From ranking algorithms to 'ranking cultures'

Investigating the modulation of visibility in YouTube search results

Rieder, B.; Matamoros-Fernández, A.; Coromina, Ò.

Link to publication

CONVERGENCE

$SAGE

From ranking algorithms to 'ranking cultures': Investigating the modulation of visibility in YouTube search results

Bernhard Rieder
University of Amsterdam, the Netherlands

Ariadna Matamoros-Fernández
Queensland University of Technology, Australia

Òscar Coromina
Universitat Autònoma de Barcelona, Spain

Abstract

Algorithms, as constitutive elements of online platforms, are increasingly shaping everyday sociability. Developing suitable empirical approaches to render them accountable and to study their social power has become a prominent scholarly concern. This article proposes an approach to examine what an algorithm does, not only to move closer to understanding how it works, but also to investigate broader forms of agency involved. To do this, we examine YouTube's search results ranking over time in the context of seven sociocultural issues. Through a combination of rank visualizations, computational change metrics and qualitative analysis, we study search ranking as the distributed accomplishment of 'ranking cultures'. First, we identify three forms of ordering over time – stable, 'newsy' and mixed rank morphologies. Second, we observe that rankings cannot be easily linked back to popularity metrics, which highlights the role of platform features such as channel subscriptions in processes of visibility distribution. Third, we find that the contents appearing in the top 20 results are heavily influenced by both issue and platform vernaculars. YouTube-native content, which often thrives on controversy and dissent, systematically beats out mainstream actors in terms of exposure. We close by arguing that ranking cultures are embedded in the meshes of mutually constitutive agencies that frustrate our attempts at causal explanation and are better served by strategies of 'descriptive assemblage'.

Corresponding author:
Bernhard Rieder, University of Amsterdam, Turfdraagsterpad 9, Amsterdam 1012TX, the Netherlands.
Email: rieder@uva.nl

Introduction

Algorithms are entangled with everyday life. We interact with them when we use technology to search for information, buy and sell products, learn or socialize. Algorithms are the product of complex processes that influence our choices, ideas and opportunities (Gillespie, 2014). However, we often know little about how they actually work and what factors inform their performativity. The frictions stemming from increased algorithmic management of information and practice, coupled with an inadequate knowledge of operating principles, have put the issue at the centre of public debates. Discussing rising political polarization, German Chancellor Angela Merkel recently argued that algorithms 'can lead to a distortion of our perception' and urged social media platforms to be more transparent about their technological mechanisms (Connolly, 2016). With the alleged role of 'fake news' in the 2016 US presidential elections, this debate has intensified and broadened beyond the scope of social media (Roberts, 2016).

However, the implications of algorithms – in particular those that filter, hierarchize or recommend – for matters of politics and culture have been addressed by scholars since the popularization of the Internet. Early work focused on search engines and examined the stakes and power relations involved in the hierarchical organization of information (Introna and Nissenbaum, 2000), leading to the emergence of the continuously active field of Web search studies (Zimmer, 2009). More recently, academic work has begun to widen its inquiry into 'the increasing prominence and power of algorithms in the social fabric' (Beer, 2009: 994). Social media platforms, in particular, have been singled out as constitutive actors of social exchange (Gillespie, 2010) that often confer to algorithms the task of modulating and curating the flow of contents, ideas and sociability (Bucher, 2012; Gillespie, 2014). Within these fields, the question of algorithmic transparency and accountability (Machill et al., 2003; Pasquale, 2015) has been of particular interest, due to possible real-world consequences such as discrimination (Edelman and Luca, 2014) and the reproduction of power structures (Edelman, 2014).

A recent priority has been to develop suitable methods to study algorithms and their outcomes empirically. Sandvig et al. (2014) have suggested different audit study designs as a starting point, advocating for 'regulation toward auditability' that would enable researchers to routinely probe platforms by impersonating users, scraping platform data or receiving direct access to algorithms. However, companies are reluctant to allow direct scrutiny of their technical infrastructures, largely because technical arrangements are tightly linked to business models (Helmond, 2015) and there is fear that systems could be 'gamed' (Crawford, 2015). Furthermore, the technical arrangements involved in algorithmic outcomes would be too complex to be meaningfully relayed to broader audiences (Nissenbaum, 2011). Scholars have therefore focused on methods for producing algorithmic accountability from the 'outside', taking the form of systematic small-scale observation (Bucher, 2012) or more substantial strategies such as 'scraping audits' (Sandvig et al., 2014) and reverse engineering (Diakopoulos, 2015). The latter can indeed bring to the surface editorial criteria, which have been demonstrated in research on Google Search's autocompletions (Diakopoulos, 2013) and autocorrections on the iPhone (Keller, 2013). However, the complexity of large-scale systems, where (probabilistic) algorithms intersect with massive amounts of possible

variables, makes it difficult to clearly unpack operating principles, and reverse engineering techniques will almost certainly miss inputs that are virtually impossible to recreate, such as built-in randomness or temporal drifts (Diakopoulos, 2015).

With these limitations in mind, we propose a method that follows a 'scraping' approach (Sandvig et al., 2014) and observes what an algorithm *does* – not only to move closer to understanding how it works, but also to investigate broader forms of agency involved. To do so, we focus on YouTube search results. YouTube, launched in 2005 and acquired by Google Inc. a year later, is the second most visited site in the world (Alexa, n.d.), with more than a billion users watching hundreds of thousands of hours of content every day (YouTube, n.d.). Early media research on YouTube focused on its potential as a tool for empowering participatory culture (Jenkins, 2006) and on the emergence of new media stars/entrepreneurs within its grassroots culture (Burgess and Green, 2009). As the platform grew in popularity, critical approaches interrogated the site's function as a database that collects, indexes and presents information (Kessler and Schäfer, 2009) and the uneven power balance between ordinary users and corporate media in the curation of YouTube content (Gehl, 2009). Our aim is to contribute to studies of YouTube as an influential source of information (Murugiah et al., 2011) by interrogating its search function, which we approach as a socio-algorithmic process involved in the construction of relevance and knowledge around a large number of issues. Previous research has examined how algorithms discover and present new content to users, but this work focuses mostly on the platform's recommender system (Airoldi et al., 2016). The search feature, however, is an area where the process of mediation or curation of content and, consequently, of perspectives or viewpoints, becomes highly explicit. The entanglement of algorithmic work in the attribution of relevance is clearly visible: the arrangement in the form of ranked lists reinforces the idea that some contents deserve more prominence than others. Through the visibility given to metadata such as view, like and comment counts, this competitive process is extended throughout the platform and becomes an influential element in information diffusion as users take such indicators into account when viewing and sharing (Shifman, 2013). This user engagement, in turn, constitutes an important part of the material inputs to YouTube's algorithms, calling for an understanding of ranking as a complex process unfolding over time.

