

Beyond collected data: Politics of APIs on social media platforms^a

Para além dos dados coletados: Políticas das APIs nas plataformas de mídias digitais

CARLOS D'ANDRÉA^b

Universidade Federal de Minas Gerais, Graduate Program in Communication Studies. Belo Horizonte – MG, Brazil

ABSTRACT

This article discusses the conceptual and methodological issues faced during empirical research based on data collected via Application Programming Interfaces (APIs) of online platforms. Based on Platform Studies, digital methods, and other approaches anchored in Science and Technology Studies (STS), this article examines the entanglements between the politics of APIs and dimensions, such as affordances, governance, datafication, and algorithmic mediations in social media platforms (e.g., Twitter and Facebook). Material, political, normative, and economic aspects are discussed by examples ranging from the first APIs implemented in the 2000s to recent challenges, especially those following the Cambridge Analytica scandal.

Keywords: Social media platform, digital methods, datafication, algorithms, APIs

^aThis article is part of a research project funded by CNPq (Edital Universal 2016 – 407462/2016) and CAPES (postdoc scholarship).

^bProfessor at the Graduate Program in Communication Studies at the Universidade Federal de Minas Gerais (UFMG), Brazil. Coordinator of the research group R-EST – studies on socio-technical networks. Orcid: <https://orcid.org/0000-0001-7328-6714>. Email: carlosfbd@gmail.com

RESUMO

O artigo discute questões conceituais e metodológicas a serem enfrentadas por pesquisas empíricas baseadas em dados obtidos através de *Application Programming Interfaces* (APIs) de plataformas *online*. Em diálogo com os Estudos de Plataforma, métodos digitais e outras abordagens ancoradas nos estudos de ciência e tecnologia (STS), o artigo volta-se para a relação entre as políticas das APIs e dimensões como *affordances*, governança, datificação e mediações algorítmicas em mídias sociais como Twitter e Facebook. Aspectos materiais, políticos, normativos e econômicos são discutidos por meio de exemplos que vão da implementação das primeiras APIs, nos anos 2000, aos desafios recentes, em especial após o escândalo Cambridge Analytica.

Palavras-chave: Plataforma de mídia social, métodos digitais, datificação, algoritmos, APIs

INTRODUCTION

PARTIAL ACCESS TO the vast number of digital traces stored by platforms such as Facebook, Twitter, and YouTube, and the diversity of applications that enable data extraction, processing, and visualization have contributed to significant growth in the use of data provided by these social media platforms in academic and market research since late 2010s. Via their Application Programming Interfaces (APIs), online platforms offer access to some of the data they generate and/or collect elsewhere, such as music genres attributed to an artist (Spotify), posts linked to a term or hashtag (Twitter), or the total number of views, likes, and comments, up to a given moment, for videos posted by a channel (YouTube). Platforms such as Instagram (in 2016) and Facebook (in 2018) have restricted the access to their public APIs¹, underline the political and epistemological implications of platform data politics, previously discussed by authors such as Bucher (2013) and Puschmann and Burgess (2014).

¹ Following authors such as Bruns (2019), in this paper we adopt the term “public API” to denote resources that allow unpaid access to data provided by private platforms.

Bodle (2011) explains that by standardizing procedures for information access and exchange between computer systems, an open API “provides information to third-party applications through ‘calls,’ a technique of retrieving data on a server in the background, without disrupting the display and function of a web page” (p. 322). According to Helmond (2015), the widespread adoption of APIs by web services throughout the 2000s is the central pillar of the process of “platformization of the web”, that is, the consolidation of online platforms as an infrastructural and economic model that, based on controlled data exchange, opposes the more decentralized model of the World Wide Web in the 1990s.

Facilitated data access via public APIs can be considered a milestone in the use of “computational methods” (Vimieiro & Bargas, 2019) by researchers interested in social media platform dynamics. In this context, Venturini and Rogers (2019) identified the emergence of “API-based research,” defined as “a type of research based on information collected by social media platforms and made available through standardized commands to query, filter, format, and download such information” (pp. 1-2).

