

# From location data alone, even low-tech snoopers can identify Twitter users' homes, workplaces

May 17 2016, by Larry Hardesty

---



“[W]hen you send location data as a secondary piece of information, it is extremely simple for people with very little technical knowledge to find out where you work or live,” Ilaria Liccardi says. Credit: Jose-Luis Olivares/MIT

Researchers at MIT and Oxford University have shown that the location stamps on just a handful of Twitter posts—as few as eight over the course of a single day—can be enough to disclose the addresses of the poster's home and workplace to a relatively low-tech snooper.

The tweets themselves might be otherwise innocuous—links to funny videos, say, or comments on the news. The location information comes from geographic coordinates automatically associated with the tweets.

Twitter's location-reporting service is off by default, but many Twitter users choose to activate it. The new study is part of a more general project at MIT's Internet Policy Research Initiative to help raise awareness about just how much privacy people may be giving up when they use social media.

The researchers describe their research in a paper presented last week at the Association for Computing Machinery's Conference on Human Factors in Computing Systems, where it received an honorable mention in the best-paper competition, a distinction reserved for only 4 percent of papers accepted to the conference.

"Many people have this idea that only machine-learning techniques can discover interesting patterns in location data," says Ilaria Liccardi, a research scientist at MIT's Internet Policy Research Initiative and first author on the paper. "And they feel secure that not everyone has the [technical knowledge](#) to do that. With this study, what we wanted to show is that when you send location data as a secondary piece of information, it is extremely simple for people with very little technical knowledge to find out where you work or live."

## Conclusions from clustering

In their study, Liccardi and her colleagues—Alfie Abdul-Rahman and

Min Chen of Oxford's e-Research Centre in the U.K.—used real tweets from Twitter users in the Boston area. The users consented to the use of their data, and they also confirmed their home and work addresses, their commuting routes, and the locations of various leisure destinations from which they had tweeted.

The time and [location data](#) associated with the tweets were then presented to a group of 45 study [participants](#), who were asked to try to deduce whether the tweets had originated at the Twitter users' homes, their workplaces, leisure destinations, or locations along their commutes. The participants were not recruited on the basis of any particular expertise in urban studies or the social sciences; they just drew what conclusions they could from location clustering.

They were also recruited in Oxford, to eliminate biasing that might result from familiarity with Boston geography. Similarly, they had no information about the content of the tweets.

The data were presented in three different forms. One was a static Google map, in which tweet locations were marked with virtual pins; one was an animated version of the same map, in which the pins appeared on-screen in chronological order; and the third—the resolutely low-tech version—was a table listing geographical coordinates, street names, and times of day.

The maps featured only street names, with no names of businesses, parks, schools, or other landmarks. Pins and table rows were, however, color coded to indicate general time of day—morning, afternoon, or evening.

The researchers also varied the volume of data that the participants were asked to consider: one day's, three days', or five days' worth. To avoid biasing, there was no overlap between data sets of different sizes.

## Bottom line

Predictably, participants fared better with map-based representations, correctly identifying Twitter users' homes roughly 65 percent of the time and their workplaces at closer to 70 percent. Even the tabular representation was informative, however, with accuracy rates of just under 50 percent for homes and a surprisingly high 70 percent for workplaces.

In general, participants also fared better with five days' worth of data than with three or one. Across all three representations, participants with five days' worth of data could correctly identify workplaces, for example, with more than 85 percent accuracy.

Interestingly, the participants' performance with three days' worth of data was generally worse than it was with only one. It could be that, while a single day's data is likely to be representative of a user's typical patterns of movement, three days' worth introduces the possibility of confounding variations, which are ironed out over five days.

"We want to investigate that," Liccardi says. "When we asked participants 'Which amount of data do you prefer?' most of them said 'medium,' even though it was the one that they got the least right. So you never know about perceptions."

"Ilaria's new paper puts two significant bricks in the wall of our privacy understanding," says Latanya Sweeney, professor of government and technology in residence at Harvard University and a former chief technology officer of the U.S. Federal Trade Commission. "First, her survey shows how people can learn sensitive information from seemingly innocuous facts, and, second, people will easily share information they believe is innocuous."

**More information:** [I Know Where You Live: Inferring Details of People's Lives by Visualizing Publicly Shared Location Data](#). DOI: [10.1145/2858036.2858272](#)

*This story is republished courtesy of MIT News ([web.mit.edu/newsoffice/](http://web.mit.edu/newsoffice/)), a popular site that covers news about MIT research, innovation and teaching.*

Provided by Massachusetts Institute of Technology

Citation: From location data alone, even low-tech snoopers can identify Twitter users' homes, workplaces (2016, May 17) retrieved 9 April 2024 from <https://phys.org/news/2016-05-low-tech-snoopers-twitter-users-homes.html>

<p>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.</p>
--