BenchCouncil Transactions on Benchmarks, Standards and Evaluations 2 (2022) 100035

Contents lists available at ScienceDirect

BenchCouncil Transactions on Benchmarks, Standards and Evaluations

journal homepage: https://www.keaipublishing.com/en/journals/benchcouncil-transactions-on-benchmarks-standards-and-evaluations/

Combating disinformation on social media: A computational perspective

Kai Shu

Department of Computer Science, Illinois Institute of Technology, Chicago 60616, United States of America

A R T I C L E I N F O

Keywords:
Disinformation
Social media mining
Social computing

A B S T R A C T

The use of social media has accelerated information sharing and instantaneous communications. The low barrier to enter social media enables more users to participate and makes them stay engaged longer, while incentivizing individuals with a hidden agenda to use disinformation to manipulate information and influence opinions. Disinformation, such as fake news, hoaxes, and conspiracy theories, has increasingly been weaponized to divide people and create detrimental societal effects. Therefore, it is imperative to understand disinformation and systematically investigate how we can improve resistance against it, taking into account the tension between the need for information and the need for security and protection against disinformation. In this survey, we look into the concepts, methods, and recent advancements of detecting disinformation from a computational perspective. We will also discuss open issues and future research directions for combating disinformation on social media.

Contents

1. Introduction

Social media has become the leading platform for individuals to communicate [1]. In particular, social media is immensely popular for news dissemination, information sharing, and event participation due to its capacity to rapidly spread information at scale [2,3]. While this massive capacity can promote social trust and enhance social connectivity, it can also facilitate the rampant propagation of disinformation [4–6]. Such rampant disinformation often leverages the trust and social connectivity of social media users, spreads manipulated information to foment hatred, and inflicts damages on individuals or groups. With disinformation growing at unprecedented volumes on social media disinformation is now viewed as one of the greatest threats to democracy, justice, public trust, freedom of expression, journalism, and economic growth [4]. Hence, there is a pressing and necessary need to tackle digital disinformation.

However, detecting disinformation and fake news with computational approaches poses unique challenges that make it nontrivial. First, the *data challenge* has been a major roadblock because the content of fake news and disinformation is rather diverse in terms of topics, styles and media platforms; and fake news attempts to distort truth with diverse linguistic styles while simultaneously mocking true news. Thus, obtaining annotated fake news data is non-scalable, and data-specific embedding methods are not sufficient for fake news detection with little labeled data. Second, the *evolving challenge* of fake news and

E-mail address: kshu@iit.edu.

https://doi.org/10.1016/j.tbench.2022.100035
Received 3 December 2021; Received in revised form 22 March 2022; Accepted 23 March 2022
Available online 7 April 2022
2772-4859/© 2022 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Table 1
The comparison with representative fake news detection methods.

Representative Methods	News Content		Social Context			
	Linguistic	Visual	User	Post	Temporal	Network
[10–16]	✓					
[17,18]		✓				
[19–21]	✓	✓				
[22,23]			✓			
[24–26]				✓		
[27]				✓		✓
[28,29]					✓	✓
[30–32]						✓

disinformation is another obstacle in this task—fake news is usually related to newly emerging, time-critical events, which may not have been properly verified by existing knowledge bases due to the lack of corroborating evidence or claims. Third, the *explainability challenge* is concerned with the development of machine learning algorithms for disinformation that are explainable. Existing disinformation detection techniques are often machine learning black boxes that provide little or no explanation on the detection process. Explainability ensures that the developed algorithms are transparent, ensuring ethically responsible and trustworthy algorithms. However, deriving algorithmic explanations useful for domain experts and enhancing explanations by understanding and incorporating prior expert knowledge has been challenging.

We present the recent advancement of learning with weak social supervision to understand and detect disinformation and fake news on social media [7]. In particular, these methods are attempting diverse challenging scenarios towards detecting fake news more effectively, with explainability, at an early stage, and across domains. We further discuss the open issues and future directions along the line of combating disinformation from a computational perspective.

