Holistic Analysis of Organised Misinformation Activity in Social Networks

Anselmo Peñas[1(✉)], Jan Deriu[2], Rajesh Sharma[3], Guilhem Valentin[4], and Julio Reyes-Montesinos[1]

[1] NLP & IR Group, UNED, Madrid, Spain
anselmo@lsi.uned.es
[2] CAI, ZHAW, Winterthur, Switzerland
jan.deriu@zhaw.ch
[3] University of Tartu, Tartu, Estonia
rajesh.sharma@ut.ee
[4] Synapse Développement, Toulouse, France
guilhem.valentin@synapse-fr.com

Abstract. To tackle the problem of disinformation, society must be aware not only of the existence of intentional misinformation campaigns, but also of the agents that introduce the misleading information, their supporting media, the nodes they use in social networks, the propaganda techniques they employ and their overall narratives and intentions. Disinformation is a challenge that must be addressed holistically: identifying and describing a disinformation campaign requires studying misinformation locally, at the message level, as well as globally, by modelling its propagation process to identify its sources and main players. In this paper, we argue that the integration of these two levels of analysis hinges on studying underlying features such as disinformation's intentionality, and benefited and injured agents. Taking these features into account could make automated decisions more explainable for end users and analysts. Moreover, simultaneously identifying misleading messages, knowing their narratives and hidden intentions, modelling their diffusion in social networks, and monitoring the sources of disinformation will also allow a faster reaction, even anticipation, against the spreading of disinformation.

Keywords: Disinformation · Social Networks · News Content Analysis · Intentions and Narratives

1 Introduction

Among the different kinds of misinformation, perhaps the most dangerous is the one created with the intention to harm, polarise, destabilise, generate distrust

or destroy reputation by means of spreading false information. In a scenario of organised intentional misinformation campaigns (also called disinformation[1]) current fact-checking strategies are not enough.

Fact Checkers need Artificial Intelligence tools to help them identify the most important claims to check (check-worthiness), detect claims that they have already checked (verified claim retrieval), and be able to check claims as soon as possible. This is important because fake news spreads 6 times faster than true ones [37], and 50% of the fake news propagation occurs in the first 10 min [39]. Disinformation is carefully constructed to behave this way, it has an intention (not always explicit) and a coordinated spreading (opportunistic most of the times).

Given this scenario of organised intentional misinformation campaigns, we need a comprehensive strategy to anticipate and mitigate the spreading of disinformation. We, as a society, must be aware not only about fake news, but also about the agents that introduce false or misleading information, their supporting media, the nodes they use in the social networks, the propaganda techniques they use, the narratives they try to push and their intentions.

Therefore, we must address this challenge in a holistic way, considering the different dimensions involved in the spreading of disinformation and bring them together to really identify and describe the orchestrated disinformation campaigns:

1. Detect misinformation: claim worthiness checking, stance detection, fake news identification and verified claim retrieval;
2. Acknowledging their organised spreading in social networks: models of disinformation propagation and source detection using social network analysis;
3. Identifying its malicious intent: narratives that are wanted to be spread, benefited and harmed agents and final goals;
4. Bring everything together: collect all the evidence and give them to final assessors and users in explainable ways, and use the aggregated information in a loop to recover in a new cycle the data missed in the previous ones.

To clarify the importance of attempting an holistic approach we need to consider the stakeholders of the technology under development. The main recipients would be content analysts that make use of services such as fact-checkers for a further analysis and better understanding of the agents and narratives involved in disinformation campaigns. For example, in electoral processes, independent observers must study disinformation campaigns in an holistic fashion to identify underlying communication intentions with specific narratives aimed to influence the elections outcome.

Tackling the hidden intention behind disinformation campaigns will help us fighting in a more efficient way. Fighting a misleading narrative should be easier than fighting all the single messages spread to promote that narrative. But for this purpose we need to move from just checking single messages or just analysing alterations in the social network to contemplating the whole picture.

---

[1] From here we use misinformation and disinformation interchangeably.

2   Previous Work

Previous works have addressed the problem of disinformation from two main different perspectives:

2.1   Content Analysis

Researchers have analysed misinformation-related tasks with various NLP-related features. The first task is to identify whether a new incoming content contains one or more claims that are worth to be checked [9,12,21,35]. Strategies to detect disinformation include the study of the correlation between psycholinguistics features and misinformation [2,4,23], usage of state of the art techniques such as knowledge graphs [18], reinforcement learning [20], context-aware misinformation detection [43] or the detection of alterations in original news [29]. Apart from exploring only text-based mechanisms, multimodal co-Attention networks (MCAN) have been used to exploit both textual and visual features for fake news detection [8,38]. Besides, recent works have improved the detection process by including non-textual features related to the user sharing the news, although there is a lack of datasets in this direction [25,28,30].

