

# Fighting fake news: a role for computational social science in the fight against digital misinformation

Giovanni Luca Ciampaglia<sup>1</sup> 

Received: 5 November 2017 / Accepted: 9 November 2017 / Published online: 29 November 2017  
© Springer Nature Singapore Pte Ltd. 2017

**Abstract** The massive, uncontrolled, and oftentimes systematic spread of inaccurate and misleading information on the Web and social media poses a major risk to society. Digital misinformation thrives on an assortment of cognitive, social, and algorithmic biases and current countermeasures based on journalistic corrections do not seem to scale up. By their very nature, computational social scientists could play a twofold role in the fight against fake news: first, they could elucidate the fundamental mechanisms that make us vulnerable to misinformation online and second, they could devise effective strategies to counteract misinformation.

**Keywords** Digital misinformation · Algorithmic bias · Fact checking

## Introduction

Information is produced and consumed according to novel paradigms on the Web and social media. The ‘laws’ of these modern marketplaces of ideas are starting to emerge [1, 2], thanks in part to the intuition that the data we leave behind as we use technology platforms reveal much about ourselves and our true behavior [3]. The transformative nature of this revolution offers the promise of a better understanding of both individual human behavior and collective social phenomena, but poses also fundamental challenges to society at large. One such risk is the massive,

---

G.L.C. is supported by the Indiana University Network Science Institute.

✉ Giovanni Luca Ciampaglia  
gciampag@iu.edu

<sup>1</sup> Indiana University Network Science Institute, 47408 Bloomington, IN, USA

uncontrolled, and oftentimes systematic spread of inaccurate and misleading information.

Misinformation spreads on social media under many guises. There are rumors, hoaxes, conspiracy theories, propaganda, and, of course, fake news. Regardless of the form, the repercussions of inaccurate or misleading information are stark and worrisome. Confidence in the safety of childhood vaccines has dropped significantly in recent years [4], fueled by misinformation on the topic, resulting in outbreaks of measles [5]. When it comes to the media and information landscape, things are not better. According to recent surveys, even though an increasing majority of American adults (67% in 2017, up from 62% in 2016) gets their news from social media platforms on a regular basis, the majority (63% of respondents) do not trust the news coming from social media. Even more disturbing, 64% of Americans say that fake news have left them with a great deal of confusion about current events, and 23% also admit to passing on fake news stories to their social media contacts, either intentionally or unintentionally [6–8].

How to fight fake news? In this essay, I will sketch a few areas in which computational social scientists could play a role in the struggle against digital misinformation.

## Social media balkanization?

When processing information humans are subject to an assortment of socio-cognitive biases, including homophily [9], confirmation bias [10], conformity bias [11], the misinformation effect [12], and motivated reasoning [13]. These biases are at work in any situations, i.e., not just online, but we are finally coming to the realization that the current structure of the Web and social media plays upon and reinforces them. At the beginning of the Web, some scholars had hypothesized that the Internet could in principle ‘balkanize’ into communities of like-minded people espousing completely different sets of facts [14–16]. Over the years, various web technologies have been tested for signs of cyber-balkanization [17] or bias, for example, the early Google search engines [18] or recommender systems [19].

Could social media provide fertile ground for cyber-balkanization? With the rise of big data and personalization technologies, one concern is that algorithmic filters may relegate us into information ‘bubbles’ tailored to our own preferences [20]. Filtering and ranking algorithms like the Facebook NewsFeed do indeed prioritize content based on engagement and popularity signals, and even though popularity in some cases may help quality contents bubble up [21], it may not be enough for our limited attention, which may explain the virality of low-quality content [22, 23], and why group conversations often turn into cacophony [24].

Combining together the effect of algorithmic bias with the homophilistic structure of social networks, one concern is that this may lead to an ‘echo chamber’ effect, in which existing beliefs are reinforced, thus lowering barriers to the manipulation of misinformation. There have been some attempts at measuring the degree of selective exposure of social media [25, 26], but as platforms continuously evolve, tune their algorithms, and grow in scope the question is far from settled.