2. Related work and foundations

Disinformation and misinformation have been an important issue and attracts increasing attention in recent years [4,8,9]. The openness and anonymity of social media make it convenient for users to share and exchange information, but also makes it vulnerable to nefarious activities. Although the spread of misinformation and disinformation has been studied in journalism, the openness of social networking platforms, combined with the potential for automation, facilitates the dis/misinformation to rapidly propagate to massive numbers of people, which brings about unprecedented challenges. Specifically, disinformation is fake or inaccurate information that is intentionally spread to mislead and/or deceive; misinformation is false content shared by a person who does not realize that it is false or misleading.

As a representative example of disinformation, we briefly introduce the related work about fake news detection on social media. Fake news detection methods generally focus on using *news content* and *social context* [8,33] (as shown in Table 1).

News content contains the clues to differentiate fake and real news. News content-based features are the most explicit clues for fake news detection, given that the social media news evaluated are primarily textual in nature. The prerequisite of news content-based fake news detection is that the content of fake news should be somewhat different from truth in some quantifiable way [10]. For news content based approaches, features are extracted as linguistic-based and visual-based. Linguistic-based features capture specific writing styles and sensational headlines that commonly occur in fake news content [11], such as lexical and syntactic features. However, these methods have difficulty not only in generalizing hand-crafted linguistic features across topics, languages, and domains but also in utilizing the rich semantic and contextual information. To address the drawbacks of linguistics-based methods, the deep neural networks-based methods such as recurrent neural network (RNN) [12,29], convolutional neural networks (CNN) [13,14], and variational autoencoders (VAEs) [15,16] have been widely explored in recent years due to their capabilities to automatically learn latent textual representation and capture complex contextual patterns of news content.

Visual-based features try to identify fake images [17] that are intentionally created or capturing specific characteristics for images in fake news. News with visual information is likely to attract much more attention from social media users and thus gains a greater range of information dissemination [18]. Combining visual and linguistic features have shown better performances than using a single modality of feature. For example, Jin et al. first proposed a RNN-based automatic multimodal fake news detection model to fuse the visual and textual information of the post using an attention mechanism [19]. In addition, Zhou et al. presented a novel fake news method considering the correlations across the modalities [20].

In addition to news content, *social context* related to news pieces contains rich information to help detect fake news. For social context based approaches, the features mainly include user-based, post-based and network-based. User-based features are extracted from user profiles to measure their characteristics and credibilities [22,34,35]. For example, Shu et al. [22] proposed to understand user profiles from various aspects to differentiate fake news. Yang et al. [23] proposed an unsupervised fake news detection algorithm by utilizing users' opinions on social media and estimating their credibilities.

Post-based features represent users' social response in term of stance [24], topics [25], or credibility [29,34]. Network-based features are extracted by constructing specific networks, such as diffusion network [27,36], bipartite user–news interaction networks [37], etc. Propagation-based models assume that the credibility of news is highly related to the credibilities of relevant social media posts, which several propagation methods can be applied [24]. Recently, geometric deep learning such as Graph Neural Networks (GNNs) have been exploited to detect fake news and shown promising performances [28,30,31]. For example, Nguyen et al. proposed an inductive heterogeneous graph representation framework, Factual News Graph (FANG), which can effectively exploit social structure and engagement patterns of users for fake news detection [32]. Deep learning models are also applied to learn the temporal and linguistic representation of news [21,38]. Research also focuses on challenging problems of fake news detection, such as fake news early detection by adversarial learning [21] and user response generating [26].

Despite the success of aforementioned fake news detection algorithms, they mostly rely on large amounts of labeled instances to train supervised models. Such large labeled training data is often difficult to obtain for disinformation and fake news. Therefore, novel algorithms to learn with weak social supervision are needed for detecting fake news effectively, with explanation, at an early stage, and across domains.

3. Advancing disinformation detection

In this section, we will detail the proposed computational approaches on disinformation detection. We will first introduce a benchmark data repository for fake news detection, then describe four computational tasks on detecting disinformation. These four tasks

Table 2
The comparison with representative fake news detection datasets [39].