One could say that the popularization of tools and procedures based on APIs affordances has opened very promising research fronts in social sciences and humanities, especially in conjunction with other areas of knowledge, such as computer science and design (data visualization). However, the relative ease of data extraction and the relief of having “the” data collection related

to the topic at hand often results in the invisibility of the technological, political, and economic aspects of an online platform.

Seeking to contribute to approaches that stress positivist analysis based on the supposed objectivity of data (van Dijck, 2017), this study assumes that APIs are “historical contingent arrangements of social and material components that coalesce to produce new realities” (Bucher, 2013, p. 1). We argue, thus, that facilitated access to a “revealing” social media dataset cannot be dissociated from how APIs produce and induce readings and interpretations, shaping both how we understand the platforms and the topics that play out on them. Bucher’s (2013) perspective and most of the bibliography used here is anchored mainly on Science and Technology Studies (STS), a transdisciplinary field that, as one of its central issues, aims to understand the *modus operandi* of technical objects from the political and epistemological conditions in which their materialities are constituted and appropriated (Law, 2017).

Resuming discussions from previous works – in particular, d’Andréa, 2018 –, in this article we turn to some conceptual and methodological issues faced during empirical research based on data provided by social media platforms. More specifically, we discuss how empirical studies based on API data can stress the logistics of datafication embedded by platforms, mainly via the normalization of users’ actions, API political and economic governance, and how recommendation algorithms hierarchize data. Aware of the risks of losing specificities, we will not address a specific API or platform. In dialogue with authors who have previously analyzed case studies, this article proposes a set of reflections and notes based on characteristics and examples of different social media platforms, especially Facebook and Twitter.

We begin by presenting the Platform Studies and its set of concepts, reflections, and methodological practices that seek to understand the articulations of the technical, political, and economic dimensions that constitute social media and other online platforms (van Dijck et al., 2018). Based mainly on researchers from the “Amsterdam School” (Gorwa, 2019)², these studies anchor themselves in the STS field to discuss how Facebook, Uber, and other online services are constituted by dimensions such as infrastructure, governance mechanisms, business models, and the possible uses of their materiality (affordances) (d’Andréa, 2020). We then present some methodological approaches that aim to explore the singularities of online environments. Based on digital methods and related approaches, we point to a reflexive perspective that highlights the commercial and computational

² Gorwa (2019) refers to a group of researchers linked to the department of Media Studies at the University of Amsterdam (UvA), some of whom participate in the Digital Methods Initiative (DMI) project. However, important Dutch researchers in the field are linked to other universities, such as the Utrecht University.

logics of online platforms, and propose paths for critical studies based on API data.

In the “(Politics of) APIs: First Initiatives” section, we examine how the widespread adoption of APIs throughout the 2000s culminated in the process of “platformization of the web” described by Helmond (2015). We list the technical characteristics of APIs and discuss the earlier political tensions regarding Internet regulation, pointing out, in a brief historical overview, how APIs are interconnected and supported the ideology of the so-called web 2.0.

“Affordances, Governance, Algorithms” examine these three aspects pertinent to data-driven research on social media platforms. First, we discuss the importance of understanding and denaturalizing how platforms measure and make available their users’ practices. Functionalities such as “liking” or “sharing” allow platforms to standardize user activities, resulting in the grammatization of action” (Gerlitz & Rieder, 2018). This standardization of practices orientates the offer of data via APIs, inducing interpretations and analysis strongly aligned to the commercial logics of popularity and engagement.

The second aspect draws attention to the intrinsic relationship between API governance and the legal and economic issues that permeate platforms’ activities. The Facebook-Cambridge Analytica scandal is briefly discussed to illustrate a recent process that, among other consequences, triggered a “crisis in social media research” (Rogers, 2018b, p. 558). The increasing restriction and deactivation of public APIs and the emergence of institutional initiatives aimed at guiding research (e.g., Facebook’s *Social Science One*), among other topics, are addressed in this section.

