

Algorithmic futures

The life and death of Google Flu Trends

Vincent Duclos

Abstract

In the last few years, tracking systems that harvest web data to identify trends, calculate predictions, and warn about potential epidemic outbreaks have proliferated. These systems integrate crowdsourced data and digital traces, collecting information from a variety of online sources, and they promise to change the way governments, institutions, and individuals understand and respond to health concerns. This article examines some of the conceptual and practical challenges raised by the online algorithmic tracking of disease by focusing on the case of Google Flu Trends (GFT). Launched in 2008, GFT was Google's flagship syndromic surveillance system, specializing in 'real-time' tracking of outbreaks of influenza. GFT mined massive amounts of data about online search behavior to extract patterns and anticipate the future of viral activity. But it did a poor job, and Google shut the system down in 2015. This paper focuses on GFT's shortcomings, which were particularly severe during flu epidemics, when GFT struggled to make sense of the unexpected surges in the number of search queries. I suggest two reasons for GFT's difficulties. First, it failed to keep track of the dynamics of contagion, at once biological and digital, as it affected what I call here the 'googling crowds'. Search behavior during epidemics in part stems from a sort of viral anxiety not easily amenable to algorithmic anticipation, to the extent that the algorithm's predictive capacity remains dependent on past data and patterns. Second, I suggest that GFT's troubles were the result of how it collected data and performed what I call 'epidemic reality'. GFT's data became severed from the processes Google aimed to track, and the data took on a life of their own: a trackable life, in which there was little flu left. The story of GFT, I suggest, offers insight into contemporary tensions between the indomitable intensity of collective life and stubborn attempts at its algorithmic formalization.

Keywords

global health, surveillance, online tracking, algorithms

On 20 August 2015, Google quietly shut down Google Flu Trends (GFT), its flagship algorithmic syndromic surveillance system, which was designed for the ‘real-time’ tracking of outbreaks of influenza. Over the preceding year or so, GFT had become the object of mounting criticism: on a number of occasions, it had heavily overestimated actual influenza activity.¹ Looking at GFT’s troubles, I ask what they can teach us about the conceptual and practical challenges raised by the recent proliferation of digital systems aimed at accounting for and protecting against uncertainty. I explore the workings, effects of, and issues raised by online tracking, specifically in relation to the emerging field of digital epidemiology.

Recent advances in algorithmic calculation, big data analytics, and artificial intelligence promise to change the way governments, institutions, and individuals understand and respond to health concerns. This is particularly the case for infectious disease surveillance. ‘Big data’ is expected to reorganize public health surveillance and to help manage global health risks of all sorts (National Academies of Sciences and Medicine 2016). The last few years have witnessed a proliferation of disease-tracking systems that harvest web data to identify trends, calculate predictions, and warn about potential epidemic outbreaks (Brownstein, Freifeld, and Madoff 2009; Schmidt 2012). Online disease-surveillance systems integrate data and digital traces, collecting information from a variety of online sources, including search queries, Twitter and Facebook feeds, online news reports, and content crowdsourced from user communities. Aggregated data are expected to be a particularly alert sentinel, whose vigilance and sensing of danger ‘can aid in preparation for an uncertain, but potentially catastrophic future’ (Keck and Lakoff 2013, 2). The tracking of online activity has become an important technique in the government of pandemic threat on a global scale (Feldman and Ticktin 2010).

GFT has been the most significant attempt by a giant data-mining corporation to transform global health. An analysis of GFT thus offers an entry point to examine how big data analytics, specifically algorithmic detection and related data-mining techniques, may intervene in population health on a global scale. GFT can be situated within a wide spectrum of analytical systems aimed at decrypting and making transparent biological and social life. These systems, and big data analytics in particular, seem to fulfill the dream of ‘a world where there is no

1 In this article, ‘flu’ and ‘influenza’ are used interchangeably.

“unknown” left to discover’ (Halpern 2015, 12). The kind of tracking performed by GFT translates assumed realities into numbers and trends (Rottenburg and Merry 2015, 2). It turns an increasingly networked world into a searchable, mappable space (Munster 2013). In this article, I understand algorithmic tracking as an ordering process aimed at narrowing an ‘unaccountably vast array of possibles’ (Peters 2015, 333). Tracking seeks to domesticate an apparently incommensurable reality marked, in the case at hand, by the threatening elusiveness of emergent flu activity and, indeed, of a coming epidemic (Caduff 2015).

GFT mined massive amounts of past data (about online search behavior and doctor visits) to extract patterns that could be used to predict future viral activity. Its predictions, it turned out, were at times rather poor. This paper focuses on GFT’s shortcomings, which were particularly severe during epidemics, in comparison to seasonal flu. The reason most commonly advanced – including by Google – to explain this failure during epidemics can be summarized *as* ‘media-stoked panic’. Put simply, during epidemics GFT’s algorithm was susceptible to unexpected surges in search queries (Copeland et al. 2013). These surges were apparently triggered by massive media coverage, which was, in turn, aggravated by GFT’s own overestimations of flu activity. Here, I examine this narrative by focusing on the difficulties presented to GFT by the behavior of the ‘digital crowd’, or, more specifically, of the ‘googling crowd’. I suggest that GFT failed to keep track of the dynamics of contagion, a contagion both biological and digital. In the case at hand, sudden increases in search traffic may reveal as much about the effect of media attention on search habits as they do about the actual presence of flu activity. I further suggest that the troubles experienced by GFT complicate the narrative of the ‘wisdom of crowds’, whose existence GFT was widely taken to demonstrate. Specifically, search behavior during epidemics points to a sort of viral anxiety not easily amenable to algorithmic anticipation, to the extent that such anticipation relies on past data and patterns. The troubles experienced by GFT are the outcome of its struggles in anticipating the behavior of the googling crowd, behavior that seemed particularly unstable during epidemics. Under these circumstances, it became evident that the behavior of Google users, which GFT trusted to be consistent over time, was greatly suggestible. GFT struggled to understand how that behavior is influenced by viral activity that was indiscernibly biological and digital.

I argue, however, that the excesses and contingencies of viral activity are not sufficient to explain the struggles of GFT. I propose a second, complementary explanation: GFT’s struggles also lie in how a certain conception of viral activity – the ways that biological and digital life are expected to translate into certain types of online search behavior – was incorporated into the design of GFT’s algorithm. That conception, for instance, guided the choice of what should or should not be considered as relevant data. In other words, to understand GFT’s troubles, we have to examine both how viral life exceeds translation into algorithmic form (explanation 1) and how such a form created the life it attempted to measure/track (explanation 2), that is, how GFT was overwhelmed by design.

The story of GFT challenges common narratives in which big data modeling and mining appear as quasi-naturalized operations. It undermines the dichotomy between all-embracing technical ordering and the apparently indomitable complexity of life itself, a dichotomy that often remains implicit in discussions of algorithmic futures. There is no clear-cut distinction, in GFT, between algorithmic formalization and viral activity. In contrast with threatening narratives of the ‘algorithmic drama’ (Ziewitz 2016), in which algorithms are cast as autonomous, opaque, yet powerful objects, GFT’s troubles have to be situated within a larger series of relations, including the interests, values, and assumptions that have been incorporated into its design (Noble 2018). To a large extent, it is in fact the fetishized conception of the algorithm, conceived as immune to the disorderly, unpredictable life of the googling crowd, that led to GFT’s troubles. Yet as this article makes clear, these troubles should not be seen as demonstrating the existence of ‘wild and distinctively human remainders that computing can supposedly never consume’ (Seaver 2018, 380). By contrast, the story of GFT highlights the need for anthropology to approach algorithmic processes on their own terms (Lowrie 2018). By focusing on how GFT was designed and functions, I underline the shakiness of any rigid boundary between the computable and the incomputable, life and form, the artificial and the biological.