Proper fact-checking the claims of a content is a task that still requires the intervention of experts, usually journalists or domain experts from the civil society [26]. Hence, this task typically consumes a large amount of resources and time, while fake news tends to spread fast and come back repeatedly, even after having been checked and debunked. Thus, the verified claim retrieval task consists in ranking verified claims that can "help verify the input claim, or a sub-claim in it" [24] to avoid a costly repeated task of fact-checking similar claims [36].

There have also been works for detecting suspicious and fake user profiles on online social media platforms often involved in spreading misinformation related news, such as Facebook [11], Twitter [1,27] or Tuenti [3]. These techniques include exploring user information such as immediate connections [1] and other meta-information such as user names [33].

Usually, disinformation is produced using propaganda techniques that help to accelerate its propagation. These techniques include specific rhetorical and psychological techniques, ranging from leveraging of emotions (such as using loaded language, flag waving, appeal to authority, slogans, and cliches) to using logical fallacies such as straw men (misrepresenting someone's opinion), red herring (presenting irrelevant data), black-and-white fallacy (presenting two alternatives as the only possibilities), and *whataboutism* [17]. A shared task was held within the 2019 on the PTC corpus [6] to identify both the specific text fragments where a propaganda technique is used and the type of technique used among 18 types. The best-performing models for both tasks used BERT based contextual representations.

2.2   Social Network Analysis

Disinformation campaigns rely nowadays on coordinated efforts to spread messages at scale. Such coordination is achieved by leveraging botnets (groups of fully automated accounts), cyborgs (partially automated) and troll armies (human-driven).

At the social network level, the current research trend is to target groups of accounts as a whole, rather than focusing on individual accounts [27]. The rationale for this choice is that malicious accounts act in coordination to amplify their effect [40]. Coordinated behaviour appears as near-fully connected communities in graphs, dense blocks in adjacency matrices, or peculiar patterns in spectral subspaces [13]. A large cluster of accounts with highly similar behaviour along time series are indications of a disinformation campaign.

The spreading of disinformation has been modelled through epidemic metaphors [7]. A (fake) piece of information, indeed, can be seen as a virus that may potentially infect people. Many SIR-based models have been proposed to model rumour spreading [19], adding forgetting and remembering mechanisms [42], sceptical agents [14], and competition among rumours [34]. [32] simulated the spreading of a hoax and its debunking at the same time taking forgetfulness into account by making a user lose interest in the fake news item with a given probability. The same authors extended their previous work comparing different fact-checking strategies on different network topologies to limit the spreading of fake news [31]. [22] studied the influence of online bots on a network through simulations, in an opinion dynamic setting. [5] studied how the presence of heterogeneous agents affects the competitive spreading of low- and high-quality information.

2.3   Multi-modal Analysis

Although there exists within the research community an awareness of the need to integrate content analysis and social network analysis to tackle misinformation [17], hitherto efforts in this direction have been limited. There are currently three approaches to combining signals from different modalities: (i) early-fusion, where features from different modalities are learned, fused, and fed into a single prediction model [8]; (ii) late-fusion, where unimodal decisions are fused via some averaging mechanism, and (iii) hybrid-fusion, where some of the features are early-fused, and other modalities are late-fused [15]. In these fusion strategies, the learning setup can also be divided into unsupervised, semi-supervised, supervised and self-supervised methods.

3   Problem Statement

There is a lack of research efforts that jointly consider the content analysis dimension and the network analysis dimension of disinformation [17].

Integrating multi-modal models for misinformation detection with network models of misinformation diffusion to identify large misinformation campaigns

and their narratives constitutes a novel, holistic view of misinformation. It poses considerable and exciting challenges both on the conceptual and technical level.

There are two main current technologies to deal with the detection of disinformation. One, related to the needs of fact-checkers, focusing on the processing and analysis of single messages. The other, related to the detection of disinformation campaigns organised to influence a social network, relies on social network analysis: highly similar behaviour of different user accounts along time series are indications of a disinformation campaign.

However, both research lines remain separate research fields, although one gives context to the other. In fact, current AI models for misinformation detection are limited in the ability to represent and consider contextual information. It is still a research frontier we want to address. We must address the integration of different technologies at both message and social network levels into a single system.

