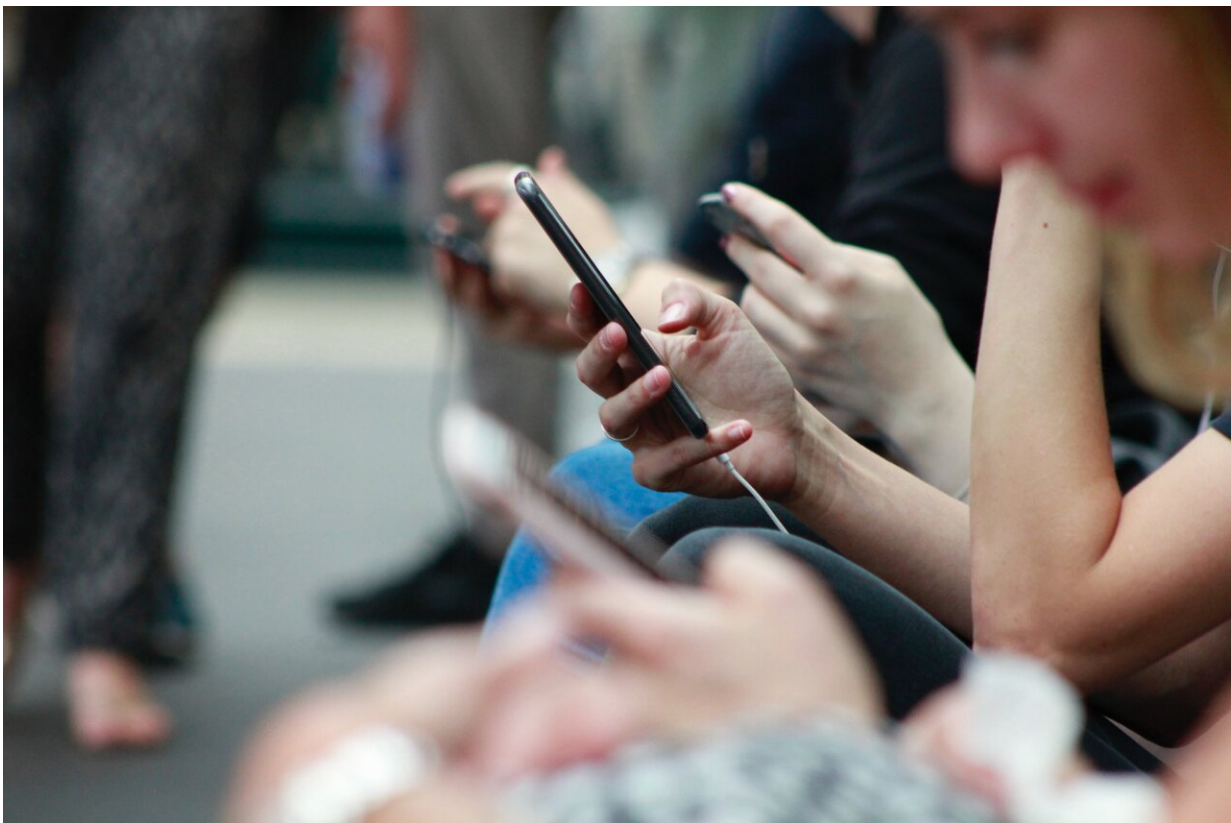# Women and lower-education users more likely to tweet personal information

July 8 2021, by Jessica Hallman

Credit: Unsplash/CC0 Public Domain

When it comes to what users share on Twitter, women and users who never attended college voluntarily disclose more personal information than users from other socioeconomic and demographic

backgrounds—potentially making these populations more susceptible to online privacy threats, according to a recent study led by the Penn State College of Information Sciences and Technology.

Additionally, the researchers unexpectedly found that neither socioeconomic status nor demographics is a significant predictor of the use of account security features such as two-factor login authentication, and that users from all backgrounds actually shared less personal information than they recalled.

