

Using Twitter to track the flu: Researchers find a better way to screen the tweets

January 25 2013

Sifting through social media messages has become a popular way to track when and where flu cases occur, but a key hurdle hampers the process: how to identify flu-infection tweets. Some tweets are posted by people who have been sick with the virus, while others come from folks who are merely talking about the illness. If you are tracking actual flu cases, such conversations about the flu in general can skew the results.

To address this problem, Johns Hopkins [computer scientists](#) and researchers in the School of Medicine have developed a new tweet-[screening method](#) that not only delivers real-time data on [flu cases](#), but also filters out online chatter that is not linked to actual [flu infections](#). Comparing their method, which is based on analysis of 5,000 publicly available tweets per minute, to other Twitter-based tracking tools, the Johns Hopkins researchers say their real-time results track more closely with government disease data that takes much longer to compile.

"When you look at Twitter posts, you can see people talking about being afraid of catching the flu or asking friends if they should get a [flu shot](#) or mentioning a public figure who seems to be ill," said Mark Dredze, an assistant research professor in the Department of Computer Science who uses tweets to monitor public health trends. "But posts like this don't measure how many people have actually contracted the flu. We wanted to separate hype about the flu from messages from people who truly become ill."

Dredze, who also is a research scientist at the Johns Hopkins Human

Language Technology Center of Excellence, led a team that in mid-2011 released one of the first and most comprehensive studies showing that Twitter data can yield useful public health information. Since then, this strategy has become so popular that the U.S. Department of Health and Human Services last summer sponsored a contest challenging researchers to design an online application that could track major [disease outbreaks](#).

This winter, as the United States entered an unusually severe and early flu season, Twitter-based flu projections have drawn increasing attention. Many public tweets, such as, "I'm so sick this week with the flu," can indicate a rise in the flu rate. Collecting enough of these tweets can help health officials gauge the scope and severity of an epidemic.

But the reliability of many computer models can be weakened by too many tweets that point to flu-related news reports and other matters not directly linked to a flu case, according to David Broniatowski, a School of Medicine postdoctoral fellow in the Department of Emergency Medicine's Center for Advanced Modeling in the Social, Behavioral, and Health Sciences. "For example," he said, "a recent spike in Twitter flu activity was caused by discussions about basketball legend Kobe Bryant's flu-like symptoms during a recent game. Mr. Bryant's health notwithstanding, such tweets do very little to help public health officials prepare our nation for the next big outbreak."

To improve their accuracy when using tweets to track the flu, the John Hopkins team developed sophisticated statistical methods based on human language processing technologies. The methods are designed to filter out the chatter. The system can distinguish, for example, between "I have the flu" and "I'm worried about getting the flu."

Another advantage of the Johns Hopkins flu projection method is that it can produce real-time results. By comparison, the U.S. Centers for Disease Control and Prevention, which record flu-related symptoms

from hospital visits, typically take two weeks to publish data on the flu's prevalence.

To check the reliability of their enhanced system, the Johns Hopkins researchers recently compared their results to CDC data for the same period. The researchers said that during November and December 2012, their system demonstrated a substantial improvement in tracking with CDC figures as compared to previous Twitter-based tracking methods. "In late December," Dredze added, "the news media picked up on the flu epidemic, causing a somewhat spurious rise in the rate produced by our Twitter system. But our new algorithm handles this effect much better than other systems, ignoring the spurious spike in tweets."

The researchers have also used their Twitter data to produce United States maps that document the stark differences between last year's mild flu season and the much higher incidence of the virus in the winter of 2012-2013.

While their new method was only recently developed, the Johns Hopkins researchers chose to release information on the flu tracking system because of the higher incidence of illness this winter. Team members hope to share the enhanced flu tracking method with leading government health agencies.

"This new work demonstrates that Twitter posts can be used to guide public health officials in their response to outbreaks of infectious diseases," Dredze said. "Our hope is that the new technology can be used track other diseases as well."

Other Johns Hopkins researchers participating in the Twitter flu project are doctoral student Michael Paul and recent bachelor's degree graduate Alex Lamb, both in the Department of Computer Science.

The Johns Hopkins researchers noted that their enhanced Twitter [flu](#) analysis system looked only at public tweets in which all user names and gender information had been removed. The system was tested only on messages from the United States. The research was funded in part by the National Institutes of Health's Models of Infectious Disease Agent Study.

Provided by Johns Hopkins University

Citation: Using Twitter to track the flu: Researchers find a better way to screen the tweets (2013, January 25) retrieved 25 April 2024 from <https://medicalxpress.com/news/2013-01-twitter-track-flu-screen-tweets.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.