



Data science, social media, and HPV vaccines: the state-of-the-art and future challenges

Adam G. Dunn

Centre for Health Informatics

Australian Institute of Health Innovation

Macquarie University

@adamgdunn

for the 2018 National HPV Vaccination Roundtable Meeting (Atlanta)



MACQUARIE
University

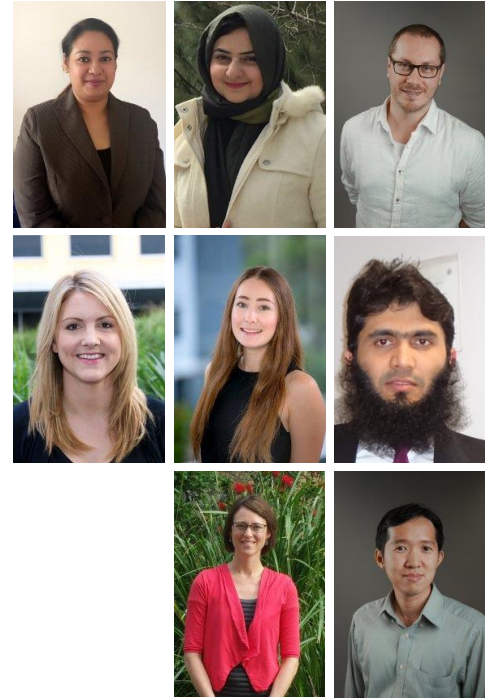
Competing interests: No financial conflicts of interest.

Funding: **NHMRC Project APP1128968**
Unrelated funding from NHMRC and AHRQ

Team: Samia Amin, Rabia Bashir; **Adam G. Dunn;**
Amalie Dyda; Paige Newman; Zubair Shah;
Maryke Steffens; Didi Surian
(team includes computer science, data science and
machine learning, software engineering, epidemiology
and public health, journalism, and clinical medicine)

Team alumni: Xujuan Zhou, Diana Arachi, Smriti Raichand

Collaborators: **Julie Leask, Kenneth D. Mandl, Aditi Dey,**
Enrico Coiera, Gilla Shapiro, Margaret Kelaher,
and many more...



Analyses of HPV vaccine information on social media:

YouTube: Ache 2008

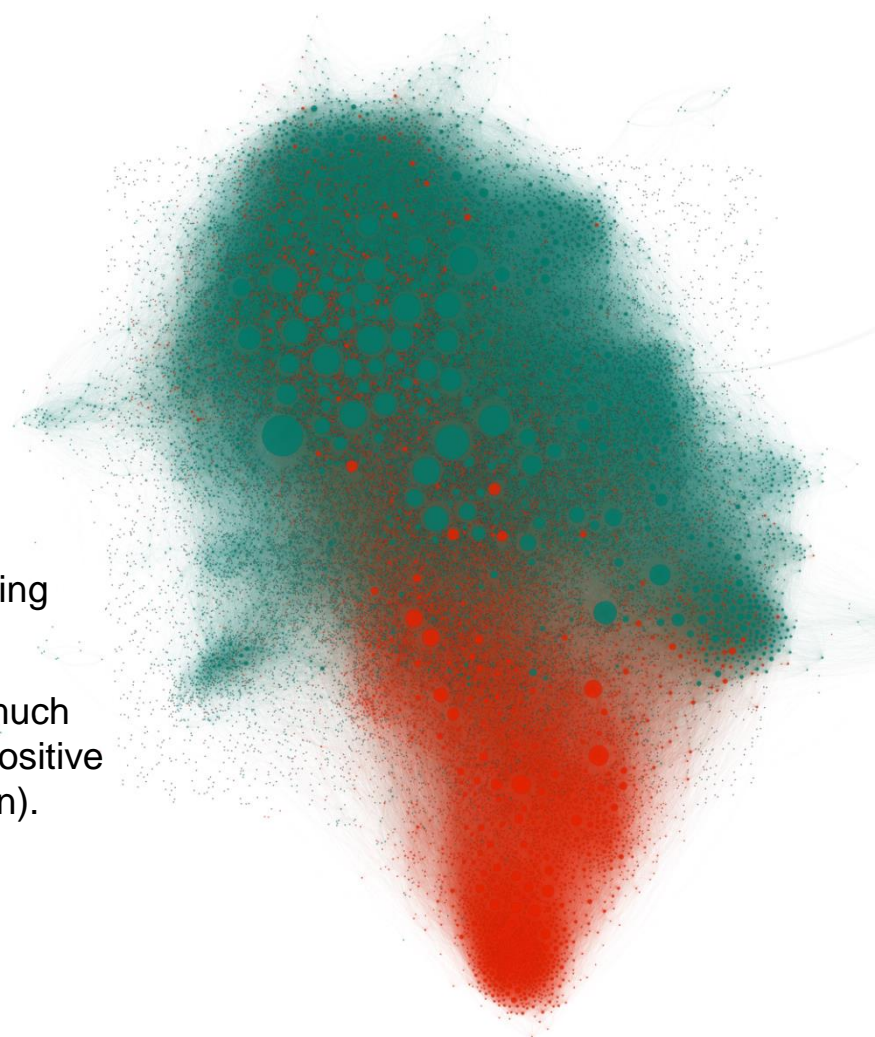
MySpace: Keelan 2010

Analysis of vaccine information on Twitter

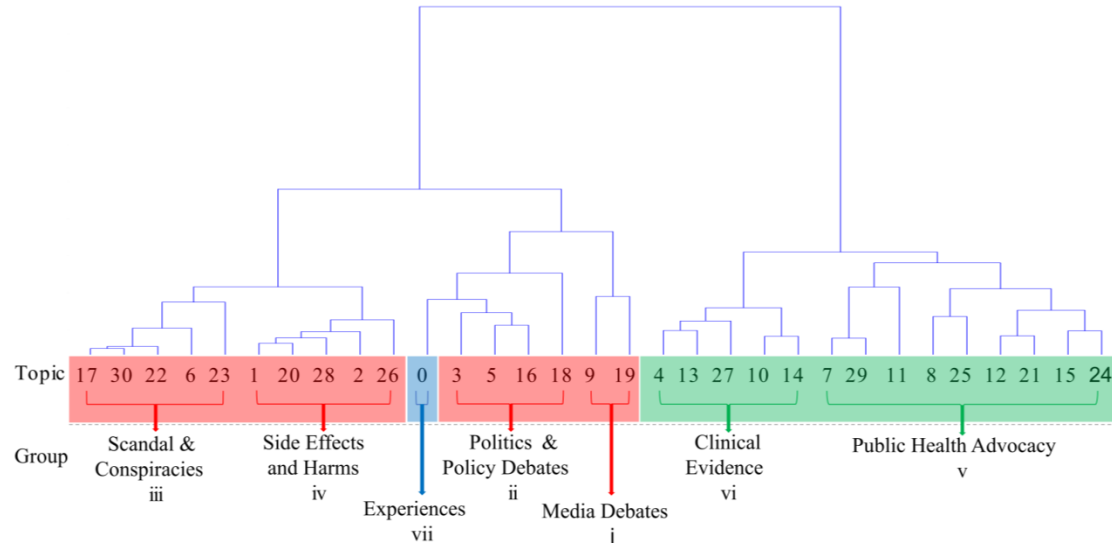
Signorini 2010

Salathe 2011

- We started collecting HPV vaccine tweets in 2013 (764,869 tweets as of 29 Jan 2018).
- Discovered we could predict negative tweets without reading them; just need to know who the users follow.
- Users mostly exposed to negative tweets (orange) were much more likely to go on to post a negative tweet rather than positive tweet about HPV vaccines compared to other users (green).



I wanted to know if Twitter could be a good enough signal of a population's **information diet** that we could *actually use* it to understand whether information was demonstrably associated with attitudes, behaviours, and HPV vaccine coverage.



“Computer scientists often brilliantly solve the wrong problems.”

Problem 1: flawed studies published with potentially dangerous conclusions.

- Call out bad epidemiology from data science;
- Contribute to pre-publication and post-publication peer review;
- Know and understand the limits and biases in social media data; and
- Real collaboration to teach computer scientists epidemiological methods.

Problem 2: robust studies that can't be used to guide policy or practice.