It worth noting here that these questions have not been explored only from an empirical point of view. There is in fact a very rich tradition of computational work that seeks to explain the evolution of distinct cultural groups using computer simulation, dating back to the pioneering work of Robert Axelrod [27–30]. Agent-based models can help us test the plausibility of competing hypotheses in a generative framework, but they often lack empirical validation [31]. The computational social science community could contribute in a unique fashion to the debate on the cyber-balkanization of social media by bridging the divide between social simulation models and empirical observation [32].

## A call to action for computational social scientists

What to do? Professional journalism, guided by ethical principles of trustworthiness and integrity, has been for decades the answer to episodes of rampant misinformation, like yellow journalism. At the moment, the fourth estate does not seem to be effective enough in the fight against digital misinformation. For example, analysis of Twitter conversations has shown that fact-checking lags behind misinformation both in terms of overall reach and of sheer response time—there is a delay of approximately 13 h between the consumption of fake news stories and that of its verifications [33]. While this span may seem relatively short, social media have been shown to spread content orders of magnitude faster [34].

Perhaps, even more troubling, doubts have been cast on the very effectiveness of factual corrections. For example, it has been observed that corrections may sometimes backfire, resulting in even more entrenched support of factually inaccurate claims [35]. More recent evidence seems to paint a more nuanced, and optimistic picture about fact-checking [36–39], showing that the backfire effect is not as common as initially alleged. Still, to this date, the debate on how to improve the efficacy of fact-checking is far from resolved [40, 41].

The debate on the efficacy of fact-checking is important also because there is a growing interest in automating the various activities that revolve around fact-checking. These include: news gathering, verification, and delivery of corrections [42–46]. These activities are already capitalizing on the growing number of tools, data sets, and platforms contributed by computer scientists to detect, define, model, and counteract the spread of misinformation [47–56]. Without a clear understanding of what are the most effective countermeasures, and of who is best equipped to deliver them, these tools may never be brought to complete fruition.

Computational social scientists could play a pivotal role in the collective effort to produce and disseminate accurate information at scale. A challenge for this type of research is how to test the efficacy of different types of corrections on large samples and possibly in the naturalistic settings of social media. Social bots, which so far have been responsible for spreading large amounts of misinformation [57–60], could be employed to this end, since initial evidence shows that they can be employed to spread positive messages [61].

## Conclusions

The fight against fake news has just started, and many answers are still open. Are there any successful examples that we should emulate? After all, accurate information is produced and disseminate on the Internet every day. Wikipedia comes to mind here, as perhaps one of the most successful communities of knowledge production. To be sure, Wikipedia is not immune to vandalism and inaccurate information, but the cases of intentional disinformation that have survived for long are surprisingly few [62]. Wikipedia is often credited for being the go-to resource of accurate knowledge for millions of people worldwide [63]. It is often said that Wikipedia only works in practice; in theory, it should never work. Computational social scientists have elucidated many of the reasons why the crowdsourcing model Wikipedia works [64], and thus could effectively contribute to a better understanding of the problem of digital misinformation.

**Acknowledgements** The author would like to thank Filippo Menczer and Alessandro Flammini for insightful conversations.