The third aspect concerns the inseparability between the datafication processes adopted by platforms and algorithmic mediations that establish personalized selection and ranking logics. By establishing singular regimes of knowledge and visibility, algorithms diversify the possible experiences of different social media users (personalization), which questions the representativeness researchers often attribute to API data. This issue is even more problematic if we consider the increasing adoption of machine learning techniques, as Mackenzie (2018) discusses.

Emphasizing our intent of expanding the dialogues with researchers interested in a critical reading of data offered by the platforms, in the final considerations we summarize our key arguments to discuss the relevance and challenges of API-based research.

ONLINE PLATFORMS: DIMENSIONS AND METHODS

In an overview study of the so-called “Platform Society,” van Dijck et al. (2018) state that “a platform is fueled by *data*, automated and organized through *algorithms* and *interfaces*, formalized through *ownership* relations driven by *business models*, and governed through *user agreements*” (p. 9). Rather than being a definition, this statement emphasizes how the agencies of an extensive and heterogeneous set of online services are organized by the articulation of computational, normative, and economic aspects.

Under development since the early 2010s by authors such as Gillespie (2010) and van Dijck (2013), Platform Studies comprise a central theoretical and methodological construct to understand a recent critical turn in Internet Studies. Previously based on concepts such as “collaboration,” “collective intelligence,” and web 2.0, current research on social media and other online platforms not only takes on the challenge of understanding online platforms as environments for “interactions,” but also considers their robust materiality and centrality in the contemporary capitalist regime.

Recuero (2019) points out that we must recognize the differences between terms such as “platform,” “social network site,” or simply “social network.” One aspect that consolidates and singularizes the idea of “online platform” is recognizing that online sociability emerges articulated with robust computational infrastructures based on connectivity and data exchange. In different ways, platforms make various social practices measurable, contributing decisively to consolidate “datafication” as an emergent form of scientific knowledge (van Dijck, 2017).

The claim to transform *everything* into storable data and manage various processes of monitoring, ranking, and prediction is a key feature of datafication. One result of this knowledge model’s widespread adoption is the advancement of dataism ideology, that is, of a “widespread belief in objective quantification” (van Dijck, 2017, p. 43) promoted by data³. In the field of academic research that appropriates API data, the excessive emphasis on the volume of data collected, the little transparency of the methodological processes used, and the indiscriminate adoption of terms such as *monitoring* or *sentiment analysis*⁴ may indicate an excessive belief in the representativeness of the results obtained.

Regarding Platform Studies, the datafication process must be understood inseparably from the extensive socio-technical networks in which data production and circulation take place. Resuming previous work (d’Andréa, 2020), five main analytical dimensions should be considered. The *datafication*

³This critical perspective stems from a set of discussions that, since the early 2010s, have emphasized the absence of objectivity or neutrality in initiatives to store, interpret and manage social data from structured data – see boyd and Crawford’s (2012) criticism of the term “big data.”

⁴Criticisms of *social media monitoring* practices question the premise that the availability of social media data can measure or even give a precise interpretation of public debates in “real time.” Similarly, we criticize *sentiment analysis* studies that assume as a premise the effective possibility of classifying or predicting emotions or other subjectivities via data processing – see van de Ven (2017).

process in online platforms, for example, cannot be separated from the selection, recommendation, or prediction operations performed by algorithms. One must also pay attention to the *infrastructural dimension* of platforms, or the different ways in which they have been progressively built based on the centralization and control of operations and information flows involving various partners and end-users. Investments made in operating systems, software development systems (SDK), application stores, and even submarine cables are some of the initiatives that illustrate how articulated the processes of platformization and infrastructuralization are (Platin et al., 2018).

The rearticulation of various sectors' commercial relations based on data exchange shows the close relationship between the datafication process and the *business models developed by online platforms*. Companies like Google and Facebook have their market value directly associated with their ability to offer services based on data, such as targeted advertising. Due to the volume and complexity of the operations and practices they host, platforms are forced to devise different *governance* mechanisms (Gorwa, 2019). This self-regulation process is implemented, managed by rules, recommendations, and other normative resources, being directly related to how data can be accessed via APIs, as discussed below. Finally, one must pay attention to the platforms' affordances, that is, the possible uses of its functionalities that condition the actions performed on interfaces or other layers (Bucher & Helmond, 2018). From the negotiations established by different types of users with the available materialities we can better understand the potential uses of an API or other datafication tools.

