



# Machine Learning Approaches for Rumor Detection in Social Media: Types, Techniques, and Opportunities

Afrin Akter Mim<sup>1</sup>, Minhajul Islam Mim<sup>2</sup>, Md Ahasan Habib<sup>3</sup>, Dipanita Mondol<sup>1\*</sup>, Md Sabbir Ahamed<sup>4</sup>, and Sudeepta Chandra Paul<sup>5</sup>

<sup>1</sup> Department of Computer Science and Engineering, Daffodil International University, Dhaka-1216, Bangladesh

<sup>2</sup> Information and Communication Engineering, Hohai University, Nanjing, P. R. China

<sup>3</sup> Computer Science and Technology, Hohai University, Nanjing, P. R. China

<sup>4</sup> Management Information Systems, Lamar University, Beaumont, United States

<sup>5</sup> Department of Computer Science and Engineering, Jagannath University, Dhaka, Bangladesh

afrin15-12862@diu.edu.bd, minhajulislam435@outlook.com,  
{ahnahid123,mondaldipanwita827}@gmail.com, mahamed1@lamar.edu,  
actualprakash@gmail.com

**Abstract.** The trend of spreading false news or rumors has kept pace with the growing utilization of social media. Hatred and fear spread through rumors, which are extremely harmful to society. As social media continues to grow, rumor detection has become an increasingly important research area. In this paper, we summarize the efforts made thus far on this topic. We provide an overview of rumors from different dimensions, existing methods, and machine learning and network-based approaches used to distinguish rumors, evaluate method performance, and, importantly, discuss several new directions for future research.

**Keywords:** Rumor detection · social media · machine learning approaches.

## 1 INTRODUCTION

A rumor is a declaration of knowledge that has been shared among people, which is unsubstantiated and has a practical application. It is well known that rumors proliferate in situations that are unclear, such as when there is a possible threat or when the situation's significance is uncertain. The potential to disseminate inaccurate or misleading information has grown along with the usage of social media. Rumor detection is one of the most widely debated research topics for preventing the spread of rumors, as it is difficult for ordinary people to determine whether the news is accurate or not. Researchers have presented different strategies to detect fake or false information from various social platforms, including Facebook, Twitter, and more. Most of the previous works focused on recognizing rumors through only highlights, substance, and the characteristics

of bloggers. Many efforts have been made with different methods to distinguish and disclose rumors from social media to analyze their content and social context. Most present rumor detection models are deep neural network-based methods [1], graph-based detectors to discover rumors on social media from a hostile perspective [2], propagation context-based deep neural networks [3], multi-view deep learning models [4], early rumor detection through recurrent neural networks based on deep attention [5], and hot topic detection [6]. The models need to predict accurately to detect rumors from social media reliably. Recently, many supervised, semi-supervised, and unsupervised algorithms have been developed to deal with the increasing rumors in data mining.

## 2 LITERATURE REVIEW

At the beginning of rumor recognition, a basic paper outlines different approaches to detecting rumors from social media. There are datasets such as Qazvinian et al., KWON Dataset, MediaEval Datasets [7], RUMDECT dataset, and more. Machine learning techniques and supervised-based methods for rumor discovery, as well as RNN-based approaches, are commonly used. All these techniques have their own unique methods [8]. Another rumor detection approach is based on a multi-view neural network model. Researchers propose a simple yet powerful three-view deep learning model designed to discover rumors based on word consideration [9]. Researchers identify implicit links between conversation fragments to classify early-stage discussions and enhance accuracy in more mature discussions. They used two credibility-based datasets to identify characteristics that can be used to categorize the reliability of news, with one being keyword-based [10]. Other researchers have developed a deep attention model for learning time-dependent hidden representations of sequential messages for spotting rumors. This model is based on a recurrent neural network (RNN)-based deep attention system [11]. Some studies focus on detecting rumors based on their spread and proliferation. Researchers proposed a novel bi-directional GCN approach to work on both propagation and dispersion. This gossip detection depends on numerous time-series information examinations for rumor discovery on social media. Their proposed approach will remove regular media triggers and distinguish gossip solely based on the data [12]. In another study, they proposed a novel hierarchical social attention model for rumor discovery (HSA-BLSTM), which effectively detects fake posts in microblogs. The researchers found that their model outperforms several state-of-the-art models by a significant margin [13].