We account for these complicated and circular instances of feedback by conceptualizing YouTube as a multisided platform orchestrating the relationships between different actors, from end users to advertisers. These relationships are enacted concretely as an 'ensemble of complex interactions between [. . .] actors taking place every time a search is launched' (Rieder and Sire, 2014) in the form of algorithmic operations on sets of data signals representing the different practices involved. We therefore consider the production of outcomes as an entanglement between platform politics (Gillespie, 2010) and cultures of use. However, our method does not propose to sort different factors into the neat categories of 'technical vs. social' or 'platform vs. users' but focuses on observing the actual outcomes of ranking and how they change over time in the context of specific sociocultural issues. We thus situate the study of 'an algorithm' in the particular setting of the YouTube platform, its use cultures and the issues it mediates. To achieve this goal, we organize our investigation around the concept of 'ranking cultures' – which we define as unfolding processes of hierarchization and modulation of visibility that call on users, content creators and a platform that intervenes and circumscribes in various ways – in order to address a number of questions. How can we make ranking observable and amendable to empirical study? How does ranking play out for specific sociocultural issues? Are there recurring 'morphologies' in ranking and how can we describe them? To what extent do YouTube search results reflect different use

cultures? We approach these questions through the lens of a set of selected queries, discussed in detail after an initial presentation of our methodology.

Approach and methodology

Our empirical approach takes YouTube as an integrated, dynamic whole where various instances of technicity, including ranking and recommendation mechanisms, take part in hosting and orchestrating massive numbers of interactions between different groups of actors. To begin delineating YouTube's 'relevance' search algorithm, which provides an ordered list of videos in response to a user query, we can think about it as a two-stage process: retrieval and ranking.[1] In the first stage, all pertinent videos for a certain query are identified. This could be based on direct text matching of video titles and descriptions, or other elements such as user comments and automatically recognized speech and image contents. Since the list of matching items can be huge, ranking is crucial. There are many different approaches for ordering result lists. But on a very basic level, one could distinguish between factors that are query independent and factors that draw on the query and its context.[2] Query independent factors could be accounted for relatively easily: Views, likes, watch time or channel subscribers, potentially on a per time frame basis, are straightforward metrics when it comes to deciding whether a video has more or less 'relevance' than others. However, even if we limit the notion of 'query' to the actual characters typed into a search box, discounting personalization or similar contextual techniques, things become more complicated. The above-mentioned metrics could be interpreted on a per query basis in the sense that only views from users searching with the same query are counted. Or channels that have previously posted videos deemed relevant for the query are given a boost. Or automated content analysis attempts to infer query relevance from the video itself. Or YouTube may actually draw videos from buckets relying on different factors, according to how it categorizes a specific query at a given point in time.

But even if one could specify these complex factors, we need to recognize that a static perspective that frames a ranking algorithm as a stable relation between inputs and outputs is not adequate for a platform of the size and technical sophistication as YouTube. On the one side, users react to rankings in various ways, by clicking, watching and – in the case of content creators – adjusting their strategies to what is considered effective at a given moment. On the other side, algorithmic techniques increasingly espouse a probabilistic and experimental outlook where parameters are constantly tested against a desired outcome and adapted according to user behaviour. An algorithm could, for example, forecast the success of a newly posted video, rank it favourably and relegate it quickly if it fails to perform. When algorithms no longer explicitly specify a decision model, but draw on user feedback to inductively generate such a model, there is a fundamental problem of 'interpretability' that persists even when full access to a system is provided (Burrell, 2016). Reverse engineering is simply not a viable strategy in such a setting.

We therefore consider ranking as a distributed and dynamic accomplishment that includes the contributions of various actors and proposes to analyze ranking *outcomes* by following a strategy of descriptive assemblage, 'where processes of creativity, conceptual innovation, and observation can be used to mobilize novel insights' (Savage, 2009: 170), resisting the urge to identify 'hard' moments of causality. This does not mean that we consider technical components to be fundamentally *unknowable*, especially since YouTube itself used to highlight watch time, recency, duplication and query pertinence as the central factors[3] for its 'relevance' ranking mechanism, but that these components need to be considered as part of a broader setting.

This understanding of ranking as a dynamic process undergirds our specific interest in change over time and in ranking 'morphologies', recurring patterns of change. The description of these patterns situates technical factors within the specific issues and publics that actual queries connect with. We thus move the focus of analysis from ranking *algorithms* to ranking *cultures*, that is, the distributed and heterogeneous agencies that converge in producing actual result lists. This includes what Gibbs et al. (2015) call 'platform vernaculars', the specific communication practices emerging within particular online services, as well as potential 'issue vernaculars' developing around specific subject matters or communities. While we would not go as far as to describe our approach as 'platform ethnography' (Geiger and Ford, 2011), we take inspiration from approaches combining computational analysis with qualitative investigation.

Data collection

Since our investigation postulates that ranking cannot be easily detached from the specificities of concrete issues, our methodology revolves around the observation of search results – and their variation over time – for a set of selected queries. To retrieve data, we made use of the YouTube Data Tools' 'Video List' module (Rieder, 2015), which uses the 'search: list' endpoint of the YouTube API v3.[4] We set the 'order' parameter to 'relevance', which is the default setting for user-facing search on YouTube's website and mobile applications – and therefore the specific ranking procedure our study investigates. Subsequently, the capture module was automatically called on 44 days at the same time to collect the 300 top results from 9 June to 22 July 2016.[5] For this article, we concentrated on the top 20 results, the same amount as the first page of a standard YouTube search on the Web. While the API approach has the advantage of rendering data capture and scheduling particularly convenient, it fails to exactly reproduce the search experience of concrete end users because it bypasses localization and personalization. Although sporadic comparisons between our API data and results obtained through several Web browsers just yielded small variations in result ordering, only a much more extensive inquiry could fully account for the issue. But problems such as this one fundamentally affect what Marres and Gerlitz (2016) call 'interface methods', that is, methods that navigate the 'uncanny' spaces between unstable, dynamic, commercially oriented and 'methodologically biased' digital platforms on the one side and traditional concepts and methods for social research on the other. The tensions stemming from this peculiar constellation will remain for the foreseeable future and researchers have to navigate the analytical possibilities and limitations creatively in the absence of easy ways to recreate methodological stability. For our case, we acknowledge that localization and personalization may play important roles on YouTube, but take the API data to be as close to a 'baseline' perspective as one can get.

Our data gathering procedure was applied to a set of manually selected queries: [gamergate], [islam australia], [islam], [trump], [sanders], [refugees] and [syria]. This selection was made to focus on contemporary and controversial issues where one can easily argue that search results pertain to politically charged debates and participate in defining how these issues are being collectively understood. This choice follows the *cartography of controversies* tradition in looking at 'situations where actors disagree' (Venturini, 2010: 261) and the production of social life can be observed through a high-contrast lens. While it would certainly be interesting to apply our method to less heated sections of YouTube, such as how-to videos, we were actively seeking queries where we could reasonably expect to find significant change over time, potentially in response to ongoing events. With the exception of Gamergate,[6] which had already stabilized before we began collecting

data, all of the selected terms denoted active controversies that were both perforated by current events and highly polarized, making the analysis of rank change particularly salient. Our selection also reflected the existing expertise of the three authors, helping us to approach the data more confidently. We recognize that our ad hoc queries do not enable us to make general claims about YouTube ranking cultures, but they allowed us to develop and test a methodology that we hope to systematically extend in future research.

In order to contextualize captured rankings with a measure of public subject interest, we relied on data from Google Trends,[7] a service that provides information about search volume over a predefined period of time for four Google services: Web Search, Product Search, News and YouTube. While the data offered are not actual search numbers, but normalized to the highest value for the period,[8] they support the analysis of variation over time. We were thus able to analyze search volume fluctuation on YouTube for each of our case studies.