Complementary to the multifaceted analytical perspective developed within Platform Studies, a set of propositions and methodological experiments also grounded in the STS field examine how platforms' communication dynamics can be studied in articulation with their materiality, political choices, and economic strategies. Possibly, the approach that best synthesizes the efforts to study platforms' singularities and transformations is the digital methods approach. In the study that introduces this perspective, Rogers (2013) points to the relevance to "follow the medium". This is a call for developing analytical features and methodological procedures sensitive to the logics that singularize each platform via its search engines, ranking models, and other online resources.

The complexification of the platforms and the maturation of research anchored on the media's specificities culminated in the proposition of complementary perspectives to digital methods, such as "interface methods" (Marres & Gerlitz, 2016), "issue mapping" (d'Andréa & Melgaço, 2019; Marres,

2015) and “programmed methods” (Borra & Rieder, 2014). Importantly, although a significant part of the empirical research, tools, and research protocols linked to digital methods are based on appropriating API data, this perspective is not limited to “API-based research”⁵ (Rogers & Venturini, 2019). Yet, not all research based on “computational methods” of data collection and treatment align with the digital methods perspective.

In a more recent study, Rogers (2018b) explains that the digital methods perspective is singularized by appropriating data obtained from platforms “for other purposes,” leading researchers to counter the hegemonic logistics adopted by tech companies. Concerned with the decentralized flows of data managed by platforms, Gerlitz and Rieder (2018) argue for a “specificity that is not reduced to a single, monolithic understanding of what a medium feature stands for, but considers digital data as outcomes and traces of distributed accomplishments materially complex and performative webs of practice” (p. 530).

The main point we would like to emphasize in this discussion is the need for placing methodological procedures not as instrumental operations that make visible foregone realities, but as procedures that enact readings and worldviews inseparable from the political and material gestures of the various actors involved. Such perspective, adopted by digital methods and related proposals, align with previous discussions on the “social life of methods,” understood in the STS field as a “materially complex and performative webs of practice” that articulate heterogeneous entities, such as subjects, objects, imaginaries, norms and institutions (Law, 2017, p. 47). In this sense, we highlight the importance of understanding API data-based research methodological procedures as part of politics and practices previously embedded by the platforms and as articulators of forms of knowledge that depend on these procedures to emerge.

(POLITICS OF THE) APIS: FIRST INITIATIVES

The centrality of the datafication process for online platforms can be understood from the development of data exchange standards and protocols. In the early 2000's, the adoption of web resource-oriented software architectures (REST style)⁶ became feasible by expanding partnerships between tech companies and external developers, which culminated in the process called “platformization of the web” by Helmond (2015), that is, the consolidation of the platform as an infrastructural and economic model that “reformats the web according to the logic of social media”

⁵ A notable exception is the use of *crawlers* to map hyperlink networks between websites (Rogers, 2013). A recent initiative to operationalize this perspective is the Hyphe tool (<https://hyphe.medialab.sciences-po.fr/>).

⁶ REST is the acronym for representational state transfer. Proposed in 2000, it allows targeted transfers of specific web resources identified by URIs (Uniform Resource Identifiers). REST architecture enables data requests made by a significant part of the APIs.

(p. 2). According to the author, this transformation is anchored on three programmability preconditions: separation of content and presentation (XML format), modularization of content and resources (allowing, for example, videos and posts to be “embedded” in external pages), and the interface with databases.

These innovations have enabled the development of services and business models based on expanding online services beyond the companies’ websites. The mixing of data provided by two or more services (mashups), the adoption of *widgets* and social plugins (Gerlitz & Helmond, 2013), and the possibility of cross-posting are some of the first applications developed by online services that soon would be characterized as platforms. The emphasis shifted to connectivity (van Dijck, 2013) between systems, which began operating on a “double logic”: the decentralization of functionalities and the re-centralization of standardized data.

