

Twitter data used to track vaccination rates and attitudes

October 13 2011



Marcel Salathé at Penn State University used Twitter data to track vaccination rates and sentiments. Credit: National Institutes of Health

A unique and innovative analysis of how social media can affect the spread of a disease has been designed and implemented by a scientist at Penn State University studying attitudes toward the H1N1 vaccine. Marcel Salathé, an assistant professor of biology, studied how users of Twitter -- a popular microblogging and social-networking service -- expressed their sentiments about a new vaccine. He then tracked how the users' attitudes correlated with vaccination rates and how microbloggers with the same negative or positive feelings seemed to influence others in their social circles. The research is considered the first case study in how social-media sites affect and reflect disease networks, and the method is expected to be repeated in the study of other diseases. The results will be

published in the journal *PLoS Computational Biology*.

Salathé explained that he chose Twitter for two reasons. First, unlike the contents of Facebook, Twitter messages, known as "tweets," are considered public data and anyone can "follow," or track, the tweets of anyone else. "People tweet because they want other members of the public to hear what they have to say," Salathé said. Second, Twitter is the perfect database for learning about people's sentiments. "Tweets are very short -- a maximum of 140 characters," Salathé explained. "So users have to express their opinions and beliefs about a particular subject very concisely." Salathé began by amassing 477,768 tweets with vaccination-related keywords and phrases. He then tracked users' sentiments about a particular new [vaccine](#) for combating H1N1 -- a virus strain responsible for swine flu. The collection process began in August 2009, when news of the new [vaccine](#) first was made public, and continued through January 2010.

Salathé explained that sorting through the enormous number of vaccination-related tweets was no simple matter. First, he partitioned a random subset of about 10 percent and asked Penn State students to rate them as positive, negative, neutral, or irrelevant. For example, a tweet expressing a desire to get the H1N1 vaccine would be considered positive, while a tweet expressing the belief that the vaccine causes harm would be considered negative. A tweet concerning a different vaccine, for example, the Hepatitis B vaccine, would be considered irrelevant. Then, Shashank Khandelwal, a computer programmer and analyst in Penn State's Department of Biology and co-author of the paper, used the students' ratings to design a computer algorithm responsible for cataloging the remaining 90 percent of the tweets according to the sentiments they expressed. "The human-rated tweets served as a 'learning set' that we used to 'teach' the computer how to rate the tweets accurately," Salathé explained. After the tweets were analyzed by the computer algorithm, the final tally, after the irrelevant ones were

eliminated, was 318,379 tweets expressing either positive, negative, or neutral sentiments about the H1N1 vaccine.

Because Twitter users often include a location in their profiles, Salathé was able to categorize the expressed sentiments by U.S. region. Also, using data from the Centers for Disease Control and Prevention (CDC), he was able to determine how vaccination [attitudes](#) correlated with CDC-estimated vaccination rates. Using these data, Salathé found definite patterns. For example, the highest positive-sentiment users were from New England, and that region also had the highest H1N1 vaccination rate. "These results could be used strategically to develop public-health initiatives," Salathé explained. "For example, targeted campaigns could be designed according to which region needs more prevention education. Such data also could be used to predict how many doses of a vaccine will be required in a particular area."

In addition, Salathé was able to construct an intricate social network by determining who followed the tweets of whom; that is, he was able to determine clusters of like-minded Twitter users. "The assumption is that people tend to communicate online almost exclusively with people who think the same way. This phenomenon creates 'echo chambers' in which dissenting opinions are not heard," Salathé said. As it turned out, that assumption was correct. Salathé found that users with either negative or positive sentiments about the H1N1 vaccine followed like-minded people. "The public-health message here is obvious," Salathé said. "If anti-vaccination communities cluster in real, geographical space, as well, then this is likely to lead to under-vaccinated communities that are at great risk of local outbreaks." He explained that, when unvaccinated individuals cluster together, herd immunity -- a population-level immunity that occurs when a critical mass has been vaccinated -- no longer affords much protection against disease. "By definition, herd immunity only works if unvaccinated, unprotected individuals are distributed sparsely throughout the population, buffered from the disease

by vaccinated individuals," Salathé said. "Unfortunately, the data from [Twitter](#) seem to indicate that the buffer of protection cannot be counted on if these clusters exist in real, geographical space."

In addition to location-related and network patterns, Salathé was able to track sentiment patterns over time. For example, he found that negative expressions spiked during the time period when the vaccine was first announced. Later, more-positive sentiments emerged when the vaccine was first shipped across the United States. Salathé also tracked spikes of negative tweets that corresponded, not surprisingly, to periods of vaccine recall.

Salathé plans to use his unique social-media analysis to study other diseases, such as obesity, hypertension, and heart disease. "We think of a disease such as obesity as noninfectious, while a disease such as the flu is clearly infectious. However, it might be more useful to think of behavior-influenced diseases as infectious, as well," Salathé said. "Lifestyle choices might be 'picked up' in much the same way that pathogens -- viruses or bacteria -- are acquired. The difference is simply that in the one instance the infectious agent is an idea rather than a biological entity."

Salathé added that, in the industrialized world, future generations will worry less about infectious diseases and more about diseases linked to lifestyle and behavior. "Behavior-influenced diseases always have existed, but, until recently, they were masked: People died of infectious diseases relatively early in their life cycles. So behavior-influenced diseases weren't really on anyone's radar," Salathé explained. "Now that heart disease -- a malady caused, at least in part, by lifestyle -- is moving to the top of the list of killers, it might be wise to focus on how social media influences behaviors such as poor diet and infrequent exercise," Salathé said.

Provided by Pennsylvania State University

Citation: Twitter data used to track vaccination rates and attitudes (2011, October 13) retrieved 11 June 2026 from

<https://medicalxpress.com/news/2011-10-twitter-track-vaccination-attitudes.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.