

# Can an Algorithm be Wrong?

How do we know if we are where it's at? **Tarleton Gillespie** explores the controversy over Twitter Trends and the algorithmic 'censorship' of #occupywallstreet.



**THROUGHOUT** the Occupy Wall Street protests, participants and supporters used Twitter (among other tools) to coordinate, debate, and publicize their efforts. But amidst the enthusiasm a concern surfaced: even as the protests were gaining strength and media coverage, and talk of the movement on Twitter was surging, the term was not “Trending.” A simple list of ten terms provided by Twitter on their homepage, Twitter Trends digests the 250 million tweets sent every day and indexes the most vigorously discussed terms at that moment, either globally or for a user’s chosen country or city. Yet, even in the cities where protests were happening, including New York, when tweets using the term #occupywallstreet seem to spike, the term did not Trend. Some suggested that Twitter was deliberately dropping the term from its list, and in doing so, preventing it from reaching a wider audience.

The charge of censorship is a revealing one. It suggests, first, that many are deeply invested in the Twitter network as a political tool, and that some worry that Twitter’s interests might be aligned with the financial and political status quo they hope

to challenge. But it reveals something else about the importance and the opacity of the algorithm that drives the identification of Trends. To suggest that the best or only explanation of #occupywallstreet’s absence is that Twitter “censored” it implies that Trends is otherwise an accurate barometer of the public discussion. For some, this glitch could only mean deliberate human intervention into what should be a smoothly-running machine.

The workings of these algorithms are political, an important terrain upon which political battles about visibility are being fought (Grimmelmann 2009). Much like taking over the privately owned Zuccotti Park in Manhattan in order to stage a public protest, more and more of our online public discourse is taking place on private communication platforms like Twitter. These providers offer complex algorithms to manage, curate, and organize these massive networks. But there is a tension between what we understand these algorithms to be, what we need them to be, and what they in fact are. We do not have a sufficient vocabulary for assessing the intervention of these algorithms. We’re not adept at appreciating what it takes to design a tool like Trends – one that appears to effortlessly identify what’s going on, yet also makes distinct and motivated choices. We don’t have a language for the unexpected associations algorithms make, beyond the intention (or even comprehension) of their designers (Ananny 2011). Most importantly, we have not fully recognized how these algorithms attempt to produce representations of the

wants or concerns of the public, and as such, run into the classic problem of political representation: who claims to know the mind of the public, and how do they claim to know it?

**THE CONTROVERSY ABOUT TWITTER TRENDS** and #occupywallstreet may be, by itself, a tempest in a teacup. But even on its face, censorship was a difficult charge to dismiss. Of course, some may have made or repeated this charge as a casual observation, as a gesture of belonging, as an expression of frustration about their political movement not being heard, or as a deep-seated anxiety about the nefarious intentions of corporations. But for those who leveled this critique with more care and conviction, the first question was, what exactly does Twitter measure when it identifies these Trending terms?

Twitter has repeatedly stated that their Trends algorithm is not a simple measure of volume (i.e. the most used terms), though the minimalist presentation of the list may suggest otherwise to some users. Some of the comments charging censorship, whether of #occupywallstreet or Justin Bieber, suggest that this list is often taken as a straightforward and unproblematic measure of pop-

standing into their everyday use of Twitter.

Until 2010, Twitter had not provided much, or any, explanation of its algorithm. What Trends represented was offered as self-evident. When similar charges of censorship were raised around #wikileaks, Twitter responded on their blog,<sup>1</sup> and there laid out some general criteria. (These criteria were corroborated and further explained by a Twitter engineer, in the comment thread of one of the more widely-read critiques.<sup>2</sup>)

Twitter explains that Trends is designed to identify topics that are enjoying a surge, not just rising above the normal chatter, but doing so in a particular way. Part of the evaluation includes: Is the use of the term spiking, i.e. accelerating rapidly, or is its growth more gradual? Are the users densely interconnected into a single cluster, or does the term span multiple clusters? Are the tweets unique content, or mostly retweets of the same post? Is this the first time the term has Trended? (If not, the threshold to Trend again is higher.) So this list, though automatically calculated in real time, is also the result of the careful implementation of Twitter's judgments as to what should count as a "trend."

## One substantial revision of the Trends algorithm occurred in May 2010 when Twitter announced it was removing Justin Bieber from the Trending Topics list.

ularity. Though a few critics went to great lengths to reverse engineer the Trends results, to uncover the criteria that enliven them (Lotan 2011), most users certainly vary in their comprehension of what Trends measures and how, and may not always incorporate that under-

Of course, these are just the publicized descriptions of what Trends looks for, and they do not offer a definitive (or satisfying) answer to critics. Trends measures something both more precise and more obscure. There are likely more—and more specific—criteria than those

described in the blog, and these criteria undoubtedly change over time. For instance, one substantial revision occurred in May 2010 when Twitter announced it was removing Justin Bieber from the Trending Topics list.<sup>3</sup> Further, their explanation does not say enough: for instance, how something like "clusters" are defined and measured remains opaque.

