http://www.boston.com/bostonglobe/ideas/articles/2009/09/20/project_gaydar_an_mit_experiment_raises_new_questions_about_online_privacy/?page=1
The Boston Globe, Johnson, C.Y., Archived column
Accessed 10-7-2009

September 20, 2009

# Project 'Gaydar'

## At MIT, an experiment identifies which students are gay, raising new questions about online privacy.

By Carolyn Y. Johnson

It started as a simple term project for an MIT class on ethics and law on the electronic frontier.

Two students partnered up to take on the latest Internet fad: the online social networks that were exploding into the mainstream. With people signing up in droves to reconnect with classmates and old crushes from high school, and even becoming online "friends" with their family members, the two wondered what the online masses were unknowingly telling the world about themselves. The pair weren't interested in the embarrassing photos or overripe profiles that attract so much consternation from parents and potential employers. Instead, they wondered whether the basic currency of interactions on a social network - the simple act of "friending" someone online - might reveal something a person might rather keep hidden.

Using data from the social network Facebook, they made a striking discovery: just by looking at a person's online friends, they could predict whether the person was gay. They did this with a software program that looked at the gender and sexuality of a person's friends and, using statistical analysis, made a prediction. The two students had no way of checking all of their predictions, but based on their own knowledge outside the Facebook world, their computer program appeared quite accurate for men, they said. People may be effectively "outing" themselves just by the virtual company they keep.

"When they first did it, it was absolutely striking - we said, 'Oh my God - you can actually put some computation behind that,' " said Hal Abelson, a computer science professor at MIT who co-taught the course. "That pulls the rug out from a whole policy and technology perspective that the point is to give you control over your information - because you don't have control over your information."

The work has not been published in a scientific journal, but it provides a provocative warning note about privacy. Discussions of privacy often focus on how to best keep things secret, whether it is making sure online financial transactions are secure from intruders, or telling people to think twice before opening their lives too widely on blogs or online profiles. But this work shows that people may reveal information about themselves in another way, and without knowing they are making it public. Who we are can be revealed by, and even defined by, who our friends are: if all your friends are over 45, you're probably not a teenager; if they all belong to a particular religion, it's a decent bet that you do, too. The ability to connect with other people who have something in common is part of the power of social networks, but also a possible pitfall. If our friends reveal who we are, that challenges a conception of privacy built on the notion that there are things we tell, and things we don't.

"Even if you don't affirmatively post revealing information, simply publishing your friends' list may reveal sensitive information about you, or it may lead people to make assumptions about you that are incorrect," said Kevin Bankston, senior staff attorney for the Electronic Frontier Foundation, a nonprofit digital rights organization in San Francisco. "Certainly if most or many of your friends are of a particular religious or political or sexual category, others may conclude you are part of the same category - even if you haven't said so yourself."

The project, given the name "Gaydar" by the students, Carter Jernigan and Behram Mistree, is part of the fast-moving field of social network analysis, which examines what the connections between people can tell us. The applications run the gamut, from predicting who might be a terrorist to the likelihood a person is happy or fat. The idea of making assumptions about people by looking at their relationships is not new, but the sudden availability of information online means the field's powerful tools can now be applied to just about anyone.

For example, Murat Kantarcioglu, an assistant professor of computer science at the University of Texas at Dallas, found he could make decent predictions about a person's political affiliation. He and a student - who later went to work for Facebook - took 167,000 profiles and 3 million links between people from the Dallas-Fort Worth network. They used three methods to predict a person's political views. One prediction model used only the details in their profiles. Another used only friendship links. And the third combined the two sets of data.

The researchers found that certain traits, such as knowing what groups people belonged to or their favorite music, were quite predictive of political affiliation. But they also found that they did better than a random guess when only using friendship connections. The best results came from combining the two approaches.

Other work, by researchers at the University of Maryland, College Park, analyzed four social networks: Facebook, the photo-sharing website Flickr, an online network for dog owners called Dogster, and BibSonomy, in which people tag bookmarks and publications. Those researchers blinded themselves to the profiles of half the people in each network, and launched a variety of "attacks" on the networks, to see what private information they could glean by simply looking at things like groups people belonged to, and their friendship links.

