Preserving the Integrity and Credibility of the Online Information Ecosystem

Matthew Sumpter and Giovanni Luca Ciampaglia*

University of South Florida

Abstract

The Internet seems awash with information that is either inaccurate or shared with malicious intents, or both. While this is not the first time that society has had to deal with this problem, certain features of our modern, fast-paced, data-driven information ecosystem seem to exacerbate it. Thus it is extremely important to equip journalists and fact-checkers, and of course the public at large, with tools to help them deal with the proliferation of false and misleading information and to promote the quality of information circulating online. In this paper we survey the state of the work in this area, focusing in particular on the challenges stemming from dealing with the peculiar nature of social media data, and discuss recent proposals to devise scalable and accurate signals of information quality.

1 Introduction

In recent years the Internet seems to have become the source and vehicle of many societal ills. The explosion of hoaxes, conspiracy theories, and state-sponsored disinformation, has given prominence and extreme urgency to the problem. But false and misleading information has existed also in the past. The style and themes of so-called "fake news"[1] websites, that came to prominence during the 2016 U.S. Presidential Elections, carry a striking resemblance to those used by "yellow journalism" outlets from the early 20$^{\text{th}}$ century [2]. Thus it is important to understand what elements of the problem are really novel and which are not.

There are a variety of classifications of information, including but not limited to: *rumor*, *gossip*, *propaganda*, *conspiracy theories*, *hoaxes*, and *satire*. The distinction between many of these terms relies on a matter of intention. For example, *satire* may be intentionally false for the purpose of amusement. However, a *rumor* can simply be a misinterpretation of the available facts with no intention to mislead. Allowing for the additional flexibility of rhetoric, such as sarcasm and persuasion, it becomes clear how difficult it can be to discern whether information is being presented with the intention to inform, persuade, or entertain.

From a quantitative point of view, social media are obviously different from the media of the past like the telegraph or the printed press: for one, they allow for a much faster and broader dissemination of information.

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

*Corresponding author. Email: glc3@usf.edu

[1]Fake news is defined as "impostor news", i.e. outlets featuring many of the trappings associated to the news publishing business without the necessary quality control mechanisms employed by professional publications [1].

But given the many striking similarities between then and now, one would be tempted to dismiss the whole phenomenon of fake news and online misinformation as the same of yellow journalism or gossip: a problem arising from fundamental human cognitive limits, and whose solution does not require any novel thinking or tools. In this paper, we argue that this is not the case. There are several novel factors at play at the nexus of the social, cognitive, and algorithmic levels, which require us to think in creative ways to address this fundamentally new societal problem. Many of these solutions could come from the field of data engineering, as we outline next.

2 The prevalence and sources of online misinformation

The introduction of the World Wide Web in the early 90s has dramatically lowered the barriers of entry to mass communication, allowing for anyone with an Internet connection to publish information affordably and quickly, regardless of its merit. Although this led to an enormous volume of information [3], it was still not easy to garner large audiences until the development of two key innovations: search engines [4], and social networking services [5], which have made it possible to retrieve information easily and have allowed for communities of homogeneous opinion to aggregate and share information.

A virtuous outcome of these innovations is the emergence of networks of like-minded individuals capable of self-organizing in novel and surprising ways [6], and to nurture strong ties that transcend the limits imposed by geographical distance or culture [7, 8], but one downside is the risk of polarization or cyberbalkanization [9–11]. Online communities are ripe for exploitation by bad actors seeking to manipulate large groups of people. Bolstered by content curation algorithms designed to increase engagement (filter bubbles [12]), these communities may fall prey to manipulation due to a mix of psychological and algorithmic biases that entrench accepted information while resisting opposing viewpoints (echo chambers [13]).

There are, unfortunately, numerous contemporary examples of widespread misinformation campaigns, including foreign influence in the 2016 election and the rise of conspiracy theories like 'Qanon', and the suspicion of widespread, yet unfounded, election fraud. However, gauging the actual prevalence of false and inaccurate information on social media is still the subject of much research [1, 14, 15], especially in the context of political communication [16]. Even though strong exposure to fake news is limited to the segment of most active news consumers, or 'superspreaders' [17], individual claims echoing the false or misleading content shared by these audiences can spread rapidly through social media [18, 19], amplified by bots [20] or other malicious actors [21], who often target elites, like celebrities, pundits, or politicians.