This analysis draws on varied sources, including technical reports, scientific scholarship, gray literature, and news reports, and engages with a wide set of arguments that I think can help us better apprehend the story at stake. The aim is not to generate a conceptual framework that could be applied to every online disease tracking system. Answering Tom Boellstorff’s (2015, 108) call in this respect, I nevertheless do try to craft theoretical tools that can address ‘patterns and dynamics beyond case study and the individual field site, even as those specificities shape the building of theory as well as its contextual modification’.

An outbreak down the street

The last decade has witnessed the rapid emergence of the field of digital epidemiology, with a proliferation of systems aimed at the early detection of disease activity (Brownstein, Freifeld, and Madoff 2009; Nuti et al. 2014). Disease detection and forecasting models are integrating data from various social media sources. The benefits often associated with such a web-based approach are speed and cost. They include early outbreak detection, reduction in the delay in response, improved geographical precision, accurate health metrics, and low cost in comparison to traditional surveillance methods. Digital technologies are, for instance, expected to fill what is considered to be a knowledge vacuum in the early stages of an outbreak or in the absence of health care-based surveillance data (Majumder et al. 2016). While these systems may make use of different data sources, they are all based on the premise of a correlation between web content (very often search or posting behavior) and trends in the

epidemiology of certain diseases. This correlation has been well documented, especially as far as web searches are concerned (Alicino et al. 2015; Hay et al. 2013). To varying degrees, web search-query data were found to be capable of tracking dengue (Chan et al. 2011), norovirus disease (Desai et al. 2012), HIV infections (Jena et al. 2013), and Lyme disease.

Launched in November 2008 by Google.org, the philanthropic arm of Google, GFT was the mother of all digital epidemiology systems.² Intended as the flagship of an emerging industry, GFT was to demonstrate the life-saving power of data and high-speed computation. It was thus cause for great optimism, including among Google's rank. 'I envision a kid (in Africa) getting online and finding that there is an outbreak of cholera down the street', explained Larry Brilliant in 2006, in an interview with *Wired News* (Zetter 2006). A few days later, Brilliant, who would soon set up Flu Trends as the executive director of Google.org, gave a TED Talk in which he expressed his wish to build a powerful 'early-warning system to protect against the things that are humanity's worst nightmare' (Brilliant 2006).³ Speaking a year after its launch, Eric Schmidt, then Google's CEO, also placed high hopes in Flu Trends' life-saving capacities: 'Many people believe that this device can save 10, 20, 30,000 lives every year just because the healthcare providers could get earlier and contain the outbreak [*sic*]. It's an example of collective intelligence of which will be [*sic*] many, many more' (Schonfeld 2009). Google's enthusiasm was also catching among public health experts and within the tech industry.⁴ GFT similarly attracted significant scholarly attention, as the initial *Nature* article describing the system has been cited more than two thousand times since 2009. It was also widely covered and celebrated in mainstream news media, including the *New York Times*, CNN, *Wall Street Journal*, and *Forbes*, among others.

The expectations created by Flu Trends are hardly surprising. After all, Flu Trends combined two contemporary technoscientific pursuits aimed at conjuring order out of disorder and safety out of a threatening future. First, interest in syndromic surveillance (surveillance using health-related data that precedes diagnosis and signals a sufficient probability of an outbreak to warrant further public health response) sharply increased after 11 September 2001, because of concern about the possibility of bioterrorist attacks (Berger, Shiao, and Weintraub 2006).

2 The US Centers for Disease Control and Prevention (CDC) has defined syndromic surveillance as 'surveillance using health-related data that precede diagnosis and signal a sufficient probability of a case or an outbreak to warrant further public health response' (Eysenbach 2006, 244).

3 For the full TED talk, see https://www.ted.com/talks/larry_brilliant_wants_to_stop_pandemics.

4 Dr. Joseph Bresee, chief of the epidemiology and prevention branch of the CDC's influenza division, commented on GFT's launch in these terms: 'We really are excited about the future of using different technologies, including technology like this, in trying to figure out if there's better ways to do surveillance for outbreaks of influenza or any other diseases in the United States' (Landau 2008).

GFT was thus launched in the midst of a decade of ‘preparedness’ efforts that, in the name of global health security, planned for future catastrophic events in the present.⁵ As has been well documented, the politics of preparedness are not per se concerned with preventing the occurrence of catastrophic events or with avoiding threats; rather, it is assumed that they will happen and thus should be considered as vulnerabilities to be mitigated (Lakoff and Collier 2010, 263). This applies to flu outbreaks. As was revealed by Carlo Caduff’s (2015) exploration of pandemic influenza preparedness in the United States, imaginaries of the pandemic as an uncertain but inevitable event draw upon contemporary sensibilities and anxieties while allowing prophetic claims to be framed in authoritative, scientific terms. In much the same way, GFT is expected to turn emergent flu outbreaks into relatively stable, predictable objects of knowledge.

The excitement around Flu Trends must also be understood in light of the relationships emerging among big data, algorithms, and health.⁶ Over the past few years, as patients and lay people increasingly have shared their health concerns, locations, and physical movements via digital media, much-lauded possibilities have emerged for the tracking and prediction of population health. This has been accompanied by a surge of interest within popular science and scholarly writings. Data-processing algorithms have found particular resonance in the field of global health, which has been for quite some time preoccupied with stabilizing ‘messy complexities of “living”’ (Adams 2016, 29) by developing indicators and quantitative metrics (Rottenburg and Merry 2015). To borrow from the biography of Christopher J. L. Murray, the director of the Institute for Health Metrics and Evaluation, the use of big data aims to generate epic measures ‘in the service of the most essential question of all: how to measure – and improve – how we live and die. And everyone, everywhere was included – now and for all time’ (Smith 2015, xvi).⁷ Producing epic measures: GFT set out to do just that.

How was GFT (not) working?

GFT monitored web searches to track flu-like illness in a population. Online search queries acted as sentinels, allowing GFT to detect surges in viral activity. Put simply, the system was

5 This has been illustrated by the establishment of the World Health Organization’s Global Outbreak Alert and Response Network and by the increasing role played by the CDC in global health.

6 It is not a coincidence that Mayer-Schönberger and Cukier (2013) open their book on big data by citing the case of GFT as prime example of the ‘datafication’ of the world.