Twitter may not be able to say much more about how Trends works. Revealing the "secret sauce" of their algorithm in greater detail risks helping those who would game the system. Everyone from spammers to marketers to activists to 4chan tricksters to narcissists might want to optimize their tweets and hashtags so as to Trend. This opacity makes the Trends results, and their criteria, deeply and fundamentally open to interpretation and suspicion.

The Trends algorithm and the data it evaluates are also the property of Twitter. Sites like Trendistic<sup>4</sup> can use the Twitter API and the corpus of public tweets to conduct more exhaustive analyses of the volume and velocity of term. But they cannot access private tweets, and they cannot know or take into account what counts as a cluster of users. Despite what Twitter is will-

ing to make known, any effort to discover the Trends criteria can only amount to sophisticated guesswork.

Trends is also part and parcel of Twitter's financial aspirations as a for-profit business. Whether or not Trends is an accurate or a deeply flawed measure of vital topics of discussion, it is

also a means to entice and retain users. This does not necessarily mean that it must squelch volatile topics like #occupywallstreet. Twitter has trumpeted its role in the democratic uprisings of recent years, pitching itself as a vital tool for political foment. If Trends is meant to high-

1. <http://blog.twitter.com/2010/12/to-trend-or-not-to-trend.html>

2. <http://studentactivism.net/2010/12/05/wikileaks-twitter-3/#comment-11619>

3. <http://mashable.com/2010/05/14/twitter-improves-trending-topic-algorithm-bye-bye-bieber/>

4. <http://trendistic.indextank.com/>

light terms that will support Twitter's self-proclaimed relevance, #occupywallstreet seems to fit the bill. But this political vitality must be balanced against the interests of other users, of advertisers, of regulators (Gillespie, 2010). Most of all, it is not as if these competing interests can be simply weighed and settled. This same algorithm must serve the desires of its users, all the while drawing new ones. The degree to which it can do all that is the most pressing criteria for Twitter.

As such, Trends is both an index of what is said and an advertisement to read further. Trends promises to bring new readers to a topic; this accounts for much of why the Occupy activists care if and when it appears. The act of measuring the phenomena, then, also changes it: Trends is both feedback and feedback loop, because something that does Trend may be discussed further.

Trends provides an algorithmic gloss of our aggregate social data practices that can always be read/misread as censorship. If #occupywallstreet is not trending, that could mean any of the following: (a) it is being deliberately censored (b) it is actually less popular than one might think (c) it is very popular but consistently so, not a spike (d) it is popular and spiking, but not in a way the algorithm is designed to measure (e) it is popular and spiking, but not as much as some pop culture phenomena that has crowded it off the list (f) it is popular and important, but not as popular as the pop culture phenomena that have been strategically gamed onto the list (g) it has not Trended because it has not Trended, thereby not enjoying the amplification Trends itself offers. However, we are unable to know for certain. Not only are the criteria opaque and the archive proprietary, we also have no benchmark against which to compare the results. Trends measures activity on Twitter, and it is only implicitly indicative of broader public concerns.

**TWITTER TRENDS IS JUST ONE OF MANY** information practices that claim to know or represent the will of the people: public opinion polls, census surveys, man-on-the-street interviews, voting mechanisms. Each employs a specific technique to assess public opinion or activity, in order to make the public will legible. With Trends, Twitter is making a claim: that surging terms, measured in a specific way, are indicative of topics of the most interest and import. But this claim is, for Twitter, caught between competing desires: reporting back what Twitter users care about most versus drawing new users into new conversations. Users too are caught between competing desires: wanting to know something accurate about the public beyond them, and wanting to be rendered visible as a part of that public. With both of these tensions at work, the politics of the artifact, i.e. the workings of the Trends algorithm, become just one piece of a greater puzzle: the politics of representation.

What's different here is that Trends promises a mathematical and an exhaustive analysis of what is being talked about, while presenting it as automatically generated and self-evident facts about the discussion. It claims to know the public through an algorithmic assessment of their complete traces, which is different than the professional judgment of a newspaper editor, the zeitgeist insight of the fashion trendspotter, or the statistical approximations of a census taker. Further, these indices are rendered in an instant and built immediately back into the service itself.

Twitter Trends is only one such tool. Search engines, while promising to provide a logical set of results in response to a query, are in fact algorithms designed to take a range of criteria into account so as to serve up results that satisfy not just the user, but the aims of the provider, their understanding of relevance or newsworthiness or public import, and the particular demands

of their business model (Granka 2010). When users of Apple's Siri iPhone tool begin to speculate that its cool, measured voice is withholding information about abortion clinics, or worse, sending users towards alternatives preferred by conservatives, they are in fact questioning the algorithmic product of the various search mechanisms that Siri consults.<sup>5</sup>

Beyond search, we are surrounded by algorithmic tools that offer to help us navigate online platforms and social networks, based not on what we want, but on what all of their users do. When Facebook, YouTube, or Digg offer to mathematically and in real time report what is "most popular" or "liked" or "most viewed" or "best selling" or "most commented" or "highest rated," they are curating a list whose legitimacy is built on the promise that it has not been curated, that it is the product of aggregate user activity itself. When Amazon recommends a book based on matching your purchases to those of its other customers, or Demand Media commissions news based on aggregate search queries (Anderson 2011), their accuracy and relevance depend on the promise of an algorithmic calculation paired with the massive, even exhaustive, corpus of the traces we all leave.