## References

1. Huberman, B. A., Pirolli, P. L. T., Pitkow, J. E., & Lukose, R. (1998). Strong regularities in world wide web surfing. *Science*, 280(5360), 95–97. <https://doi.org/10.1126/science.280.5360.95>.
2. Ciampaglia, G. L., Flammini, A., & Menczer, F. (2015). The production of information in the attention economy. *Scientific Reports*, 5, 9452. <https://doi.org/10.1038/srep09452>.
3. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., et al. (2009). Computational social science. *Science*, 323(5915), 721–723. <https://doi.org/10.1126/science.1167742>.
4. Wooley, M. (2015). *Childhood vaccines*. Presentation at the workshop on Trust and Confidence at the Intersections of the Life Sciences and Society, Washington D.C. [http://nas-sites.org/publicinterfaces/files/2015/05/Wooley\\_PILS\\_VaccineSlides-3.pdf](http://nas-sites.org/publicinterfaces/files/2015/05/Wooley_PILS_VaccineSlides-3.pdf).
5. Hotez, P. J. (2016). Texas and its measles epidemics. *PLoS Medicine*, 13(10), 1–5. <https://doi.org/10.1371/journal.pmed.1002153>.
6. Barthel, M., Mitchell, A., & Holcomb, J. (2016). *Many Americans believe fake news is sowing confusion*. Online, Pew Research. <http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/>.
7. Barthel, M., & Mitchell, A. (2017). *Americans' attitudes about the news media deeply divided along partisan lines*. Online, Pew Research. <http://www.journalism.org/2017/05/10/americans-attitudes-about-the-news-media-deeply-divided-along-partisan-lines/>.
8. Gottfried, J., & Shearer, E. (2017). *News use across social media platforms 2017*. Online, Pew Research. <http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/>.
9. McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1), 415–444. <https://doi.org/10.1146/annurev.soc.27.1.415>.
10. Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>.
11. Asch, S. E. (1961). Effects of group pressure upon the modification and distortion of judgements. In M. Henle (Ed.) *Documents of Gestalt psychology* (pp. 222–236). Oakland: University of California Press.
12. Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12(4), 361–366. <https://doi.org/10.1101/lm.94705>.
13. Kahan, D. M. (2013). A risky science communication environment for vaccines. *Science*, 342(6154), 53–54. <https://doi.org/10.1126/science.1245724>.

14. Van Alstyne, M., & Brynjolfsson, E. (1996). Could the internet balkanize science? *Science*, 274(5292), 1479–1480. <https://doi.org/10.1126/science.274.5292.1479>.
15. Katz, J. E. (1998). Fact and fiction on the world wide web. *The Annals of the American Academy of Political and Social Science*, 560(1), 194–199. <https://doi.org/10.1177/0002716298560001015>.
16. Van Alstyne, M., & Brynjolfsson, E. (2005). Global village or cyber-balkans? modeling and measuring the integration of electronic communities. *Management Science*, 51(6), 851–868. <https://doi.org/10.1287/mnsc.1050.0363>.
17. Kobayashi, T., & Ikeda, K. (2009). Selective exposure in political web browsing. *Information, Communication & Society*, 12(6), 929–953. <https://doi.org/10.1080/13691180802158490>.
18. Fortunato, S., Flammini, A., Menczer, F., & Vespignani, A. (2006). Topical interests and the mitigation of search engine bias. *Proceedings of the National Academy of Sciences*, 103(34), 12684–12689. <https://doi.org/10.1073/pnas.0605525103>.
19. Hosanagar, K., Fleder, D., Lee, D., & Buja, A. (2014). Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation. *Management Science*, 60(4), 805–823. <https://doi.org/10.1287/mnsc.2013.1808>.
20. Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. UK: Penguin.
21. Nematzadeh, A., Ciampaglia, G. L., Menczer, F., & Flammini, A. (2017). How algorithmic popularity bias hinders or promotes quality. *CoRR*. [arXiv:1707.00574](https://arxiv.org/abs/1707.00574)
22. Weng, L., Flammini, A., Vespignani, A., & Menczer, F. (2012). Competition among memes in a world with limited attention. *Scientific Reports*, 2, 335. <https://doi.org/10.1038/srep00335>.
23. Qiu, X., Oliveira, D. F. M., Sahami Shirazi, A., Flammini, A., & Menczer, F. (2017). Limited individual attention and online virality of low-quality information. *Nature Human Behavior*, 1, 0132. <https://doi.org/10.1038/s41562-017-0132>.
24. Nematzadeh, A., Ciampaglia, G. L., Ahn, Y. Y., & Flammini, A. (2016). Information overload in group communication: From conversation to cacophony in the twitch chat. *CoRR*. [arXiv:1610.06497](https://arxiv.org/abs/1610.06497)
25. Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4), 1799–1839. <https://doi.org/10.1093/qje/qjr044>.
26. Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239), 1130–1132. <https://doi.org/10.1126/science.aaa1160>.
27. Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, 41(2), 203–226. <https://doi.org/10.1177/0022002797041002001>.
28. Centola, D., González-Avella, J. C., Eguíluz, V. M., & San Miguel, M. (2007). Homophily, cultural drift, and the co-evolution of cultural groups. *Journal of Conflict Resolution*, 51(6), 905–929. <https://doi.org/10.1177/0022002707307632>.
29. Iñiguez, G., Kertész, J., Kaski, K. K., & Barrio, R. A. (2009). Opinion and community formation in coevolving networks. *Physical Review E*, 80, 066119. <https://doi.org/10.1103/PhysRevE.80.066119>.
30. Flache, A., & Macy, M. W. (2011). Local convergence and global diversity: From interpersonal to social influence. *Journal of Conflict Resolution*, 55(6), 970–995. <https://doi.org/10.1177/0022002711414371>.
31. Ciampaglia, G. L. (2013). A framework for the calibration of social simulation models. *Advances in Complex Systems*, 16, 1350030. <https://doi.org/10.1142/S0219525913500306>.
32. Ciampaglia, G. L., Ferrara, E., & Flammini, A. (2014). Collective behaviors and networks. *EPJ Data Science*, 3(1), 37. <https://doi.org/10.1140/epjds/s13688-014-0037-6>.
33. Shao, C., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2016). Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th International Conference Companion on World Wide Web, WWW '16 Companion* (pp. 745–750). Republic and Canton of Geneva: International World Wide Web Conferences Steering Committee. <https://doi.org/10.1145/2872518.2890098>
34. Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10* (pp. 851–860). New York: ACM. <https://doi.org/10.1145/1772690.1772777>
35. Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330. <https://doi.org/10.1007/s11109-010-9112-2>.
36. Bode, L., & Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication*, 65(4), 619–638. <https://doi.org/10.1111/jcom.12166>.
37. Wood, T., & Porter, E. (2016). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *SSRN*. <https://ssrn.com/abstract=2819073>. Accessed 26 Nov 2017.