7 The use of algorithmic modeling is central in the calculation and visualization of health trends and forecasts by the IHME. For a short story on how Netflix’s algorithm was imported into the global health field, see Paulson (2013).

based on a correlation between search behavior and flu activity. This correlation can be traced back to the design of GFT in 2008. To design its algorithms, Google took the most common fifty million search terms that Americans type and compared them with data from the US Centers for Disease Control and Prevention (CDC) on the spread of seasonal flu between 2003 and 2008 (Ginsberg et al. 2009). Google then identified a combination of forty-five items – words such as ‘headache’ and ‘runny nose’ and queries related to influenza complications – that had a strong correlation with the official CDC figures of doctor visits for influenza-like illness. Both past online search data and CDC data were region sensitive, with GFT identifying IP addresses associated with each search to determine the state in which the query was entered. Using this correlation between localized searches and doctor visits, it developed an algorithmic model aimed at anticipating influenza activity based on people’s web searches. And, importantly, it was able to do so one to two weeks ahead of national health agencies’ estimates. GFT used aggregated historical logs of online web search queries and CDC reports on actual flu activity to anticipate an outbreak of the disease in a particular region: as Thomas (2014, 294) suggests, in GFT the ‘future is anticipated and surveilled using past data’. GFT did not represent present flu activity. Rather, it forecasted future CDC reports in the present (Arthur 2014). By the time it was shut down, GFT was active in twenty-nine countries worldwide.⁸

For some time after the launch of its operations, GFT was relatively successful. But it soon had its ups and downs. In February 2013, an article published in *Nature* reported that it had drastically overestimated the peak of influenza during the 2012–2013 season in the United States (Butler 2013). The primary reason for this, the article argued, was the widespread media coverage of a particularly severe flu season, in which viral activity – as established by conventional epidemiological methods – rose quickly to well above average. This eventually led to the declaration of a public health emergency by the state of New York. The related media coverage then triggered unanticipated online behavior: in early 2013, there were more flu-related searches in the United States than ever before. This, Google later acknowledged, provoked GFT’s overestimation, which reached more than twice the CDC-reported incidence (Copeland et al. 2013). Google further concluded that GFT’s algorithm was susceptible to heightened media coverage and to the related sudden surges in flu-related search queries, which the algorithm erroneously counted as ‘positives’ although they in fact were ‘false positives’.

Adding to this, GFT itself contributed to aggravating media reports of the flu season, since its own publicly available predictions of a record-breaking severe flu season were used in reports

8 GFT also led to the launch, in 2011, of Google Dengue Trends, which was active in ten countries before also being closed, in August 2015.

providing an amplified picture of the flu season. GFT data thus found its way into widespread media coverage, with titles such as ‘Going Viral: Google Searches for Flu Symptoms Are at an All-Time High. Is It Time to Panic?’ (Oremus 2013) or ‘The Google Flu Chart That’s Blowing Everyone’s Mind’ (Boesler 2013). This led to a vicious cycle in which media coverage, surging numbers of internet searches, and GFT reports seemed to validate but actually exacerbated one another, provoking a general sense of panic (Flahault et al. 2017).

How did GFT actually work? Specifically, in GFT’s tracking algorithm, what was the relationship between signal and noise? Flu Trends aimed to detect influenza-like illness (ILI) before it was diagnosed. To do so, it sought to correlate online search queries with actual flu activity. But, of course, not all flu-related queries signal disease. Flu-related queries also include misspellings, people feeling anxious about reading the news, searches for unrelated information, and so on. The challenge, then, was to differentiate queries that signaled viral activity – either affecting the Google user or a relative – from ambient noise.⁹ The inconsistency of online health-related search behavior during an outbreak or pandemic was an issue Google had been well aware of since the early days of GFT (Eysenbach 2006). In a paper published in 2009 in *Nature*, GFT engineers acknowledged this potential limitation:

In the event that a pandemic-causing strain of influenza emerges, accurate and early detection of ILI percentages may enable public health officials to mount a more effective early response. Although we cannot be certain how search engine users will behave in such a scenario, affected individuals may submit the same ILI-related search queries used in our model. Alternatively, panic and concern among healthy individuals may cause a surge in the ILI-related query fraction and exaggerated estimates of the ongoing ILI percentage. (Ginsberg et al. 2009, 1014)

In other words, Google was worried that pandemic influenza might cause healthy but panicked individuals to change their search behavior in an erratic way. The question raised was not trivial: would the efficacy of GFT be limited to seasonal flu, or could it account for pandemic influenza? Difficulties experienced by GFT during the initial wave of pH1N1, in the spring of 2009, raised further doubts; Google then reworked the GFT model and obtained better results

9 This challenge is not unique to Flu Trends. As was noted by Dion, AbdelMalik, and Mawudeku (2015, 212), writing in the context of the Global Public Health Intelligence Network: ‘One of the primary challenges of Big Data in general and social media content in particular, is the “signal-to-noise” ratio which can significantly increase the potential for false positives and false negatives. With the influx of discussions and tweets surrounding the Ebola outbreak in West Africa, for example, it was difficult to distinguish between actual signals of concern and the plethora of messages that would otherwise be expected during such an event’.

during the second wave of pH1N1, later in 2009 (Cook et al. 2011). The 2012–2013 influenza season in the United States further suggested that while GFT may not lack sensitivity (it did not miss signals), there should be concerns over its specificity (it picked up false signals). GFT struggled to estimate flu activity properly when there were unanticipated changes in online search behaviors triggered by extraordinary exposure to epidemic-related information. As a result, GFT failed the most during epidemics.

Panic in the digital crowd

At least since they became an object of analysis in the late nineteenth century, crowds have been considered dumber than their smartest members. Crowds, suggests this dominant narrative, are irrational, stupid, regressive, and potentially dangerous.¹⁰ The last two decades, however, have witnessed the formation of an apparently new language of crowds – ‘crowdsourcing’, ‘crowdfunding’, etc. – that ‘has become one of the dominant modes of figuring the collective at the heart of new information technologies today’ (Kelty 2012, 6). What has emerged is a concept of the ‘digital crowd’ that could not be further away from crowd theory’s notion of chaotic, suggestible masses. Decentralized and self-organizing, the digital crowd has come to stand for a democratizing force by which connected citizens participate – knowingly or not – in various forms of collaborative work that produce different kinds of value.¹¹

This newfound potential of the digital crowd was compellingly elaborated and popularized by James Surowiecki in his best-selling 2005 book *The Wisdom of Crowds*. Under the right conditions, suggests Surowiecki, large groups of people can be smarter than the smartest people in them. The argument goes like this: if you put a large enough group of diverse people in just about any problem-solving situation – for instance, asking them to make a prediction or estimate a probability – the average of all responses will be more accurate than any given individual response. The reason for this, explains Surowiecki (2005, 10), is quite simple: the errors people make in coming up with an answer will cancel themselves out. The average judgment of the crowd converges on the right solution. Google, Surowiecki notes, was built

10 Gustave Le Bon perhaps best captures this collective dumbing down in his classic *The Crowd: A Study of the Popular Mind* ([1895] 2009, 205): ‘For instance, a gathering of scientific men or of artists, owing to the mere fact that they form an assemblage, will not deliver judgments on general subjects sensibly different from those rendered by a gathering of masons or grocers’.

11 The value produced by digital crowds can obviously be commercial, for example, crowdsourcing marketplaces such as Amazon’s Mechanical Turk, but also moral or biopolitical, as in the case of Google Flu Trends, in which more precise epidemiological information was expected to help tailor public health interventions.

on such wisdom, as its PageRank algorithm – and thus its whole searching experience – encapsulates the knowledge of web users: the more links to a page, the most valuable it is estimated to be, and the higher its ranking will be. A similar logic applied to the design of GFT. As was noted by Cukier and Mayer-Schönberger (2013, 33), the sheer size of the data set mined by GFT's algorithm was to compensate for its messiness and noise, for things like misspellings and searches for unrelated information. GFT promised to provide a vivid illustration of how the wisdom of the crowd could be tapped into for social good.