We might, then, pursue the question of the algorithm's politics further. The Trends algorithm does have criteria built in: criteria that help produce the particular Trends results we see, criteria that are more complex and opaque than some users take them to be, criteria that could have produced the absence of the term #occupywallstreet that critics noted. But further, the criteria that animate the Trends algorithm also presume a shape and character to the public they intend to measure, and in doing so, help to construct publics in that image.

A term that has trended before has a higher threshold before it can trend again. The implication is that the algorithm prefers novelty

5. <http://www.rawstory.com/rs/2011/11/29/10-things-the-iphone-siri-will-help-you-get-instead-of-an-abortion/>

in public discourse over phenomena with a longer shelf-life. This is a longstanding critique of broadcast journalism, reappearing in social media. Perhaps we could again make the case that this choice fosters a public more attuned to the “new” than to the discussion of persistent problems, to viral memes more than to slow-building political movements. Trends also measures the velocity of a term, but within Twitter’s network; it does not compare this surge of interest with matching attention on Facebook, say, or on the blogosphere, or in traditional news coverage. So this public is understood to be platform specific, though we know that users participate in and manage overlapping networks of information and people. With a different commitment to understanding public discourse, one might design an algorithm that gives greater recognition to a topic surging across multiple platforms than one that only surges inside a single network. When Twitter restricts its attention to Twitter, though it is plain why they might want to do so, they put their finger on the scale of a debate about how political discourse does and should function online.

But here’s a harder problem: Twitter takes into account whether a term is circulating within a pre-existing cluster of users (users who are already interconnected, following each other, regionally co-located, etc), or spans clusters. Trends presumes that topics are more important if they exceed clusters. The choice of how to know a

trend matters. Twitter could have designed Trends to weigh heavily a term that does not span clusters of users but instead powerfully ignites a single cluster of users. This kind of “trend” might spotlight issues and concerns discussed intensely by a small but engaged community. It certainly would have regarded #occupywallstreet more highly. Putting such terms in front of all Twitter users via the Trends list would intervene not by showing users what lots of people are talking about, but about what some are talking about hotly together. It might add issues to the public docket rather than rehearsing them. Choosing instead to value terms that span clusters is a choice, and a political one—more Habermas (1989) than Mouffe (2000) in its implicit theory of political dialogue, for example.

These algorithms produce not barometric readings but hieroglyphs. At once so clear and so opaque, they beg to be read as reliable measures of the public mind, as signs of “us.” But the shape of the “us” on offer is by no means transparent. Social media tools like Twitter may be adept at mapping networks of people, if only because they provide the substrate within which these networks form and interact. Even if they cannot as easily capture the human networks that extend beyond their own services, they certainly can claim to have scrutinized the part that is rendered on and by their system. However, though they aspire to with algorithmic tools like Trends, they

may not be as adept at identifying or forging the publics that emerge from those networks. □

---

**TARLETON GILLESPIE** is an associate professor in the Department of Communication and the Department of Information Science at Cornell University.

#### WORKS CITED

- Ananny, Mike. 2011. “The Curious Connection Between Apps for Gay Men and Sex Offenders.” *The Atlantic*, April 14: <http://www.theatlantic.com/technology/archive/2011/04/the-curious-connection-between-apps-for-gay-men-and-sex-offenders/237340/>
- Gillespie, Tarleton. 2010 “The Politics of ‘Platforms.’” *New Media & Society* 12(3): 347-364: <http://hdl.handle.net/1813/12774>
- Granka, Laura. 2010. “The Politics of Search: A Decade Retrospective.” *The Information Society* 26(5), 364-374.
- Grimmelmann, James. 2009 “The Google Dilemma” *New York Law School Law Review* 53(4): [http://works.bepress.com/james\\_grimmelmann/19](http://works.bepress.com/james_grimmelmann/19)
- Habermas, Jürgen. 1989. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*, trans. Thomas Burger with Frederick Lawrence. Cambridge, MA: MIT Press.
- Lotan, Gilad. 2011. “Data Reveals that Occupying Twitter Trending Topics is Harder Than it Looks.” *Social Flow*, October 12: <http://blog.socialflow.com/post/7120244374/data-reveals-that-occupying-twitter-trending-topics-is-harder-than-it-looks>
- Mouffe, Chantal. 2000. *The Democratic Paradox*. London: Verso.

**NEXT PAGE:** Over 5,000 tweets using the #needsoftheoccupiers hashtag (used by occupations to list current needs) were collected from October to December of last year, then geo-located and sorted into top unique needs per occupation site. Tweeted needs included books for New York, garden supplies for San Francisco, Kool-Aid and Crystal Light in Orlando, and in Seattle, after police used pepper spray on protestors, Maalox, which can help neutralize and relieve pepper spray symptoms. Weather patterns across regional areas also affected various needs across the country. Data collection and illustration by J.R. Baldwin.