

# Researchers use Twitter to track the flu in real time

May 9 2017, by Thea Singer



Researchers led by Northeastern's Alessandro Vespignani have developed a computational model to project the spread of the flu using Twitter posts in combination with key parameters of each season's epidemic. Credit: Greg Grinnell

"This flu is horrendous. Can't breathe, can't sleep or eat. Muscles ache, fever 102. Should have gotten the shot. Time for a movie marathon."

The above tweet looks like 140 characters of misery. But in the hands of Northeastern's Alessandro Vespignani and his colleagues, it is so much more.

An international team led by Vespignani has developed a [unique computational model](#) to project the spread of the seasonal flu in real time. It uses posts on Twitter in combination with key parameters of each season's epidemic, including the incubation period of the disease, the immunization rate, how many people an individual with the virus can infect, and the viral strains present.

Tested against official influenza surveillance systems, the model has been shown to accurately forecast the disease's evolution up to six weeks in advance—significantly earlier than other models. It will enable public health agencies to plan ahead in allocating medical resources and launching campaigns that encourage individuals to take preventative measures such as vaccination and increased hand washing.

"In the past, we had no knowledge of initial conditions for the flu," says Vespignani, who is also director of the Network Science Institute at Northeastern. The initial conditions—which show where and when an epidemic began as well as the extent of infection—function as a launching pad for forecasting the spread of any disease.

To ascertain those conditions, the researchers incorporated Twitter into their parameter-driven model. "This kind of integration has never been done before," says Vespignani. "We were not looking for the number of people who were sick because Twitter will not tell you that. What we wanted to know was: Do we have more flu at this point in time in Texas or in New Jersey, in Seattle or in San Francisco? Twitter, which includes GPS locations, is a proxy for that. By looking at how many people were tweeting about their symptoms or how miserable they were because of the flu, we were able to get a relative weight in each of those areas of the

U.S."

The paper on the novel model received a coveted Best Paper Honorable Mention award at the 2017 International World Wide Web Conference last month following its presentation. It was one of only four papers out of more than 400 presented to be selected for an award.

## **A work in progress**

The researchers' work began when the Centers for Disease Control and Prevention announced the "Predict the Influenza Season Challenge" in November 2013, an invitation to external researchers to advance the science of forecasting infectious diseases. Vespignani and his team have been participating ever since, with the new paper covering their projections for the 2014-15 and 2015-16 flu seasons in the U.S., Italy, and Spain.

Over those time periods, they applied forecasting and other algorithms week by week to the key parameters informed by the Twitter data. "This gave us a large number of possible ways the disease might evolve," says Vespignani. They then matched the resulting simulations with the surveillance data generated by the CDC and clinical and personal reports of influenza-like illnesses from the three countries. "The [surveillance data](#) tells us the ground truth for the past four weeks, but it is always delayed by about one week because you need to get the report from the doctor," he says. By analyzing the evolving dynamics revealed in the past data, they were able to select the model that would most likely forecast the future.

The explicit modeling of the disease's parameters—information about the dynamics of the disease itself—set Vespignani's model apart from others in the challenge. For example, they could identify the week when the epidemic would reach its peak and the magnitude of that peak with

an accuracy of 70 to 90 percent six weeks in advance of the event.

"By capturing the key parameters, we could track how serious the flu was each year compared with every other year and see what was driving the spread," says first author Qian Zhang, PhD'14, associate research scientist at Northeastern. "That is what the public health agencies and the epidemiologists really care about. We are not just playing a game of numbers, which is what straightforward statistical models do."

While the paper reports results using Twitter data, the researchers note that the model can work with data from many other digital sources, too, as well as online surveys of individuals such as [influenzanet](#), which is very popular in Europe.

"Our model is a work in progress," emphasizes Vespignani. "We plan to add new parameters, for example, school and workplace structure. This is not a challenge in the sense that you want to win. This is a science challenge in which you want to learn—to see that there is not a single [model](#) but a portfolio of models that will tell us new things."

Provided by Northeastern University

Citation: Researchers use Twitter to track the flu in real time (2017, May 9) retrieved 25 April 2024 from <https://phys.org/news/2017-05-twitter-track-flu-real.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.