GFT's troubles, however, evidently suggest that even very large crowds can go wrong. The digital crowd, as Surowiecki himself has specified, is far from infallible. The most probable reason GFT started to integrate data it should not have points to what Surowiecki considers to be the main paradox of the intelligent crowd: it is imagined not merely as a collection of individuals but as independently deciding individuals who should affect one another as little as possible.¹² The wisdom of the crowd apparently depends on avoiding the 'herding effect', in the form, for example, of social influence: 'One key to successful group decisions is getting people to pay much less attention to what everyone else is saying' (Surowiecki 2005, 65). The most intelligent crowd is thus composed of members in a state of connected, or 'shared', isolation.¹³ It is connected but not influenced. For the stupidity of the old crowd to give way to the intelligence of the digital crowd, it has to rid itself of its most basic feature, which made it suspect in the first place: its affectability. It has to restore the autonomy of the liberal subject: smarter in a group but only if thinking on their own. The intelligent digital crowd very much looks like a human group deprived of its emergent, vital energies.

It is, I suggest, this very act of domestication that GFT's troubles expose. In times of epidemics, there is a rupture of the equilibrium between the apparent independence of the individual members of the crowd and the affective pressure exerted by the crowd on its members. News gathers momentum, but rumors and fake reports go viral too. Faraway funeral rites, traveling bodily fluids, maps of contagion risk, and projections of how the virus might spread all converge into a sense of being overexposed to viruses, to others, to the world. During epidemics, the conditions for independent thinking, if they ever applied to GFT (more on this later), are most explicitly compromised. In times of epidemics, the human body appears more porous than usual. This is a body whose relation to the world is made up of a continuous

12 Over the past decade, many experiments were conducted about the 'social influence bias' and how it apparently undermines collective intelligence. See Muchnik, Aral, and Taylor (2013) and Lorenz and colleagues (2011).

13 Surowiecki (2005, 41) does not speak of 'isolation' but rather of 'independence' and of 'relative freedom from the influence of others'.

and largely involuntary process of encounter (Thrift 2006), a body marked by the irruption of the ‘much-too-much’, which arrives from without (Sloterdijk 1989, 69). A body for which the ‘outside world’, replete with stimulation of all kinds, suddenly appears all too intimate. This is a mobilized body, caught out in the global speeding up and intensification of events, exposed to ambient noise and information overload (Duclos 2017).¹⁴ Epidemics reveal a body open to suggestion, participating in a self-spreading tendency, a quasi-autonomous movement of imitative contagion (Sampson 2012). This movement also alters the course of the virus, as changes in behavior affect the rates of transmission and infection.

When the apparently unpredictable course of viral life meets the contagiousness of public hysteria, whole populations start googling symptoms they do not have. Massive media coverage may contribute to generating web searches in numbers totally disproportionate to actual disease activity or to the real threat to public health (Towers et al. 2015). Online traffic gets denser, generating more noise and making it harder to translate search activity into meaningful signals. During October 2014, the month following the first case of Ebola in the United States, there were more than twenty-one million tweets about Ebola within the country. Data collected in this period also demonstrate a dramatic increase in web searches. Under such circumstances, online data no longer translate into relatively predictable patterns, comparable, for instance, to data from previous years. During epidemics, noise-signal ratios run amok: ‘The event exceeds its representation, making its meaning uncertain’ (Caduff 2014, 41). It then becomes clear that algorithmic anticipation struggles to exhaust the possibilities of a virtual, ‘unspecifiable’ future-to-come (Braun 2011, 402). The correlation between online search trends and actual viral activity is compromised. Ultimately, GFT provided a map of collective anxieties and how they translated into search practices (Tausczik et al. 2012). Under particular circumstances, the digital crowd apparently became the old crowd again, immersed into and participating in the complications of the world.

A trackable life

Google Flu Trends could not distinguish the actual spread of flu within a population from the manifestation of collective anxieties, which is what it ended up mapping. However, GFT’s struggles should not be reduced to some irrational, fundamentally indomitable behavior of Google users, of Google’s crowd. By contrast, I suggest that they are also largely attributable to how GFT constituted the dynamics of flu activity, both apparent and real. The main issue

14 The tagline ‘This is the app you want around soon to keep track of this deadly outbreak’, for instance, advertised the Ebola Tracker iPhone application. More than one hundred similar phone applications were launched in response to the Ebola epidemics. Most of them can be divided into three categories: Ebola news and information, Ebola gaming, and Ebola diagnosing.

with GFT appears to lie in how it created the very realities it attempted to measure and track. That is, in how it performed ‘epidemic reality’.

To understand this critical distinction better, it is perhaps worth insisting that GFT had no access to viral life, that is, to the ‘reality’ of influenza-like illness. It did not, for instance, use remote diagnostic solutions or monitor doctor visits in a particular region. What Flu Trends had access to was searching behavior and specifically search queries submitted by millions of users. Yet GFT was intended to track flu activity. Its designers did claim that, by monitoring health-seeking behavior in the form of queries to online search engines, it could actually detect influenza epidemics (Ginsberg et al. 2009). To do so, GFT relied on a correlation of the relative frequency of certain search queries and the percentage of physician visits in which a patient presented influenza-like symptoms, visits whose count represents a classical public health method to assess viral activity. Only by assuming that this correlation would be stable over time could GFT be expected to track not only search behavior but also outbreaks themselves in near real-time. But this correlation turned out not to be stable over time. Specifically, the data on which GFT’s algorithm was trained (past search queries and epidemiological data) often did not match ‘the data being operated on in the wild’ (Gillespie 2016, 21).

When considering how Flu Trends struggled to track changes in search behavior (in the wild) over time, two of its features are particularly worthy of attention. First, Flu Trends’ algorithm did not take into consideration the potential effect of alterations that were made, in the years following its launch, to Google’s own search environment. A notorious critique of Flu Trends published in *Science* suggested that, by making some search terms more prevalent, Google contributed to GFT’s overestimation issue (Lazer et al. 2014). Another critique, this time largely based on Flu Trends engineers’ narrative of their own difficulties, concurred: ‘Perhaps Google adding more suggestions for users threw off the baseline. Or perhaps the underlying media ecosystems or user search behavior are changing more quickly than anticipated’ (Madrigal 2014).

Google’s designers do indeed regularly re-engineer the interface of the world’s largest online search engine. Modifications include such things as the introduction of new health-based add-ons, navigation controls, search suggestions, and controls for refining queries and amplifying result information. Such changes seek to improve the user experience by making the search environment more appealing or easier to use. With Google generating most of its revenue from search, it also adapts its search environment in order to promote advertising revenue. Central to this operation is Google’s PageRank algorithm, which extensively governs the traffic to a website (Pasquinelli 2009, 7). PageRank ranks the position of a website within the search engine’s results primarily on its popularity, rather than, say, its relevance. PageRank is widely cited as a prime example of an algorithm that feeds on the collective intelligence of the digital

crowd. But this also takes us back to the great paradox of the digital crowd: the very networks that allow its ‘intelligence’ to express itself are constantly making it harder for independent thinking to be sustained, thus raising the specter of the monstrous, untamed force of the susceptible crowd. Given its compounding effect, Google’s search algorithm is particularly conducive to online imitation or contagion.

