes.com. (1994).

Starbird, K. (2017). Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter. In *Eleventh international AAAI conference on web and social media (Icwsm)* (pp. 230–239). Montreal, Quebec, Canada.

Starbird, K., Maddock, J., Orand, M., Achterman, P., & Mason, R. M. (2014). Rumors, false flags, and digital vigilantes: misinformation on Twitter after the 2013 Boston Marathon Bombing. *I Conference* Berlin, Germany. doi:10.9776/14308.

Tacchini, E., Ballarin, G., Vedova, M. L. Della, Moret, S., & Alfaro, L. De (2017). Some like it hoax : Automated fake news detection in social networks. 1–12 arXiv:1704.07506.

Talkwalker. (2009).

Thota, A., Tilak, P., Ahluwalia, S., & Lohia, N. (2018). Fake news detection : A deep learning approach. *SMU Data Science Review, 1*(3), 1–20.

Truth or Fiction - Fact Check. (2019).

Tschiatschek, S., Singla, A., Rodriguez, M. G., Merchant, A., & Krause, A. (2018). Fake news detection in social networks via crowd signals. In *The web conference 2018* (pp. 517–524). International World Wide Web Conferences Steering Committee, 2018. Lyon, France.

Turenne, N. (2018). The rumour spectrum. *PLoS ONE, 13*(1), 1–27. doi:10.1371/journal.pone.0189080.

Sivasangari, V., Pandian, V. A., & Santhya, R. (2018). A Modern approach to identify the fake news using machine learning. *International Journal of Pure and Applied Mathematics, 118*(20), 3787–3795.

Vaghela, D. B., & Patel, D. M. (2018). Rumor detection with Twitter and news channel data using sentiment analysis and classification. *International Journal of Advance Engineering and Research Development, 5*(2), 924–930.

Varol, O., Ferrara, E., Menczer, F., & Flammini, A. (2017). Early detection of promoted campaigns on social media. *EPJ Data Science, 6*(1). doi:10.1140/epjds/s13688-017-0111-y.

Vishwakarma, D. K., Varshney, D., & Yadav, A. (2019). Detection and veracity analysis of fake news via scrapping and authenticating the web search. *Cognitive Systems Research, 58*, 217–229. doi:10.1016/j.cogsys.2019.07.004.

Viviani, M., & Pasi, G. (2017). Credibility in social media: Opinions, news, and health information—a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 7*(5), 1–25. doi:10.1002/widm.1209.

Vosoughi, S., Deb, R., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151.

Vosoughi, S., Mohsenvand, M. 'Neo', & Roy, D. (2017). Rumor gauge: Predicting the veracity of rumors on Twitter. *ACM Transactions on Knowledge Discovery from Data, 11*(4) 50:1–50:36. doi:10.1145/3070644.

Wang, F., Moreno, Y., & Sun, Y. (2006). Structure of peer-to-peer social networks. *Physical Review E, 73*(3), 1–7. doi:10.1103/PhysRevE.73.036123.

Wang, W. Y. (2017). "Liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv:1705.00648.

Was Hillary Clinton photographed with Osama Bin Laden? (2017).

Was Kamal Nath the driver of Rajiv Gandhi? (2018).

Wen, S., Jiang, J., Xiang, Y., Yu, S., Zhou, W., & Jia, W. (2014). To shut them up or to clarify: Restraining the spread of rumors in online social networks. *IEEE Transactions on Parallel and Distributed Systems, 25*(12), 3306–3316. doi:10.1109/TPDS.2013.2297115.

Wu, K., Yang, S., & Zhu, K. Q. (2015). False rumors detection on sina weibo by propagation structures. In *IEEE 31st international conference on data engineering* (pp. 651–662). Seoul, South Korea. doi:10.1109/ICDE.2015.7113322.

Wu, L., & Liu, H. (2018). Tracing fake-news footprints: Characterizing social media messages by how they propagate. In *Eleventh ACM international conference on web search and data mining* (pp. 637–645). Los Angeles, California, USA. doi:10.1145/3159652.3159677.

Wu, Y., Huang, H., Zhao, J., Wang, C., & Wang, T. (2018). Using mobile nodes to control rumors in big data based on a new rumor propagation model in vehicular social networks. *IEEE Access, 6*, 62612–62621. doi:10.1109/ACCESS.2018.2876394.

Yang, Y., Zheng, L., Zhang, J., Cui, Q., Zhang, X., Li, Z., & Yu, P. S. (2018). TI-CNN : Convolutional neural networks for fake news detection. (2018) arXiv:1806.00749.

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. In *2016 conference of the north american chapter of the association for computational linguistics: human language technologies* (pp. 1480–1489). San Diego, California, USA.

Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2018). The web of false information : rumors, fake news, hoaxes, clickbait, and various other shenanigans. arXiv:1804.03461.

Zhang, J., Cui, L., Fu, Y., & Gouza, F. B. (2018). Fake news detection with deep diffusive network model. (2018) arXiv:1805.08751.

Zhang, Q., Lipani, A., Liang, S., & Yilmaz, E. (2019). Reply-aided detection of misinformation via bayesian deep learning. (February), 2333–2343. The World Wide Web Conference, (pp. 2333–2343). San Francisco, USA. doi:10.1145/3308558.3313718.

Zhang, X., Zhang, Y., Lv, T., & Yin, Y. (2016). Identification of efficient observers for locating spreading source in complex networks. *Physica A: Statistical Mechanics and Its Applications, 442*, 100–109. doi:10.1016/j.physa.2015.09.017.

Zhang, Z., Zhang, Z., & Li, H. (2015). Predictors of the authenticity of Internet health rumours. *Health Information and Libraries Journal, 32*(3), 195–205. doi:10.1145/3308558.3313718.

Zhao, J., Cao, N., Wen, Z., Song, Y., Lin, Y. R., & Collins, C. (2014). #FluxFlow: Visual analysis of anomalous information spreading on social media. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 1773–1782. doi:10.1109/TVCG.2014.2346922.

Zhao, L., Wang, J., Chen, Y., Wang, Q., Cheng, J., & Cui, H. (2012). SIHR rumor spreading model in social networks. *Physica A: Statistical Mechanics and Its Applications, 391*(7), 2444–2453. doi:10.1016/j.physa.2011.12.008.

Zhao, L., Wang, Q., Cheng, J., Chen, Y., Wang, J., & Huang, W. (2011). Rumor spreading model with consideration of forgetting mechanism: A case of online blogging LiveJournal. *Physica A: Statistical Mechanics and Its Applications, 390*(13), 2619–2625. doi:10.1016/j.physa.2011.03.010.

Zhou, X., & Zafarani, R. (2018). Fake news: A survey of research, detection methods, and opportunities. arXiv:1812.00315.

Zhou, X., Zafarani, R., Shu, K., & Liu, H. (2019). Fake news: Fundamental theories, detection strategies and challenges. In *Welfth ACM international conference on web search and data mining* (pp. 836–837). ACM. Melbourne, Australia.

Zhu, K., & Ying, L. (2016). Information source detection in the SIR model: A sample-path-based approach. *IEEE/ACM Transactions on Networking(TON), 24*(1), 408–421. doi:10.1109/TNET.2014.2364972.

Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR), 51*(2), 1–36. doi:10.1145/3161603.