Some researchers approach rumor detection by separating the process into two stages. In the first stage, they design a unique model to identify rumors, and in the second stage, they analyze ongoing posts. This method achieved excellent results in detecting rumor patterns in trending topic data [14]. Four key research lines are examined in the paper: exploratory data analysis, rumor detection, epidemiological modeling, and multi-agent-based social simulation [15]. A systematic survey of fake news detection on social media and benchmark datasets is also

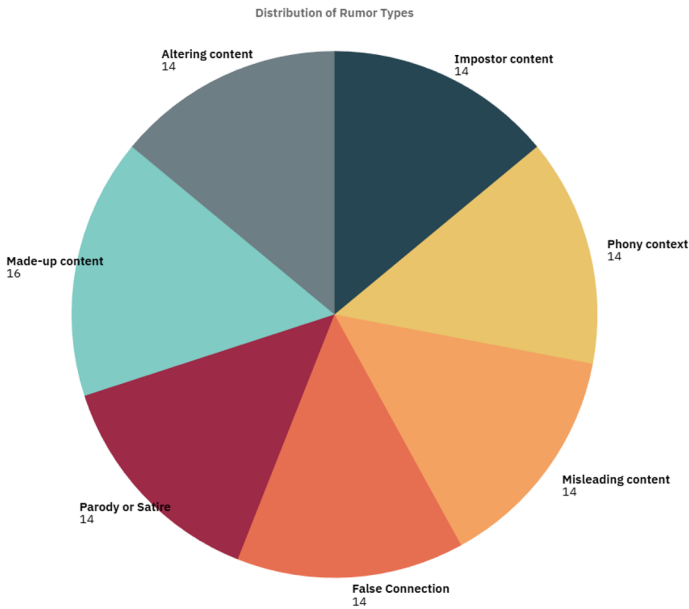
discussed [16]. A formal definition of rumors, compared with other definitions used in the literature, is provided, along with a summary of studies on automatic rumor detection. The paper presents three distinct paradigms for rumor detection: the feature-based classification approach, the credibility propagation approach, and neural network approaches [17].

**Table 1.** Summary of Methods, Datasets, and Performance Metrics in Rumor Detection Studies

Study	Method	Dataset	Performance Metric(s)
Study 1	Supervised Learning (e.g., SVM)	Qazvinian et al. Dataset	Accuracy, Precision, Recall
Study 2	Graph-Based Detector	KWON Dataset	F1-Score, Precision
Study 3	Deep Learning (RNN)	MediaEval Dataset	AUC, F1-Score
Study 4	Multi-View Neural Network	RUMDECT Dataset	Accuracy, Recall
Study 5	Propagation-Based Approach	Twitter 15/16	Precision, Recall, F1-Score

### 3 RUMORS

#### 3.1 Types of Rumor



**Fig. 1.** Distribution of rumor types..

Given that a rumor is a narrative or fact that might or might not be accurate, there are numerous types of rumors. According to various typologies, rumors have been classified into categories based on their subject matter, such as racial, disaster, organizational, political, and product rumors. They have also been categorized according to the collective concern of the public. For example, pecking order rumors relate to changes in management structure and how these changes may affect job responsibilities and salaries. Motivational tension is perhaps the most common rumor typology. After research, researchers identified seven types of rumors or false information:

**Parody or Satire:** Information of this kind is not dangerous, but it has the potential to deceive.

**False Connection:** This type of rumor has no connection between the title and its context.

**Misleading Content:** Its primary purpose is to use false facts to frame a problem or person.

**Phony Context:** Sometimes legitimate content is disseminated with deceptive context.

**Impostor Content:** This type of rumor involves fake sources impersonating authentic ones.

**Altering Content:** When actual data or visuals are altered to deceive, such as in a photoshopped image.

**Made-up Content:** This refers to entirely untrue information designed to influence and damage people.

### 3.2 Types of data in the rumor

News is made up of data, and in rumor detection, data can be classified into three types:

**Visual-based:** A rumor that is visual and incorporates various forms of media is known as visual-based fake news. Among the most popular social media platforms for sharing information online are Facebook, YouTube, and Twitter. The majority of people use these platforms daily to acquire the latest news, making them a major source of information.