Yet the digital crowd, here, should not be confused with pure affective immediacy. Its vital energy is at once emergent and shaped by algorithmic inscription. Put bluntly, Google’s business model depends on its capacity to affect online behavior in specific ways. Although Google prides itself on the distinction between search results and advertising – as users should be influenced as little as possible by exterior biases and interventions – PageRank is tweaked on average four hundred times every year.¹⁵ And modifications to PageRank’s algorithm do influence search results and the ranking of websites in huge ways. Search results are not neutral and they affirmatively control their users’ experiences, for better or for worse (Goldman 2006; Vaidyanathan 2012). Google tweaks its algorithms in certain directions and not others (Peters 2015, 357). Modifications in Google’s search environment are relevant to Flu Trends to the extent that Google does not simply channel and track user behavior. It also produces new triggers and shapes specific behaviors. In deciding not to tweak Flu Trends in accordance with or in response to changes in its own search environment, Google intriguingly assumed that flu-related queries would not be affected by tweaks aimed precisely at influencing web search behaviors, not to mention by the overall susceptibility of its search environment to contagion or herding effects. In other words, Google remained insensitive to the ‘internal’ conditions – the modifications in the very working of the search platform – affecting the generation of the data it mined. As a result, Flu Trends was left to predict the behavior of a deceptive, erratic crowd.

In addition to this neglect of these ‘internal’ conditions, a related design feature that contributed to undermining Flu Trends’ algorithm has to do with how it related to whatever was ostensibly ‘outside’ the data it mined. Put simply, Flu Trends relied only on data from web searches. Its algorithm aggregated multiple query terms into a single explanatory variable, namely, the main trend for the terms with the strongest correlation with CDC data from previous years of flu activity. It has been suggested that this use of a single explanatory variable

15 In a letter to its future stockholders in 2004, Google’s founders Larry Page and Sergey Brin write: ‘Google users trust our systems to help them with important decisions: medical, financial and many others. Our search results are the best we know how to produce. They are unbiased and objective, and we do not accept payment for them or for inclusion or more frequent updating. We also display advertising, which we work hard to make relevant, and we label it clearly’. For the full letter, see <https://abc.xyz/investor/founders-letters/2004/ipo-letter.html>.

largely contributed to Flu Trends' difficulties in tracking changes in people's internet search behavior over time (Yang, Santillana, and Kou 2015). Most importantly, the use of a single explanatory variable has been criticized for preventing newly available flu activity reports, produced by the CDC as the season evolved, to be integrated into the model (Yang, Santillana, and Kou 2015). This refusal to incorporate lagged CDC data allowed GFT to detect sudden changes in flu activity faster than if it had periodically had recent CDC numbers fed into the algorithm. GFT could thus almost instantaneously translate dynamics of search behavior into perceptible, recognizable patterns. This was only achieved, however, by overlooking delayed yet relevant information about the actual evolution of viral activity.

GFT wanted to keep its signal pure. Its algorithm was designed to isolate anticipated viral activity from the behavior changes that affected both its digital (the search queries) and biological (its transmission through contact, etc.) dynamics (Mackenzie 2014). GFT's tracking of viral activity was based on a correlation between past search activity and past viral activity, which meant that its capacity to make viral activity intelligible was entirely dependent on past behavior patterns. What became most evident during epidemics is the contingency of these patterns and the increasingly unsuccessful work by GFT to subsume viral activity into these patterns nevertheless. This work by which data are abstracted from their biological, social, and material substrate is instrumental in attempting to represent (to stabilize) the volatile processes and relations of viral activity, including things like the factors affecting how people feel, think, and behave about a potential flu infection at a certain point in time. But the extent to which GFT's data were severed from the processes they aimed to represent allowed them to take on a life of their own: a trackable life, in which there was very little actual flu.¹⁶

Into the wild: Algorithmic biopolitics

Even though it failed to predict the next epidemic outbreak, GFT hasn't necessarily been a failure. As was noted by Theresa MacPhail (2015), many within the epidemiological community saw it as a step in the right direction, as it provided insight into predictive models in general. More than ever, the shutdown of GFT now seems to be little more than a distant memory. Advances in big data analytics, crowdsourcing, and machine learning promise to improve epidemiological surveillance. The consensus seems to be rapidly building around the

16 When asked, in retrospect, what lessons he drew from Flu Trends' 'parable', one of two engineers that spearheaded its creation makes this interesting statement: "There isn't a ground truth for what exists. [Flu tracking systems] are all proxies for flu illness or influenza-like illness, and there is great power that comes from combining these signals. ... There's huge promise with these techniques, but you have to understand how they should be used" (Madrigal 2014).

fact that many of the troubles experienced by GFT are solvable by, for instance, combining data sources and better updating algorithms. What is announced is nothing less than a sea change in global health surveillance (National Academies of Sciences and Medicine 2016).

This enthusiasm has translated into a recent proliferation of digital platforms aimed at mapping, monitoring, and predicting infectious diseases. Take the example of the CDC. Only a few months after GFT's most flagrant error (during the 2012–2013 flu season), the CDC launched the Predict the Influenza Season Challenge. It announced that the research team that most successfully used digital surveillance data (Twitter, internet search data, web surveys, etc.) to predict the timing, peak, and intensity of the upcoming 2013–2014 flu season would receive an award of US\$75,000. Eleven teams completed the challenge, using a variety of data sources. A team from the Mailman School of Public Health at Columbia University, led by Dr. Jeffrey Shaman, won the contest. Put bluntly, the winning system used Google Flu Trends' data, which was available online, conjointly with weekly region-specific reports from the CDC on verified cases of flu. It thus addressed what was earlier identified as a key issue in the design of GFT: its overreliance on online data alone. In following years, the flu-forecasting challenge was won by a team led by Dr. Roni Rosenfeld (Carnegie Mellon University), using a combination of machine learning and crowdsourcing (Biggerstaff et al. 2016).

As an upshot of its annual competition, the CDC has worked closely with different teams to foster innovation in flu-activity modeling and prediction. This led to the launch of the FluSight website in January 2016, only a few months after GFT was shut down, to provide forecasts of flu activity. As was explained in the statement announcing the launch, while the CDC does track flu activity through its nationwide flu surveillance system, this system 'lags behind real-time flu activity'. The goal of infectious disease forecasting, the statement continued, 'is to provide a more-timely and forward-looking tool that predicts rather than monitors flu activity so that health officials can be more proactive in their prevention and mitigation responses'.

The CDC is not alone. Global health researchers and institutions are eagerly investing in similar solutions, to be used in a vast array of situations. Digital communication, modeling, and visualization technologies are being rapidly integrated in epidemiology and health logistics, thus reconfiguring the practices and scope of global health (Peckham and Sinha 2017; Parks 2009). The growing role of digital technologies in the government of epidemiological threat is also illustrated by their spread in humanitarian settings, leading to the emergence of the field of 'digital humanitarianism' (Read, Taithe, and Mac Ginty 2016; Duffield 2016). What we are witnessing is a broad movement in which experimenting with big data is increasingly being heralded as critical to global public health efforts to detect and contain disease (Erikson 2018). As a result, insecurity, poverty, and precarity are increasingly framed as informational issues, that is, as issues amenable to data-based modeling techniques and interventions.