A straightforward approach would be to run all involved systems separately and then compare and combine their output. However, they don't leverage each other's signals and, in fact, the current state of the art achieves rather low performance.

Thus, the alternative is what we call an "holistic" approach, where all tasks are considered simultaneously by one integrated system. Our position here is that, in order to integrate these two signal sources we must take advantage of the hidden variable they share: the intentionality of the communication. Following this perspective, many research questions arise and have to be addressed.

This resembles the end-to-end approach with neural networks which has replaced component-based architectures for several NLP tasks. Apart from solving the "whole" task—i.e. detection and description of organised disinformation campaigns—we also see a great potential to improve each single subtask, since they have access to much more data and insights. This hope is motivated by the success of multi-task learning, where additional unrelated subtasks help each other [41]. Messages that would be missed by local analysis could be uncovered at this deeper latent level if they are strongly connected to an identified potential harmful network and, provided with contextual information to better interpret their intention, eventually bring them to the attention of analysts.

4   Towards a Holistic Methodology for Disinformation Analysis

Our position is that we need methodologies that gather evidence from the message and social network levels and try to integrate both by inferring the narratives and intentions behind their spreading. In the following subsections we describe in more detail some of the core elements such methodologies must integrate.

4.1 Disinformation Detection at the Message Level in Multiple Modalities

Tackling disinformation at the local level of individual messages has been extensively reviewed in literature, specially with respect to the identification of fake news. However, moving beyond twitter that used to provide a network context for the message, and a user profile also as context, the task is still unsolved. There are several reasons for this, such as the combination of images and text, the lack of broader communication contexts, and the pragmatic use of language where implicatures are raised into the receptor by means of humor, irony or misleading reasoning.

The reconstruction of the communicative context justify the need of holistic methodologies but, still, there are some signals to be recovered from single messages both related to the semantic content and to the communication style.

Stylometric Analysis. Current systems for disinformation identification, such as fake-news checkers, usually rely on text. Under the assumption that the text content might use specific writing styles focused on convincing readers, we can conduct stylometric analysis of this content.

Studying the Use of Propaganda Techniques. While studies about disinformation detection often employ definitions of the term that differ on the conditions of untruthfulness and harmfulness, some widely employed rhetorical and psychological devices are more stably defined and therefore allow more straightforward approaches to disinformation at the message level. For instance, harmful content often makes use of well-defined propaganda techniques [17], which we can leverage to detect common patterns in disinformation writing.

Multi-modal Content Analysis. Multi-modal content analysis needs to account for each of the modalities present in the message, as well as for the interactions between these different modalities. Another clear challenge in multimodal content is that audio and video posts feature spoken language, while current language models have been trained on written language. The use of speech-to-text technology to transcribe audio and video posts requires taking into account those models for misinformation detection that perform well on written text but still have to be adapted to robustly handle the repetitions, stutters, and interjections present in spoken language transcripts.

Addressing Content in Low-Resource Languages. Automatic disinformation classifiers at message-level are by nature limited for low-resource languages. Introducing multilingual and cross-lingual language models would be a significant improvement for the verified claim retrieval and message clustering sub tasks, especially for languages with less support in terms of labelled data and limited verified claim knowledge base [16]. Such approach has demonstrated its

value as for fact checking [10], but it has not yet been exploited to perform verified claim retrieval.

4.2 Disinformation Detection at Social Network Level

User Profile Features. It is difficult to verify new information as it spreads quickly through social networks. Thus, we must consider features related to user profile such as followers with the goal of modelling the behaviour of disinformation spreaders. To this end, different ways of combining textual and non-textual features - still an open research question - must be explored.

Leveraging the Diffusion Network to Model Communities and Echo Chambers. Social metadata attached to messages often describes a network by listing all the nodes involved in the propagation of a certain message. In combination with the network's structure, this information allows describing the role of different nodes in the disinformation diffusion network in terms of their structural position in such network.

By studying the spread of multiple (clustered) messages we can identify and model communities, determining whether nodes are part of multiple communities or to single ones. The presence of multiple nodes that remain stuck to a single community can be an indication of polarisation and reveal that such community is an echo chamber.

Sources and Means of Diffusion. To identify misinformation sources, we can take advantage of the techniques developed in the previous step at message level. Subsequently, network science techniques can be applied to identify the sources of disinformation. Nodes in the network represent the individuals involved in disinformation propagation (edges) by forwarding, retweeting or reposting.