Datasets	News Content		Social Context				Spatiotemporal	
	Linguistic	Visual	User	Post	Response	Network	Spatial	Temporal
BuzzFeedNews	✓							
LIAR	✓							
BS Detector	✓							
CREDBANK	✓		✓	✓			✓	✓
BuzzFace	✓			✓	✓			✓
FacebookHoax	✓		✓	✓	✓			
NELA-GT-2018	✓							
FakeNewsNet	✓	✓	✓	✓	✓	✓	✓	✓

address the important aforementioned challenges of data, detection and evolving, and establish a principled way of learning with weak social supervision for social computing. Specifically, we aim to answer the following questions. First, when we have enough labeled data of news content and social context, how to detect fake news more effectively? Second, to involve and benefit domain experts, such as fact-checkers and journalists, how can we make fake news detection results more understandable? Third, social context may be useful to help detect fake news while it takes a long time to aggregate; can we detect fake news at an early stage and how early can it be? At last, fake news has diverse topics and obtaining labels for each domain is costly; would auxiliary information across domains be helpful in detecting fake news? For each of the task, we will present the technical details of the research issues and proposed research.

3.1. FakeNewsNet: A Benchmark data repository

The first and most important step to detect fake news is to collect a benchmark dataset. Despite several existing computational solutions on the detection of fake news, the lack of comprehensive and community-driven fake news datasets has become one of major roadblocks. Therefore, we create, and curate a multi-dimensional data repository *FakeNewsNet*,[1] which contains two datasets with news content, social context, and spatiotemporal information [39]. It currently contains two datasets with (23K) news pieces annotated from fact-checking sites, rich user engagements (691K users, 2M tweets and 2B network followers) from Twitter.

From Table 2, we observe that no existing public dataset can provide all possible features of news content, social context, and spatiotemporal information. The constructed FakeNewsNet repository has the potential to boost the study of various open research problems related to fake news study. First, the rich set of features in the datasets provides an opportunity to experiment with different approaches for fake new detection, understand the diffusion of fake news in social network and intervene in it. Second, temporal information enables the study of early fake news detection. Third, we can investigate the fake news diffusion process by identifying provenances, persuaders, and developing better fake news intervention strategies [40].

3.2. Effective fake news detection

In our recent work [41], we investigate effective fake news detection with social context. The basic idea is that the news ecosystem on social media provides abundant social context information, which involves three basic entities, i.e., publishers, news pieces, and users. Fig. 1 gives an illustration of such ecosystem. In Fig. 1, p_1, p_2 and p_3 are news publishers who publish news $a_1, \ldots, a_4$ and $u_1, \ldots, u_6$ are users who have engaged in sharing these news pieces. In addition, users tend to form social links with like-minded people with similar interests. The *tri-relationship*, the relationship among publishers, news pieces, and users, contains additional information to help detect fake news.

[1] https://github.com/KaiDMML/FakeNewsNet.

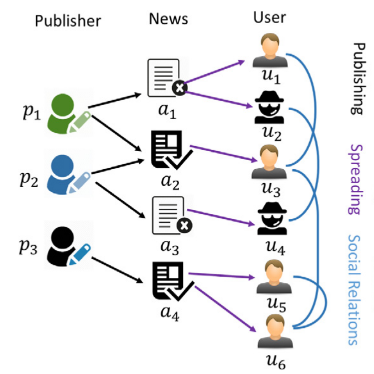

Fig. 1. An illustration of tri-relationship among publishers, news pieces, and users, during the news dissemination process [41].

