TIM ELLIS

Why do our social networks know so much more about us than we care to reveal, asks Jim Giles

The oracle of Facebook

I'M A Facebook curmudgeon. The world's most popular social network is all about sharing, but my profile doesn't give much away. The interests field is blank. Ditto for educational history. I have listed neither the sports teams I support nor the political views I hold.

My reticence isn't confined to Facebook – I don't reveal much on Twitter, either. I recently deleted my account for location-sharing service Foursquare. I doubt anyone wants to view my Flickr photos, but they are password protected anyway. And I'm not alone: more than half of people who use social networks try to control the information others see about them, by not filling it in or by adjusting their privacy settings, according to one recent Pew Research survey.

Here's the thing about our antisocial behaviour: it might not matter. Even the little information I post online could be enough for someone to guess where I live. In fact, my sexual orientation, mood and personality type can all be worked out from snippets of online data. Throw in some cellphone records and algorithms have enough information to predict where I will go tomorrow and who I will be with. The data can even be used to gauge whether my lack of sociability is run-of-the-mill grumpiness, or a sign of impending depression.

Though many people might find this loss of privacy disturbing, advocates point to the many benefits we accrue from allowing algorithms to sift through our data. They might have a point. Allowing websites and phone companies to peer into our minds could allow social networks to put us in touch with like-minded people that we would never otherwise befriend, enriching our offline lives. It could even give our phones the ability to warn our loved ones when we need help. How much are we willing to share in order to reap these rewards? And do we still have any choice in the matter?

It is a fact of connected online life that privacy concerns flare up from time to time: Facebook has come under fire for changing its privacy settings on a dime, and search engines have been accused of harvesting too much data. However, if you were vigilant about keeping up with the fine print, you could take some comfort in the fact that your data was only available to designated people.

The new mind-reading algorithms are changing that. They make use of machine learning techniques and network theory, a field at the intersection of computer science and mathematics, whose roots go back to the 18th century. Network theory is the study of the structure and dynamics of networks, ranging from gene regulatory systems to the internet. In particular, it can reveal unseen relationships between a network's nodes. The mathematicians of the 1700s could not have imagined the era of Facebook, or that their theory would be able to reveal so much about individuals. And yet that is exactly what data scientists at Facebook and elsewhere use network theory to do.

Facebook's algorithms, for example, know more about your social life than you post on the site. I have friends I am not linked to online whom the site can nonetheless identify. At the simplest level, the algorithms know that most new friendships close a triangle: if I am friends with Alice, and Alice is friends with Bob, ›

there is a good chance that I am friends with Bob, too. The algorithm refines its decisions with supporting information, such as when Alice became friends with both of us. Once it is satisfied that I know Bob, the site will suggest that we connect. I can choose not to declare a friendship, but the site still knows.

What about other things I choose to keep to myself, such as the university I attended? There is an algorithm for that, too. To tease out the answer, the method relies on a basic human trait: we tend to befriend people that are a similar age and have similar interests to us. If just a fraction of my university friends publish the name of their university and graduation date, the right algorithm can fill in the blanks in my profile with an accuracy of 80 per cent (*Proc. Third ACM Int. Conf. on Web Search and Data Mining*, p 251). A similar algorithm for guessing sexual orientation also boasts an 80 per cent accuracy (*First Monday*, vol 14, no 10).

The more information an algorithm has about the friends, family and colleagues that populate your social network, the more it knows about you. Take your address. If you post it on Facebook, you probably don't make it openly available for everyone to see. I don't share mine at all. But 6 per cent of the site's users do. That probably includes some of your friends, and because we tend to live close to them, the internet's crystal ball can use friends to take a competent stab at where you live. In 2010, Lars Backstrom, a researcher at Facebook, showed that he could locate two-thirds of the site's users to within 40 kilometres by identifying where their friends live (*Proc. 19th Int. Conf. on World Wide Web*, p 61).

Add the location information that those friends make public on Twitter and it often becomes possible to home in on someone's exact location. This February, Adam Sadilek at the University of Rochester in New York showed that if just nine of a person's friends attach GPS tags to their tweets, around half of the time it is possible to pinpoint that person to within a 100-metre radius (*Fifth ACM Int. Conf. on Web Search and Data Mining*, p 723). While neither technique is perfect, both invalidate the idea that not publishing your location is the same as not revealing it.

Getting to know you

Is this necessarily a bad thing? With the ability to pinpoint your location, it might be possible for social networks to predict not just the friends you have now, but tell you who you might become friends with in the future.