These developments carry significant implications as far as a (bio)politics of algorithms is concerned. Algorithmic tracking seeks to enclose the future in the present through anticipation. GFT's aptitude at early threat detection relied upon the automatic recognition of particular scenarios – for instance, what is a normal search behavior and what is not – themselves based on the mining of past data. It thus was based on the assumption that big data analytics can harvest patterns of life – how things have been in the past and are thus supposed to be in the future – to detect potential threat. Algorithmic tracking produces the sense that certain possible futures are inevitable in the present (Adams, Murphy, and Clarke 2009). This may constrict the imagination of other possible futures (Duclos, Sánchez-Criado, and Nguyen 2017), of futures not already mapped out by and contained in past patterns. For that reason, the troubles experienced by GFT may instill a sense of relief. They hint at how the messiness of life may disrupt past patterns and exceed technical capacities at predicting their occurrence. They may gesture toward the presence of an unthinkable remainder of realities that remain opaque to or exceed algorithmic tracking. Even in the midst of ubiquitous computing, GFT's troubles suggest, there remains a lurking sense of the force of indeterminacy (Munster 2013, 141). In the context of the ubiquitous influence of online tracking on most dimensions of individual and collective life, the failure by one of the largest data-mining companies to anticipate the behavior of the digital crowd may feel reassuring.¹⁷ It may serve as a reminder not to magnify the threat posed by algorithmic capacities – the same applies to artificial intelligence – for blurring the distinction between the calculation of a possible future and the intensity of its unfolding. To a certain extent, then, it is possible that the death of GFT contributes in moderating a growing sense of moral panic often associated with the increasing influence of algorithmic tracking and artificial intelligence on the ordering of social life.

But taking into account the intensity of the 'googling crowds' is not sufficient to understand the troubles experienced by GFT. Unpredictable, viral anxiety in web searching was obviously a key issue for GFT, since it did not correspond to past online search patterns. As I have suggested here, online search behavior during epidemics resembled that of a suggestible crowd, one exposed to, reacting to, and participating in the complications of the world. But the problem with GFT was not only how the self-intensifying dynamics of online contagion affected the algorithm's search behavior. The problem was also, I argue, that the autonomy of these dynamics was taken for granted by the very design of the system. While it did

17 Examples are too many to enumerate. Suffice it to mention that they include speculator demonstrations such as the role played by data tracking in the recent US presidential election and the establishment of the Social Credit System in China but also more normalized forms of tracking enabling targeted content and advertisement.

acknowledge that the affectability of its users could lead to ‘mass-panic searching’ (Ginsberg et al. 2009), Google did not take that into consideration when designing GFT’s algorithm. GFT did not account for the many conditions, including within the Google search environment, that may affect user behavior in new ways. As a result, the conditions – put simply, the Google search engine – under which data mined by GFT was generated differed substantially from the ones under which the algorithm was initially trained, namely, data from previous years, upon which the correlation established by GFT was based. As mentioned earlier, this was exacerbated by the fact that GFT also did not include data about viral activity from the CDC when it became available as the flu season evolved. This led to prolonged misdetections and overestimations of flu activity.

The troubles experienced by GFT raise questions that apply to recent developments in machine learning and artificial intelligence, and addressing these may help us better apprehend the future effects of digital technologies on global health. For the most fervent prophets of machine learning, GFT may raise the question of the experimental conditions under which an algorithm could be trained so that they ‘learn to learn’ in ways that make them less susceptible to the kind of human biases on which GFT was still fatally reliant. Skeptics of machine learning might find comfort in the notion that, even as methods improve and tracking becomes more accurate, there will always remain a remainder, indeterminate, in excess. These viewpoints are situated at the ends of a spectrum oscillating, we may say, between technological utopianism and romantic humanism. But is this excess in and of itself something with which to be content? Is it, for instance, ethically and politically satisfactory to feel reassured by the apparent limits of tracking? Or could it hint at a strategic retreat into more opaque forms of life, or, conversely, at the emergent potentials of the digital crowd? These questions are important. But to deal with them better, it is important to insist that the underlying challenge posed by GFT is to be able to think about tracking without foundering on dichotomous thinking: technical ordering or untamed human nature, technical function or social behavior, formal abstraction or overflowing life. The challenge is to equip ourselves with a language that allows us to critically address tracking processes by which the future is enclosed while nevertheless insisting on what will not be contained. It is, in other words, to engage with algorithmic systems as spaces we craft and inhabit, in which we shape our lives, with their own combination of openness and finitude.

About the author

Vincent Duclos is Assistant Professor at Drexel University, teaching in the Center for Science, Technology & Society, and the Department of Global Studies & Modern Languages. He is an anthropologist of medicine, writing about the deepening enmeshment of digital infrastructures and biomedicine, including in digital health networks and disease-tracking platforms. He has

carried out research in India and West Africa. His writing is inspired by work in cultural anthropology, science and technology studies, and the philosophy of science and technology, and has been published in *Cultural Anthropology*, *Medical Anthropology Quarterly*, and *New Formations*, among other venues. <https://www.mcgill.ca/globalhealth/funding/steinberg-global-health-postdoctoral-fellowship/our-fellows#Vince>.

References

- Adams, Vincanne. 2016. 'Metrics of the Global Sovereign: Numbers and Stories in Global Health'. In *Metrics: What Counts in Global Health*, edited by Vincanne Adams, 19–54. Durham, NC: Duke University Press.
- Adams, Vincanne, Michelle Murphy, and Adele E. Clarke. 2009. 'Anticipation: Technoscience, Life, Affect, Temporality'. *Subjectivity* 28 (1): 246–65.
- Alicino, Cristiano, Nicola Luigi Bragazzi, Valeria Faccio, Daniela Amicizia, Donatella Panatto, Roberto Gasparini, Giancarlo Icardi, and Andrea Orsi. 2015. 'Assessing Ebola-Related Web Search Behaviour: Insights and Implications from an Analytical Study of Google Trends–Based Query Volumes'. *Infectious Diseases of Poverty* 4 (1): 54. <https://doi.org/10.1186/s40249-015-0090-9>.
- Arthur, Clarke. 2014. 'Google Flu Trends Is No Longer Good at Predicting Flu, Scientists Find'. *Guardian*, 27 March. <https://www.theguardian.com/technology/2014/mar/27/google-flu-trends-predicting-flu>.
- Berger, Magdalena, Rita Shiao, and June M. Weintraub. 2006. 'Review of Syndromic Surveillance: Implications for Waterborne Disease Detection'. *Journal of Epidemiology and Community Health* 60 (6): 543–50. <https://doi.org/10.1136/jech.2005.038539>.
- Biggerstaff, Matthew, David Alper, Mark Dredze, Spencer Fox, Isaac Chun-Hai Fung, Kyle S. Hickmann, Bryan Lewis, Roni Rosenfeld, Jeffrey Shaman, and Ming-Hsiang Tsou. 2016. 'Results from the Centers for Disease Control and Prevention's Predict the 2013–2014 Influenza Season Challenge'. *BMC Infectious Diseases* 16 (1): 357.
- Boellstorff, Tom. 2015. 'Making Big Data, in Theory'. In *Data, Now Bigger and Better!*, edited by Tom Boellstorff and Bill Maurer, 87–108. Chicago: Prickly Paradigm.
- Boesler, Matthew. 2013. 'The Google Flu Chart That's Blowing Everyone's Mind'. *Business Insider*, 11 January. <http://www.businessinsider.com/google-flu-chart-blowing-everyones-mind-2013-1>.
- Braun, Bruce. 2011. 'Governing Disorder: Biopolitics and the Molecularization of Life'. In *Global Political Ecology*, edited by Richard Peet, Paul Robbins, and Michael Watts, 389–411. London: Routledge.