When it comes to the ethical dimension of our research, we decided not to anonymize channel names, neither in our basic data nor for the purpose of publication. Contrary to Facebook, YouTube does not mandate the use of real names; more importantly, the queries we chose did not open onto spaces of intimate conversation, where expectations of privacy raise legitimate concerns, but onto channels with often large numbers of subscribers[9] addressing a broad audience. While YouTube certainly retains aspects of a social networking service, our queries highlighted a large ecosystem of channels adhering to a logic of broadcasting reminiscent, in certain ways, of talk radio. There is considerable public interest in understanding this emergent media constellation.

Data analysis

The collected data provided a daily ranking of videos for each query and a number of metadata for each video, including its title, channel, description, category, publishing date and length, as well as counts for views, likes, dislikes and comments. In order to analyze these data, we combined descriptive and investigative strategies. Using the RankFlow tool (Rieder, 2016), which creates flow diagrams that make changes in rank easily analyzable, we were able to inspect the data in a form that is detailed and non-reductive, since no data aggregation is performed, but also conducive to synthetic apprehension, because patterns can be identified and examined quickly from the provided overview. These visualizations provided both the means to identify specific morphologies of change on a higher level and the starting point for in-depth investigations around particular moments in time, allowing us to 'navigate' (Latour et al., 2012) through the 7262 videos that appeared in the top 20 results for our seven queries during the time of observation. Following Tukey's (1977: 1) famous statement that '[e]xploratory data analysis is detective work – numerical detective work – or counting detective work – or graphical detective work', we iterated between the visualizations, where we mostly focused on view counts as quantitative components, and a subject-centred and temporally contextualized scrutiny of actual contents. The former provided a broader context highlighting instances and patterns of change and stability, whereas the latter allowed us to understand and qualify the meaning of these changes in cultural terms. While the controversies behind our queries often neatly arranged around a 'pro' and 'con' stance, our goal was not to measure this relationship quantitatively, heeding Latour et al.'s counsel to not 'go from the particular to the general, but from particular to *more* particulars' (2012: 599). Rather, we were interested in characterizing the specificities of stability and change. Which videos stay on top over longer periods? What kind of content comes in during moments of change? Can these changes be related to specific events? In answering these questions, we did not follow a comprehensive

classificatory scheme, but engaged in 'a concertina-like process of ever more elaborate description' (Savage 2009, 156) that would reflect the particularities of the specific case. This iterative form of descriptive assemblage allowed us to take algorithmic outcomes as the starting point for establishing connections between platform technicities, user practices and creator tactics – central elements in what we address as *ranking cultures*.

On a second level, we explored possibilities to study variations in rank in more formal ways to facilitate comparison between different queries and to prepare for the scaling-up of our methodology to a much larger set of queries, where the investigation of individual videos would no longer be practical. Limiting ourselves, in this article, to a contained and in-depth approach allowed us to test the methodological potential and feasibility of a purely computational strategy. While the field of information science, in particular, provides many techniques for quantifying change, mainly to study the effects of technical adjustments on information systems, we were interested in approaches that measure changes in rank, not just presence/absence, and weigh changes at the top of the list more heavily than changes further down. The rank-biased distance (RBD) metric (Webber et al., 2010) satisfies these criteria and provides a parameter to adjust how heavily variation at the top of the list should be counted.[10] We implemented the algorithm in the RankFlow tool used to produce our visualizations, resulting in a quantitative description of change between every two lists – every 2 days, in our case – in the form of a single metric d ($0 \leq d \leq 1$). To measure the 'changiness' for a query over the full 44 days, we calculated both the average and variance of all d values. While the former yields a broad appreciation of how much overall change in rank there was, the latter indicates whether we are dealing with a more evenly distributed movement of reordering (low variance) or rather with a combination of moments of stability interrupted by radical shifts (high variance).

Although we did not correlate RBDs systematically with search volume gathered from Google Trends, due to doubts concerning temporal congruence that we discuss further down, we confronted the two variables around specific moments in time, adding the number of videos published on a specific day as a third factor. While we did not expect to find hard causal links in a context where many variables interact dynamically, we were hoping to begin to understand relational patterns on a broader level.

Findings

Relying on a combination of visual inspection and average RBD (avRBD) values, we distinguished three general types or morphologies of rank change over time among our case studies: stability in ranking over the entire period studied, stability with 'newsy' interruptions and fully 'newsy' patterns, where results change at very high speed. We found, however, that platform and issue cultures associated with each case left strong marks on result lists and the more qualitative examination of videos and actors proved to be highly instructive for understanding *what* was actually being ranked in the first place. In the following sections, we describe the results for each of our seven queries from more to less stable morphologies, followed by a discussion of the different agencies that played out in these ranking cultures.

Stable morphologies over long periods of time

The rank morphology for the query [islam australia] (Figure 1) was the most stable of our case studies. The video ranking first for almost the entire period contained bootlegged content from the
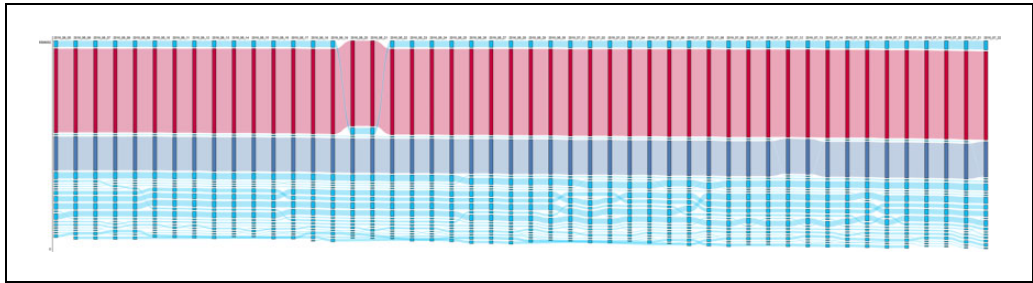
Figure 1. RankFlow visualization of [islam australia], which shows very little variation over time (avRBD 0.02). In a RankFlow chart, columns represent days and blocks individual videos, with the top ranking video on top of the column. Colour (blue to red) and bar height both indicate the view count for each video. avRBD: average RBD; RBD: rank-biased distance.

Australian TV channel 7 News, depicting violent riots against an anti-Islam video that was uploaded by an 'amateur'[11] channel – now suspended for repeated copyright infringement – dedicated to posting anti-Islamic content. While the second video, a social experiment showing how public hatred against Muslims was systematically opposed by random bystanders, had many more views (2,655,456 vs. 204,577 on 9 June), it had fewer comments (2113 vs. 2616 on 9 June), a difference that grew significantly during our observation period (2069 vs. 4100 on 22 July).[12] Overall, there was a balance between anti- and pro-Islam sources, with most of the more obviously partisan content coming from amateur – but highly followed – channels (e.g. *AustralianNeoCon1*, *NewMan2015*). The presence of official news sources was minimal, with the exception of a number of informative videos from the *Australian Broadcasting Corporation* (*ABC*) *News*, *Al Jazeera* and *Russia Today* (*RT*). While a series of videos with high view, like or comment counts appeared in the top ranks, none of these factors are clearly correlated with position in the result list. The same goes for recency, since all of the videos that divide the top five positions among themselves were published in 2015 or earlier.