We propose to use non-negative matrix factorization (NMF) to learn the news representations by encoding the relations in the heterogeneous network with the following objective function:

$$\begin{aligned} \min_{\mathbf{D},\mathbf{U},\mathbf{V},\mathbf{T}\geq 0,\mathbf{p},\mathbf{q}} & \|\mathbf{X}-\mathbf{D}\mathbf{V}^T\|_F^2 + \alpha\|\mathbf{Y}\odot(\mathbf{A}-\mathbf{U}\mathbf{T}\mathbf{U}^T)\|_F^2 + \beta\mathrm{tr}(\mathbf{H}^T\mathbf{L}\mathbf{H}) \\ & + \gamma\|\mathbf{e}\odot(\bar{\mathbf{B}}\mathbf{D}\mathbf{q}-\mathbf{o})\|_2^2 + \eta\|\mathbf{D}_L\mathbf{p}-\mathbf{y}_L\|_2^2 + \lambda R \end{aligned} \tag{1}$$

The proposed framework consists of three major components. First, the news content embedding learns news representation $\mathbf{D}$ by factorizing the bag-of-word matrix $\mathbf{X}$. Second, the social context embedding includes three parts, a user–user embedding from social relations $\mathbf{A}$, a user–news embedding from spreading relations $\mathbf{H}$, and a publisher–news embedding from publishing relation $\mathbf{B}$. Specifically, we decompose the matrix $\mathbf{A}$ into a user representation matrix $\mathbf{U}$, and $\mathbf{T}$ is the user–user correlation matrix, $\mathbf{Y}$ controls the contribution of $\mathbf{A}$, and $\odot$ denotes the Hadamard product operation. For the user–news embedding, we optimize the graph Laplacian of a matrix $\mathbf{H}$ which consists of user and news representation, and $\mathbf{L}$ is the Laplacian matrix. For the publisher–news embedding, the idea is to utilize publisher partisan labels vector $\mathbf{o}$ and the normalized publisher–news matrix $\bar{\mathbf{B}}$ to optimize the news feature matrix $\mathbf{D}$, where $\mathbf{e}$ is a binary matrix that indicates the label availability of publishing relations. Third, we further incorporate a semi-supervised linear classifier to map news representation $\mathbf{D}$ to news label $\mathbf{y}_L$ with the mapping function $\mathbf{p}$; R is a regularization terms to avoid overfitting, and α, β, γ, η, and λ are controlling the importance of each component. Experimental results show that TriFN has better fake news detection performance than state-of-the-art approaches. More details can be found in [41].

3.3. Explainable fake news detection

In [42,43], we study the explainable fake news detection. Despite the promising results of existing fake news detection methods, however, the majority of these methods focus on *detecting* fake news effectively with latent features but cannot explain "why" a piece of news was detected as fake news, which like a black-box. Being able to *explain*

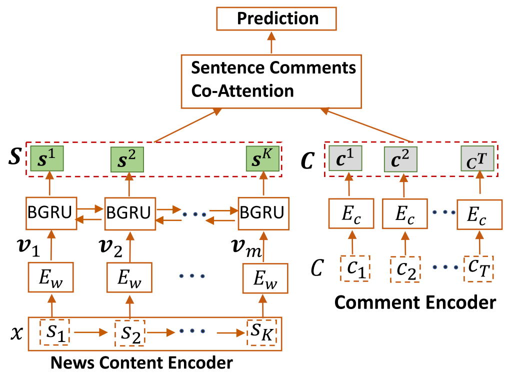

Fig. 2. The proposed framework dEFEND for explainable fake news detection [42].

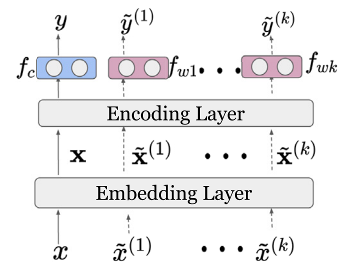

Fig. 3. Learning with weak signals from users' comments to detect disinformation [44].

