OFFICIAL ABSTRACT and CERTIFICATION

Categorizing geocoded anti-vaccination tweets in urban areas using Latent Dirichlet Allocation (LDA) and dictionary based modeling

Jaime Levin
Paul D. Schreiber High School, Port Washington, New York, USA

The anti-vaccination movement, involving an expansion of vaccine hesitancy and vaccine refusal among parents, poses a public health threat by increasing the possible spread of contagious and preventable diseases, including measles, mumps, and pertussis. In recent years, social media platforms such as Twitter have provided a channel for the dissemination of vaccine misinformation. Geographic correlates of particular anti-vaccination arguments remain to be unknown.

This study aimed to identify the differences in anti-vaccination rhetoric expressed in Twitter posts within urban cities on opposite coasts of the United States. Using novel streamed Twitter data from the summer of 2019, this study identified the most prominent topics of anti-vaccination Twitter arguments in northern and southern urban areas in California and New York by creating a topic model with Latent Dirichlet Allocation (LDA). The LDA model pinpointed four distinct reasons why users are against vaccinations in the Twitter data studied, each with a unique set of keywords. Using topical key word frequencies, I created a novel measurement (Topic Density Proportion) to find the density of the four topics in each geographic region analyzed. Residual chi-squared tests revealed that there was a significant difference between the topics of anti-vaccination tweets expressed among the four geographic locations, ranging from concerns about the physical dangers of vaccines to questions of government conspiracy. The ability to pinpoint the most prevalent anti-vaccination arguments in distinct areas is immensely impactful, enabling public health officials and policy makers to design tailored interventions to combat specific vaccine misinformation and prevent mass outbreaks.

Category
Pick one only — mark an "X" in box at right

- Animal Sciences
- Behavioral & Social Sciences
- Biochemistry
- Biomedical & Health Sciences
- Biomedical Engineering
- Cellular & Molecular Biology
- Chemistry
- Computational Biology & Bioinformatics
- Earth & Environmental Sciences
- Embedded Systems
- Energy: Sustainable Materials and Design
- Engineering Mechanics
- Environmental Engineering
- Materials Science
- Mathematics
- Microbiology
- Physics & Astronomy
- Plant Sciences
- Robotics & Intelligent Machines
- Systems Software
- Translational Medical Sciences

1. As a part of this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

   ☐ human participants ☐ potentially hazardous biological agents

   ☐ vertebrate animals ☐ microorganisms ☐ rDNA ☐ tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting: ■ Yes ☐ No

3. This project is a continuation of previous research. ☐ Yes ■ No

4. My display board includes non-published photographs/visual depictions of humans (other than myself): ☐ Yes ■ No

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only ■ Yes ☐ No

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. ■ Yes ☐ No

This stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.