

Sites want your (anonymized) social media data

6 June 2018, by Anthony Sanford



Balancing personal privacy with detailed insights. Credit: Dawn Hudson/Shutterstock.com

Social media sites' responses to the Facebook-Cambridge Analytica scandal and new European privacy regulations have given users [much more control](#) over who can access their data, and [for what purposes](#). To me, as a social media user, these are positive developments: It's scary to think what these platforms could do with the troves of data available about me. But [as a researcher](#), increased restrictions on data sharing worry me.

I am among the many scholars who depend on [data](#) from [social media](#) to gain insights into people's actions. In a rush to protect individuals' privacy, I worry that an unintended casualty could be knowledge about human nature. My most recent work, for example, [analyzes feelings people express on Twitter](#) to explain why the stock market fluctuates so much over the course of a single day. There are applications well beyond finance. Other scholars have studied [mass transit rider satisfaction](#), [emergency alert systems' function during natural disasters](#) and how online interactions [influence people's desire to lead healthy lifestyles](#).

This poses a dilemma – not just for me personally, but for society as a whole. Most people don't want [social media platforms](#) to share or sell their personal information, unless specifically [authorized by the individual user](#). But as members of a collective society, it's useful to understand the social forces at work influencing everyday life and long-term trends. Before the recent crises, Facebook and other companies had already been making it hard for [legitimate researchers to use their data](#), including by [making it more difficult and more expensive](#) to download and access data for analysis. The renewed public pressure for privacy means it's likely to get even tougher.

Using social media data in research

It's definitely alarming to consider the prospect that people or companies might analyze my data and find ways to influence me to make decisions I might not otherwise – or that are even counter to my own best interests. I need think only of the number of times I've seen a TV ad for pizza during a sporting event and ordered a pizza.

That's the point of marketing, of course – but social media is different because the information is about me specifically. And using that information can [affect much more than what food I buy](#), such as whom I vote for. However, as a researcher in finance, I also recognize that the same data can be used to help us understand collective behaviors that are otherwise impossible to explain.

Some of my research, for example, explores short-term trends in stock prices. Financial experts have found that over the long term, a company's stock prices are driven by the firm's future value. Yet over the course of any single day, stock prices can vary widely. Many finance researchers and financial analysts will tell you that these movements are meaningless noise, seemingly random pieces of information about companies influencing investors' perceptions and causing [stock prices](#) to vary

constantly.

But by analyzing social media data, I can actually understand what that noise is, where it comes from and what it means. For instance, what people write on Twitter about the new iPhone will affect Apple's stock price, sometimes within minutes – but even over the course of days. The speed of the effect depends on the importance or prominence of the person sending the tweet, as well as how quickly others – including the media – pick up the message.

Results from my research can help investors fine-tune when and how they enter the market. If, for example, social media users believe that the newest iPhone will not be as good as expected, investors might hold off on their investment in Apple stock. That could free them up to invest in something else with better buzz, in hopes of higher returns.

Anonymizing data

It's true – and concerning – that some presumably unethical people have tried to use social media data for their own benefit. But the data are not the actual problem, and cutting researchers' access to data is not the solution. Doing so would also deprive society of the benefits of [social media analysis](#).

Fortunately, there is a way to resolve this dilemma. Anonymization of data can keep people's individual privacy intact, while giving researchers access to collective data that can yield important insights.

There's even a strong model for how to strike that balance efficiently: the U.S. Census Bureau. For decades, that government agency has collected extremely personal data from households all across the country: ages, employment status, income levels, Social Security numbers and political affiliations. The results it publishes are very rich, but also not traceable to any individual.

It often is technically possible to [reverse anonymity protections on data](#), using multiple pieces of anonymized information to identify the person they all relate to. The Census Bureau takes steps to

prevent this.

For instance, when members of the public access [census data](#), the [Census Bureau restricts information](#) that is likely to identify specific individuals, such as reporting there is just one person in a community with a particularly high- or low-income level.

For researchers the process is somewhat different, but provides significant protections both in law and in practice. Scholars have to pass the Census Bureau's vetting process to make sure they are legitimate, and must undergo training about what they can and cannot do with the data. The penalties for violating the rules include not only being barred from using census data in the future, but also civil fines and even criminal prosecution.

Even then, what researchers get comes without a name or Social Security number. Instead, the Census Bureau uses what it calls "[protected identification keys](#)," a [random number that replaces data](#) that would allow researchers to identify individuals.

Each person's data is labeled with his or her own identification key, allowing researchers to link information of different types. For instance, a researcher wanting to track how long it takes people to complete a college degree could follow individuals' education levels over time, thanks to the identification keys.

Social media platforms could implement a similar anonymization process instead of increasing hurdles – and cost – to access their data. They could assign users identification numbers instead of sharing their real identities, and could agree to government regulations defining who could get access to what data, including real penalties for violating the rules. Then researchers could discover the insights offered by social [media](#) use, just like they do with census data, without threatening people's privacy.

This article was originally published on [The](#)

Provided by The Conversation

APA citation: Sites want your (anonymized) social media data (2018, June 6) retrieved 23 May 2019 from <https://phys.org/news/2018-06-sites-anonymized-social-media.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.