**Linguistically-based:** Fake news in this category is composed of text and language, and is typically evaluated through linguistic analysis of written content. Its primary focus is on text as a communication system, with aspects such as style, syntax, and pragmatics playing key roles in conveying meaning.

**Graph-based:** Graphs can be an excellent tool for disseminating information to the general audience. However, not all graphs are created equally. Using incorrect graph types, mislabeling axes, or using improper or inconsistent scales can alter the appearance of data, causing users to misinterpret the information and spread rumors.

### 3.3 C. Cases for spreading rumor

A rumor's primary purpose is to make sense of something that is already unclear. Rumors have the potential to illuminate a baffling element of life. For example, in the case of the "soda rumor," confrontations between groups of people are terrifying and baffling; this ambiguity leads to rumors, which people grasp in an attempt to understand why things are the way they are. Rumors circulate for a variety of reasons:

**Uncertainty:** We cannot be sure of everything because we are human. With nearly everyone now using the internet, people may encounter a lot of news on the same topic, but they often have no idea what is correct or incorrect. In this state of confusion, people upload and circulate fake news, causing rumors to spread.

**Belief:** Many people hold the same beliefs, whether they are correct or incorrect. If they believe something false, it is difficult to persuade them to believe the truth. This belief, in turn, propagates false information.

**Self-image:** People are generally concerned with their public image. They occasionally share incorrect information to maintain a positive image in the eyes of a large audience.

**Social Status:** Sometimes, in an effort to maintain their social standing, people intentionally propagate rumors to preserve their social power or prestige. This can sometimes result in significant harm to a large number of people.

## 4 WAYS OF DETECTING RUMOR

Rumor detection in traditional news media typically focuses on media content. However, detecting rumors on social media can be enhanced by utilizing additional social context and auxiliary information. Consequently, this section will explore various techniques for extracting and expressing key elements from news content and the social environment. Some common methods for detecting fake news include:

### 4.1 User-Based

Non-human accounts, such as automated bots or media platforms, are more likely to generate and propagate rumors. Therefore, user-based features are used to record user profiles and assist in identifying rumors by providing pertinent information. These characteristics define the traits of individuals engaging with content on social media. User-based characteristics can be divided into two categories: individual and collective. To determine the accuracy and authenticity of each user, personal information such as their age, number of followers, and frequency of posts is collected.

## 4.2 Network-Based

Users create various networks on social media based on their contacts, interests, and topics. Network-based feature extraction is crucial for identifying misinformation, as the spread of false news often creates echo chambers. By analyzing networks among users who share similar posts, network-based features can help identify the spread of rumors. There are several types of networks that can be constructed for this purpose.

## 4.3 Stance Detection

This technique analyzes users' perspectives or stances in relation to similar posts to assess the authenticity of original news articles. It is possible to detect both implicit and explicit viewpoints expressed in users' posts. Explicit sentiments, such as Facebook's "Like" or "Sad" reactions, are direct expressions of emotion or opinion. Stance detection refers to automatically determining whether a person is supportive, indifferent, or opposed to a given event, factor, or concept based on their post.

## 4.4 Filtering by Keywords to Gather Information on an Event

Keywords can be extremely helpful in rumor detection. Rumors can spread quickly during live online events or sessions, and by analyzing the comments and posts related to a particular event, key keywords can be identified to make spotting rumors easier.

### **Providing a List of Users to Follow in Order to Monitor Their Posts:**

People often express their opinions regarding false news on social media, including skepticism, emotional responses, and a range of other reactions. Therefore, it is possible to extract post-based attributes to help detect likely fake news through public reactions. Post-based features are used to gather relevant information from various important social media posting characteristics in order to determine the credibility of news.

**Propagation-Based:** To predict the plausibility of news, propagation-based fake news detection techniques examine the connections between important social media posts. The core idea is that the credibility of an event is closely linked to the authenticity of relevant social media posts. It is possible to create both homogeneous and heterogeneous credibility networks to analyze how information spreads.

**Style-Based:** To appeal to a wide audience, fake news producers often use persuasive and manipulative techniques to spread inaccurate and distorted information. This requires a distinct writing style not typically found in legitimate news items. Style-based techniques aim to identify fake news by analyzing the authoring style of the content. Two types of style-based methods are commonly distinguished: Deception-Focused and Objective-Focused.