There are two inherent limitations to many of these studies. The first is that they typically rely on active engagement (e.g. likes and comments) as a proxy for exposure, but miss data about impressions, thus neglecting exposure due to "passive" forms of consumption such as scrolling and reading. The second stems from the impossibility of verifying each individual piece of content shared online. To get around this limitation, scientists have resorted to coarse-grained content annotation schemes to reduce sparsity and increase coverage. Source reliability ratings are a prime example of this approach [14, 17, 20, 22], as they typically identify sources by their web domain name. Lists of domain ratings compiled by expert journalists or fact-checkers have allowed to measure the prevalence of online misinformation, as well as the impact of social bots in spreading content from low-reliability sources. There are two advantages to this approach.

First, due to the highly skewed character of online popularity, one advantage of annotating sources at the web domain level is that it is possible to attain a high coverage by focusing only on the top most popular domains. Second, since web domains are the main form of identity on the web, source-level ratings are an effective way for social media platforms and search engines to implement content exclusion lists. Blocking or deprioritizing domains with low reliability introduces strong reputational costs for any agent who wishes to spread misinformation.

However source-level reliability ratings also come with limitations and challenges, which are especially relevant for data engineers. Maintaining such lists is typically time consuming, and authoring tools that facilitate such annotations are of great help. On top of that, source-level ratings may miss important heterogeneity *within* individual sources. This is especially true of larger outlets, that may employ different journalistic standards between the newsroom and their editorial desk. Finally, when lists are used to enforce moderation standards, they cause misinformation agents to update their domain name frequently in an attempt to circumvent detection, requiring more frequent updates on the part of the reliability annotators.

3 Verifying content and fact-checking claims

The prevalence of misinformation has led to the rise of the fact-checking industry. Fact-checking has been recognized as a solution to misinformation, as those who risk being fact-checked are less likely to share misinformation in the first place [23]. Unfortunately, claims online are generated and spread at a rate far too quickly for human journalists to keep up with. This is because fact-checking is a non-trivial task requiring the claim to be researched and then the fact-check written, published, and distributed. The window of time between initial claim and published fact-check can be significant (an estimate places it at in the 10–20h range [24]), allowing the claim to have reached too wide of an audience for the fact-check to be effective.

In addition to the time requirement of fact-checking, there is the confounding human component. Although fact-checkers are typically trained journalists, they are still susceptible to both mistakes and bias. These issues have revealed an opportunity for computer scientists to get involved. Computational approaches to fact-checking have the potential to address both the time-dependent and human-dependent issues of manual fact-checking.

ClaimBuster [25], developed in 2014, is the first end-to-end fact-checking system. Using a variety of machine learning and NLP techniques, media sources are monitored for statements that match an existing repository of professional fact-checks so that they may be served expressly to an audience. On unseen claims, it attempts to frame the claim as a question for the purpose of querying knowledge bases and question answering engines, such as Wolfram Alpha and Google search results.

Another approach engages with the content in a more structured way. It does so by means of knowledge representation techniques such as knowledge graphs [26]. A knowledge graph aims to store and provide knowledge of the real world, where nodes are entities and the edges are the relations that exist between these entities. Knowledge graphs have a number of advantages over relational and non-relational models for the massive, open-world domain that “real-world” knowledge implies. The basic unit of knowledge contained in a knowledge graph is typically the semantic triple, composed of two entities (eg. a subject and an object) and a predicate relation that exists between them. An example of a semantic triple is `<Tallahassee, capitalOf, Florida>`. Compiling many series of semantic triples results in a knowledge graph which can be traversed to gain insights regarding the semantic relationships between entities.

There exist several general-purpose knowledge graphs, including Yago [27], DBpedia [28], and Wikidata [29]. It is notable that, although these contain lots of overlapping and related information, they lack consolidation and techniques must be developed to map entities and relations between these various knowledge graphs. This is a similar task to that of ontology alignment [30].

Fact-checking has been performed using existing knowledge graphs by receiving a claim as a semantic triple and checking its validity based on the sets of existing triples in a knowledge base that connect the subject and object of the claim triple [31–34]. Although this approach has proven promising, it is limited by its input because reducing a complex claim into a semantic triple is a nontrivial task, akin to the NLP task of relation extraction [35]. Fact-checking a political claim requires identifying an ontology best suited to modeling this type of claims and then developing a tool that can reduce claims into triples using the selected ontology. To address this task, we have built a pipeline which focuses on extracting relations by modeling a network of claims as a graph network using sentence dependency trees [36]. By modeling input claims in a format more analogous to the verification method (knowledge graphs), we were able to successfully extract relations from real-world claims and use them as input into fact-checking algorithms.

