Monitoring Health Misinformation: An Early Detection Methodology for Newsrooms

Jenna Sherman
Harvard School of Public Health
Boston, Massachusetts
jsherman@cyber.harvard.edu

Nat Gyenes
Meedan
San Francisco, California
nat@meedan.com

Scott A. Hale
Meedan and Oxford Internet Institute
San Francisco, California
scott@meedan.com

CCS CONCEPTS

• **Information systems → Social networking sites**; *Similarity measures*; • **Human-centered computing** → *Social networking sites*; • **Computing methodologies** → *Natural language processing*.

KEYWORDS

health, misinformation, word embeddings

ACM Reference Format:
Jenna Sherman, Nat Gyenes, and Scott A. Hale. 2019. Monitoring Health Misinformation: An Early Detection Methodology for Newsrooms. In *Computation + Journalism 2020, March 20–21, 2020, Boston, MA*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/xxxxxx.xxxxxx

EXTENDED ABSTRACT

How humans use words reveals a lot about their biases and implicit associations. Modern national language processing techniques such as neural word embeddings for the representation of words capture the implicit meaning of words based on how words co-occur together in large volumes of text. Understanding how word associations can serve as proxies for health misinformation, and the ways in which word associations can be used by newsrooms in covering crisis and prolonged epidemics, is crucial given that online health misinformation is becoming increasingly ubiquitous, in turn yielding adverse health impacts globally. The World Health Organization lists "vaccine hesitancy" as one of the top ten threats to global health in 2019 [8], measles rates in the U.S. are the highest since 1992—after declared eradication in 2000—due to a lack of vaccination compounded by misinformation about the measles, mumps, and rubella (MMR) vaccine [1], and the spread of substandard or false information amid health crises has been shown to increase fatality rates, such as in West Africa amidst the 2013–2016 Ebola epidemic [6]. While the public health community has identified the proliferation of online health misinformation as a public health threat, what has yet to be thoroughly researched and developed are clear methods to identify, track, and mitigate such misinformation that continues to lead to increased morbidity and mortality, and applications of these methods to building epidemic and misinformation response capacities within newsrooms.

Word embeddings have been successfully used to capture semantic change, or change in a words' meanings [e.g., 4, 5, 7], as well as reveal implicit biases associated with words. Garg et al. 2018, for instance, show how certain professions (e.g., the word "teacher" or "doctor") are more or less associated with male/female words (e.g., he/him/man and she/her/woman). While these methods have been used extensively in communications research to assess changes in semantic trends overtime, they have the potential to be used as early detection techniques for misinformation by identifying large shifts in the associations between words. We are specifically exploring the associations between health-related words in our collaborations with newsrooms.

In developing this early monitoring system, we measure the strength of associations between various health terms and alleged symptoms, contraindications, and outcomes, which include word pairs such as "HPV" and "autism," or "abortion" and "cancer". By including a large set of key terms for monitoring, we can identify strong associations between unexpected terms as they emerge, and discover new or strengthening associations between words that may indicate the beginning of a misinformation threat. Equitable processes for discovery and selection of health claims to both report on and fact-check is a challenging area [3] where we believe this methodology will assist journalists. By monitoring all of the content on a platform, this process can be applied towards generating 'tips' that can be used by newsrooms.

Our system aims to address the challenges around identifying and tracking health misinformation by (i) exploring a potential new framework for surveilling the proliferation

C+J '20, March 20–21, 2020, Boston, MA

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/xxxxxx.xxxxxx

of online health information across digital contexts and (ii) applying this early detection methodology in our collaborations with health journalists. We trained word embeddings on 5.5 years of Twitter data in monthly intervals and developed a list of 57 health-related terms with input from journalists, public health scholars, and social media scholars. Our preliminary results show our methods are sensitive to emerging health information. We, for example, detect a large increase in the association between "mosquito" and "abortion" following the 2015–16 Zika virus epidemic (Figure 1).

Our methodology can provide newsrooms with ability to track such associations in real time, and across languages to indicate topics that may be of emerging interest or importance. By training our word embeddings on different corpora—including text from Twitter, Wikipedia, news media, and scientific abstracts—and in different languages—including English and Spanish—, we will be able to compare the strength of associations across languages and platforms, enabling early detection of patterns and the proliferation of problematic associations across the digital media ecosystem.

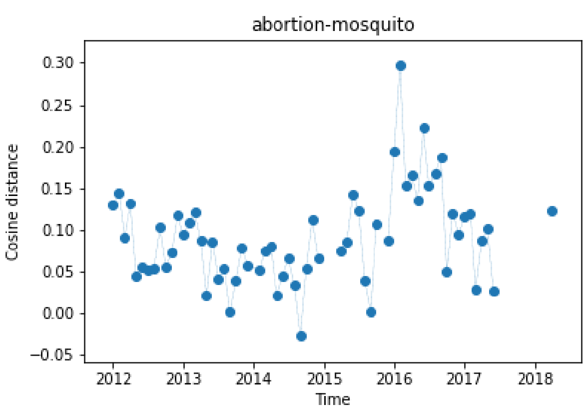

Figure 1: The strength of the association between "abortion" and "mosquito" as measured by the cosine distance of word embeddings trained on monthly Twitter data rises markedly at the beginning of 2016.

REFERENCES

[1] Centers for Disease Control and Prevention. 2019. Measles Cases and Outbreaks, Measles Cases in 2019. Measles (Rubeola). https://www.cdc.gov/measles/cases-outbreaks.html

[2] Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences* 115, 16 (2018), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

[3] Nat Gyenes and Megan Marrelli. 2019. Health Equity Through Health Fact-checking: A Primer. https://health.meedan.com/#misinfodemic-response-toolkit

[4] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Austin, Texas, 2116–2121. https://doi.org/10.18653/v1/D16-1229

[5] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, 1489–1501. https://doi.org/10.18653/v1/P16-1141

[6] Allgaier Joachim and Svalastog Anna Lydia. 2015. The communication aspects of the Ebola virus disease outbreak in Western Africa–do we need to counter one, two, or many epidemics? *Croatian medical journal* 56, 5 (oct 2015), 496–499. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4655935/

[7] Philippa Shoemark, Farhana Ferdousi Liza, Dong Nguyen, Scott Hale, and Barbara McGillivray. 2019. Room to Glo: A Systematic Comparison of Semantic Change Detection Approaches with Word Embeddings. 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)*. Association for Computational Linguistics, Hong Kong, China, 66–76. https://doi.org/10.18653/v1/D19-1007

[8] World Health Organizaton. 2019. Ten threats to global health in 2019. https://www.who.int/emergencies/ten-threats-to-global-health-in-2019