Location data on two apps enough to identify someone, says study

April 13 2016, by Kim Martineau

Time-stamped locations in just two social media apps are enough to link accounts held by the same person and identify him or her, says a new study. Credit: Kim Martineau

Stripping a big data set of names and personal details is no guarantee of privacy. Previous research has shown that individual shoppers, Netflix subscribers and even taxicab riders are identifiable in heaps of supposedly anonymous data.

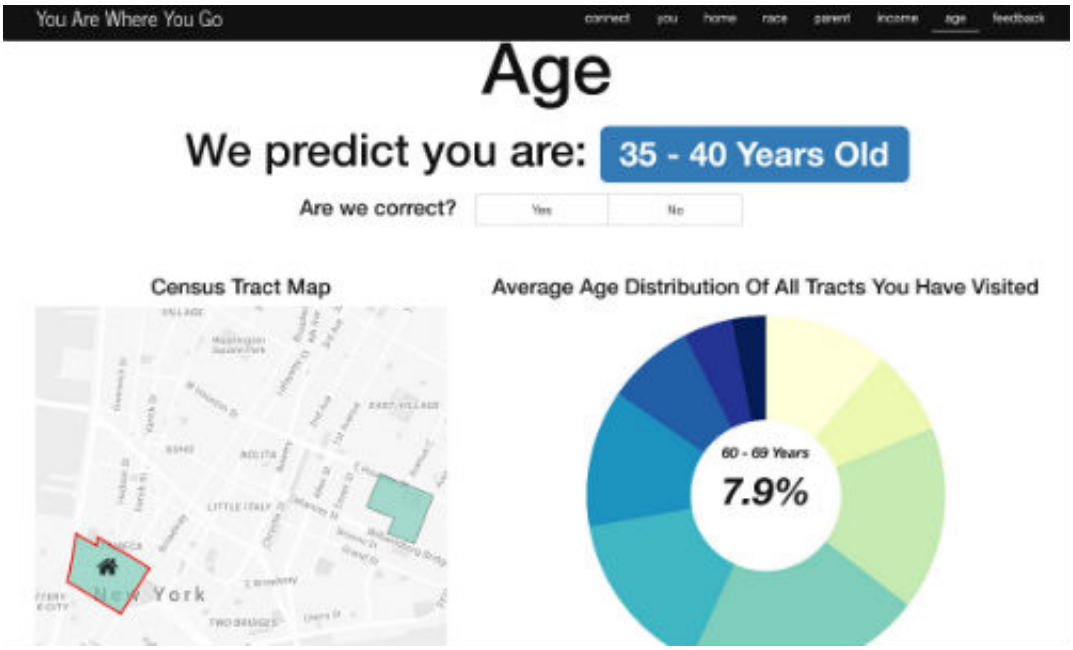
Now, a team of computer science researchers at Columbia University

and Google has identified new privacy concerns by demonstrating that geotagged posts on just two social media apps are enough to link accounts held by the same person. The team will present its results at the World Wide Web conference in Montreal on April 14.

Of the many digital traces we leave in daily life, location metadata may be the most revealing. Our real world movements are so distinctive that most people can be identified from a few data points within a single data set. With as little as four credit card purchases, individual shoppers can be picked out from among millions of other credit card users.

The new study takes these previous findings a step further by showing that individuals can be identified with a high degree of confidence by matching their movements across two data sets. "If you look unique in how you make phone calls, it is possible to connect that to where you've made credit card purchases," said study coauthor Chris Riederer, a graduate student in computer science at Columbia Engineering.

The team developed an algorithm that compares geotagged posts on Twitter with posts on Instagram and Foursquare to link accounts held by the same person. It works by calculating the probability that one person posting at a given time and place could also be posting in a second app, at another time and place. The Columbia team found that the algorithm can also identify shoppers by matching anonymous credit card purchases against logs of mobile phones pinging the nearest cell tower. This method, they found, outperforms other matching algorithms applied to the same data sets.



A related app developed by the researchers lets individuals query their own social media accounts to see what personal information they may be inadvertently leaking. Credit: Chris Riederer

Location tracking is now embedded in phones and many apps precisely because it's so useful. It's what allows you to get directions from here to there, learn that a friend is unexpectedly nearby, or that a store in the neighborhood is offering a promotion. These perks, however, come with large privacy risks that remain poorly understood.

Privacy rights and protections vary from country to country, and have evolved with advances in information technology, from the rise of photography to telephones to new recording technologies and computing. Privacy is considered by some to be necessary for self-development and creativity, and the ability to speak freely and criticize powerful institutions.

"Many people choose not to identify themselves online," said study

author Augustin Chaintreau, a computer science professor at Columbia Engineering and a member of the Data Science Institute. "If I now tell you that your location data makes you recognizable across all of your accounts, how does that change your behavior? This is a question we now have to answer."

Not only is mobility data unique, it may also reveal sensitive information including someone's age, gender and ethnicity, which could be used to discriminate in housing, lending, health care and other areas.

"What this really shows is that simply removing identifying information from large-scale data sets is not sufficient," said Yves-Alexandre de Montjoye, a research scientist at the MIT Media Lab who was not involved in the study. "We need to move to a model of privacy-through-security. Instead of anonymizing data and making it public, there should be technical controls over who gets access to the data, how it is used, and for what purpose."

One problem with the current system is how opaque it is to the people unknowingly leaking their data simply by carrying around a phone. With two Columbia undergraduates, Danny Echikson and Stephanie Huang, Riederer built a related tool, You Are Where You Go, to let individuals audit their social media trail.

With a few clicks, the tool retraces your steps on Twitter, Instagram and Foursquare. A few simple algorithms process this information to make relatively accurate inferences about your age, ethnicity, income, and whether you have kids.

"People are now sharing their location on a growing number of apps, often without realizing it," said Riederer. "Companies no longer have to be very sophisticated to access this data and use it for their own purposes."

Riederer says he hopes to add more social networks and predictions to the tool, and to create ways for users to donate their data to science or give recommendations.

Provided by Columbia University School of Engineering and Applied Science