

Scientists find social media provides proxy measurement of pollution

26 September 2016, by David Ruth

灰霾	dust-haze
雾霾	haze
雾蒙	misty
灰蒙	dusky
气污	air pollution
度污	degree of pollution
很蓝	very blue
明媚	bright
好天	good day
晴朗	sunny

Credit: Rice University

Residents of China's megacities who post comments about air quality to social media can give environmental scientists a window into pollution levels there.

A multidisciplinary study by Rice University researchers showed that the frequency of key words like dust, cough, haze, mask and blue sky can be used as a proxy measurement of the amount of [airborne particulate matter](#) in the country's urban centers at any given time.

The words were culled from millions of posts to China's Weibo, a popular microblogging platform. The posts were collected by Rice computer scientists for a study on Chinese censorship of

[social media](#) three years ago.

The research led by Rice computer scientist Dan Wallach and environmental engineer Daniel Cohan appears this month in the open-access journal *PLOS One*.

"The big takeaway is that people grouse about [air quality](#), and as it gets worse, people complain more," said Wallach, a professor of computer science and electrical and computer engineering, whose lab collected the publicly available posts.

"When it's really bad, it flattens out," he said. "They're as complained-out as they're going to be. And if it gets good enough, few people complain. But there's a zone in the middle where people really grouse, and we can measure that.

"A city the size of Beijing has air-quality meters, but not many," Wallach said. "But if you have millions of people, you potentially have millions of meters. It's a way of adding extra data."

The researchers came up with a metric, the Air Discussion Index (ADI), based on the frequency with which pollution-related terms appeared in 112 million posts from 2011 to 2013 by residents of Beijing, Shanghai, Guangzhou and Chengdu, where pollution is thought to be most troublesome in China.

"We looked at what words correlated with the pollution-level data we had," Wallach said. "Some words that came out were nonsense. But others, like cough or wheeze, clearly had something to do with the conditions. Others, like blue sky, inversely correlated with the weather or pollution."

"There's a lot of discussion about censorship in Chinese media, including in Dan Wallach's work, but one of the things we like about this particular study is that it relies on data that are almost never censored, the most innocuous terms of all," said co-

author Aynne Kokas, an assistant professor of media studies at the University of Virginia and an affiliate of Rice University's Baker Institute for Public Policy.

"These terms are almost impossible to censor because of how common they are," she said. "As a result, we think this method is really effective not only in China but could also work in other contexts where there are heavily regulated social-media environments."

The most accurate ADI readings were those for Beijing. When matched to hourly sensor readings from the U.S. Embassy there, the researchers found the technique analyzed [pollution levels](#) with an accuracy of 88.2 percent. ADI performance for the other cities where the pollution isn't as severe and Weibo posts not as plentiful wasn't as accurate: 63 percent for Shanghai, 42 percent for Guangzhou and 36 percent for Chengdu.

Particulate matter measuring less than 2.5 microns in diameter—about 30 times less than the diameter of the average human hair—is known to permanently damage the lungs. The United States' air-quality standard for concentrations of this size of particulate matter is no more than 35 micrograms (millionths of a gram) per cubic meter over any 24-hour period and an annual average of no more than 12 micrograms per cubic meter.

Cohan said Chinese air pollution standards aren't vastly different from those in the U.S., but the pollutant concentrations are. "Particulate matter levels in Beijing are often 10 times as high as we typically observe in U.S. cities," he said.

Wallach said he was surprised by the level of air-quality information that was found in the Weibo posts—data that he and colleagues had collected for a 2013 study on social media censorship.

"I was chatting with Dan Cohan, and I said, 'Hey, I've got all this data about China. Do you think we could measure something about pollution from all this data?'" Wallach recalled. "We all got together to see if the Weibo data told a story, and it turns out it did."

Cohan said, "China is an ideal testbed, because the pollutant levels are so high and so variable that you can literally see the difference day to day. Still, I was surprised that social media posts could correlate so strongly with air-quality conditions."

Wallach said it was interesting to note that the U.S. Embassy measurements correlated well with the Chinese government's own ground-level reporting on urban pollution. "Some people in China think their government might be lying to them about air quality, but based on what we found, that isn't the case," he said.

More information: Zhu Tao et al. Inferring Atmospheric Particulate Matter Concentrations from Chinese Social Media Data, *PLOS ONE* (2016). [DOI: 10.1371/journal.pone.0161389](https://doi.org/10.1371/journal.pone.0161389)

Provided by Rice University

APA citation: Scientists find social media provides proxy measurement of pollution (2016, September 26) retrieved 25 May 2019 from <https://phys.org/news/2016-09-scientists-social-media-proxy-pollution.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.