4.3 Studying the Context Behind Intentional Spreading of Misleading Information

Disinformation campaigns consist of multiple messages and multiple related claims which spread in a coordinated way through multiple pathways in social networks. We can only can observe the messages and their spreading in the network, but their occurrence is due to some intention or goal which is pushed by means of a set of narratives. Narratives and intentions are the primary communication context we need to infer in order to correctly analyse the content of individual messages.

Modelling Intentionality. The integration of evidence coming from the message and network levels can be articulated around the idea of disinformation intentionality: agents that create and introduce disinformation in the social media networks carefully select narratives aimed to have a concrete impact such

as influencing the outcome of elections by discrediting political adversaries, influence financial markets, polarise and destabilise society, generate distrust, destroy reputation, etc. This adversarial game has, at the end, benefited and injured agents.

Modelling Narratives. Our hypothesis is that, given a scenario (e.g. a political election process), the set of intentions at play will be finite (e.g. destroying an opponent's reputation), and the narratives used to achieve it (e.g. X has money overseas) will be limited and predictable according to some general taxonomies.

Real scenarios of disinformation such as political elections can be seen as event-type instances from which to build these taxonomies of intents and narratives in disinformation.

4.4   Holistic Integration and Prediction

The holistic integration is an important stepping stone for modelling and detecting organised misinformation campaigns. We have so far described which steps would be taken at the local (individual message) and global (diffusion network) levels, and how these would be combined at a third level (intentions and narratives). New findings at this third level should improve detection at the former levels, in a virtuous loop. In other words, once the hidden intent is detected, we would come back to the message and network levels, this time bringing the aggregated evidence from all three levels. In this way, we will find new opportunities to capture the items that local approaches missed in the first pass, and new disinformation propagation paths.

The reconstruction of a broader communicative context will also give us the chance to find patterns that enable some kind of prediction power. For example, we could observe that in the scenario of a political election there will be some agents with the goal of delegitimizing the outcome of the process, so we can expect narratives questioning the counting process, so we can predict messages discrediting the agents involved in this process. This prediction power at the narrative level could help us to rise mitigation actions even before the disinformation comes into play.

5   Risks and Challenges

A general challenge for computational approaches to disinformation mitigation stems from the lack of agreement on a definition of disinformation. Studies differ on whether they require disinformation to be both untrue and harmful, and some authors propose that these two dimensions are not binary. For instance, information can be true but misleading. Another problem with disinformation's definition is that it is usually done in reactive terms, meaning that by the time of detection, the damage is already done. We pose that focusing on modelling underlying features such as intentionality helps to circumvent these problems.

A significant challenge for the holistic approach arises from the divergent focuses of the technologies involved, namely Natural Language Processing, Social Network Analysis, Epidemic Modelling, and Agent-based Simulation. Leaving aside the ambitious goal of a complete integration of signals from both the message and the network levels, a holistic approach can still advance the current state of the art in various tasks even if only partial integrations can ultimately be realized.

For instance, considering implicit narratives can help us cluster messages and therefore capture those misleading messages that a local approach would miss. This, in turn, will help identify undetected social network nodes involved in the spreading of disinformation. In the opposite direction, clusters of similar misleading messages can help us build language models that identify implicit narratives and hidden intentions. Such an approach would help addressing the issue with mitigation action's reactive definition.

However, there are some risks also related to the interpretation of the scenarios where disinformation campaigns occur. For example, we must be very careful about making explicit intentions or benefited agents of disinformation campaigns due to the subjective nature of their inference. Although we could talk about the effects or injured agents, it might be better to model these ideas as hidden variables that allocate more evidence in the holistic approach, instead of working in the goal of making them explicit.

To this moment, there is also a lack of established methodologies for evaluating disinformation detection systems able to consider all these levels of analysis. Developing such methodologies and organizing evaluation campaigns in different languages could be a first step in fostering a stronger interdisciplinary research community in Europe around the field of misinformation.

6   Conclusion

In this paper we have motivated the need of holistic methodologies for the identification of disinformation campaigns. The holistic integration of signals coming from message and network levels requires the modelling of implicit or hidden variables such as the types of narratives and their intentions or goals. At the end, what we need is to advance the state of the art towards methodologies aimed to reconstruct the communicative context of these campaigns.

Towards this goal, we really need to involve other disciplines such as journalism and communication, politics and sociology, or psychology.

We claim that the multiple interactions that take place within a holistic approach will give us the opportunity to evolve and achieve new results even if some attempts don't succeed.