38. Nyhan, B., Porter, E., Reifler, J., & Wood, T. (2017). Taking corrections literally but not seriously? the effects of information on factual beliefs and candidate favorability. *SSRN*. <https://ssrn.com/abstract=2995128>. Accessed 26 Nov 2017.
39. Vraga, E. K., & Bode, L. (2017). I do not believe you: how providing a source corrects health misperceptions across social media platforms. *Information, Communication & Society*, 1–17. <https://doi.org/10.1080/1369118X.2017.1313883>.
40. Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131.
41. Ecker, U. K., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, 6(2), 185–192. <https://doi.org/10.1016/j.jarmac.2017.01.014>.
42. Wu, Y., Agarwal, P. K., Li, C., Yang, J., & Yu, C. (2014). Toward computational fact-checking. *Proceedings of the VLDB Endowment*, 7(7), 589–600. <https://doi.org/10.14778/2732286.2732295>.
43. Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F., & Flammini, A. (2015). Computational fact checking from knowledge networks. *PLoS One*, 10(6), 1–13. <https://doi.org/10.1371/journal.pone.0128193>.
44. Shiralkar, P., Flammini, A., Menczer, F., & Ciampaglia, G. L. (2017). Finding streams in knowledge graphs to support fact checking. In *Proceedings of the 2017 IEEE 17th International Conference on Data Mining, Extended Version*. Piscataway, NJ: IEEE.
45. Wu, Y., Agarwal, P. K., Li, C., Yang, J., & Yu, C. (2017). Computational fact checking through query perturbations. *ACM Transactions on Database Systems*, 42(1), 4. <https://doi.org/10.1145/2996453>.
46. Hassan, N., Arslan, F., Li, C., & Tremayne, M. (2017). Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17* (pp. 1803–1812). New York, NY: ACM. <https://doi.org/10.1145/3097983.3098131>
47. Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2011). Truthy: Mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th International Conference Companion on World Wide Web, WWW '11* (pp. 249–252). New York, NY: ACM. <https://doi.org/10.1145/1963192.1963301>
48. Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Flammini, A., & Menczer, F. (2011). Detecting and tracking political abuse in social media. In *Proc. International AAAI Conference on Web and Social Media* (pp. 297–304). Palo Alto, CA: AAAI. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2850>.
49. Liu, X., Nourbakhsh, A., Li, Q., Fang, R., & Shah, S. (2015). Real-time rumor debunking on twitter. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15* (pp. 1867–1870). New York, NY: ACM. <https://doi.org/10.1145/2806416.2806651>.
50. Metaxas, P. T., Finn, S., & Mustafaraj, E. (2015). Using twittertrails.com to investigate rumor propagation. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing, CSCW'15 Companion* (pp. 69–72). New York, NY: ACM. <https://doi.org/10.1145/2685553.2702691>
51. Mitra, T., & Gilbert, E. (2015). Credbank: A large-scale social media corpus with associated credibility annotations. In *Proc. International AAAI Conference on Web and Social Media* (pp. 258–267). Palo Alto, CA: AAAI. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10582>.
52. Zubiaga, A., Liakata, M., Procter, R., Bontcheva, K., & Tolmie, P. (2015). Crowdsourcing the annotation of rumourous conversations in social media. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion* (pp. 347–353). New York, NY: ACM. <https://doi.org/10.1145/2740908.2743052>.
53. Davis, C. A., Ciampaglia, G. L., Aiello, L. M., Chung, K., Conover, M. D., Ferrara, E., et al. (2016). Osome: The iuni observatory on social media. *PeerJ Computer Science*, 2, e87. <https://doi.org/10.7717/peerj-cs.87>.
54. Declerck, T., Osenova, P., Georgiev, G., & Lendvai, P. (2016). Ontological modelling of rumors. In D. TrandabăŢ, D. Gifu (Eds.) *Linguistic Linked Open Data: 12th EUROLAN 2015 Summer School and RUMOUR 2015 Workshop, Sibiu, Romania, July 13–25, 2015, Revised Selected Papers* (pp. 3–17). Berlin: Springer International Publishing. [https://doi.org/10.1007/978-3-319-32942-0\\_1](https://doi.org/10.1007/978-3-319-32942-0_1).