**Knowledge-Based:** One of the simplest ways to spot rumors is by assessing the veracity of the main claims in a news article to determine its reliability. Rumors often attempt to propagate unfounded accusations in the media. To fact-check claims and verify their truthfulness, knowledge-based approaches are used. Fact-checking involves assigning a truth value to a claim within a certain context. Recently, fact-checking has gained significant attention, and several attempts have been made to develop automated fact-checking systems. These systems currently rely on methods such as crowdsourcing, machine learning, and expert reviews.

## 5 APPROACHES OF DETECTING RUMOR

We are all greatly affected by rumors. Therefore, it is imperative to detect them as soon as possible. The approach must be logical and effective. Detecting rumors can be challenging, especially as they spread rapidly on the internet, with social media platforms being highly active. An effective strategy should detect rumors as infrequently as possible.

### 5.1 Machine Learning Approaches

Machine learning approaches do not rely on a fixed equation but learn from data individually. These algorithms use computational strategies and search for recurring patterns in data that provide insights, which in turn lead to more informed decisions and predictions. The more samples available for learning, the more adaptively the models improve. Machine learning techniques are generally divided into two categories: supervised and unsupervised approaches.

Supervised Learning prepares a model for use with known inputs and outputs:

**Classification Mining:** Classification mining determines a new data category based on the training set. This strategy is valuable in pattern recognition.

**Regression Analysis:** This evaluates the relationship between dependent and independent variables, primarily used in model and framework development.

**SVMs (Support Vector Machines):** SVMs are supervised learning techniques that use classification and regression analysis to separate data.

**Text Mining:** This method extracts meaningful content from text. It is widely used in natural language processing and information retrieval, as well as in statistical pattern mining.

Unsupervised Learning uncovers hidden patterns or inherent structures in input data:

**Association Rule Mining:** A data mining technique used to extract behavioral data patterns.

**Cluster Analysis:** Cluster analysis divides data into subsets based on certain criteria, although the specific notations cannot always be fully defined.

**Decision Tree:** A decision tree is a tool for decision-making, using a tree-like structure to determine the best option out of several. It is often used in operations research.

**Factor Analysis:** Factor analysis identifies the variability of observable and related data.

## 5.2 Deep learning Approaches

Deep learning is a more specialized form of machine learning, involving teaching computers to perform tasks typically done by humans. A computer model gains the ability to categorize items directly from images, text, or sound. State-of-the-art (SOA) precision, sometimes exceeding human-level performance, can be achieved with deep learning models. These models are trained using large sets of labeled data and neural network structures with multiple layers.

**Sequence Mining:** A statistically based pattern mining technique for extracting commonly occurring patterns.

**Neural-Network:** A neural network, based on biological brain networks, is used to estimate unknown patterns.

## 5.3 RNN based Approaches

Recurrent Neural Networks (RNNs) are deep learning models used for modeling sequences of arbitrary length by applying transition functions and recursive states. RNNs are well-suited for sequence modeling related to time-variant and time-invariant inputs. While RNNs have a fixed structure, they are differentiable at every node, overcoming long-term dependency issues by relying on the current input and previous outputs.

## 5.4 CNN-based Approaches

Convolutional Neural Networks (CNNs) are among the most well-known types of deep neural networks. CNNs eliminate the need for manual feature extraction, which is typically required for image classification. CNNs extract features directly from images, learning important highlights as the network trains on a set of images. This automatic feature extraction makes deep learning models highly accurate for computer vision tasks such as object classification.

## 5.5 Hybrid Approaches

Hybrid approaches combine different strategies to create solutions that integrate the strengths of individual methods and mitigate their weaknesses. Bimodal and research development are common hybrid approaches in data science.

## 5.6 Propagation/Network-based approaches

**Fake news detection using networks:** Networks can be used to build agile systems that track the indirect spread of false information. The network may be hierarchical, heterogeneous, or homogeneous.

**Homogeneous Network:** A homogeneous network consists of a single type of node and edge. The support network is a classic homogeneous network, where nodes represent user messages, and edges represent the positive or negative relationships between messages, assessing the reliability of news-related posts.