References

1. Benevenuto, F., Magno, G., Rodrigues, T., Almeida, V.: Detecting spammers on Twitter. In: Collaboration, Electronic Messaging, Anti-abuse and Spam Conference (CEAS), vol. 6, p. 12 (2010)

2. Butt, S., Sharma, S., Sharma, R., Sidorov, G., Gelbukh, A.: What goes on inside rumour and non-rumour tweets and their reactions: a psycholinguistic analyses. Comput. Hum. Behav. **135**, 107345 (2022)

3. Cao, Q., Sirivianos, M., Yang, X., Pregueiro, T.: Aiding the detection of fake accounts in large scale social online services. In: Presented as part of the 9th {USENIX} Symposium on Networked Systems Design and Implementation, {NSDI} 2012, pp. 197–210 (2012)

4. Chulvi, B., Toselli, A., Rosso, P.: Fake news and hate speech: language in common. arXiv preprint arXiv:2212.02352 (2022)

5. Cisneros-Velarde, P., Oliveira, D.F.M., Chan, K.S.: Spread and control of misinformation with heterogeneous agents. In: Cornelius, S.P., Granell Martorell, C., Gómez-Gardeñes, J., Gonçalves, B. (eds.) CompleNet 2019. SPC, pp. 75–83. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-14459-3_6

6. Da San Martino, G., Yu, S., Barrón-Cedeno, A., Petrov, R., Nakov, P.: Fine-grained analysis of propaganda in news article. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP, pp. 5636–5646 (2019)

7. Daley, D.J., Kendall, D.G.: Epidemics and rumours. Nature **204**, 1118 (1964)

8. Dhawan, M., Sharma, S., Kadam, A., Sharma, R., Kumaraguru, P.: Game-on: graph attention network based multimodal fusion for fake news detection. arXiv preprint arXiv:2202.12478 (2022)

9. Gencheva, P., Nakov, P., Màrquez, L., Barrón-Cedeño, A., Koychev, I.: A context-aware approach for detecting worth-checking claims in political debates. In: Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, pp. 267–276 (2017)

10. Ghanem, B., Glavaš, G., Giachanou, A., Ponzetto, S.P., Rosso, P., Rangel, F.: UPV-UMA at CheckThat! Lab: verifying Arabic claims using a cross lingual approach. In: CEUR Workshop Proceedings, vol. 2380, pp. 1–10. RWTH Aachen (2019)

11. Gupta, A., Kaushal, R.: Towards detecting fake user accounts in Facebook. In: 2017 ISEA Asia Security and Privacy (ISEASP), pp. 1–6. IEEE (2017)

12. Hassan, N., Li, C., Tremayne, M.: Detecting check-worthy factual claims in presidential debates. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 1835–1838 (2015)

13. Jiang, M., Cui, P., Beutel, A., Faloutsos, C., Yang, S.: Inferring lockstep behavior from connectivity pattern in large graphs. Knowl. Inf. Syst. **48**, 399–428 (2016)

14. Jin, F., Dougherty, E., Saraf, P., Cao, Y., Ramakrishnan, N.: Epidemiological modeling of news and rumors on Twitter. In: Proceedings of the 7th Workshop on Social Network Mining and Analysis, pp. 1–9 (2013)

15. Jin, Z., Cao, J., Guo, H., Zhang, Y., Luo, J.: Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In: Proceedings of the 25th ACM International Conference on Multimedia, pp. 795–816 (2017)

16. Kazemi, A., Garimella, K., Gaffney, D., Hale, S.A.: Claim matching beyond English to scale global fact-checking. arXiv preprint arXiv:2106.00853 (2021)

17. Martino, G.D.S., Cresci, S., Barrón-Cedeño, A., Yu, S., Di Pietro, R., Nakov, P.: A survey on computational propaganda detection. arXiv preprint arXiv:2007.08024 (2020)

18. Mayank, M., Sharma, S., Sharma, R.: DEAP-FAKED: knowledge graph based approach for fake news detection. In: 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 47–51 (2022)