As “border resources” (Helmond et al., 2019) that connect data input and output, APIs have decisively contributed to the platformization of the web. E-commerce companies such as eBay (2000) and Amazon (2002) adopted the first generation of applications-oriented web pages. In 2004, the adoption of an API by the photo-sharing service Flickr started an intense adoption process by *social networking sites*. Not coincidentally, the following year publisher Tim O’Reilly drew the attention of a wider audience to the “web as a platform” and the potential for a new generation of services he named Web 2.0. The widespread adoption of this term was related to a new cycle of business appropriations of technologies rhetorically oriented to *collaboration* or *participation* practices.

As Bodle (2011) points out, the search for new business opportunities based on API interoperability is directly related to the “bursting of the bubble” of .com companies in the late 1990s. This search can also be attributed to the antitrust lawsuit filed by the US government against Microsoft. By reorganizing the logistics of data exchange, this early platformization seemed to indicate an openness of tech companies to a more decentralized performance, leveraging a process that ultimately had the opposite effect. APIs are, thus, computational infrastructures that reorganized economic performance, the practices of developers and users, and the normative political relations of online services.

Regarding social media, some landmarks help us pinpoint the migration from the “social networking sites” model to “social media platforms” (Helmond, 2015). Helmond, Nieborg, and van der Vlist (2019), for example, point out that “Facebook became programmable” by incorporating external developers

through the “Facebook Development Platform” service, launched in 2006. In a pioneering study that examined APIs beyond the computational perspective, Bodle (2011) highlights how, for Facebook, the progressive adoption of different apps⁷ authorized more expanded access to user data. Amid the great diversity of third-party *apps*, the author points to an “avalanche” of social games and tests offered to users. One such test – “This Is Your Digital Life” – would be at the center of the controversy that emerged years later, featuring Facebook and Cambridge Analytica.

⁷ Besides Facebook Developer (2006), the author mentions: Facebook Platform (2007), Facebook Connect (2008), Open Stream (2009) and Open Graph with Instant Personalization (2010). For a more up-to-date list, see Helmond *et al.* (2019)

The centrality of APIs for consolidating platforms is even more evident in the case of Twitter. Research on the topic (Ahmed et al., 2017) seems to be unanimous in reinforcing how the model of data decentralization and recentralization adopted by Twitter since 2006 was decisive for its relevance in changing the ecosystem of online platforms. Puschmann and Burgess (2014), for example, highlight the strong association between the popularization of *Streaming* API, which allows to collect a stream of tweets as soon as they are posted, and Twitter’s business positioning as an “irreplaceable source of real-time information” (p. 47). Different appropriations of these data by journalistic companies, marketing and academic research have led Twitter to be seen as a “thermometer” or a “sensor” of public discussions on different contemporary issues, consolidating methodologies and analyses anchored in the widespread belief of the full explanatory capacity of data (van Dijck, 2017).

The importance, for Twitter, of governing its data became increasingly evident after 2010, when the company adopted the first in a series of data exchange restrictions, which included a ban on third-party managed ads and the possibility to acquire data via companies like Gnip (which would then be acquired by Twitter in 2014). These restrictions are directly associated with a change in the platform’s business model, which started to consider itself an “information company” (van Dijck, 2013).

Although the historical nuances of APIs are not our focus⁸, it is worth briefly retracing the somewhat anecdotal launch of Twitter’s first API in 2006. Generically referred to as the “Twitter API” (renamed “The Twitter REST API” in 2007), the service was briefly introduced in a post signed by one of Twitter’s founders (Stone, 2006). The note defined the Internet as a “series of pipes,” referencing an expression used three months earlier by then US Senator Ted Stevens during a speech on the topic of *net neutrality*. By minimizing Internet service providers’ commercial and political actions, the senator’s speech limited the Internet to a simplistic concept of infrastructure and was repeatedly criticized and derided (Schneider, 2006).

