

# Your Google searches and tweets might help forecast the next disease outbreak

18 May 2022, by Eva Botkin-Kowacki



People leave a trail of breadcrumbs when they navigate the digital world, offering clues about what is happening in their lives—including their health. Northeastern's Mauricio Santillana is using machine learning algorithms to turn those clues into an early warning system for disease outbreaks. Credit: Matthew Modoono/Northeastern University

It seems like yet another punchline for anyone joking about the past two years of pandemic life. But to scientists forecasting future disease outbreaks, it's important data.

Scented candles began receiving an influx of [negative reviews](#) online in 2020. Dissatisfied customers proclaimed that some of the most fragrant, most popular products from famous companies like Yankee Candle had "no smell" or even smelled bad.

This wasn't just a few bad reviews. The most popular scented candles sold on Amazon were receiving an average of 4 to 4½ stars before 2020, but over the course of that first year of the pandemic, the reviews fell by about a full star. Social media users mused about a link between these negative reviews and the loss of the sense

of smell associated with COVID-19 infections.

When COVID-19 cases rose again at the end of 2021 due to the omicron variant, researchers noted another uptick in those negative "no smell" reviews.

Those negative online reviews are what Mauricio Santillana calls "breadcrumbs." As people navigate the [digital world](#), they leave traces of what is going on in their offline lives, explains the director of the Machine Intelligence Group for the betterment of Health and the Environment (MIGHTE) in the Network Science Institute at Northeastern. Those "breadcrumbs" leave a trail for researchers like Santillana to follow as they project potential future outbreaks of COVID-19 and other diseases.

If there are anomalies in online trends—a spike in Google searches for shops that deliver chicken noodle soup, a sudden flurry of Tweets about navigating a quarantining family member, or bad reviews on scented candles—it could indicate that trouble is brewing. So Santillana is creating [machine-learning models](#) to spot the anomalies, make sense of these clues, and create an early warning system for disease outbreaks.

By adding [human behavior](#) to the mix, "we're creating an observatory of disease activity using different telescopes," says Santillana, a professor of physics and of electrical and computer engineering who recently joined Northeastern from Harvard University.

Santillana is teaming up with Alessandro Vespignani, director of the Network Science Institute and Sternberg Family Distinguished Professor at Northeastern, who leads a team of infectious-disease modelers that have been developing a set of projections about the possible futures of the COVID-19 pandemic since the crisis began.

Vespignani's models integrate details such as case

counts, hospitalizations, deaths, human mobility patterns, how often humans interact, how the virus transmits and more data focused on the disease spread itself. Santillana says his research adds a different sort of thermometer by looking at digital traces of human behaviors that are a step removed from the epidemiological data.

"In a way, we're trying to bring together these two perspectives to provide a more whole picture of outbreaks like COVID-19," Santillana says.

Santillana and Vespignani have already been collaborating, combining this digital behavioral data with epidemiological data in their modeling work. In a paper published in *Science Advances* last year, they showed that such a harmonized early warning system could anticipate a surge in COVID-19 cases and deaths by two to three weeks. With Santillana joining the Network Science Institute, the pair will work together to further develop this early-warning system for disease outbreaks—and not just for COVID-19.

The data that Santillana gathers encompasses a vast, diverse collection of information—not just Google search trends, social media posts, and online shopping reviews or orders. He has also used anonymized smart thermometer data to identify when some sort of illness might be ticking up in a region, anonymized mobility data from smartphones that illustrates when more people might be staying home sick, as well as trends in clinician searches for certain kinds of treatments or symptoms.

Even the Google searches and social media posts encompass a wide range of data. People could be searching for more information about their symptoms or quarantine recommendations, or they could simply be trying to figure out where to buy cough syrup or soup.

An uptick in just one of these behaviors in a region might indicate that COVID-19 or another infectious disease is sweeping into a community, or it might just be that there was a new sci-fi film that came out and piqued people's curiosity about pandemics more generally. That's why Santillana says it's important for his models to take into account many

different data sources. The machine learning models are also designed to figure out whether a rise in certain Google searches, for example, actually correlates with a rise in infections and hospitalizations in order to determine if it is worth considering as a harbinger of a disease outbreak.

This new type of "telescope," as Santillana termed it, will be a component of the U.S.'s new disease forecasting initiative, the Center for Forecasting and Outbreak Analytics (CFA). Santillana is part of a team of experts advising that effort.

"In the same way that the weather forecasting systems around the world work," he explains, "the idea is to contribute different ways to look at information that is being produced in real time and design systems that will recognize when something anomalous happens."

Like weather forecasting agencies, the CFA will essentially be an [early warning system](#), identifying when and where disease outbreaks might occur so that public-health officials can take action to prevent them from becoming devastating.

**More information:** Nicole E. Kogan et al, An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time, *Science Advances* (2021). [DOI: 10.1126/sciadv.abd6989](https://doi.org/10.1126/sciadv.abd6989)

Provided by Northeastern University

APA citation: Your Google searches and tweets might help forecast the next disease outbreak (2022, May 18) retrieved 8 August 2022 from <https://techxplore.com/news/2022-05-google-tweets-disease-outbreak.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.*