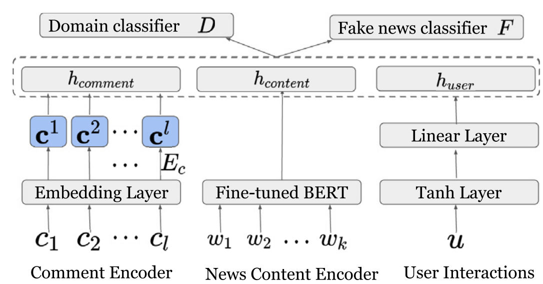

Fig. 4. Cross-domain Fake News Detection [45].

why news was determined as fake from news contents and auxiliary information is much desirable because: (1) the derived explanation can provide new insights and knowledge originally hidden to practitioners; and (2) extracting explainable features from noisy auxiliary information can further help improve fake news detection performance. We aim to tackle the following challenges: (1) how to perform explainable fake news detection that can improve detection performance and explainability simultaneously; (2) how to extract explainable comments without the ground truth during training; and (3) how to model the correlation between news contents and user comments jointly for explainable fake news detection. The basic idea is that news content may contain information that is verifiably false and user comments express users' opinions, stances, and sentiment towards news content. News contents and user comments are inherently *related* to each other and can provide important clues to detect and explain why a given news article is fake. Thus, we simultaneously capture top-k check-worthy news sentences and user comments for fake news detection and explanation.

Fig. 2 shows the proposed dEFEND. In the figure, $x = \{s_1, \dots, s_K\}$ is the content of a news. $C = \{c_1, \dots, c_T\}$ is the set of comments on the news. dEFEND consists of news content encoder, comment encoder, sentence-comment co-attention layer and the prediction layer. The *news content encoder* designs an attentive hierarchical GRU to learn representations of each sentences, denoted as $\mathbf{S} = \{\mathbf{s}^1, \dots, \mathbf{s}^K\}$. Similarly, *comment encoder* uses LSTM to learn representations of comments, denoted as $\mathbf{C} = \{\mathbf{c}^1, \dots, \mathbf{c}^T\}$. With $\mathbf{S}$ and $\mathbf{C}$, the *sentence-comment co-attention layer* models the mutual influences between the news sentences and user comments to assign larger attention-weights to sentences and comments that can be used to detect and explain why a piece of news is fake. The final representation by aggregating representations of important news sentences and user comments is used for classification. Experimental results show that dEFEND has better fake news detection performance than state-of-the-art approaches. Meanwhile, dEFEND can derive meaningful interpretations from news content and comments. More details can be found in [42].

3.4. Early fake news detection

In our recent study, we exploit limited information from multiple sources on user comments (i.e., *weak signals*), along with text content, to detect fake news [44]. We observe that users' posts and comments carry rich crowd information including opinions, stances, and sentiments that can help detect fake news for various reasons. First, previous work has shown that conflicting sentiments among spreaders may indicate fake news [8,24]. Second, different users have different levels of credibility. These findings have great potential to provide additional signals for early detection of fake news. Thus, one can exploit multiple sources of weak social supervision simultaneously (in the form of weak labels) from social media to detect fake news.

To that end, we have developed a framework MWSS jointly exploit Multiple sources of Weak Social Supervision (see Fig. 3). It utilizes a label weighting network (LWN) to model the weight of these weak labels $(\tilde{y}^{(1)}, \dots, \tilde{y}^{(k)})$ to regulate the learning of the fake news classifier f_x. The LWN is a meta-model to produce weights for the weak labels and can be trained by back-propagating the validation loss of a trained classifier on a separate dataset. LWN is suitable for early detection of fake news as LWN only requires news content for prediction. Our results show that weak supervision from multiple sources provides complementary information on top of news content; hence, can significantly improve fake news detection. More details can be found in [44].

