

Hate speech on Twitter predicts frequency of real-life hate crimes

June 24 2019



Credit: CC0 Public Domain

According to a first-of-its-kind study, cities with a higher incidence of a certain kind of racist tweets reported more actual hate crimes related to race, ethnicity, and national origin.

A New York University research team analyzed the location and linguistic features of 532 million tweets published between 2011 and 2016. They trained a machine learning model—one form of artificial intelligence—to identify and analyze two types of tweets: those that are targeted—directly espousing discriminatory views—and those that are self-narrative—describing or commenting upon discriminatory remarks or acts. The team compared the prevalence of each type of discriminatory [tweet](#) to the number of actual hate crimes reported during that same time period in those same cities.

The research was led by Rumi Chunara, an assistant professor of computer science and engineering at the NYU Tandon School of Engineering and biostatistics at the NYU College of Global Public Health, and Stephanie Cook, an assistant professor of biostatistics and social and behavioral sciences at the NYU College of Global Public Health.

"We found that more targeted, discriminatory tweets posted in a city related to a higher number of hate crimes," said Chunara. "This trend across different types of cities (for example, urban, rural, large, and small) confirms the need to more specifically study how different types of discriminatory speech online may contribute to consequences in the physical world."

The analysis included cities with a wide range of urbanization, varying

degrees of population diversity, and different levels of [social media](#) usage. The team limited the dataset to tweets and bias crimes describing or motivated by race, ethnic or national origin-based discrimination. Hate crimes are categorized and tracked by the Federal Bureau of Investigation, and crimes motivated by race, ethnicity, or national origin represent the largest proportion of hate crimes in the nation. Statistics for sexual orientation crimes were not available in all cities, although the researchers previously studied this form of bias.

The group also identified a set of discriminatory terms and phrases that are commonly used on social media across the country, as well as terms specific to a particular city or region. These insights could prove useful in identifying groups that may be likelier targets of racially motivated crimes and types of discrimination in different places. While most tweets included in this analysis were generated by actual Twitter users, the team found that an average of 8% of tweets containing targeted discriminatory language was generated by bots.

There was a negative relationship between the proportion of race/ethnicity/national-origin-based discrimination tweets that were self-narrations of experiences and the number of crimes based on the same biases in cities. Chunara noted that while experiences of discrimination in the [real world](#) are known psychological stressors with health and social consequences, the implications of online exposure to different types of online discrimination—self-narrations versus targeted, for example—need further study.

These results represent one of the largest, most comprehensive analyses of discriminatory social media posts and real-life bias crimes in this country, although the researchers emphasize that the specific causal mechanisms between social media [hate speech](#) and real-life acts of violence need to be explored.

More information: Kunal Relia et al. Race, Ethnicity and National Origin-based Discrimination in Social Media and Hate Crimes Across 100 U.S. Cities. arXiv:1902.00119 [cs.CY]. arxiv.org/abs/1902.00119

Provided by NYU Tandon School of Engineering

Citation: Hate speech on Twitter predicts frequency of real-life hate crimes (2019, June 24)
retrieved 3 May 2024 from
<https://techxplore.com/news/2019-06-speech-twitter-frequency-real-life-crimes.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.