Certain places, for example, define our social dynamics: offices are fertile ground for new friendships; airports are not. To quantify this effect, Salvatore Scellato and colleagues at the University of Cambridge ranked locations obtained from around 300,000 people using Gowalla, a now-defunct location-tracking service that allowed cellphone users to announce their presence at restaurants, bars, airports and the like. A low score – indicating that the place was frequented by a small number of people – suggested high friendship-forming potential. By watching for pairs of users who simultaneously visited low-scoring locations, Scellato was able to identify two-thirds of the friendships that users subsequently declared (*Proc. 17th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, p 1046). Though many of these bonds probably existed before these people joined Gowalla, Scellato's method almost certainly predicted some genuinely new friendships.

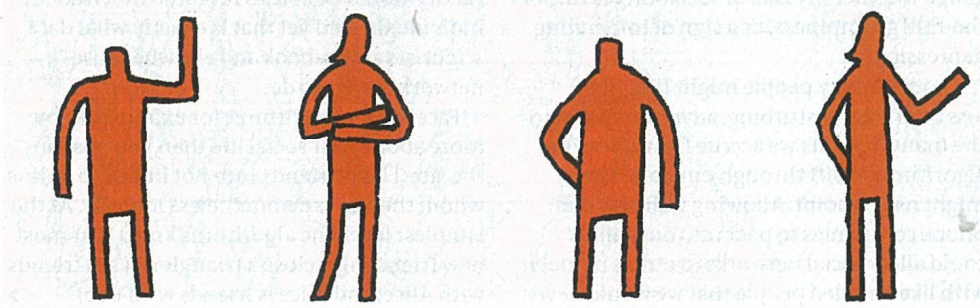
It is not just future friendships that can be deciphered; the algorithms can also provide insights into someone's personality. Mine is certainly not apparent from my Facebook profile, which is populated chiefly by tumbleweed. I don't keep a personal blog. When I do use Twitter, I am conservative both about my language and the topics I cover. And yet the 715 tweets I have sent may be enough to gauge my personality with unexpected precision.

The language we use in tweets – and in many other forms of writing – can provide a surprisingly clear window into our personality. Jennifer Golbeck at the University of Maryland in College Park scored 50 people on the five traits that psychologists use to define personality: extroversion, agreeableness, conscientiousness, neuroticism and openness, assigning each participant a 1-to-5 ranking in these qualities. When she looked at their tweets, she found some unsurprising correlations: agreeable types tended to talk to others more often, as revealed by their frequent use of "you" and "your". Conscientious folk avoided words with negative connotations, such as "kill". But there were also some odd surprises. For some reason, conscientious people used more colons. People who tweeted at length about eating were found to be agreeable; those who peppered their speech with achievement-related terms such as "earning" and "winning" were not. Combining these and other correlations, she was able to predict volunteers' trait scores to within almost 10 per cent. (*Proc. 3rd IEEE Int. Conf. Social Computing*, p 149).

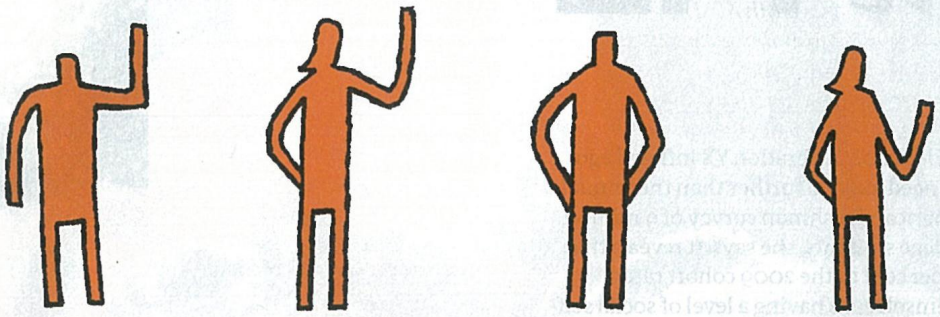
Facebook and Twitter aren't the only technologies that can tap into your social network to figure you out: your cellphone also knows your quirks. Daniel Gatica-Perez, a researcher at the Idiap Research Institute in Martigny, Switzerland, and colleagues spent 18 months tracking 117 volunteers using an app that logged the numbers they called and how often they did so. Then the team looked for correlations between this data and the volunteers' personalities.

