

Studying social media can give us insight into human behaviour—it can also give us nonsense

July 12 2021, by Jason Burton, Nicole Cruz and Ulrike Hahn



Credit: AI-generated image ([disclaimer](#))

Since the early days of social media, there has been [excitement](#) about how data traces left behind by users can be exploited for the study of human behaviour. Nowadays, researchers who were once restricted to surveys or experiments in laboratory settings have access to huge

amounts of "real-world" data from social media.

The research opportunities enabled by [social media](#) data are undeniable. However, researchers often analyse this data with tools that were not designed to manage the kind of large, noisy observational sets of data you find on social media.

We explored problems that researchers might encounter due to this mismatch between data and methods.

What we [found](#) is that the methods and statistics commonly used to provide evidence for seemingly significant scientific findings can also seem to support nonsensical claims.

Absurd science

The motivation for our paper comes from a series of research studies that deliberately present absurd [scientific results](#).

[One brain imaging study](#) appeared to show the neural activity of a dead salmon tasked with identifying emotions in photos. An [analysis of longitudinal statistics from public health records](#) suggested that acne, height, and headaches are contagious. And an [analysis of human decision-making](#) seemingly indicated people can accurately judge the population size of different cities by ranking them in alphabetical order.

Why would a researcher go out of their way to explore such ridiculous ideas? The value of these studies is not in presenting a new substantive finding. No serious researcher would argue, for example, that a dead salmon has a perspective on emotions in photos.

Rather, the nonsensical results highlight problems with the methods used to achieve them. Our research explores whether the same problems can

afflict studies that use data from social media. And we discovered that indeed they do.

Positive and negative results

When a researcher seeks to address a research question, the method they use should be able to do two things:

- reveal an effect when there is a meaningful effect
- show no effect when there is no meaningful effect.

For example, imagine you have chronic back pain and you take a [medical test](#) to find its cause. The test identifies a misaligned disc in your spine. This finding might be important and inform a treatment plan.

However, if you then discover the same test identifies this misaligned disc in a large proportion of the population who do not have [chronic back pain](#), the finding becomes far less informative for you.

The fact the test fails to identify a relevant, distinguishing feature of negative cases (no back pain) from positive cases (back pain) does not mean the misaligned disc in your spine is non-existent. This part of the finding is as "real" as any finding. Yet the failure means the result is not useful: "evidence" that is as likely to be found when there is a meaningful effect (in this case, back pain) as when there is none is simply not diagnostic, and, as result, such evidence is uninformative.

XYZ contagion

Using the same rationale, we evaluated commonly used methods for analysing [social media data](#)—called "null hypothesis significance testing" and "correlational statistics"—by asking an absurd research question.

Past and current studies have tried to identify what factors influence Twitter users' decisions to retweet other tweets. This is interesting both as a window into human thought and because resharing posts is a key mechanism by which messages are amplified or spread on social media.

So we decided to analyse Twitter data using the above standard methods to see whether a nonsensical effect we call "XYZ contagion" influences retweets. Specifically, we asked, "Does the number of Xs, Ys, and Zs in a tweet increase the probability of it being spread?"

Upon analysing six datasets containing hundreds of thousands of tweets, the "answer" we found was yes. For example, in a dataset of 172,697 tweets about COVID-19, the presence of an X, Y, or Z in a tweet appeared to increase the message's reach by a factor of 8%.

Needless to say, we do not believe the presence of Xs, Ys, and Zs is a central factor in whether people choose to retweet a message on Twitter.

However, like the medical test for diagnosing back pain, our finding shows that sometimes, methods for social media data analysis can "reveal" effects where there should be none. This raises questions about how meaningful and informative results obtained by applying current social science methods to social media data really are.

As researchers continue to analyse social media data and identify factors that shape the evolution of public opinion, hijack our attention, or otherwise explain our behaviour, we should think critically about the methods underlying such findings and reconsider what we can learn from them.

What is a 'meaningful' finding?

The issues raised in our paper are not new, and there are indeed many

[research practices](#) that have been developed to ensure results are meaningful and robust.

For example, researchers are encouraged to pre-register their hypotheses and analysis plans before starting a study to prevent a kind of data cherry-picking called "[p-hacking](#)". Another helpful practice is to check whether results are stable after removing outliers and controlling for [covariates](#). Also important are [replication studies](#), which assess whether the results obtained in an experiment can be found again when the experiment is repeated under similar conditions.

These practices are important, but they alone are not sufficient to deal with the problem we identify. While developing standardised research practices is needed, the research community must first think critically about what makes a finding in social [media](#) data meaningful.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Studying social media can give us insight into human behaviour—it can also give us nonsense (2021, July 12) retrieved 2 May 2024 from <https://phys.org/news/2021-07-social-media-insight-human-behaviourit.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--