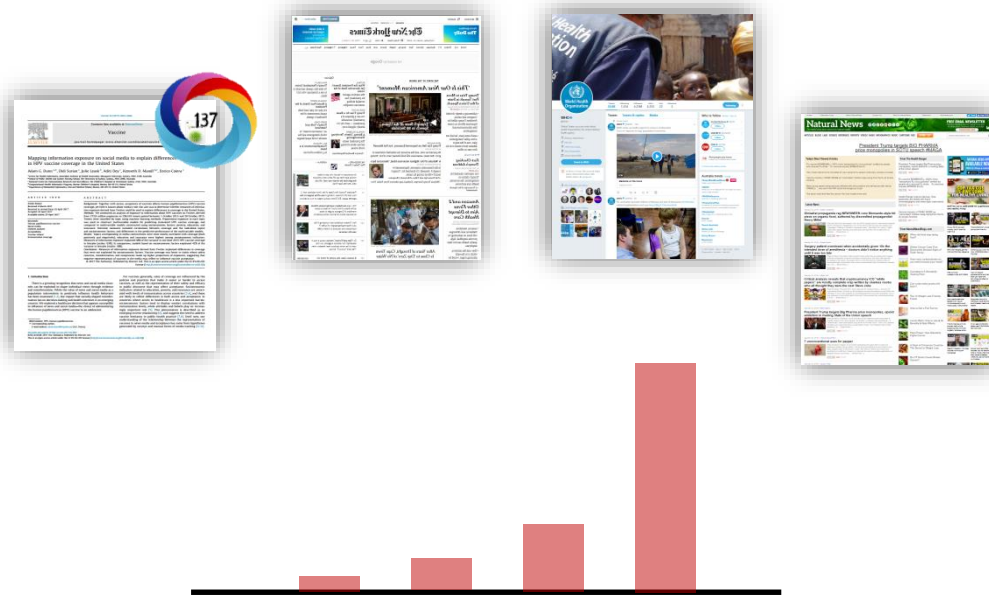
- Ask data scientists the *“why are we doing this?”* question more often; and
- Help translate research findings into real-time news/misinformation reporting and surveillance of attitudes and behaviours.

A poor literature review can be a good indicator of less robust research in this area.



MACQUARIE
University

I wanted to be able to connect the work back to how people and communities *actually engage* with **evidence and misinformation**, because these are where we can provide evidence for how to manipulate the information diets of at-risk populations.



We already have the tools we need to **identify individuals** with opinions and attitudes that put them at risk of unhealthy decisions and behaviours.

With automated personalisation of advertising and chatbots we are now able to deploy **autonomous behavioural interventions**; digital mass persuasion with no consent, and no humans involved in the diagnosis or delivery of the intervention.

If we choose to go down this path as a field of research, it will be vital that we get the ethics and implementation right.

OPEN ACCESS | freely available online

Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach

M. Andrew Schwartz^{1,2}, Johannes C. Eichstaedt¹, Margaret L. Kern¹, Lukasz Duzvarynski¹, Stephanie M. Ramones¹, Megha Agrawal^{1,3}, Ashal Shah¹, Michal Kosinski¹, David Stillwell¹, Martin P. Sellman¹, Lyle H. Ungar¹

Abstract

We analyzed 700 million words, phrases, and text instances collected from the Facebook messages of 75,000 volunteers who also provided personality tests, and found striking correlations in language with personality, gender, and age in our open-vocabulary technique. The data itself offers a comprehensive registration of language that distinguishes people. Reading connections are not captured with traditional closed-vocabulary word-category analyses. Our analyses show new links to psychosocial processes including trust, but also find that, contrary to high theoretical but untested expectations, we do not see other research findings, including gender segregation on the internet and the well-documented link between age and personality. We also find that personality and gender are related to the same linguistic features, suggesting that personality is not as distinct from gender as we have assumed. We also find that personality and gender are related to the same linguistic features, suggesting that personality is not as distinct from gender as we have assumed. We also find that personality and gender are related to the same linguistic features, suggesting that personality is not as distinct from gender as we have assumed.