- Brilliant, Larry. 2006. 'My Wish: Help Me Stop Pandemics'. Filmed February 2006 in Monterey, California. TED video, 25:47. https://www.ted.com/talks/larry_brilliant_wants_to_stop_pandemics.
- Brownstein, John S., Clark C. Freifeld, and Lawrence C. Madoff. 2009. 'Digital Disease Detection: Harnessing the Web for Public Health Surveillance'. *New England Journal of Medicine* 360 (21): 2153–57. <https://doi.org/10.1056/NEJMp0900702>.
- Butler, Declan. 2013. 'When Google Got Flu Wrong: US Outbreak Foxes a Leading Web-Based Method for Tracking Seasonal Flu'. *Nature* 494: 155–56.
- Caduff, Carlo. 2014. 'Sick Weather Ahead: On Data-Mining, Crowd-Sourcing, and White Noise'. *Cambridge Journal of Anthropology* 32 (1): 32–46.
- Caduff, Carlo. 2015. *The Pandemic Perhaps: Dramatic Events in a Public Culture of Danger*. Oakland: University of California Press.
- CDC. 2016. 'Flu Activity Forecasting Website Launched'. Centers for Disease Control and Prevention. <https://www.cdc.gov/flu/news/flu-forecast-website-launched.htm> (link defunct).
- Chan, Emily H., Vikram Sahai, Corrie Conrad, and John S. Brownstein. 2011. 'Using Web Search Query Data to Monitor Dengue Epidemics: A New Model for Neglected Tropical Disease Surveillance'. *PLoS Neglected Tropical Diseases* 5 (5): e1206. <https://doi.org/10.1371/journal.pntd.0001206>.
- Cook, Samantha, Corrie Conrad, Ashley L. Fowlkes, and Matthew H. Mohebbi. 2011. 'Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic'. *PLoS ONE* 6 (8): e23610. <https://doi.org/10.1371/journal.pone.0023610>.
- Copeland, Patrick, Raquel Romano, Tom Zhang, Greg Hecht, Dan Zigmond, and Christian Stefansen. 2013. 'Google Disease Trends: An Update'. *International Society of Neglected Tropical Diseases*.
- Cukier, Kenneth, and Viktor Mayer-Schoenberger. 2013. 'Rise of Big Data: How It's Changing the Way We Think about the World'. *Foreign Affairs* 92: 28.
- Desai, Rishi, Aron J. Hall, Benjamin A. Lopman, Yair Shimshoni, Marcus Rennick, Niv Efron, Yossi Matias, Manish M. Patel, and Umesh D. Parashar. 2012. 'Norovirus Disease Surveillance Using Google Internet Query Share Data'. *Clinical Infectious Diseases* 55 (8): e75–78.
- Dion, M., P. AbdelMalik, and A. Mawudeku. 2015. 'Big Data and the Global Public Health Intelligence Network (GPHIN)'. *Canada Communicable Disease Report (CCDR)* 41 (9): 209–214.
- Duclos, Vincent. 2017. 'Inhabiting Media: An Anthropology of Life in Digital Speed'. *Cultural Anthropology* 32 (1): 20–26.
- Duclos, Vincent, Tomás Sánchez-Criado, and Vinh-Kim Nguyen. 2017. 'Speed: An Introduction'. *Cultural Anthropology* 32 (1): 1–11. <https://doi.org/10.14506/ca32.1.01>.

- Duffield, Mark. 2016. 'The Resilience of the Ruins: Towards a Critique of Digital Humanitarianism'. *Resilience* 4 (3): 147–65.
<https://doi.org/10.1080/21693293.2016.1153772>.
- Erikson, Susan L. 2018. 'Cell Phones ≠ Self and Other Problems with Big Data Detection and Containment during Epidemics'. *Medical Anthropology Quarterly* 32 (3): 315–39.
- Eysenbach, Gunther. 2006. 'Infodemiology: Tracking Flu-Related Searches on the Web for Syndromic Surveillance'. *AMIA Annual Symposium Proceedings* 2006: 244–48.
- Feldman, Ilana, and Miriam Ticktin, eds. 2010. *In the Name of Humanity: The Government of Threat and Care*. Durham, NC: Duke University Press.
- Flahault, Antoine, Antoine Geissbuhler, Idris Guessous, Philippe Guérin, Isabelle Bolon, Marcel Salathé, and Gérard Escher. 2017. 'Precision Global Health in the Digital Age'. *Swiss Medical Weekly* 147: w14423.
- Gillespie, Tarleton. 2016. 'Algorithm'. In *Digital Keywords: A Vocabulary of Information Society and Culture*, edited by Ben Peters, 18–30. Princeton, NJ: Princeton University Press.
- Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. 2009. 'Detecting Influenza Epidemics Using Search Engine Query Data'. *Nature* 457 (7232): 1012–14.
http://www.nature.com/nature/journal/v457/n7232/supinfo/nature07634_S1.html.
- Goldman, Eric. 2006. 'Search Engine Bias and the Demise of Search Engine Utopianism'. *Yale Journal of Law & Technology* 8: 188–200.
- Halpern, Orit. 2015. *Beautiful Data: A History of Vision and Reason since 1945*. Durham, NC: Duke University Press.
- Hay, Simon I., Dylan B. George, Catherine L. Moyes, and John S. Brownstein. 2013. 'Big Data Opportunities for Global Infectious Disease Surveillance'. *PLoS Medicine* 10 (4): e1001413.
- Jena, Anupam B., Pinar Karaca-Mandic, Lesley Weaver, and Seth A. Seabury. 2013. 'Predicting New Diagnoses of HIV Infection Using Internet Search Engine Data'. *Clinical Infectious Diseases* 56 (9): 1352–53.
- Keck, Frédéric, and Andrew Lakoff. 2013. 'Preface (Sentinel Devices)'. *Limn* 3: 2–3.
<https://limn.it/articles/preface-sentinel-devices-2/>.
- Kelty, Christopher M. 2012. 'Preface (Clouds and Crowds)'. *Limn* 2: 4–7.
<https://limn.it/articles/preface-crowds-and-clouds/>.
- Lakoff, Andrew, and Stephen J. Collier. 2010. 'Infrastructure and Event: The Political Technology of Preparedness'. In *Political Matter: Technoscience, Democracy, and Public Life*, edited by Bruce Braun and Sarah J. Whatmore, 243–66. Minneapolis: University of Minnesota Press.
- Landau, Elizabeth. 2008. 'Google Tool Uses Search Terms to Detect Flu Outbreaks'. *CNN*, 9 December.