4 Toward signals of quality

Fact-checking has become a critical component of the socio-technical infrastructure devoted to preserving the integrity of the online information ecosystems. Social media platforms, despite initial reluctance, have embraced this valuable resource. Facebook, for example, partners with accredited fact-checking organization to review potentially misleading content in circulation on its platform. If a piece of content is flagged by a fact-checker as false, Facebook automatically reduces its visibility across the platform, reducing the chances that users may be exposed to it [37].

Of course, given the scale of social media, fact-checking every piece of content in circulation on it is likely an unfeasible task. Therefore, social media platforms are seeking to identify signals of credibility, or news quality, that could help them promote trustworthy and reliable information. These signals should rely on information about the content that is readily accessible to the platform or at least easy to estimate, without having to inspect the content itself or having to bring in a human to review it.

There is a vast literature devoted to identifying the reputation of content and actors on the Web [38–44], but many of these approaches are either hard to scale or make restrictive assumptions about the type of metadata available. These assumptions often reflect the specific Web platforms for which they were originally developed, for example the requirement that content is organized following wiki principles and that there is a full history of all user actions available [38]. More recently work in the context of political news consumption has proposed to use crowdsourcing to evaluate the reliability of news sources [45]. Finally, other work proposes instead to promote content that produces engagement within a politically diverse audience [46].

5 Discussion and future challenges

When it comes to the modern information ecosystem, misinformation and disinformation (in all its related forms) poses strong threats to the integrity of public discourse. Here we have outlined two areas where data engineering could help conduct the fight against fake news. The first is about measuring the actual prevalence of misinformation, the amount of exposure it gets, and the main actors responsible for producing and disseminating it. Source-level reliability ratings are the current standard used in much of the literature and form the backbone of several content moderation schemes used by real platforms [47, 48]. Lists curated by fact-checkers and other third-party organizations are thus a valuable tool, but present several challenges and limitations. First of all, the content of these lists is often static while the Web and social media are dynamic environments. How to update these lists? One possibility could be to look into the tail of the popularity distribution for potential candidates of up-and-coming misinformation producers. Another approach could be to focus on users that are routinely exposed to known low-credibility sources (or that share content from them) and identify what other sources they are routinely exposed to (co-exposure) or they share content from (co-sharing). This could reveal novel unrated sources that could be passed to expert journalist for rating and further evaluation (see Fig. 1). More broadly, combining data about co-exposure (or co-sharing) of multiple users it could be possible to define networks of low-credibility sources with overlapping audiences, potentially revealing the broader ecosystem of online misinformation. Several interesting questions arise: can we look into niche communities to see what content they are sharing? Can we identify collusion rings of sources pushing similar content in an orchestrated fashion? To support this endeavor, there is an urgent need for novel technical standards and methods. Novel Web standards could help provide more meaningful annotation for Internet sources, while novel methods to link different domains operated by the same organizations could mitigate the aforementioned problems with churn and active avoidance by malicious actors.

The second major area of activity is the quest to automate fact-checking or at least help support human fact-checkers in their task of verifying claims. Considering the unstructured nature of online discourse, there is in this case too a need for novel standards of annotation of online content. The development of a standardized