The results for the query [gamergate] show similar stability, but skew even further toward 'alternative' or YouTube-native content (Figure 2). The persistently top-ranked video featured former Breitbart technology editor Milo Yiannopoulos on *The Rubin Report*, a fan-funded[13] talk show for those who 'care about free speech' and are 'tired of political correctness'.[14] Below, videos by well-known 'alt-right' YouTube stars,[15] such as *Amazing Atheist*, *Thunderf00t* or *Sargon of Akkad*, 'explaining' the Gamergate controversy dominate the rankings. The overwhelmingly pro-Gamergate results are complemented with videos from more traditional conservative sources, such as the American Enterprise Institute, a conservative think tank. One exception is the video with the highest view count, which features Anita Sarkeesian, the preferred target of Gamergaters, on Stephen Colbert's show. Interestingly, this video never ranks above place 14. Similarly to [islam australia], mainstream media had only minimal presence in this query, with the isolated example of a video from *ABC News* also featuring Sarkeesian.

Stable morphologies with 'newsy' interruptions

For the query [refugees], YouTube returned a relatively stable morphology in respect to videos with the highest view counts (Figure 3). However, it presented more variance with regard to videos with fewer views; some of them quickly appeared, disappeared and changed positions over time.
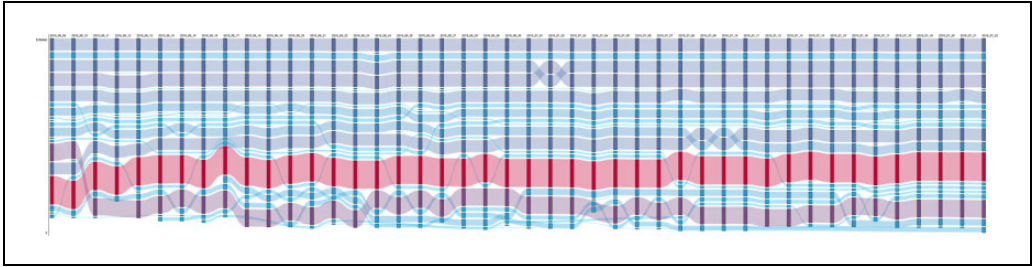
Figure 2. RankFlow visualization of [gamergate] again showing very little variation (avRBD 0.03). avRBD: average RBD; RBD: rank-biased distance.
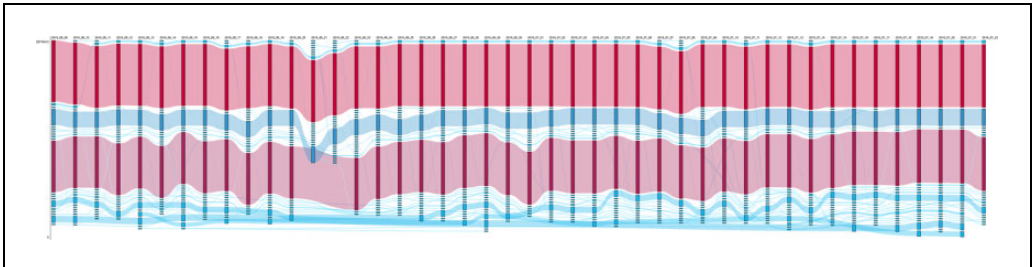

Figure 3. RankFlow visualization of [refugees] showing a general pattern of stability, with a significant perturbation around 21 June (avRBD 0.31). avRBD: average RBD; RBD: rank-biased distance.

The two top ranking videos for almost the entire period were uploaded by the independent broadcasting channel Journeyman Pictures in April 2016 and by the very popular 'explainer' channel *Kurzgesagt – In a Nutshell*, which creates a video per month about current affairs, in September 2015. These top videos were briefly pushed down, for example around World Refugee Day (20 June), where videos uploaded 1 or 2 days before made it briefly to the top, such as a video of Pope Francis talking about refugees that occupied the first position on 21 June, but already left the top 22 days later. The video with the second highest view count featuring John Oliver never made it above ninth place. Ranking higher, occupying the fourth position for almost the entire period, was a video stereotyping refugees as violent, uploaded by successful amateur content creator *Black Pigeon Speaks*. In comparison to the queries [islam australia] and [gamergate], this query shows a strong presence of both mainstream and fringe media channels (e.g. *Al Jazeera English*, *VICE News*, *Ruptly TV*, *BBC News*, *Rebel Media* and *The Alex Jones Channel*) and a higher variety of actors uploading content.

In the query [syria] (Figure 4), we observed even more variation involving videos with fewer views, which often reach top positions. The two videos with the highest view count ranked rather low (generally below sixth position) and presented less stability, sometimes even disappearing from the top 20. Ranking high were videos from nontraditional, non-western news sources that produce content between information, propaganda and what we refer to as 'warporn': sparingly edited footage of war machinery in operation. These videos often – but not always – received relatively few views and were uploaded by channels called *WarLeaks*, *R&U Vid* or *SouthFront*. Videos with higher views were didactic explanations about the Syrian war, from official media channels like *VOX* and *VICE*. Ranking also changed in response to issue events, with explainer
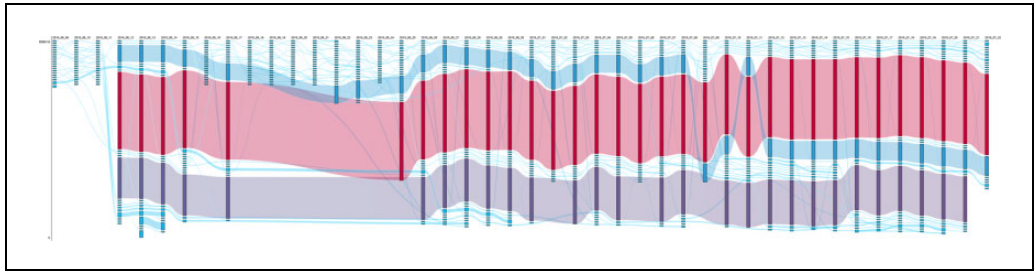
Figure 4. RankFlow visualization of [syria], characterized by a balance between stability and variation (avRBD 0.47). avRBD: average RBD; RBD: rank-biased distance.
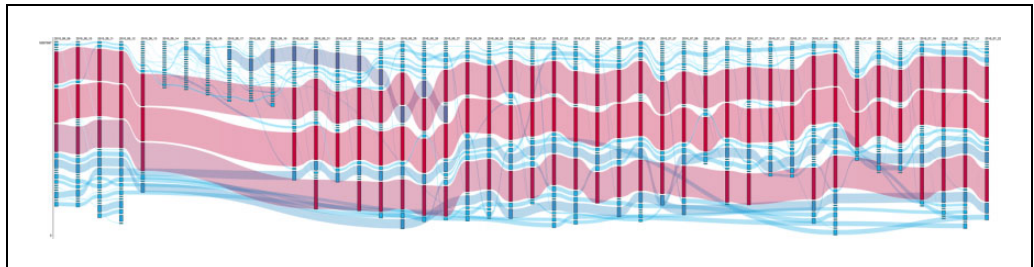

Figure 5. RankFlow visualization of [islam] showing already much higher levels of variation (avRBD 0.71), although a number of videos still show up regularly. The period after 13 June is particularly turbulent. avRBD: average RBD; RBD: rank-biased distance.

videos ranking lower, replaced by videos from both western (*Fox News*, *CNN*, *BBC News*) and non-western (*Press TV*, *RT*, *Al Jazeera*, *Ruptly TV*) news channels. These news videos, however, did not necessarily get more views than the many warporn videos populating the top of the rankings in more quiet periods, which indicates how a quite specific YouTube subculture can dominate a query as broad as this one.

