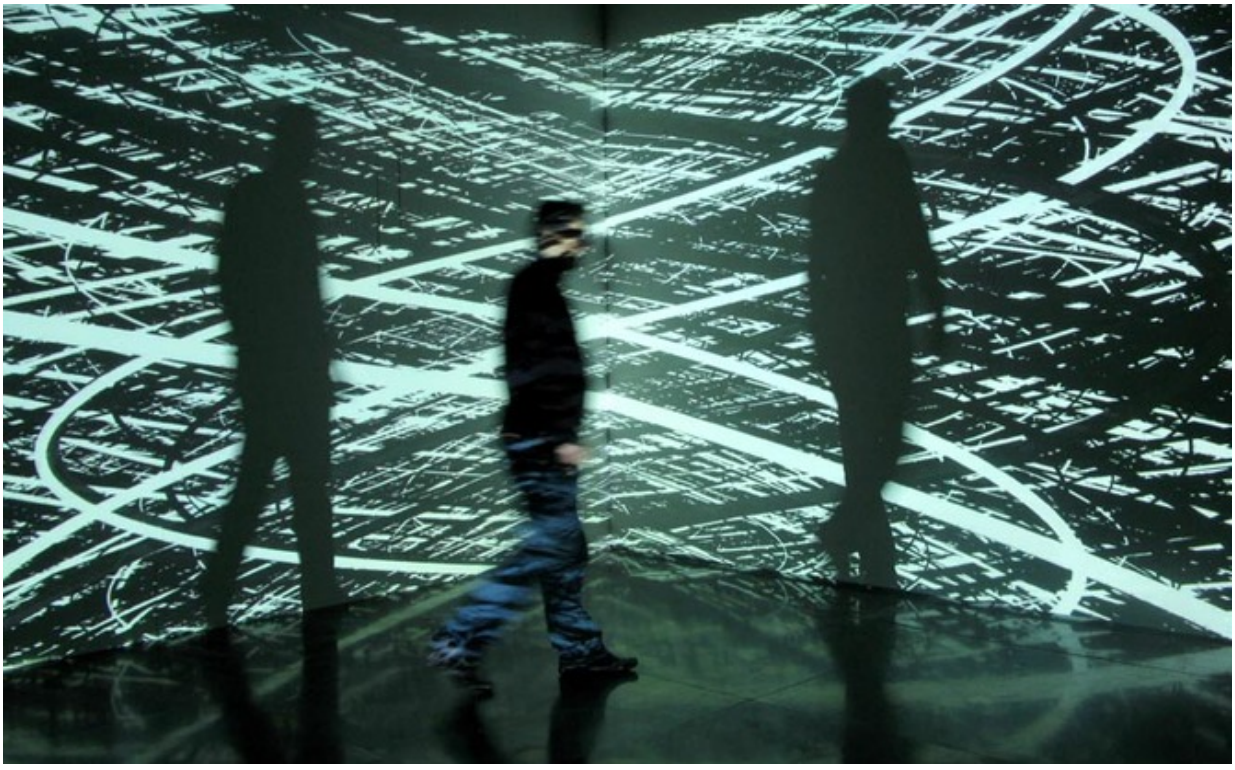
It's time to shine a light on the unseen algorithms that power 'Big Brother'

January 13 2016, by Philip Garnett, University Of York



We know what we look like, but how do algorithms see us? Credit: Cracksinthestreet

Society seems set on a course to a point where our lives are subject to the scrutiny of computer algorithms. The data we generate is pored over and analysed, whether by governments for national security or

companies for profit, and this is unlikely to change – the power and appeal of data analysis, once found, will not be given up easily.

But in truth I wonder whether I'm concerned more that our data is being collected or by the fact that we know nothing about the algorithms that pronounce judgement upon us.

The level of detail about our lives and habits that can be unpicked from the data we leave behind has been discussed before, and is getting a fresh airing as part of the debate around the UK draft Investigatory Powers Bill. We know at least something about what data is collected and how long it is stored for, some of which is governed by UK and European law.