On each network, at least one attack worked. Researchers could predict where Flickr users lived; Facebook users' gender, a dog's breed, and whether someone was likely to be a spammer on BibSonomy. The authors found that membership in a group gave away a significant amount of information, but also found that predictions using friend links weren't as strong as they expected. "Using friends in classifying people has to be treated with care," computer scientists Lise Getoor and Elena Zheleva wrote.

The idea behind the MIT work, done in 2007, is as old as the adage that birds of a feather flock together. For years, sociologists have known of the "homophily principle" - the tendency for similar people to group together. People of one race tend to have spouses, confidants, and friends of the same race, for example. Jernigan and Mistree downloaded data from the Facebook network, choosing as their sample people who had joined the MIT network and were in the classes 2007-2011 or graduate students. They

were interested in three things people frequently fill in on their social network profile: their gender, a category called "interested in" that they took to denote sexuality, and their friend links.

Using that information, they "trained" their computer program, analyzing the friend links of 1,544 men who said they were straight, 21 who said they were bisexual, and 33 who said they were gay. Gay men had proportionally more gay friends than straight men, giving the computer program a way to infer a person's sexuality based on their friends.

Then they did the same analysis on 947 men who did not report their sexuality. Although the researchers had no way to confirm the analysis with scientific rigor, they used their private knowledge of 10 people in the network who were gay but did not declare it on their Facebook page as a simple check. They found all 10 people were predicted to be gay by the program. The analysis seemed to work in identifying gay men, but the same technique was not as successful with bisexual men or women, or lesbians.

"It's just one example of how information could be inadvertently shared," said Jernigan. "It does highlight risks out there."

The researchers treated their data anonymously, never using names except to validate their predictions during data analysis. The only copy of the data is on an encrypted DVD they gave to a professor, and they said they got the approval of an ethical review board at MIT. The students, who have since graduated, discussed the paper with the Globe, but did not provide a copy of it because they are hoping to have it published in a journal.

Facebook spokesman Simon Axten could not respond to Jernigan and Mistree's analysis, since it is not public, but pointed out that it is something that happens every day.

"In general, it's not too surprising that someone might make inferences about someone else without knowing that person based on who the person's friends are. This isn't specific to Facebook and is entirely possible in the real world as well," Axten wrote in an e-mail. "For example, if I know that someone has certain political views because that person makes them known in some way (say, by putting a bumper sticker on his car), and then I see the person walking out of a movie with friends I don't know, I might assume those friends also have those political views."

Privacy has become a growing and evolving concern as social networks learn how to deal with the fact that they provide a resource that brings people together, but also may endanger privacy in ways they did not anticipate. Social networks like Facebook already give people power over that information, with privacy features that allow people to hide their profiles, and even make their list of friends invisible to outsiders, as well as from select friends.

Because the features and services offered on social networks are new, they also evolve in response to user demand that may not always be anticipated by the company. In 2007, for example, Facebook introduced Beacon, a feature that broadcasted friends' activities - such as buying movie tickets on a specific website - like targeted advertisements. That drew an angry response from users concerned about privacy, and prompted an apologetic blog posting from Facebook cofounder Mark Zuckerberg, along with modifications that meant people could opt out.

Computer scientists are identifying the ways in which anyone from a potential employer to an advertiser might be able to make informed guesses about a person. But there are limits to online privacy, and ultimately, say some experts, people will simply have to weigh the costs and benefits of living online.

"You can do damage to your reputation with social networking data, and other people can do damage to you. I do think that there's been a very fast learning curve - people are quickly learning the dos and don'ts of Internet behavior," said Jason Kaufman, a research fellow at the Berkman Center for Internet and Society at Harvard University who is studying a set of Facebook data. "Potentially everything you ever do on the Internet will live forever. I like to think we'll all learn to give each other a little more slack for our indiscretions and idiosyncrasies."