The rank morphology for the query [islam] (Figure 5) shows even more variation over time, although a number of videos with high views still appeared regularly, even if they often remain below the 10th position. In general, videos ranking first were uploaded by a wider variety of sources, from parody channels to news channels, Islamic channels and antagonistic voices. During news events, however, videos ranking first were not from mainstream media channels, as with [syria], but recently uploaded clips by YouTube stars with anti-Islam views (e.g. *Sargon of Akkad, Steven Crowder*, *Cult of Dusty* or *Paul Joseph Watson*). For example, after the Orlando shooting on 12 June, YouTube-native content equating Islam with danger and LGBTQ hate dominated. But from 20 June onwards, videos debunking stereotypes about Islam, recently uploaded by Muslim YouTube stars and educational Islamic channels (e.g. *Joey Salads*, *Anes Tina, Kuku, FreeQuranEducation* or *The Review of Religions*), briefly occupied top positions. A controversial video uploaded by *Joey Salads* on 16 June about a social experiment showing how people were more scared about 'radical' Islam than 'radical' Christianity accumulated a high number of views and comments and held a prominent position (up to third) until disappearing on 27 June. The popularity of this video was noticed by the press, which pointed out the author's strategy to tag the video with 'Orlando shooting', 'Orlando gay club' and 'Orlando massacre'
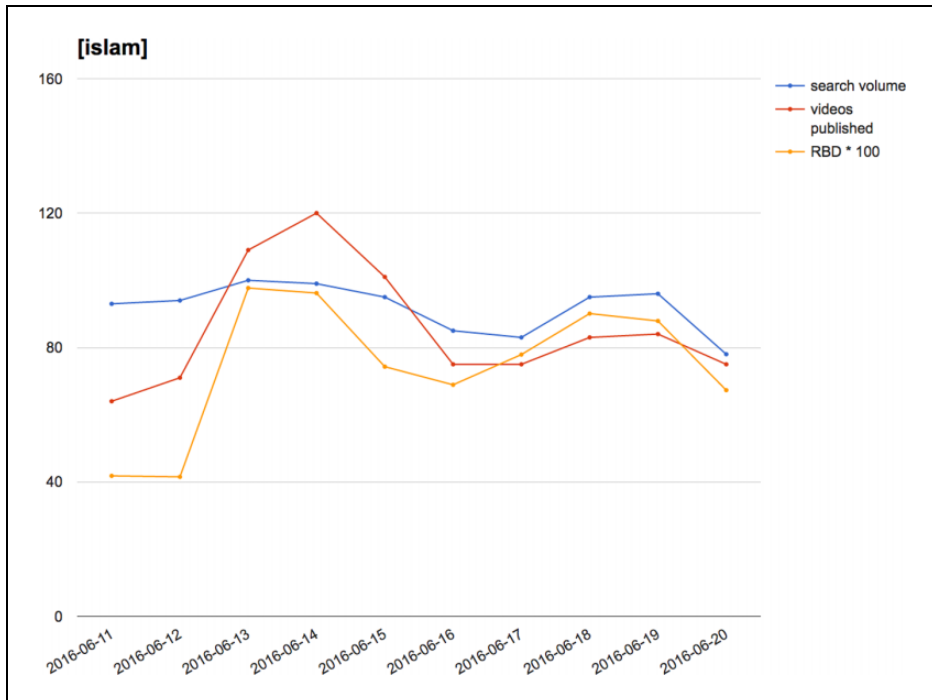
Figure 6. Search volume, published video count and RBD for [islam], 11–20 June. RBD: rank-biased distance.

(Scott, 2016). While we did not systematically investigate such forms of 'issue hijacking', this example illustrates the strategies content creators employ to gain views. Videos from mainstream media also appeared during these periods, but with a more timid presence and ranking lower than in the [refugees] and [syria] cases. During quiet periods, three videos with high views featuring more conciliatory content remained stable over time (e.g. '10 Lies You Were Told About Islam'), although the top of the list was still dominated by antagonistic voices (e.g. *The Rubin Report*, *Milo Yiannopoulos*, *Steven Crowder*). Moreover, this query featured another specific subculture during less turbulent phases, namely Islamic religious channels from different parts of the world (e.g. *Lampu Islam*, *Islamic Productions*) and Muslim YouTube stars (e.g. *Adam Saleh Vlogs*).

Since these morphologies seemed to switch between stable and newsy periods, we decided to dig deeper into two moments where such a shift happened. We selected one particularly significant period from both [islam] and [refugees] and compared the relationship between search volume (via Google Trends), the number of videos published and the RBD per day.

For [islam] (Figure 6), we found that RBD correlated significantly with search volume (0.337) and even more so with the number of videos published for the query (0.756). We can thus infer that current events drive both viewer interest and video production. When an event occurs, users not only search for relevant terms, but also producers make event-related videos that users receive through their channel subscriptions and actually watch due to the heightened attention. The search algorithm then picks up on this cumulative increase in interest, leading to changes in search ranking that privilege recently uploaded items.
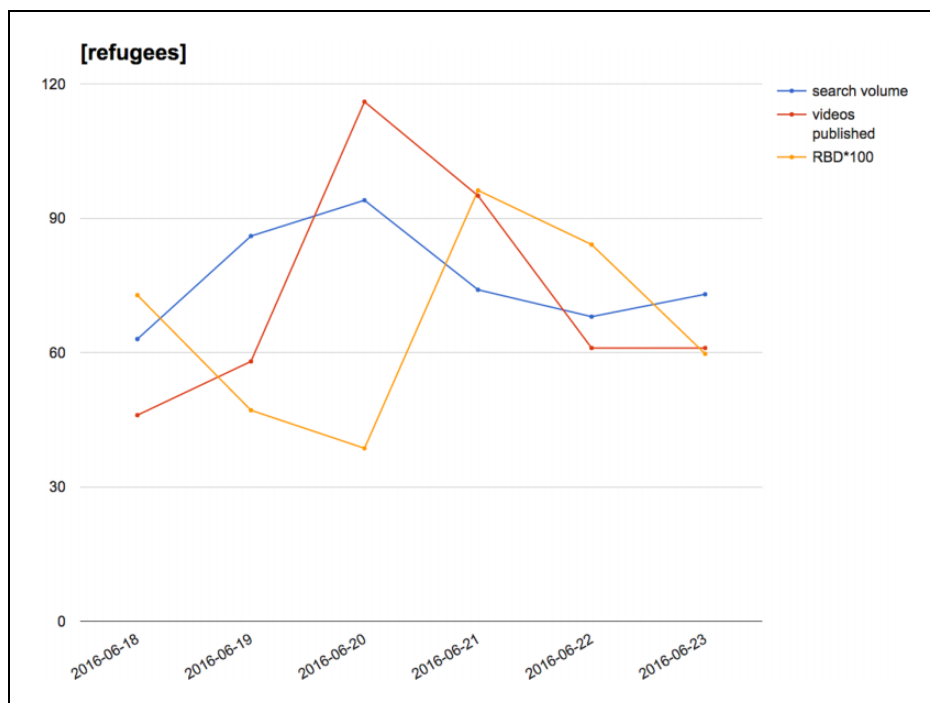
Figure 7. Search volume, published video count and RBD for [refugees], 18–23 June. RBD: rank-biased distance.