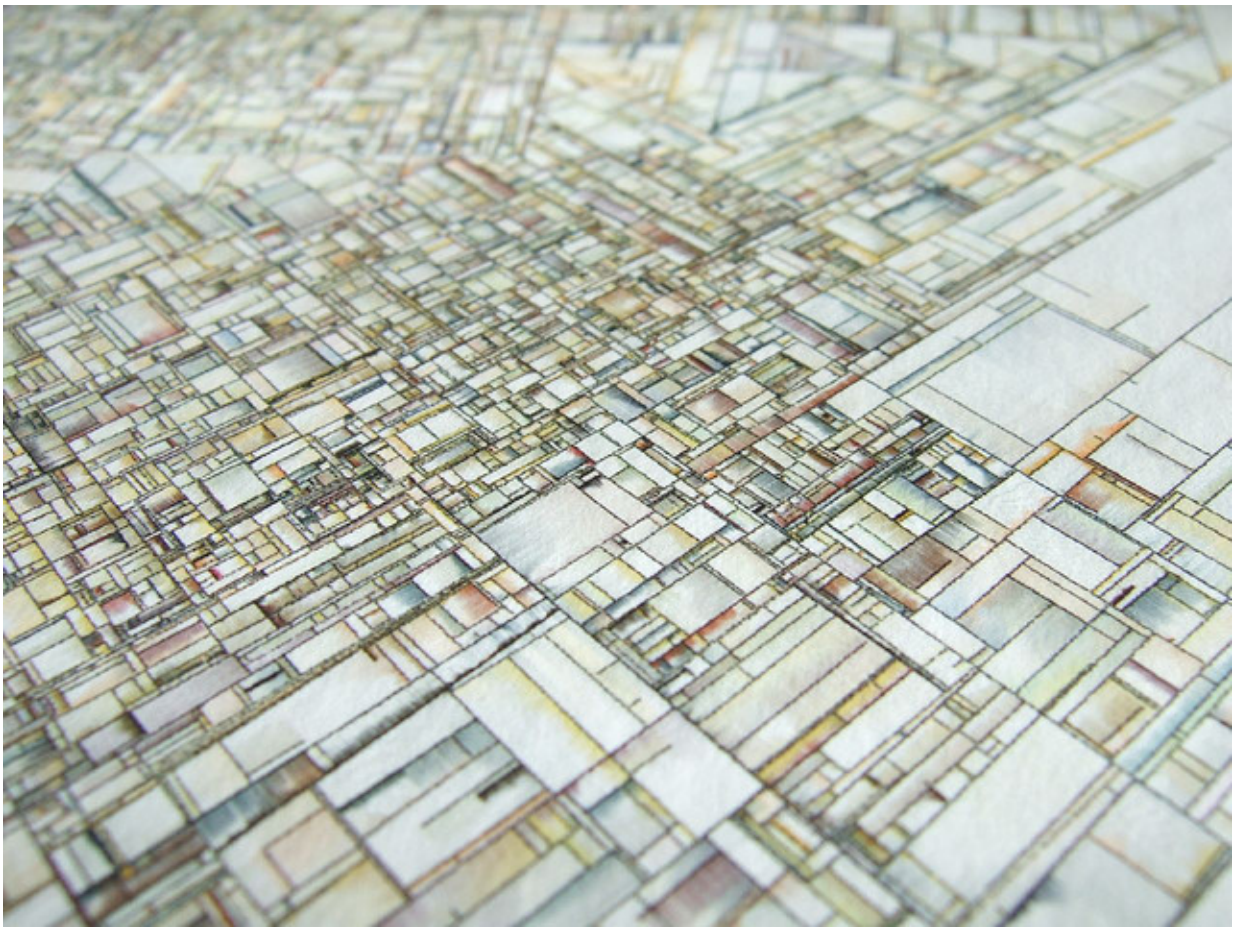
In the text of the draft bill, for example, we know that the UK government will "only" demand (unwarranted) access to communications metadata, the headers and subjects of emails, and phone call records. But we also know just how revealing metadata alone can be: have a look at the MIT Media Lab's Immersion project for a powerful example of just how much detail can be ascertained from it. It's certainly not at all comparable to an itemised phone bill, as claimed.

So for better or worse we, the public, have some clue as to what's being recorded. But we have absolutely no idea what analytical tools and techniques are being applied to this data – and the significance of this should not be underestimated.

What crunches the numbers?

We can make educated guesses. National security agencies probably use our metadata to generate social networks between people and places, among other things, linking us together. These relationship networks will then be analysed to determine if we are a person of interest, determined

by how you compare to other persons of interest, and how you connect to existing persons of interest or those related to them.



Sorting by algorithms puts us in boxes. How do we know they’re the correct ones? generated, CC BY

Researchers who use these techniques understand their limitations, and that the algorithms that power them may contain errors or underlying assumptions that have a profound effect on their output. In this case, that may mean whether you're labelled a terrorist or not, or whether you qualify for a loan or mortgage.

It's also not exactly clear where in the fuzzy border areas the existence of relationship is defined. Does simply visiting the same website as a terrorist imply shared values, or riding the same bus route every day suggest you regularly converse with terrorists? It is quite possible to visit sites frequented by known terrorists for many legitimate reasons. If you get your news from the same websites as terrorists are you more likely to be a terrorist? Discrimination and bias can be introduced at the point of data collection, and then again when decisions are made about how to analyse that data. Algorithms can discriminate, too.

Blurred boundaries

The possibility that algorithms introduce undesirable bias is a very real one. For example, those used by the security services are trained on datasets of known terrorists and known non-terrorists. Does this mean that, as most known terrorists are males aged 20-30, you're more likely to be classified as a terrorist for merely being male and aged roughly 20-30, regardless of your other attributes?. If so, does this have a significant effect on how the data is used?

The problem stems from the fact that I and other academic researchers using complex network analysis, machine learning, pattern matching, or artificial intelligence techniques have our use of those techniques publically peer reviewed to determine the strength of the techniques and the validity of the conclusions; government security services and private sector organisations do not. We have no idea of the quality of their methods and how they deploy them. Is there a solution to this?

Those from another field of security, cryptography, learned long ago that the best way to improve the quality, and therefore security, of its algorithms was to make them public. Cryptographic implementations and ciphers are published, and researchers encouraged to try to find errors or flaws, in doing so improving security for all who use them.

Additionally, any implementation of closed-source (non-public) cryptogaphic algorithms is generally regarded with suspicion. If they are to pronounce life-changing judgements upon us – whether we are labelled as terrorists or financially unworthy – the same model should be applied to security algorithms.

An argument against such a move is that open and transparent algorithms might lead terrorists to modify their real-world behaviour in order to avoid being detected. This would mean changing things like their interactions, associations, browsing habits, and potentially movements. But this, if the algorithms are working properly, would mean they essentially cease to act like terrorists. If our future security, freedom, and safety are going to be dependant on these algorithms, we must be assured exactly how – and that – they work.

This article was originally published on The Conversation. *Read the* original article.

Source: The Conversation

Citation: It's time to shine a light on the unseen algorithms that power 'Big Brother' (2016, January 13) retrieved 23 June 2024 from https://phys.org/news/2016-01-unseen-algorithms-power-big-brother.html