3.5. Cross-domain fake news detection

In [45], we study cross-domain fake news detection. We address the challenge of limited within-domain labeled data for high-performance fake news detection by leveraging on cross-domain knowledge transfer and within-domain joint learning. We introduce a principled framework, CrossFND (Cross-domain Fake News Detection) (see Fig. 4). First, we utilize the embedding layer to encode each comment $c_1, \dots, c_l$ to obtain the comment representation $h_{comment} = [\mathbf{c}^1; \cdots; \mathbf{c}^l]$, fine-tuning BERT with the training data in the source and target domain such that it learns news representation $h_{content}$, and pass the user–news interaction vector u to a nonlinear and linear layer and get h_{user}. Then these representation vectors are concatenated and perform two tasks: (i) predicting whether a piece of news is fake or not with F; and (2) predicting which domain is the news coming from (i.e., source or target) with D. The final objective is a minmax function: $\min_F \max_D L(F) - \alpha L(D)$; where $L(D)$ ($L(F)$) is the cross-entropy loss to predict domain (news) label. Because we require the model to learn domain-independent features, we maximize domain loss $L(D)$. This forces the model to create a representation for news contents that represses domain-specific features; thus, it tries to deceive the classifier. To meet the objective of accurately classifying news as fake or authentic, we minimize the fake news classifier's loss function $L(F)$. More details can be found in [45].

4. Open issues and future research

Combating disinformation is an ongoing and continuous challenge in society. We discuss several emerging open issues for developing computational methods to detect and mitigate disinformation, highlighting the pressing need for interdisciplinary research.

4.1. Trustworthy AI for combating disinformation

People are reluctant to believe the results of AI-powered disinformation detection tools as these techniques are often like a black-box and lack of transparency. However, most of the current methodologies are data-driven and conducted in a passive fashion, and thus are inadequate to apprehend knowledge from human intents and demands. Thus, it is important to adapt learning strategies and acquire knowledge from human feedback. To establish the explainable methods for combating disinformation, it is important to investigate: (1) how to transform the meaningful human cognition into knowledge? (2) how to incorporate noisy, incomplete, and complicated human feedback for better representation learning? (3) how to interpret prediction results with knowledge reasoning and causal discovery. In addition, it is also critical to build equitable AI algorithms for detecting disinformation. For example, disinformation has been targeting marginalized groups, and it is important to understand how to measure the different types biases are and their degree in disinformation. Moreover, building a robust disinformation detector can increase its trustworthy and proactively considering adversarial attacks on disinformation detection methods can better guide the development of robust detection models.

4.2. Neural disinformation generation and detection

With the recent development of neural generation models such as GPT-3 [46] and BART [47], it becomes possible to generate realistic long documents such as news. There is a growing concern on these powerful language models being utilized by malicious users to generate disinformation. Detecting these neural fake news pieces firstly requires us to understand the characteristic of these news pieces and the detection difficulty. Some recent work propose to generate topic-preserving [48] and factual-enhanced [49] synthetic news pieces to understand the characteristics of machine-generated news. Though recent advancements on computationally detecting human-written disinformation show some promising results, it is largely unclear whether these models can distinguish machine-written disinformation effectively. It is important to explore: (1) how to generate realistic fake news with neural generative models to better understand neural-generated fake news? and (2) how is the capacity of human and machine on differentiating human-generated and machine-generated fake/real news?

4.3. Online disinformation and its offline impact

Disinformation is widely spread in online social networks and has been shown to result in offline events. Existing social network studies are mostly separately conducted to understand user online behaviors and offline activities in cyber and physical space. However, there is a growing need to integrate studies of these online and offline space and investigate their interdependencies. There is a gap to identify and assess how an online disinformation piece triggers actual behavior changes in the offline space. To establish the fundamental understanding of this gap, it is essential to explore: (1) how to discover the behavioral dynamics that are significant for inducing behavior change among different types of social networks; (2) how to develop high-fidelity models to understand the behavioral dynamics of transfer and spread among online and offline communities; and (3) how to build connections and causal analysis to understand the effects of online disinformation to the offline real-world events. Interdisciplinary research is much desired from disciplines such as social science, psychology, and computer science to deeply understand the aforementioned questions.

5. Conclusion

With the increasing popularity of social media, more and more people consume news from social media instead of traditional news media. However, social media has also enabled the wide dissemination of disinformation. In this article, we explored the fake news problem by reviewing existing literature, and discuss recent advancements for computationally detecting fake news with social media data and analysis. We also further discuss the promising future directions for combating disinformation and the pressing need for interdisciplinary research.

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.

Acknowledgments

This material is in part supported by XCMG American Research Corporation, United States of America and Google Cloud Research Credits Award, United States of America.