55. Sampson, J., Morstatter, F., Wu, L., & Liu, H. (2016). Leveraging the implicit structure within social media for emergent rumor detection. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM '16* (pp. 2377–2382). New York, NY: ACM. <https://doi.org/10.1145/2983323.2983697>.
56. Wu, L., Morstatter, F., Hu, X., & Liu, H. (2016). Mining misinformation in social media. In M. T. Thai, W. Wu, H. Xiong (Eds.) *Big Data in Complex and Social Networks, Business & Economics* (pp. 125–152). Boca Raton, FL: CRC Press.
57. Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 U.S. presidential election online discussion. *First Monday*, 21(11). <https://doi.org/10.5210/fm.v21i11.7090>. <http://firstmonday.org/ojs/index.php/fm/article/view/7090>.
58. Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104. <https://doi.org/10.1145/2818717>.
59. Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday*, 22(8). <https://doi.org/10.5210/fm.v22i8.8005>. <http://firstmonday.org/ojs/index.php/fm/article/view/8005>.
60. Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. In *Proc. International AAAI Conference on Web and Social Media* (pp. 280–289). Palo Alto, CA: AAAI. <https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587>
61. Mønsted, B., Sapieyski, P., Ferrara, E., & Lehmann, S. (2017). Evidence of complex contagion of information in social media: An experiment using twitter bots. *PLoS One*, 12(9), 1. <https://doi.org/10.1371/journal.pone.0184148>.
62. Kumar, S., West, R., & Leskovec, J. (2016). Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In *Proceedings of the 25th International Conference on World Wide Web, WWW '16* (pp. 591–602). Republic and Canton of Geneva: International World Wide Web Conferences Steering Committee. <https://doi.org/10.1145/2872427.2883085>.
63. Singer, P., Lemmerich, F., West, R., Zia, L., Wulczyn, E., Strohmaier, M., & Leskovec, J. (2017). Why we read wikipedia. In *Proceedings of the 26th International Conference on World Wide Web, WWW '17* (pp. 1591–1600). Republic and Canton of Geneva: International World Wide Web Conferences Steering Committee. <https://doi.org/10.1145/3038912.3052716>.
64. Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F. Å., & Lanamäki, A. (2015). The sum of all human knowledge: A systematic review of scholarly research on the content of wikipedia. *Journal of the Association for Information Science and Technology*, 66(2), 219–245. <https://doi.org/10.1002/asi.23172>.