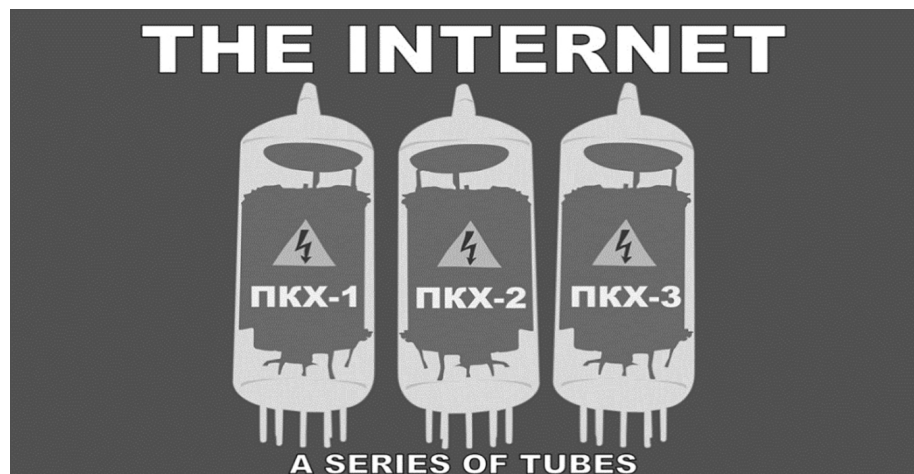
⁸ For an introductory chronology, see “History of APIs” (n.d.).

Stone's (2006) metaphor is even more curious when looking at the image illustrating the post (Figure 1). It reproduces a shirt print initially shared on the Boing Boing website (Doctorow, 2006) as part of the "memefication" of Senator Ted Stevens' slogan, where "A Series of Tubes" is associated with objects similar to bags used in hospitals to inject patients with saline solution and other substances. The labels NKX-1, NKX-2, and NKX-3 seem to refer to proteins that act as regulators of organ development in the human body⁹.

⁹ We thank researcher Daniel Loyola for insights that helped refine this description. Any factual errors are the responsibility of the author.

Figure 1

Image used by Twitter in the post announcing the launch of its API (2006)



Note. Stone (2006).

Beyond a *geek* reference, the post announcing Twitter's first API illustrates how, from the beginning, API adoption was intrinsically linked to the platforms' strategic positioning. Appropriating the informality that would later be recognized as part of a *memetic culture*, the founder of Twitter suggests having the ability to reinvent the infrastructure that would, from then on, guide the data flow on an Internet challenged by intense regulatory discussions and the commercial euphoria of web 2.0. In this sense, the reference to medical instruments can be associated with a supposed ability of APIs to "heal" a sick patient (.com companies); but above all, it reinforces the argument that data exchange applications are techno-scientific constructs that produce knowledge through data "distillation" (Marres & Weltevrede, 2013).

AFFORDANCES, GOVERNANCE AND ALGORITHMIC MEDIATIONS

Operating via an arrangement that involves infrastructure, protocols, terms of use, economic interests, and the practices of different types of users, APIs collaborate to produce new realities (Bucher, 2013). Based on the theoretical and methodological premises discussed so far, the datafication of empirical research on online platforms are connected to three key aspects: the grammatization of actions, the recent changes in data access policies (governance), and the centrality of the regimes of visibility and hierarchization articulated by various algorithmic mediations.

A key aspect of API-based research is to pay attention to what aggregate information the platforms provide and the conditions under which this information was generated, organized, and distributed. First we must acknowledge platforms' effort to standardize user activities (what can and cannot be done) on the interface through features such as liking and sharing – process called “grammatization of action” (Gerlitz & Rieder, 2018). Using Twitter's technogrammar as an example, Gerlitz and Rieder (2018) point out that forms such as tweets, retweets, replies, mentions, or hashtags “allows platforms to collapse... action, grammar, and data capture, inscribing user activities directly into highly formalized units” (p. 531).