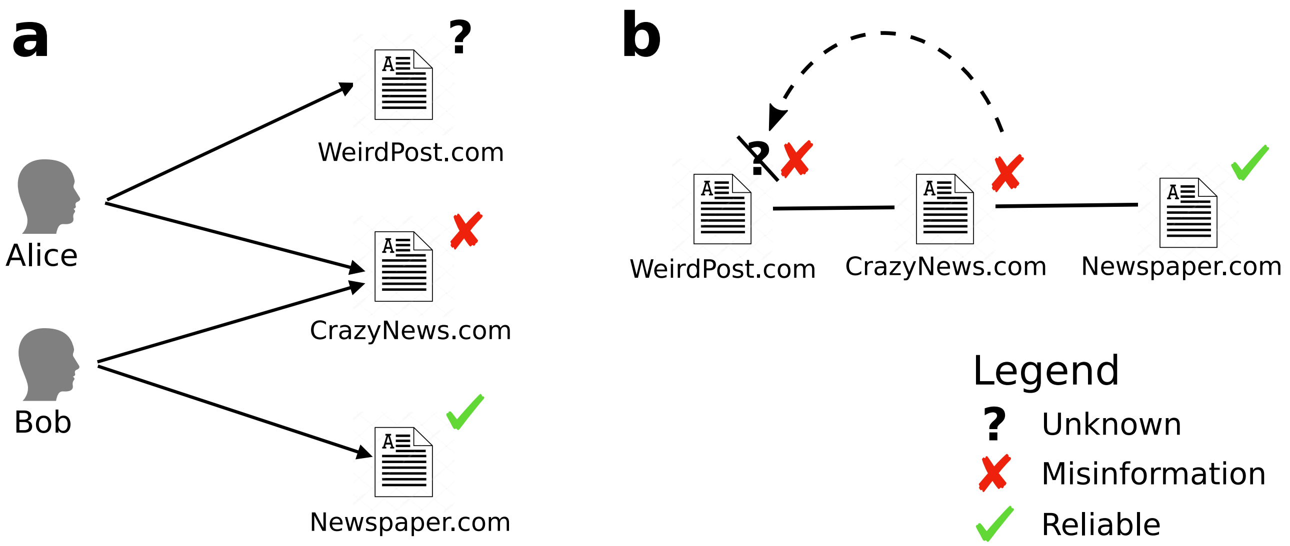

Figure 1: **Example of a co-exposure network**. **(a)** Alice and Bob follow three sources: a known misinformation source (CrazyNews.com), a known reliable source (Newspaper.com) and a source for which no rating is available (WeirdPost.com). **(b)** The co-exposure network is a graph whose nodes are sources and there is an edge (solid lines) between two sources if they share some users (i.e., they *co-expose*). Since WeirdPost.com as a co-exposure with CrazyNews.com due to Alice, it is possible to propagate the rating of 'misinformation source' to its neighbor with unknown rating (dashed arrow). Note that the propagation does not apply to neighbors with a known rating: Newspaper.com is already a known reliable source so its rating does not change.

schema and ontology for content annotation would provide a means by which multimedia content on the web could begin to properly be aggregated and cross-referenced in a dependable way. Journalists and publishers could be provided education and tools on how to annotate their own work such that computer scientists would not need to retroactively mine and process media on the web.

A good example is the ClaimReview schema [49], a content annotation format developed by researchers at Duke University. ClaimReview provides an important first step in the direction of content annotation as it relates to claims in news media. This schema allows professional fact checkers to annotate their fact-checks with distinct properties, such as the claim reviewed and the rating decision. This allows claims to be aggregated and queried by search engines. There has been some preliminary research regarding necessary schemas for annotating news media. Arslan et al. [50] have identified a set of 20 semantic frames (11 novel frames and 9 exisiting Berkeley FrameNet [51] frames) that can be used to model factual claims. Ciampaglia and Licato [52] identified the need for argumentation schemes to capture rhetorical methods present among claims found in news media.

References

[1] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, *et al.*, "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.

[2] S. McQueen, "From yellow journalism to tabloids and clickbait: The origins of fake news in the united states," in *Information Literacy and Libraries in the Age of Fake News* (D. Agosto, ed.), pp. 12–35, ABC-CLIO, 2018.

[3] M. Hilbert and P. Lopez, “The world's technological capacity to store, communicate, and compute information,” *Science*, vol. 332, pp. 60–65, Feb. 2011.

[4] L. Page, S. Brin, R. Motwani, and T. Winograd, “The pagerank citation ranking: Bringing order to the web.,” Technical Report 1999-66, Stanford InfoLab, Nov. 1999. Previous number = SIDL-WP-1999-0120.

[5] H. Kwak, C. Lee, H. Park, and S. Moon, “What is twitter, a social network or a news media?,” in *Proceedings of the 19th international conference on World wide web - WWW '10*, ACM Press, 2010.

[6] Z. Tufekci, *Twitter and tear gas : the power and fragility of networked protest*. New Haven: Yale University Press, 2017.

[7] D. Mocanu, A. Baronchelli, N. Perra, B. Gonçalves, Q. Zhang, and A. Vespignani, “The twitter of babel: Mapping world languages through microblogging platforms,” *PLoS ONE*, vol. 8, p. e61981, Apr. 2013.

[8] P. S. Park, J. E. Blumenstock, and M. W. Macy, “The strength of long-range ties in population-scale social networks,” *Science*, vol. 362, pp. 1410–1413, Dec. 2018.