PubMed: 24086296

Psychological targeting as an effective approach to digital mass persuasion

S. C. Mead^{1,2}, M. Kosinski^{1,3}, G. Haver¹, and D. S. Stillwell^{1,4}

Abstract

People are exposed to persuasive communication every time they click on the internet. This information is targeted to them based on their personality, gender, and age. We found that this targeting is effective, and that it is possible to predict the effectiveness of this targeting. We found that this targeting is effective, and that it is possible to predict the effectiveness of this targeting. We found that this targeting is effective, and that it is possible to predict the effectiveness of this targeting.

PubMed: 29133409

Exposure to ideologically diverse news on Facebook

David Bakula¹, Yoonsook Shin¹, Leah A. Sidelinger¹

Abstract

Exposure to news and other information increasingly occurs through social media. How do these news networks influence exposure to perspectives that are ideologically diverse? We used a naturalistic experiment to investigate this question. We found that exposure to ideologically diverse news on Facebook is associated with a more liberal political orientation. We found that exposure to ideologically diverse news on Facebook is associated with a more liberal political orientation.

PubMed: 25953820

The Spread of Behavior in an Online Social Network Experiment

David Corley

Abstract

How does one person's behavior spread to others in a social network? We conducted an experiment on Facebook to investigate this question. We found that behavior spreads through social networks in a way that is consistent with theoretical models of diffusion. We found that behavior spreads through social networks in a way that is consistent with theoretical models of diffusion.

PubMed: 20813952

PHOTO: GETTY IMAGES

Twitter: [@adamgdunn](https://twitter.com/adamgdunn)
Email: adam.dunn@mq.edu.au

These slides, with links, will be made available online at:
www.adamgdunn.net/speaking/
with thanks to @parryville for video production

1. **AG Dunn**, D Surian, J Leask, A Dey, KD Mandl, E Coiera (2017) Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States, *Vaccine*, 35(23):3033-3040.
2. G Shapiro, D Surian, **AG Dunn**, R Perry, M Kelaher (2017) Comparing human papillomavirus vaccine concerns on Twitter: A cross-sectional study of users in Australia, Canada, and the UK, *BMJ Open*, 7:e016869.
3. D Surian, DQ Nguyen, G Kennedy, M Johnson, E Coiera, **AG Dunn** (2016) Characterizing Twitter discussions about HPV vaccines using topic modelling and community detection. *Journal of Medical Internet Research*, 18(8):e232.
4. **AG Dunn**, J Leask, X Zhou, KD Mandl, E Coiera (2015) Associations between exposure and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. *Journal of Medical Internet Research*, 17(6):e144.
5. X Zhou, E Coiera, G Tsafnat, D Arachi, M-S Ong, **AG Dunn** (2015) Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter, *Studies in Health Technology and Informatics*, 216:761-765.

Other examples of *data-driven* studies using Twitter data related to HPV vaccines:

1. J Du, J Xu, H-Y Song, **C Tao** (2017) Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data, *BMC Medical Informatics and Decision Making*, 17(Suppl2):69.
2. J Du, J Xu, H-Y Song, **C Tao** (2017) Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets, *Journal of Biomedical Semantics*, 8:9.
3. J Du, J Huang, R Duan, Y Chen, **C Tao** (2017) Comparing the human papillomavirus vaccination opinions trends from different Twitter user groups with a machine learning based system and semiparametric nonlinear regression, *MEDINFO 2017*.
4. P Chakraborty P, **J Colditz**, A Silvestre, M Friedman, K Bogen, B Primack (2017) Observation of public sentiment toward human papillomavirus vaccination on Twitter. *Cogent Medicine*, 4(1):1390853.
5. Wheldon C, **Bian J**, Moser R. Surveillance of HPV-related vaccination discourse on Twitter in the United States: A case study of Kansas and Rhode Island. *2016 International Conference on Social Media & Society*. July 11-13, 2016. London, UK.
6. P Massey, A Leader, **E Yom-Tov**, A Budenz, K Fisher, A Klassen. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. *Journal of Medical Internet Research*, 2016;18(12):e318.
7. L Mahoney, T Tang, K Ji, J Ulrich-Schad. The digital distribution of public health news surrounding the human papillomavirus vaccination: a longitudinal infodemiology study. *JMIR Public Health and Surveillance*, 2015;1(1):e2.

Names in bold are good contacts for more information.