For [refugees] (Figure 7), however, we found *inverse* correlations between RBD with both search volume (−0.750) and number of videos published (−0.188). But this impression turned around completely when we shifted RBD values by a day: search volume (0.445) and number of videos published (0.947) suddenly correlated very strongly. How can we explain this apparent delay in rank change? First, we are confronted with the complicated question of the temporal synchronization between data sets that span different time zones. Second, whereas the case of [islam] is clearly affected by the Orlando shootings, the perturbation in [refugees] most probably responds to 20 June being World Refugee Day, an anticipated event that content creators engage with scheduled videos. Third, both search volume and videos published represent daily counts, whereas RBD is a measure of change between 2 days, each captured at a specific moment in time. While the adjusted data seem to confirm the correlation between rank change and both search interest and video production, we are reminded that the investigation of global platforms will necessarily run into seemingly trivial issues that can have important consequences. A more formalized methodology that simply relies on multivariate correlation or aggregate analysis would very probably not be able to pick up on these subtleties.

Fully 'newsy' morphologies

The results for the query [sanders] represented a fully 'newsy' morphology with constant changes over time (Figure 8). In the top ranks, we find recently uploaded videos featuring promotional
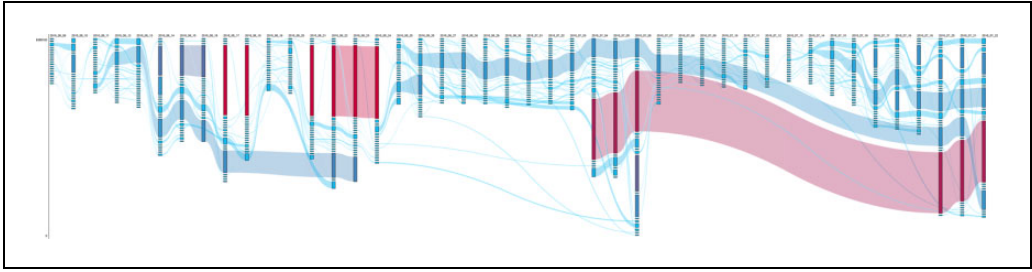
Figure 8. RankFlow visualization of [sanders], where variation reaches very high levels (avRBD 0.74). avRBD: average RBD; RBD: rank-biased distance.
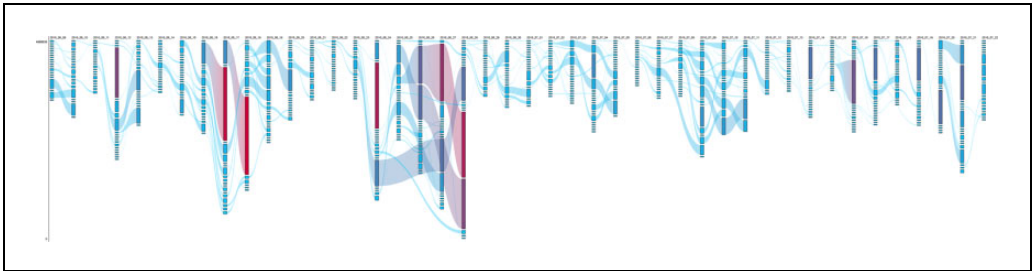

Figure 9. RankFlow visualization of [trump], with the highest level of variation among our case studies (avRBD 0.8). avRBD: average RBD; RBD: rank-biased distance.

material, such as press conferences, official speeches and interviews from news channels (e.g. *MSNBC, Fox News, CNN, The Young Turks*), entertainment shows (e.g. Jimmy Fallon and Stephen Colbert) and activist 'alternative' media leaning to the left (e.g. *Democracy Now!*, *The Humanist Report*). Although pro-Sanders voices largely influenced this morphology, channels with a clear pro-Trump agenda appear among the first positions (e.g. *wwwMOXNEWScom*). YouTube-native channels (e.g. *The Young Turks*, with more than three million subscribers) often dominated the top position by constantly uploading new content. The videos with the highest view counts were live streaming videos (uploaded by channels like *Live Streaming News* and *Fox 10 Phoenix*) that are no longer available, although the video with the second most views was again Stephen Colbert: his interview with Sanders appeared several times in the rankings but always at low positions.

The query [trump], finally, represented the most unstable morphology, with the top positions being highly influenced by recency (Figure 9). There was constant change and many different videos from a wide variety of actors appeared and disappeared quickly. Channels like *Fox News*, which dominated the rankings, and Trump-affiliated *Right Side Broadcasting* constantly uploaded new content and no video stayed in the top 20 longer than 4 days. Live streams, mostly published by *Right Side Broadcasting*, again accumulated high numbers of views but were not archived, making them unavailable to researchers. The same goes for bootlegged TV content, which often received considerable views before getting deleted. Parodies from known entertainers and YouTubers also received many views, but still left the rankings very quickly. Various videos (broadcast speeches, news items, talk shows, comedic content) came in sporadically. The morphology was largely influenced by news events (e.g. Brexit, Nice terrorist attacks) and generally right-wing dominated.

Table 1. Comparison between queries.

	islam australia	gamergate	refugees	syria	islam	sanders	trump
avRBD	0.02	0.03	0.31	0.47	0.71	0.74	0.8
Videos	496	483	969	1068	1317	1152	1777

avRBD: average RBD; RBD: rank-biased distance.

Discussion

When we interrogate these findings in terms of how they shed light on YouTube's 'relevance' algorithm, we find few direct links between basic variables such as view, like or comment counts and search results. There is little reason to think that a more thorough multivariate analysis would yield stronger connections, since (a) the one variable YouTube used to explicitly identify as a factor – watch time – is not available via the API and (b) query-dependent statistical factors such as 'videos that users searching for a term actually ended up watching' are impossible to reconstruct from the outside. Although an investigation of the considerable practical knowledge collected by content creators would certainly be worthwhile in that regard, our mostly descriptive method yielded nonetheless a number of interesting results, even if strong causality remains elusive.

First, the distinction of stable, newsy and mixed morphologies provides a way of thinking about forms of change over time, and the combination of visual inspection and change metrics (avRBD) provides a pathway toward both in-depth and comparative analysis. The most distinctive difference between periods of stability and change is the age of videos: in the former we often find videos that are several years old, whereas the latter is characterized by newly published clips that often quickly disappear from the top results. Our tentative analysis of the factors at play shows that search volume and, in particular, publication volume seem to be crucial in driving variation in rankings. Unsurprisingly, there is a close relationship between avRBD values and the absolute number of unique videos appearing for each query (Table 1): during our 44 days of observation over three times as many videos for [trump] made it into the top 20 than for [gamergate]. While we cannot infer precise mechanics from our findings, it is clear that YouTube's search mechanism is designed to pick up on what we have called 'newsy' moments, and whenever it does, results can change drastically from one day to the next. One could say that the algorithm switches from a Google Search mode to a Google News mode and back, following a different rationale for each. Practices like the instance of issue hijacking we mentioned when discussing [islam] show that content creators are at least partially aware of this dynamic and adapt their strategies accordingly.