**Heterogeneous Network:** This type of network is made up of multiple types of edges and nodes. It is usually a hybrid framework with three components: entity representation and injection, relationship modeling, and semi-supervised learning. An example of a heterogeneous network could be the relationships between news publishers and their reports.

**Hierarchical Network:** A hierarchical network is composed of various types of edges and nodes that form relationships between sets and subsets. Checking the message in this structure becomes a graphics optimization problem.

## 6 FUTURE RESEARCH SCOPE

Rumor detection is a vast concept. Due to the increasing reliance on social media, the spread of rumors will likely worsen over time. As a result, it presents an extensive area for research. The following are some promising future research directions:

**Feature-Oriented:** The goal of feature-oriented rumor detection research is to discover relevant traits for detecting rumors from various information sources. The two key data sources are internet content and contextual factors. Linguistic and visual-based rumor detection methods extract features from textual and visual information, respectively, focusing on news content. It is important to note that linguistic-based features have already been thoroughly explored for basic natural language processing (NLP) tasks, including clustering, message classification, and specific uses like policy implementation.

**Application-Oriented:** Application-oriented research in rumor detection extends beyond merely detecting false information. It focuses on two main approaches: fake news intervention and fake news diffusion. "Fake news diffusion" refers to the methods and trends by which fake news spreads on social media platforms. According to early studies, trustworthy information and misinformation propagate through online social networks in distinct ways, which requires further investigation.

**Data-Oriented:** Data-driven research on fake news examines a variety of data characteristics, such as temporal patterns, psychological aspects, and dataset quality. This paper aims to highlight the lack of a comprehensive benchmark

dataset that includes all required characteristics for effective rumor detection. A promising strategy would be to create a robust, large-scale misinformation dataset that researchers can use to support further studies in this area.

**Model-Oriented:** The development of model-oriented rumor detection algorithms will pave the way for more effective and practical fake news detection systems. Most current approaches focus on extracting various features and incorporating them into supervised classification models. However, more research is needed to develop complex and realistic models that integrate better-extracted features, such as aggregation, statistical methods, collective intelligence, and extrapolation techniques.

## 7 CONCLUSION

This paper presents a survey on the detection of social media rumors. Fake news is spreading rapidly across social networks due to its increasing ubiquity, which has detrimental effects on both individuals and society as a whole. By providing a comprehensive explanation of the deep learning models, dataset attributes, and evaluation metrics used in the existing works, this survey offers valuable insights for researchers in the field. Furthermore, it highlights potential research directions for improving the identification of false news, making this survey a useful resource for future studies and comparisons with existing approaches.

## 8 ACKNOWLEDGMENT

The authors would like to acknowledge the valuable insights and contributions from previous research studies that laid the foundation for this work. We are also grateful to our academic mentors and peers for their constructive suggestions and continuous encouragement throughout the preparation of this paper. Their guidance has been instrumental in shaping the analysis and discussions presented in this study.

## References

1. Song, C., Yang, C., Chen, H., Tu, C., Liu, Z., Sun, M.: CED: Credible Early Detection of Social Media Rumors. *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1 (2015).
2. Yang, X., Lyu, Y., Tian, T., Liu, Y., Liu, Y., Zhang, Y., Xi, X.: Rumor Detection on Social Media with Graph Structured Adversarial Learning. In: Proceedings of IJCAI 2020, pp. 1417–1423 (2020). <https://doi.org/10.24963/ijcai.2020/197>
3. Gao, J., Han, S., Song, X., Ciravegna, F.: RP-DNN: A Tweet Level Propagation Context Based Deep Neural Networks for Early Rumor Detection in Social Media. In: Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020), pp. 6094–6105 (2020). Preprint available at: <https://doi.org/10.48550/arXiv.2002.12683>