APIs thus have affordances that not only dictate which data can or cannot be accessed by third parties, but also predefine what and how actions are inscribed. Gerlitz and Rieder (2018) further explains that this normative force is visible at four articulated moments: when specifying entities and relationships in a database, in the possible actions established in infrastructures (e.g., back-end and middleware), in the API governance of “inputs and outputs”, and how tweets are displayed and published in different interfaces (p. 53). This multidimensionality helps us understand how API affordances are closely articulated to procedures that precede and follow data requests.

API-based research is always challenged to find ways to address the interpretations and analyses induced by the grammatization of actions. When, for example, only the total of likes, comments, and shares are emphasized, research tend to naturalize readings based on the platforms' commercial logic and automatically adhere to vague notions of popularity or engagement (i.e., in “vanity metrics” [Rogers, 2018a]). Often, the result is overly uncritical analyses that do not underline the platforms' logic. This risk is even more evident when studies use services that purport to facilitate access to and synthesize data provided by platforms, such as Netlytic (<https://netlytic.org/>)¹⁰.

¹⁰ A list of “social media research tools” is available in Ahmed (2019).

In free or paid versions, this type of service facilitates data interpretation and analysis by “packaging” them in files such as “top hashtags,” “top users,” or “co-hashtags,” accentuating a process of blackboxing the affordances and other dimensions of platforms.

In the second key aspect of data-based research, we again focus on the intrinsic relationship between the politics of APIs and the legal, political, and economic issues that permeate platform performance. Governance concerns the platforms’ ability to propose procedures, norms, and innovations that enable their self-regulation in a global capitalist market. But reconciling the interests and demands of diverse groups of users and partners appears to be an increasingly complicated task.

The scandal involving the misuse of data obtained via Facebook by Cambridge Analytica can be taken as a milestone for discussions on violation of privacy, the tremendous economic power of platforms and, by extension, the governance of public APIs. The first revelations about the case emerged in December 2015, and drew attention to the use of data obtained via “psychological tests” by US Senator Ted Cruz’s campaign for the Republican Party seat for the following year’s presidential election (Davies, 2015). The definitive outbreak took place in May 2018, when revelations made by the English newspaper *The Guardian* (Cardwallard & Graham-Harrison, 2018) and other publications pointed out that Cambridge Analytica had collected data from tens of millions of Facebook users from a personality test (“This Is Your Digital Life”), which took advantage of the platform’s openness to access third-party data. In other words, Facebook’s policy of sharing data with third parties not only allowed millions of citizens to be exposed to ads and other hyper-personalized actions, but also made clear to a wider public how the “collaboration” practices between users of the *global community* controlled by Mark Zuckerberg were strongly oriented towards optimizing the commercial use of datafication.

Although the scandal was clearly associated with abusive business practices adopted by commercial companies, the argument of privacy protection took center stage and culminated in consecutive restrictions on access to public API data. Some key events in this process were the accelerated restriction of apps associated with Instagram’s API (2016), the restrictions on access to data related to Facebook events, groups, and pages (April 2018), and Twitter’s adoption of a stricter governance of its data, for example, by requiring detailed information to be provided by developers (June 2018).

For Rogers (2018b), the process triggered a “crisis in social media research,” while Bruns (2019) is more emphatic and names it an “APIcalypse”.

The fact is that the increasing restriction on data access has resulted in a significant reorganization of data-driven research and sparked discussions about alternatives to public APIs. Freelon (2018), for example, points out the need to discuss “post-APIs research”, which would entail a return to data scraping techniques available on the platforms’ interface and a deeper ethical and legal understanding of the consequences of violating the “terms of use” of platforms that, by default, prohibit this kind of practice.

Concurrently, calls for more transparency have led platforms to propose institutional initiatives that, in contrast to the (controlled) openness that guided their previous public APIs, privileged a narrow and targeted set of academic research. Twitter, for example, launched a call for research proposals that should help “measure the health” of the platform. In sum, 230 proposals were submitted and two projects approved¹¹. Facebook, in turn, teamed up with major partners to launch *Social Science One* in 2018, defined by its coordinators as “a new model of industry-academia partnerships designed to span the divide between the needs of Internet technology companies, academic researchers, and the public, even in highly politicized environments” (King & Persily, 2020, p. 704). Targeted at funding research on elections and democracy, the first call for proposals announced 12 approved projects in April 2019. According to the project page, research groups will have access to billions of Facebook and Instagram posts and profiles (via the *CrowdTangle* tool and the APIs linked to it)¹² and millions of ads displayed in different countries, including Brazil (by the *Ad Library* API)¹³. One of Bruns’ (2019) criticisms to this initiative is the prior definition of the issues that should be studied.