References

[1] A. Perrin, Social media usage, Pew Res. Cent. 125 (2015) 52–68.
[2] A. Al-Rawi, Viral news on social media, Digit. J. 7 (1) (2019) 63–79.
[3] E. Shearer, A. Mitchell, News Use Across Social Media Platforms in 2020, Pew Research Center, 2020.
[4] K. Shu, H. Liu, Detecting fake news on social media, Synth. Lect. Data Min. Knowl. Discov. 11 (3) (2019) 1–129.
[5] H. Allcott, M. Gentzkow, Social Media and Fake News in the 2016 Election, Tech. Rep., National Bureau of Economic Research, 2017.
[6] A. Guess, J. Nagler, J. Tucker, Less than you think: Prevalence and predictors of fake news dissemination on facebook, Sci. Adv. 5 (1) (2019) eaau4586.
[7] K. Shu, Understanding Disinformation: Learning with Weak Social Supervision (Ph.D. thesis), Arizona State University, 2020.
[8] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: A data mining perspective, ACM SIGKDD Explor. Newsl. 19 (1) (2017) 22–36.
[9] K. Shu, A. Bhattacharjee, F. Alatawi, T.H. Nazer, K. Ding, M. Karami, H. Liu, Combating disinformation in a social media age, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 10 (6) (2020) e1385.
[10] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, Y. Liu, Combating fake news: A survey on identification and mitigation techniques, ACM Trans. Intell. Syst. Technol. (TIST) 10 (3) (2019) 1–42.
[11] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, B. Stein, A stylometric inquiry into hyperpartisan and fake news, 2017, arXiv preprint arXiv:1702.05638.
[12] T. Chen, X. Li, H. Yin, J. Zhang, Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2018, pp. 40–52.
[13] W.Y. Wang, "Liar, liar pants on fire": A new benchmark dataset for fake news detection, 2017, arXiv preprint arXiv:1705.00648.
[14] M.H. Goldani, R. Safabakhsh, S. Momtazi, Convolutional neural network with margin loss for fake news detection, Inf. Process. Manage. 58 (1) (2021) 102418.
[15] M. Cheng, S. Nazarian, P. Bogdan, Vroc: Variational autoencoder-aided multi-task rumor classifier based on text, in: Proceedings of the Web Conference 2020, 2020, pp. 2892–2898.
[16] D. Khattar, J.S. Goud, M. Gupta, V. Varma, Mvae: Multimodal variational autoencoder for fake news detection, in: The World Wide Web Conference, 2019, pp. 2915–2921.
[17] A. Gupta, H. Lamba, P. Kumaraguru, A. Joshi, Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy, in: WWW'13.
[18] P. Qi, J. Cao, T. Yang, J. Guo, J. Li, Exploiting multi-domain visual information for fake news detection, in: 2019 IEEE International Conference on Data Mining, ICDM, IEEE, 2019, pp. 518–527.
[19] Z. Jin, J. Cao, H. Guo, Y. Zhang, J. Luo, Multimodal fusion with recurrent neural networks for rumor detection on microblogs, in: Proceedings of the 25th ACM International Conference on Multimedia, 2017, pp. 795–816.
[20] X. Zhou, J. Wu, R. Zafarani, [... Formula...]: Similarity-aware multi-modal fake news detection, Adv. Knowl. Discov. Data Min. 12085 (2020) 354.
[21] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, J. Gao, EANN: Event adversarial neural networks for multi-modal fake news detection, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, 2018, pp. 849–857.