4. Geng, Y., Lin, Z., Fu, P., Wang, W.: Rumor Detection on Social Media: A Multi-view Model Using Self-attention Mechanism. In: *Lecture Notes in Computer Science (LNCS)*, Springer (2019).
5. Li, X., Yin, H., Zhang, J., Chen, T., Wu, L., Wang, Y.: Call Attention to Rumors: Deep Attention Based Recurrent Neural Networks for Early Rumor Detection. In: *Lecture Notes in Computer Science*, vol. 11154, pp. 40–52. Springer, Heidelberg (2018). <https://doi.org/10.1007/978-3-030-04503-6-4>
6. Bian, T., Xiao, X., Xu, T., Zhao, P., Huang, W., Rong, Y., Huang, J.: Rumor Detection on Social Media with Bi-Directional Graph Convolutional Networks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 549–556 (2020). <https://doi.org/10.1609/aaai.v34i01.5393>
7. Li, Q., Zhang, Q., Si, L., Liu, Y.: Rumor Detection on Social Media: Datasets, Methods, and Opportunities. In: Feldman, A., Da San Martino, G., Barrón-Cedeño, A., Brew, C., Leberknight, C., Nakov, P. (eds.) *\*Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda\* (NLP4IF)*, LNCS, vol.?, pp.66–75. Association for Computational Linguistics, HongKong (2019). <https://doi.org/10.18653/v1/D19-5008>
8. Pathak, A.R., Mahajan, A., Singh, K., Patil, A., Nair, A.: Analysis of Techniques for Rumor Detection in Social Media. *\*Procedia Computer Science\* \*\*167\*\**, 2286–2296 (2020). <https://doi.org/10.1016/j.procs.2020.03.281>
9. Geng, Y., Lin, Z., Fu, P., Wang, W.: Rumor Detection on Social Media: A Multi-view Model Using Self-attention Mechanism. In: *\*International Conference on Computational Science\* (ICCS)*, LNCS, vol.11536, pp.300–314. Springer, Cham (2019). <https://doi.org/10.1007/978-3-030-22734-0-25>
10. Sampson, J., Morstatter, F., Wu, L., Liu, H.: Leveraging the Implicit Structure within Social Media for Emergent Rumor Detection. In: *\*Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM 2016)\**, pp.2377–2382. ACM (2016). <https://doi.org/10.1145/2983323.2983697>
11. Chen, T., Li, X., Yin, H., Zhang, J., Wu, L., Wang, Y.: Call Attention to Rumors: Deep Attention Based Recurrent Neural Networks for Early Rumor Detection. In: *\*Trends and Applications in Knowledge Discovery and Data Mining – PAKDD 2018 Workshops\**, pp.40–52. Springer, Melbourne, Australia (2018). Preprint: [arXiv:1704.05973](https://arxiv.org/abs/1704.05973). <https://arxiv.org/abs/1704.05973>
12. Kotteti, C.M.M., Dong, X., Qian, L.: Multiple Time-Series Data Analysis for Rumor Detection on Social Media. In: *\*2018 IEEE International Conference on Big Data (BigData)\**, pp.4413–4419. IEEE (2018). <https://doi.org/10.1109/BigData.2018.8622631>
13. Guo, H., Cao, J., Zhang, Y., Guo, J., Li, J.: Rumor Detection with Hierarchical Social Attention Network. In: *\*Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM 2018)\**, pp.943–951. ACM (2018). <https://doi.org/10.1145/3269206.3271709>
14. Yang, Z., Wang, C., Zhang, F., Zhang, Y., Zhang, H.: Emerging Rumor Identification for Social Media with Hot Topic Detection. In: *\*2015 12th Web Information System and Application Conference (WISA)\**, pp.53–58. IEEE (2015). <https://doi.org/10.1109/WISA.2015.19>
15. Serrano, E., Iglesias, C.A., Garijo, M.: A Survey of Twitter Rumor Spreading Simulations. In: *\*Computational Collective Intelligence\* (ICCCI 2015)*, LNCS, vol.9329, pp.113–122. Springer, Cham (2015). <https://doi.org/10.1007/978-3-319-24069-5-11>
16. Elhadad, M.K., Li, K.F., Gebali, F.: Fake News Detection on Social Media: A Systematic Survey. In: *\*2019 IEEE Pacific Rim Conference on Communi-*

- cations, Computers and Signal Processing (PACRIM)\*, pp.1–8. IEEE (2019). <https://doi.org/10.1109/PACRIM47961.2019.8985062>
17. Cao, J., Guo, J., Li, X., Jin, Z., Guo, H., Li, J.: Automatic Rumor Detection on Microblogs: A Survey. arXiv preprint arXiv:1807.03505 (2018). <https://arxiv.org/abs/1807.03505>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