[9] J. E. Katz, “Struggle in cyberspace: Fact and fiction on the world wide web,” *The Annals of the American Academy of Political and Social Science*, vol. 560, pp. 194–199, Nov. 1998.

[10] M. Van Alstyne and E. Brynjolfsson, “Global village or cyber-balkans? modeling and measuring the integration of electronic communities,” *Management Science*, vol. 51, no. 6, pp. 851–868, 2005.

[11] M. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, F. Menczer, and A. Flammini, “Political polarization on Twitter,” in *International AAAI Conference on Web and Social Media*, ICWSM ’11, (Palo Alto, CA, USA), pp. 89–96, AAAI, 2011.

[12] E. Pariser, *The filter bubble : how the new personalized web is changing what we read and how we think*. New York: Penguin Press, 2011.

[13] C. R. Sunstein, *#Republic*. Princeton University Press, Apr. 2018.

[14] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” *Journal of economic perspectives*, vol. 31, no. 2, pp. 211–36, 2017.

[15] A. Bovet and H. A. Makse, “Influence of fake news in twitter during the 2016 us presidential election,” *Nature Communications*, vol. 10, p. 7, Jan. 2019.

[16] A. M. Guess, B. Nyhan, and J. Reifler, “Exposure to untrustworthy websites in the 2016 us election,” *Nature human behaviour*, vol. 4, no. 5, pp. 472–480, 2020.

[17] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, “Fake news on twitter during the 2016 us presidential election,” *Science*, vol. 363, no. 6425, pp. 374–378, 2019.

[18] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[19] Zhao, Zilong, Zhao, Jichang, Sano, Yukie, Levy, Orr, Takayasu, Hideki, Takayasu, Misako, Li, Daqing, Wu, Junjie, and Havlin, Shlomo, “Fake news propagates differently from real news even at early stages of spreading,” *EPJ Data Sci.*, vol. 9, no. 1, p. 7, 2020.

[20] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, “The spread of low-credibility content by social bots,” *Nature communications*, vol. 9, no. 1, pp. 1–9, 2018.

[21] J. Weedon, W. Nuland, and A. Stamos, "Information operations and facebook," tech. rep., Facebook, Inc., 2017.

[22] M. Zimdars, *Fake News: Understanding Media and Misinformation in the Digital Age*, ch. Viral "Fake News" Lists and the Limitations of Labeling and Fact-Checking, pp. 361–372. MIT Press Ltd, Feb. 2020.

[23] B. Nyhan and J. Reifler, "The effect of fact-checking on elites: A field experiment on u.s. state legislators," *American Journal of Political Science*, vol. 59, no. 3, pp. 628–640, 2015.

[24] C. Shao, G. L. Ciampaglia, A. Flammini, and F. Menczer, "Hoaxy: A platform for tracking online misinformation," in *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW '16 Companion, (Republic and Canton of Geneva, Switzerland), pp. 745–750, International World Wide Web Conferences Steering Committee, 2016.

[25] N. Hassan, G. Zhang, F. Arslan, J. Caraballo, D. Jimenez, S. Gawsane, S. Hasan, M. Joseph, A. Kulkarni, A. K. Nayak, V. Sable, C. Li, and M. Tremayne, "Claim buster: The first ever end-to-end fact checking system," *Proceedings of the VLDB Endowment*, vol. 10, no. 12, pp. 1945–1948, 2017.

[26] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, J. E. L. Gayo, S. Kirrane, S. Neumaier, A. Polleres, R. Navigli, A.-C. N. Ngomo, S. M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, and A. Zimmermann, "Knowledge graphs," tech. rep., CoRR, Mar. 2020.

[27] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: A core of semantic knowledge," in *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, (New York, NY, USA), pp. 697–706, Association for Computing Machinery, 2007.

[28] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "Dbpedia: A nucleus for a web of open data," in *The Semantic Web* (K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, and P. Cudré-Mauroux, eds.), (Berlin, Heidelberg), pp. 722–735, Springer Berlin Heidelberg, 2007.

[29] D. Vrandečić, "Wikidata: A new platform for collaborative data collection," in *Proceedings of the 21st International Conference on World Wide Web*, WWW '12 Companion, (New York, NY, USA), pp. 1063–1064, Association for Computing Machinery, 2012.