Some findings were unsurprising: extroverts made longer calls; introverts preferred to text. When combined, these seemingly obvious correlations proved more powerful than the sum of their parts: the data

"The language you use in tweets can provide a surprisingly clear window into your personality"



"No matter how we feel about privacy, we want to make our own decisions. The new algorithms take the choice away"



could identify whether a person would score above or below average on the five personality traits (*Personal and Ubiquitous Computing*, DOI: 10.1007/s00779-011-0490-1).

Combine enough of this data and you get a crystal ball that could yield some useful insights. You could figure out, for example, whether a friend is feeling depressed. Using a similar phone app, Anmol Madan and his colleagues at the Massachusetts Institute of Technology tracked 70 residents of a US university dormitory. For about two years, the volunteers filled in regular health surveys, reporting back on everything from their mood to the presence of a sniffle. After crunching the numbers, Madan found that depressed students tended to communicate less and spend less time with friends. That is not a surprise: a well-known and perverse effect of depression is that people who suffer from it tend to isolate themselves, eschewing the very help they need. The surprise was that this information could now be determined from phone data alone. If your cellphone knew you were depressed, it could send an automatic alert to a friend or family member – even when you're too low to do it yourself.

Similarly, what if an app could alert your loved ones to an unusually prolonged period of inactivity? Ingrid Burbey, who studies location prediction at Virginia Tech in Blacksburg, thinks such algorithms could monitor the movements of older people and send an alert when they sense a problem.

The predictions of the social oracle could also usher in smaller conveniences for the rest of us. Location prediction could lead to more useful advertising, enabling stores to send discounts to consumers heading their way. Websites and cellphone apps might use personality profiling to tailor interfaces to specific users.

And yet, despite the conveniences they offer, many people consider such algorithms deeply unsettling.

Shifting ground

We all have very different feelings about privacy: some of us over-share, others play their cards close to their chest. Nearly all of us want to remain in charge of our decisions. The new algorithms are taking away the choice. By inferring information that we have chosen not to reveal, they take privacy decisions out of our hands, including our control over some very sensitive data. I am not sure I want my cellphone provider to know I am feeling depressed. After all, how can I be sure that information won't get into the hands of, say, my insurance company?

For now, the algorithms' power is limited by the segregation of our digital data. Facebook guards its data jealously; the company will not send your entire private profile data to Starbucks. Neither will cellphone companies let Starbucks know when you're next likely to pass by. But there is no guarantee that this will remain the status quo.

On occasion, the companies that control our data have been less than transparent about how they share it. In 2010, *The Wall Street Journal* revealed that Facebook was passing the personal details of its users – including those with strict privacy settings – to advertisers. More recently, cellphone companies have come under similar scrutiny. Last year, independent computer programmers discovered cellphone software that, unknown to users, relays data about phone use to providers. In a separate investigation, programmers discovered that iPhones and iPads log users' locations, although Apple insists that the data is not used to track people.

"The tension between utility and privacy is a genuinely hard problem," says Jon Kleinberg, who studies social and information networks at Cornell University in Ithaca, New York. It is not yet clear how to strike the right balance. Individuals, governments and companies are taking different approaches to the problem. Some are creating tools that help people fine-tune their privacy preferences: for example, programmer Tea Vui Huang created software that removes the GPS traces that are automatically added to most smartphone photos. Others have begun to demand better privacy controls: the US government, for example, is a supporter of the Do Not Track tool, which lets users instruct websites not to track them as they move across the web. New algorithms can analyse personal data without exposing private information (*Proc. 41st Ann. ACM Symp. Theory of Computing*, p 169). "That said," Kleinberg admits, "it's going to be hard for anyone to anticipate the new ways in which conclusions can be drawn from your data. After all, these algorithms can often pick up statistical patterns in your behaviour that even you are not aware of."

Whether we like it or not, we have made a deal with these companies. As the saying goes: if you are not paying for it, the product is you. This oft-repeated mantra sets out the internet's social contract. If you want to use the free services out there, you must accept the conditions – in this case, the ability of companies to track your movements, activities, relationships and perhaps, soon, your emotional state.

We are all free, of course, to simply delete our accounts, switch off our smartphones and become Luddites. But despite my grumbles, I like knowing what my friends are up to on social networks. And if I ever want to meet up with them again I need my cellphone. I just wish I could use these convenient technologies without granting marketers a window into my mind. ■

Jim Giles is a correspondent for *New Scientist* in an undisclosed location