19. Moreno, Y., Nekovee, M., Pacheco, A.F.: Dynamics of rumor spreading in complex networks. Phys. Rev. E **69**(6), 066130 (2004)
20. Nikopensius, G., Mayank, M., Phukan, O.C., Sharma, R.: Reinforcement learning-based knowledge graph reasoning for explainable fact-checking. In: 2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (2023)
21. Patwari, A., Goldwasser, D., Bagchi, S.: TATHYA: a multi-classifier system for detecting check-worthy statements in political debates. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 2259–2262 (2017)
22. Ross, B., Pilz, L., Cabrera, B., Brachten, F., Neubaum, G., Stieglitz, S.: Are social bots a real threat? An agent-based model of the spiral of silence to analyse the impact of manipulative actors in social networks. Eur. J. Inf. Syst. **28**(4), 394–412 (2019)
23. Schütz, M., Schindler, A., Siegel, M., Nazemi, K.: Automatic fake news detection with pre-trained transformer models. In: Del Bimbo, A., et al. (eds.) ICPR 2021. LNCS, vol. 12667, pp. 627–641. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-68787-8_45
24. Shaar, S., Martino, G.D.S., Babulkov, N., Nakov, P.: That is a known lie: detecting previously fact-checked claims. arXiv preprint arXiv:2005.06058 (2020)
25. Sharma, S., Agrawal, E., Sharma, R., Datta, A.: FaCov: Covid-19 viral news and rumors fact-check articles dataset. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 16, pp. 1312–1321 (2022)
26. Sharma, S., Datta, A., Shankaran, V., Sharma, R.: Misinformation concierge: a proof-of-concept with curated Twitter dataset on Covid-19 vaccination. In: CIKM (2023)
27. Sharma, S., Sharma, R.: Identifying possible rumor spreaders on Twitter: a weak supervised learning approach. In: 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2021)
28. Sharma, S., Sharma, R., Datta, A.: (Mis) leading the Covid-19 vaccination discourse on Twitter: an exploratory study of infodemic around the pandemic. IEEE Trans. Comput. Soc. Syst. (2022)
29. Shu, K., Cui, L., Wang, S., Lee, D., Liu, H.: dEFEND: explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 395–405 (2019)
30. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: FakeNewsNet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data **8**(3), 171–188 (2020)
31. Tambuscio, M., Ruffo, G.: Fact-checking strategies to limit urban legends spreading in a segregated society. Appl. Netw. Sci. **4**, 1–19 (2019)
32. Tambuscio, M., Ruffo, G., Flammini, A., Menczer, F.: Fact-checking effect on viral hoaxes: a model of misinformation spread in social networks. In: Proceedings of the 24th International Conference on World Wide Web, pp. 977–982 (2015)
33. Thomas, K., McCoy, D., Grier, C., Kolcz, A., Paxson, V.: Trafficking fraudulent accounts: the role of the underground market in Twitter spam and abuse. In: USENIX Security Symposium, pp. 195–210 (2013)
34. Trpevski, D., Tang, W.K., Kocarev, L.: Model for rumor spreading over networks. Phys. Rev. E **81**(5), 056102 (2010)
35. Vasileva, S., Atanasova, P., Màrquez, L., Barrón-Cedeño, A., Nakov, P.: It takes nine to smell a rat: neural multi-task learning for check-worthiness prediction. arXiv preprint arXiv:1908.07912 (2019)

36. Vo, N., Lee, K.: Where are the facts? Searching for fact-checked information to alleviate the spread of fake news. arXiv preprint arXiv:2010.03159 (2020)
37. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018)
38. Wu, Y., Zhan, P., Zhang, Y., Wang, L., Xu, Z.: Multimodal fusion with co-attention networks for fake news detection. In: Findings of the Association for Computational Linguistics, ACL-IJCNLP 2021, pp. 2560–2569 (2021)
39. Zaman, T., Fox, E.B., Bradlow, E.T.: A Bayesian approach for predicting the popularity of tweets. Ann. Appl. Stat. **8**(3), 1583–1611 (2014). https://doi.org/10.1214/14-AOAS741
40. Zhang, J., Zhang, R., Zhang, Y., Yan, G.: The rise of social botnets: attacks and countermeasures. IEEE Trans. Depend. Secure Comput. **15**(6), 1068–1082 (2016)
41. Zhang, Y., Yang, Q.: A survey on multi-task learning. IEEE Trans. Knowl. Data Eng. **34**(12), 5586–5609 (2021)
42. Zhao, L., Qiu, X., Wang, X., Wang, J.: Rumor spreading model considering forgetting and remembering mechanisms in inhomogeneous networks. Phys. A **392**(4), 987–994 (2013)
43. Zubiaga, A., Liakata, M., Procter, R.: Learning reporting dynamics during breaking news for rumour detection in social media. arXiv preprint arXiv:1610.07363 (2016)