[22] K. Shu, S. Wang, H. Liu, Understanding user profiles on social media for fake news detection, in: MIPR, 2018.
[23] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, H. Liu, Unsupervised fake news detection on social media: A generative approach, in: AAAI, 2019.
[24] Z. Jin, J. Cao, Y. Zhang, J. Luo, News verification by exploiting conflicting social viewpoints in microblogs, in: AAAI, 2016.
[25] J. Ma, W. Gao, Z. Wei, Y. Lu, K.-F. Wong, Detect rumors using time series of social context information on microblogging websites, in: CIKM, 2015.
[26] F. Qian, C. Gong, K. Sharma, Y. Liu, Neural user response generator: Fake news detection with collective user intelligence, in: IJCAI, 2018.
[27] Y. Dou, K. Shu, C. Xia, P.S. Yu, L. Sun, User preference-aware fake news detection, 2021, arXiv preprint arXiv:2104.12259.
[28] C. Song, K. Shu, B. Wu, Temporally evolving graph neural network for fake news detection, Inf. Process. Manage. 58 (6) (2021) 102712.
[29] L. Wu, H. Liu, Tracing fake-news footprints: Characterizing social media messages by how they propagate, in: WSDM, 2018.
[30] Y.-J. Lu, C.-T. Li, GCAN: Graph-aware co-attention networks for explainable fake news detection on social media, 2020, arXiv preprint arXiv:2004.11648.
[31] A. Silva, Y. Han, L. Luo, S. Karunasekera, C. Leckie, Propagation2Vec: Embedding partial propagation networks for explainable fake news early detection, Inf. Process. Manage. 58 (5) (2021) 102618.
[32] V.-H. Nguyen, K. Sugiyama, P. Nakov, M.-Y. Kan, Fang: Leveraging social context for fake news detection using graph representation, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 1165–1174.
[33] X. Zhou, R. Zafarani, K. Shu, H. Liu, Fake news: Fundamental theories, detection strategies and challenges, in: WSDM, 2019.
[34] C. Castillo, M. Mendoza, B. Poblete, Information credibility on twitter, in: WWW, 2011.
[35] S. Kwon, M. Cha, K. Jung, W. Chen, Y. Wang, Prominent features of rumor propagation in online social media, in: ICDM, 2013.
[36] K. Shu, D. Mahudeswaran, S. Wang, H. Liu, Hierarchical propagation networks for fake news detection: Investigation and exploitation, in: Proceedings of the International AAAI Conference on Web and Social Media, Vol. 14, 2020, pp. 626–637.
[37] X. Zhou, R. Zafarani, Network-based fake news detection: A pattern-driven approach, ACM SIGKDD Explor. Newsl. 21 (2) (2019) 48–60.
[38] H. Karimi, P. Roy, S. Saba-Sadiya, J. Tang, Multi-source multi-class fake news detection, in: COLING, 2018.
[39] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, FakeNewsNet: A data repository with news content, social context and dynamic information for studying fake news on social media, 2018, arXiv preprint arXiv:1809.01286.
[40] K. Shu, H.R. Bernard, H. Liu, Studying fake news via network analysis: Detection and mitigation, 2018, CoRR, arXiv:1804.10233.
[41] K. Shu, S. Wang, H. Liu, Beyond news contents: The role of social context for fake news detection, in: WSDM, 2019.
[42] K. Shu, L. Cui, S. Wang, D. Lee, H. Liu, Defend: Explainable fake news detection, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 395–405.
[43] L. Cui, K. Shu, S. Wang, D. Lee, H. Liu, Defend: A system for explainable fake news detection, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 2961–2964.
[44] K. Shu, G. Zheng, Y. Li, S. Mukherjee, A.H. Awadallah, S. Ruston, H. Liu, Leveraging multi-source weak social supervision for early detection of fake news, 2020, arXiv:2004.01732.
[45] K. Shu, M. Ahmadreza, H. Liu, Cross-domain fake news detection on social media: A context-aware adversarial approach, in: Frontiers in Fake Media Generation and Detection, Springer, 2022.
[46] T.B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, 2020, arXiv preprint arXiv:2005.14165.
[47] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, 2019, arXiv preprint arXiv:1910.13461.
[48] A. Mosallanezhad, K. Shu, H. Liu, Topic-preserving synthetic news generation: An adversarial deep reinforcement learning approach, 2020, arXiv preprint arXiv:2010.16324.
[49] K. Shu, Y. Li, K. Ding, H. Liu, Fact-enhanced synthetic news generation, in: Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence, 2021.