"We didn't find a strong correlation between people's stated attitudes and their observed behaviors, which is pretty contradictory to what privacy literature has explained about people's digital inequality and privacy divide," said Jooyoung Lee, doctoral student of information sciences and technology and lead author on the research paper.

In the exploratory study, the researchers set out to understand whether socio-demographic factors impact the usage of login verification, a user's likeliness to share personal information online, and whether topics of self-disclosure vary across socio-demographic groups.

"There is a robust literature on self-disclosure, but purely data-driven approaches typically don't allow us access to users' gender, education, occupation, race and other sensitive information," said Sarah Rajtmajer, assistant professor of information sciences and technology. "At the same time, there is growing concern about the inequitable distribution of privacy risk amongst different socio-demographic groups with respect to online information sharing. The experimental approach taken in this work allowed us a first attempt to bridge the gap."

According to Shomir Wilson, assistant professor of information sciences and technology, the researchers were motivated to expand on past work that indicated that people in lower socio-economic brackets had more

difficulty understanding online privacy controls.

"The original thing we were expecting to see based on the survey methods and prior work actually didn't bear out in that we got negative results on the socio-economic status," said Wilson. "But we got some other results that surprised us and are leading us into next steps."

The Penn State study is novel in that it explores the contents of personal information in self-disclosure along socio-demographic lines. In prior work, only gender and age variables have been primarily explored.

The researchers surveyed 110 active Twitter users and monitored their posting behaviors in more than 6,900 tweets over the course of a month. Then, using statistical analysis methods, they examined the tweets for mentions of topics in 12 categories of self-disclosure—such as marital status, or location—and labeled which of the categories, if any, the tweet fit.

Those categories were then measured against six socio-demographic factors—income, gender, age, education level, race/ethnicity and occupation—to analyze users' login verification settings, quantity of self-disclosure, and self-disclosure by topic. Finally, a post-study survey was sent to participants to collect their recollection of self-disclosure, which the researchers measured against their actual posts.

"A key distinction between our work and prior work was that prior work surveyed people for their attitudes and beliefs," said Wilson. "We took this a step further: we not only gave people surveys, but we followed them on Twitter to see how they were behaving and if their behaviors actually correlated with what they thought they were doing. And we found that people were sharing less than they thought they were sharing."

Added Lee, "People don't always remember what they share on social

media, which could be a really big problem. Reminding people of their sharing behaviors could be a good solution to help them keep track of what kind of data they're sharing publicly."

Rajtmajer added that this is particularly true about the combined information of what they've shared over time, which led the researchers to ask survey participants whether they remembered sharing specific pieces of personal information.

"We know that, most often, the critical worries derive from inferences about an individual made possible by the aggregation of all the various, and often seemingly harmless, details they share," she said. "These inferences can be used to profile, monetize, manipulate and surveil. Already-vulnerable groups in many cases are most at risk."

According to Wilson, there are also scenarios where users don't realize that they are sharing posts containing personal information with an audience that includes their co-workers or the general public. Conversely, there are cases where people might not share enough, not realizing that there are certain pieces of information that their friends and followers might want to know.

"Aligning those two things helps people better understand their public persona and gives people a greater sense of security when they use online social networks," Wilson said. "And that in itself is valuable."

The study unveils that users often can't accurately construct a mental model of their sharing behaviors over a month-long period, which could potentially lead to design updates for social networks to implement features that help users keep track of their sharing behaviors.

"This provides context to how people use these tools, both for the users and for the people creating them," said Wilson.

Eesha Srivatsavaya, an undergraduate data sciences student at the College of IST, was also involved with the project. The team's paper appears in the *July 2021 Proceedings on Privacy Enhancing Technologies.* The work was supported in part by an Accelerator Award from the Center for Social Data Analytics at Penn State.

Provided by Pennsylvania State University