- <http://www.cnn.com/2008/HEALTH/conditions/11/11/google.flu.trends/#cnnSTCOther2>.
- Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. 'The Parable of Google Flu: Traps in Big Data Analysis'. *Science* 343 (6176): 1203–1205. <https://doi.org/10.1126/science.1248506>.
- Le Bon, Gustave. (1895) 2009. *The Crowd: A Study of the Popular Mind*. Portland, OR: Floating Press.
- Lorenz, Jan, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. 2011. 'How Social Influence Can Undermine the Wisdom of Crowd Effect'. *Proceedings of the National Academy of Sciences* 108 (22): 9020–25.
- Lowrie, Ian. 2018. 'Algorithms and Automation: An Introduction'. *Cultural Anthropology* 33 (3): 349–59.
- Mackenzie, Adrian. 2014. 'Multiplying Numbers Differently: An Epidemiology of Contagious Convolution'. *Distinktion: Scandinavian Journal of Social Theory* 15 (2): 189–207.
- MacPhail, Theresa. 2015. 'Data, Data Everywhere'. *Public Culture* 27 (2 [76]): 213–19.
- Madrigal, Alexis C. 2014. 'In Defense of Google Flu Trends'. *Atlantic*, 27 March. <https://www.theatlantic.com/technology/archive/2014/03/in-defense-of-google-flu-trends/359688/>.
- Majumder, Maimuna S., Mauricio Santillana, Sumiko R. Mearu, Denise P. McGinnis, Kamran Khan, and John S. Brownstein. 2016. 'Utilizing Nontraditional Data Sources for Near Real-Time Estimation of Transmission Dynamics During the 2015–2016 Colombian Zika Virus Disease Outbreak'. *JMIR Public Health and Surveillance* 2 (1): e30. <https://doi.org/10.2196/publichealth.5814>.
- Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt.
- Muchnik, Lev, Sinan Aral, and Sean J. Taylor. 2013. 'Social Influence Bias: A Randomized Experiment'. *Science* 341 (6146): 647–51.
- Munster, Anna. 2013. *An Aesthesia of Networks: Conjunctive Experience in Art and Technology*. Cambridge, MA: MIT Press.
- National Academies of Sciences, Engineering, and Medicine. 2016. *Big Data and Analytics for Infectious Disease Research, Operations, and Policy: Proceedings of a Workshop*. Washington, DC: National Academies Press.
- Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.
- Nuti, Sudhakar V., Brian Wayda, Isuru Ranasinghe, Sisi Wang, Rachel P. Dreyer, Serene I. Chen, and Karthik Murugiah. 2014. 'The Use of Google Trends in Health Care Research: A Systematic Review'. *PLoS ONE* 9 (10): e109583.
- Oremus, Will. 2013. 'Going Viral: Google Searches for Flu Symptoms Are at an All-Time High. Is It Time to Panic?' *Slate*, 9 January 2013.

- http://www.slate.com/articles/technology/technology/2013/01/flu_shot_time_google_flu_trends_predicts_worst_season_on_record.html.
- Parks, Lisa. 2009. 'Digging into Google Earth: An Analysis of "Crisis in Darfur"'. *Geoforum* 40 (4): 535–45. <https://doi.org/10.1016/j.geoforum.2009.04.004>.
- Pasquinelli, Matteo. 2009. 'Google's PageRank Algorithm: A Diagram of the Cognitive Capitalism and the Rentier of the Common Intellect'. In *Deep Search: The Politics of Search*, edited by Konrad Becker and Felix Stalder, 152–62. London: Transaction.
- Paulson, Tom. 2013. 'Visualize Global Health'. *Humanosphere*, 5 March. <http://www.humanosphere.org/global-health/2013/03/visualizing-global-health-or-how-netflix-and-burger/>.
- Peckham, Robert, and Ria Sinha. 2017. 'Satellites and the New War on Infection: Tracking Ebola in West Africa'. *Geoforum* 80: 24–38.
- Peters, John Durham. 2015. *The Marvelous Clouds: Toward a Philosophy of Elemental Media*. Chicago: University of Chicago Press.
- Read, Róisín, Bertrand Taithe, and Roger Mac Ginty. 2016. 'Data Hubris? Humanitarian Information Systems and the Mirage of Technology'. *Third World Quarterly* 37 (8): 1314–31. <https://doi.org/10.1080/01436597.2015.1136208>.
- Rottenburg, Richard, and Sally Engle Merry. 2015. 'A World of Indicators: The Making of Governmental Knowledge through Quantification'. In *A World of Indicators: The Making of Governmental Knowledge through Quantification*, edited by Richard Rottenburg, Sally Engle Merry, Sung-Joon Park, and Johanna Mugler, 1–33. Cambridge: Cambridge University Press.
- Sampson, Tony D. 2012. *Virality: Contagion Theory in the Age of Networks*. Minneapolis: University of Minnesota Press.
- Schmidt, Charles W. 2012. 'Trending Now: Using Social Media to Predict and Track Disease Outbreaks'. *Environmental Health Perspectives* 120 (1): a30.
- Schonfeld, Eric. 2009. 'Eric Schmidt Tells Charlie Rose Google Is "Unlikely" to Buy Twitter and Wants to Turn Phones into TVs'. *TechCrunch*, 7 March. <https://techcrunch.com/2009/03/07/eric-schmidt-tells-charlie-rose-google-is-unlikely-to-buy-twitter-and-wants-to-turn-phones-into-tvs/>.
- Seaver, Nick. 2018. 'What Should an Anthropology of Algorithms Do?' *Cultural Anthropology* 33 (3): 375–85.
- Seifter, Ari, Alison Schwarzwald, Kate Geis, and John Aucott. 2010. 'The Utility of "Google Trends" for Epidemiological Research: Lyme Disease as an Example'. *Geospatial Health* 4 (2): 135–37.
- Sloterdijk, Peter. 1989. *Thinker on Stage: Nietzsche's Materialism*. Translated by Jamie Owen Daniel. Minneapolis: University of Minnesota Press.
- Smith, J. N. 2015. *Epic Measures: One Doctor. Seven Billion Patients*. New York: HarperCollins.
- Surowiecki, James. 2005. *The Wisdom of Crowds*. New York: Anchor.

- Tausczik, Yla, Kate Faasse, James W. Pennebaker, and Keith J. Petrie. 2012. 'Public Anxiety and Information Seeking Following the H1N1 Outbreak: Blogs, Newspaper Articles, and Wikipedia Visits'. *Health Communication* 27 (2): 179–85.
- Thomas, Lindsay. 2014. 'Pandemics of the Future: Disease Surveillance in Real Time'. *Surveillance & Society* 12 (2): 287.
- Thrift, Nigel. 2006. 'Space'. *Theory, Culture & Society* 23 (2–3): 139–46. <https://doi.org/10.1177/0263276406063780>.
- Towers, Sherry, Shehzad Afzal, Gilbert Bernal, Nadya Bliss, Shala Brown, Baltazar Espinoza, Jasmine Jackson, et al. 2015. 'Mass Media and the Contagion of Fear: The Case of Ebola in America'. *PLoS ONE* 10 (6): e0129179. <https://doi.org/10.1371/journal.pone.0129179>.
- Vaidhyanathan, Siva. 2012. *The Googlization of Everything (and Why We Should Worry)*. Berkeley: University of California Press.
- Yang, S., M. Santillana, and S. C. Kou. 2015. 'Accurate Estimation of Influenza Epidemics Using Google Search Data via ARGO'. *Proceedings of the National Academy of Sciences* 112 (47): 14473–78. <https://doi.org/10.1073/pnas.1515373112>.
- Zetter, Kim. 2006. 'Brilliant's Wish: Disease Alerts'. *Wired Magazine*, 23 February. <http://archive.wired.com/science/discoveries/news/2006/02/70280>.
- Ziewitz, Malte. 2016. 'Governing Algorithms: Myth, Mess, and Methods'. *Science, Technology, & Human Values* 41 (1): 3–16. <https://doi.org/10.1177/0162243915608948>.