Second, we have to recognize that search rankings may be heavily influenced by the site's other features for content discovery, namely recommendations and, most importantly, subscriptions. Although YouTube does not divulge how video traffic distributes over different pathways, the platform's interface now heavily privileges channel subscriptions as a means for content creators to build and address an audience. We know little about how this affects ranking, but this should be a reminder that any single algorithm operates in the context of a larger system it cannot easily be abstracted from.

Third, and maybe most importantly, we found that search result rankings may well oscillate between stability and change, but *what* is actually ranked is heavily influenced by both issue and platform vernaculars. There were important differences between the seven queries, but overall we were surprised by the strong presence of YouTube-native content, which ranges from bootleggers

to grassroots content networks, such as the highly active 'warporn' channels in [syria], to YouTube stars of various persuasions appearing across issues, to 'explainer' videos that are remarkably persistent, to professionally produced original shows like *The Young Turks* or *The Rubin Report*. These channels systematically beat established news media in terms of both subscriber count and exposure in the search rankings. While YouTube may not conform to dreams of horizontal communication and flat hierarchies, sitting on top, more often than not, is a new elite that thrives on controversy and dissent. The most striking exception from this rule are late night comedy shows, such as *The Colbert Report* or John Oliver's *Last Week Tonight*, that succeed in garnering large audiences when they address specific issues. While we cannot exclude that YouTube privileges native content in its ranking procedures, any deeper investigation into the broader hierarchies characterizing YouTube today would not only require an integrated perspective on ranking but also a way to reconstruct the historical growth of the channel ecosystem to its current form. Many mainstream media producers were latecomers on the platform and competing with channels having been active for up to a decade is proving difficult.

Taking these three elements together, we find that our mixed methodology, initially designed to study algorithmic ranking procedures, does indeed highlight the integrated nature of ranking *cultures*: when focusing on the actual videos returned as search results, the method almost flips on its head and, instead of investigating quantitative factors, we end up studying the peculiar content and authors through a lens that highlights their respective relevance or resonance (cf. Rogers et al., 2009).

Conclusion

The workings and specifications of algorithms involved in social, cultural and political decisions have recently moved to the centre of attention in areas such as media studies, philosophy and legal scholarship. The ability to audit algorithms (Sandvig et al., 2014), in order to understand and critique how they work and what they do, is being demanded not only by scholars, but also by a growing array of social actors. In this article, we have developed a methodology that first of all makes algorithmic *outputs* visible and offers a pathway toward describing them through a combination of visual analysis, computational measures and in-depth investigation of actual contents. Our main findings concern, on the one side, the striking shift between stability and variation over time, which shows that YouTube's search function is highly reactive to attention cycles, and, on the other side, the dominance of YouTube-native contents. Cases like the 'warporn' videos in the [syria] query and the overall presence of often (far) right leaning YouTube personalities show that the video platform arranges search ranking in a way that allows highly active 'niche entrepreneurs' to gain exceptional levels of visibility. Feeding on controversy and loyal audiences, these channels consistently appear in top positions, even if their videos most often receive fewer views than more mainstream or conciliatory voices. Contrary to the idea that simple popularity metrics and the search for the lowest common denominator drive visibility online, we were able to get a glimpse at complex ranking cultures that reward platform-specific strategies and audience activation through strongly opinionated expression.

Throughout the article, we have emphasized two connected methodological perspectives or caveats that we want to quickly restate. First, the search ranking algorithm may be a contained technical procedure, but it makes little sense to abstract it from the larger system it is embedded in. And, second, platform and issue vernaculars, content strategies and attention cycles flow through these complex front-end and back-end arrangements in ways that make them part of YouTube's

functioning in ways that go far beyond merely producing data for algorithmic input. Looking at a single instance of algorithmic work forced us to confront the realities of an intricate mesh of mutually constitutive agencies that frustrate our desire for causal explanation. Access to the mythical source code would not solve this problem. It would also do little to alleviate the even more fundamental problem that the ranking of viewpoints is not a problem to be solved but an ongoing conundrum and site of struggle. Our decision to examine and discuss YouTube's search function in the context of concrete sociocultural issues or controversies is – at least in part – inspired by doubts about forms of disembodied technological ethics that consider ranking to be something that can be discussed, in normative terms, independently from the very matters it acts upon.

To move forward, we propose to develop this work along three related axes. First, we hope to apply our methodology to a wider array of queries. There are many sections of YouTube, such as music or how-to videos, that are highly relevant in terms of view numbers and commercial significance. This would shed light on ranking in a wholly different context, where different social dynamics may yield different morphologies. Second, the formal part of our methodology, based on quantitative measures of change, could be easily applied to thousands of (popular) search queries, allowing for an extensive analysis of ranking patterns on YouTube. This could go as far as providing ways to delineate and map subject areas according to how prone to change they are. Third, our approach could be further extended in the direction of descriptive assemblage by including other constitutive elements of the YouTube platform, such as recommendation patterns, subscription dynamics, channel networks, production schemes and optimization strategies. Such a broader perspective would have to incorporate the company's strategic shaping of their platform, including advertisement policy, copyright enforcement and video deletion, for example, around issues such as sex, racism or self-harm. Only a perspective that situates the technical deeply in the social will be able to account for the effects of platform politics.

Notes

1. This discussion draws on standard concepts from the field of *Information Retrieval*. For an excellent overview, see Manning et al. (2008).
2. This could be elements such as time, location or operating system used, going as far as including entire user profiles. Personalization and localization are ways to 'enrich' the query with such contextual elements.
3. YouTube took this information offline in June 2015. An archived version can be found here: https://web.archive.org/web/20150329045851/https://support.google.com/youtube/answer/6046978?hl=en&ref_topic=6046759
4. https://developers.google.com/youtube/v3/docs/search/list
5. YouTube delivers up to 500 results for the 'search: list' endpoint, but since we were planning to focus on the top results, 300 results seemed largely sufficient. It is worth mentioning that YouTube's API has very few limitations and would allow for capturing daily search results for tens of thousands of queries. There are no guarantees concerning the sustainability of this considerable openness, but compared to other services, the YouTube API has been remarkably stable in recent years.
6. According to the two sides of the controversy, gamergate is either an organized online harassment campaign against women within the game industry or a movement preoccupied by a perceived lack of ethics in gaming journalism.
7. https://www.google.com/trends/
8. For the exact normalization procedure, see https://support.google.com/trends/answer/4365533?hl=en
9. Every channel we mention by name in this article has at least 10,000 subscribers.

10. Since users tend to stick to the first few search results, we selected a value of 0.8 for p, which means that changes in the first five results will represent 86% of the RBD value.

11. We define amateur channels as those lacking markings of corporate professionalism, such as a clear corporate identity or high production values.

12. The lower comment count on the later date reminds us of an important platform feature that allows content creators to delete comments from their videos.

13. Fan-funding platforms such as Patreon increasingly represent an important income stream. The Rubin Report received a monthly stipend of $24,263 from its Patreon supporters as of early February 2017.

14. https://www.youtube.com/user/RubinReport/about

15. We define YouTube stars as individuals or channels with very high numbers of subscribers (500,000+) that focus on YouTube for publishing content.

References

Alexa (n.d.) The top 500 sites on the web. Available at: http://www.alexa.com/topsites (accessed 31 January 2017).