[30] F. M. Suchanek, S. Abiteboul, and P. Senellart, "PARIS," *Proceedings of the VLDB Endowment*, vol. 5, pp. 157–168, nov 2011.

[31] N. Lao and W. W. Cohen, "Relational retrieval using a combination of path-constrained random walks," *Machine Learning*, vol. 81, no. 1, pp. 53–67, 2010.

[32] G. L. Ciampaglia, P. Shiralkar, L. M. Rocha, J. Bollen, F. Menczer, and A. Flammini, "Computational fact checking from knowledge networks," *PLOS ONE*, vol. 10, pp. 1–13, June 2015.

[33] P. Shiralkar, A. Flammini, F. Menczer, and G. L. Ciampaglia, "Finding streams in knowledge graphs to support fact checking," in *2017 IEEE International Conference on Data Mining (ICDM)*, (Piscataway, NJ), pp. 859–864, IEEE, Nov. 2017. Extended Version.

[34] B. Shi and T. Weninger, "Discriminative predicate path mining for fact checking in knowledge graphs," *Knowledge-Based Systems*, vol. 104, pp. 123–133, July 2016.

[35] A. Smirnova and P. Cudré-Mauroux, "Relation extraction using distant supervision," *ACM Computing Surveys*, vol. 51, pp. 1–35, jan 2019.

[36] M. Sumpter and G. L. Ciampaglia, "REMOD: Relation extraction for modeling online discourse," in *Proceedings of the 1st International Workshop on Knowledge Graphs for Online Discourse Analysis*, 2021.

[37] Full Fact, "Report on the Facebook Third-Party Fact-Checking programme," tech. rep., Full Fact, 2020.

[38] B. T. Adler and L. de Alfaro, "A content-driven reputation system for the Wikipedia," in *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, (New York, NY, USA), pp. 261–270, ACM, 2007.

[39] J.-H. Cho, K. Chan, and S. Adali, "A survey on trust modeling," *ACM Comput. Surv.*, vol. 48, pp. 28:1–28:40, Oct. 2015.

[40] J. A. Golbeck, *Computing and applying trust in web-based social networks*. PhD thesis, University of Maryland at College Park, 2005.

[41] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier, *TweetCred: Real-Time Credibility Assessment of Content on Twitter*, pp. 228–243. Cham: Springer International Publishing, 2014.

[42] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, (Copenhagen, Denmark), pp. 2931–2937, Association for Computational Linguistics, Sept. 2017.

[43] S. Jiang and C. Wilson, "Linguistic signals under misinformation and fact-checking: Evidence from user comments on social media," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, Nov. 2018.

[44] S. Jiang, S. Baumgartner, A. Ittycheriah, and C. Yu, *Factoring Fact-Checks: Structured Information Extraction from Fact-Checking Articles*, pp. 1592–1603. New York, NY, USA: Association for Computing Machinery, 2020.

[45] G. Pennycook and D. G. Rand, "Fighting misinformation on social media using crowdsourced judgments of news source quality," *Proceedings of the National Academy of Sciences*, vol. 116, no. 7, pp. 2521–2526, 2019.

[46] S. Bhadani, S. Yamaya, A. Flammini, F. Menczer, G. L. Ciampaglia, and B. Nyhan, "Political audience diversity and news reliability in algorithmic ranking," tech. rep., CoRR, July 2020.

[47] G. Morrow, B. Swire-Thompson, J. Polny, M. Kopec, and J. Wihbey, "The emerging science of content labeling: Contextualizing social media content moderation," *SSRN Electronic Journal*, 2020.

[48] J. Nassetta and K. Gross, "State media warning labels can counteract the effects of foreign disinformation," *Harvard Kennedy School Misinformation Review*, oct 2020.

[49] Schema.org, "ClaimReview." online at `https://schema.org/ClaimReview`, 2020.

[50] F. Arslan, J. Caraballo, D. Jimenez, and C. Li, "Modeling factual claims with semantic frames," in *Proceedings of the 12th Language Resources and Evaluation Conference*, (Marseille, France), pp. 2511–2520, European Language Resources Association, May 2020.

[51] C. F. Baker, C. J. Fillmore, and J. B. Lowe, "The berkeley framenet project," in *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pp. 86–90, 1998.

[52] G. L. Ciampaglia and J. Licato, "Fact-checking, false narratives, and argumentation schemes," in *Proceedings of the 1st International Workshop on Knowledge Graphs for Online Discourse Analysis*, 2021.