Airoldi M, Beraldo D and Gandini A (2016) Follow the algorithm: An exploratory investigation of music on YouTube. *Poetics* 57: 1–13.

Beer D (2009) Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media and Society* 11(6): 985–1002.

Bucher T (2012) Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media and Society* 14(7): 1164–1180.

Burgess J and Green J (2009) *YouTube: Online Video and Participatory Culture*. Cambridge: Polity.

Burrell J (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data and Society* 3(1): 1–12.

Connolly K (2016) Angela Merkel: Internet search engines are 'distorting perception'. *The Guardian*. Available at: https://www.theguardian.com/world/2016/oct/27/angela-merkel-internet-search-engines-are-dis torting-our-perception (accessed 31 January 2017).

Crawford K (2015) Can an algorithm be agonistic? Ten scenes from life in calculated publics. *Science, Technology & Human Values* 41(1): 77–92.

Diakopoulos N (2013) Algorithmic accountability reporting: On the investigation of black boxes. In: *Nick Diakopoulos. Musing Media*. Available at: http://www.nickdiakopoulos.com/wp-content/uploads/2011/07/Algorithmic-Accountability-Reporting_final.pdf (accessed 31 January 2017).

Diakopoulos N (2015) Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism* 3(3): 398–415.

Edelman BG (2014) Leveraging market power through tying and bundling: Does Google behave anti-competitively? *Harvard Business School NOM Unit Working Paper No*. 14-112.

Edelman BG and Luca M (2014) Digital discrimination: The case of Airbnb.com. *Harvard Business School NOM Unit Working Paper No*. 14-054.

Gehl R (2009) YouTube as archive: Who will curate this digital Wunderkammer? *International Journal of Cultural Studies* 12(1): 43–60.

Geiger RS and Ford H (2011) Participation in Wikipedia's article deletion processes. In: *Proceedings of the 7th international symposium on Wikis and open collaboration*, Mountain View, CA, 3–5 October 2011, pp. 201–202. New York: ACM.

Gibbs M, Meese J, Arnold M, et al. (2015) #Funeral and Instagram: Death, social media, and platform vernacular. *Information, Communication & Society* 18(3): 255–268.

Gillespie T (2010) The politics of 'platforms'. *New Media and Society* 12(3): 347–364.

Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski P and Foot K (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: MIT Press, pp. 167–193.

Helmond A (2015) The platformization of the Web: Making web data platform ready. *Social Media and Society* 1(2).

Introna LD and Nissenbaum H (2000) Shaping the Web: Why the politics of search engines matters. *The Information Society* 16: 169–185.

Jenkins H (2006) *Convergence Culture*. New York: NYU Press.

Keller M (2013) The Apple 'kill list': What your iPhone doesn't want you to type. *The Daily Beast*. Available at: www.thedailybeast.com/articles/2013/07/16/the-apple-kill-list-what-your-iphone-doesn-t-want-you-to-type.html (accessed 31 January 2017).

Kessler F and Schäfer MT (2009) Navigating YouTube: Constituting a hybrid information management system. In: Snickars P and Vonderau P (eds) *The YouTube Reader*. Stockholm: National Library of Sweden, pp. 275–291.

Latour B, Jensen P, Venturini T, et al. (2012) The whole is always smaller than its parts. *British Journal of Sociology* 63(4): 590–615.

Machill M, Neuberger C and Schindler F (2003) Transparency on the Net: Functions and deficiencies of Internet search engines. *Info* 5(1): 52–74.

Manning C, Raghavan P, and Schtze H (2008) *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Marres N and Gerlitz C (2016) Interface methods: renegotiating relations between digital social research, STS and sociology. *The Sociological Review* 64(1): 21–46.

Murugiah K, Vallakati A, Rajput K, et al. (2011) YouTube as a source of information on cardiopulmonary resuscitation. *Resuscitation* 82(3): 332–334.

Nissenbaum H (2011) A contextual approach to privacy online. *Daedalus* 140(4): 32–48.

Pasquale F (2015) *The Black Box Society*. Harvard: Harvard University Press.

Rieder B (2015) *YouTube Data Tools*. Computer software. N.p. Web. Available at: https://tools.digital methods.net/netvizz/youtube/ (accessed 5 September 2017).

Rieder B (2016) *RankFlow*. Computer software. Available at: N.p. Web. http://labs.polsys.net/tools/rankflow/ (accessed 5 September 2017).

Rieder B and Sire G (2014) Conflicts of interest and incentives to bias: A microeconomic critique of Google's tangled position on the Web. *New Media and Society* 16(2): 195–211.

Roberts H (2016) Google made changes to its search algorithm that unintentionally made it vulnerable to the spread of fake news, sources say. *Business Insider*. Available at: http://www.businessinsider.com/google-algorithm-change-fake-news-rankbrain-2016-12 (accessed 31 January 2017).

Rogers R, Jansen F, Stevenson M, et al. (2009) Social research with the Web. In: *Association for Progressive Communications and Humanist Institute for Cooperation with Developing Countries (2009) Global Information Society Watch 2009*. Report, 49–56.

Sandvig C, Hamilton K, Karahalios K, et al. (2014) Auditing algorithms: Research methods for detecting discrimination on Internet platforms. In: *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*, Seattle, USA, 22 May 2014, pp. 1–23.

Savage M (2009) Contemporary sociology and the challenge of descriptive assemblage. *European Journal of Social Theory* 12(1): 155–174.

Scott F (2016) 'Disgusting' YouTube prank testing public reactions to 'Islamic vs Christian' terrorists. *The Daily Mail*. Available at: www.dailymail.co.uk/news/article-3641998/Joey-Salads-social-experiment-video-measuring-Islamic-vs-Christian-terrorists-uploaded-day-Orlando-shootings.html (accessed 31 January 2017).

Shifman L (2013) *Memes in Digital Culture*. Cambridge, MA: MIT Press.

Tukey J (1977) *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Venturini T (2010) Diving in magma: How to explore controversies with actor-network theory. *Public Understanding of Science* 19(3): 258–273.

Webber W, Moffat A and Zobel J (2010) A similarity measure for indefinite rankings. *ACM Transactions on Information Systems* 28(4): 1–38. New York: ACM.

YouTube (n.d.) Statistics. Available at: https://www.youtube.com/yt/press/statistics.html (accessed 31 January 2017).

Zimmer M (2009) Web search studies: Multidisciplinary perspectives on Web search engines. In: Hunsinger J, Klastrup L and Allen M (eds) *International Handbook of Internet Research*. Dordrecht: Springer Netherlands, pp. 507–521.

Author biographies

Bernhard Rieder is an Associate Professor of New Media and Digital Culture at the University of Amsterdam and a collaborator with the Digital Methods Initiative. His research focuses on the history, theory, and politics of software and in particular on the role algorithms play in social processes and in the production of knowledge and culture.

Ariadna Matamoros-Fernández is a PhD candidate in the Digital Media Research Centre (DMRC) at the Queensland University of Technology. She holds an MA from the Digital Methods Initiative, University of Amsterdam, and a BA in Journalism from the Autonomous University of Barcelona. Her research explores the entanglement between technology and users' practices in the cultural dynamics of race and racism online.

Òscar Coromina is an Associate Professor and researcher at the Department of Audiovisual Communication and Advertising at Universitat Autònoma de Barcelona. He holds a PhD in media content in the digital age and his research is focused on the mediation of political disputes through social media.