¹¹For more information, see Gadde and Gasca (2018).

¹²For more information, see Shiffman (n.d.).

¹³For more information, see King and Persily (2019).

Our third and final question concerns the different ways in which API data are intertwined with algorithmic mediations that, among other operations, hierarchize and customize information flows on platforms. To this end, one must focus on the interdependence between the datasets produced, stored and (eventually) made available by online platforms and the computational routines designed to produce relationships and hierarchies between them. Often defined as a sequence of instructions written to accomplish predetermined tasks, here algorithms are taken as “instances that enable an *interested reading* of an empirical and ‘data-fictionalized’ reality” (Rieder, 2018, p. 127).

Often labeled as opaque or black boxes, algorithms have assumed, in recent years, a visible prominence in Internet studies, mainly to discuss phenomena around political polarization and expansion of power asymmetries. When invoking algorithmic power, however, one should contemplate that what

singularizes an algorithm are not only the instructions materialized in its code, but mainly how it articulates, in practice, the associations between a heterogeneous set of actors (Introna, 2016). The visible agency of algorithms in interfaces should be understood in articulation with the infrastructures that host them, with governance mechanisms, with users' agencies, and with other dimensions of online platforms.

By their performances, algorithms stand out as socio-technical constructs that establish particular logics of selection, hierarchization, recommendation, and control of information flows. Articulated with APIs (Bucher, 2013), they establish regimes of knowledge and visibility that appropriate available data to identify patterns, trends, and, increasingly often, to make predictions (Bucher, 2018; Gillespie, 2018). This dynamic is especially key to the "ranking cultures" (Rieder et al., 2018) embedded by platforms. Ranking the most commented themes as "trending topics" and organizing diverse content in personalized feeds are possibly the most evident practices of how platforms appropriate data to, articulated with each user's actions, mediate, via algorithms, personalized experiences.

Our main argument here is that the complexification of mediation processes exercised by algorithms has culminated in a significant disconnect between the standardization of data provided by public APIs and the variety of personalized data-driven contents offered to each user by the platforms' feeds or other interfaces. In general, the datasets collected induce a static reading about the interplay between a given topic and the affordances of a platform. In other words, the collected data invisibilize both the way a discussion (or conversation) has changed during a given period and the possible changes in governance policies (e.g., criteria for content moderation). This issue is more prominent when a research is analyzing an event that, mediated by data politics, emerges from methods that privilege the common sense notion of "real-time" (Marres & Weltevrede, 2013).

The challenges posed by algorithmic mediations to social media research are even more pressing if we consider how performance has been guided by artificial intelligence principles and methods (Mackenzie, 2018). As the author argues, the way platforms operate is less and less aligned with API-based data exchange models. They focus is now on the programmability centered on machine learning processes, that is, on techniques that leverage training data to adjust and recreate analysis models (Mintz, 2019).

If the API-based model focuses on automating data flow between different platforms, the increasing adoption of machine learning techniques culminates by implementing *workflows* and *pipelines* that allow programmers and other

partners to experiment with the efficiency of predictive models. In Mackenzie's (2018) words, "the platform itself becomes an experimental system for observing the world and testing how the world responds to changes in the platform" (p. 2003). In this model, the author argues that platforms take a step forward to consolidate themselves as infrastructures, questioning van Dijck's (2013) argument about the centrality of "connectivity" to online platforms.

FINAL CONSIDERATIONS

This article discussed a set of conceptual and methodological issues to be addressed by research based on data provided by APIs of social media platforms. Based on Platform Studies and the digital methods approach, we discuss how empirical data-based studies should consider the material, political, economic, and normative dimensions that permeate the performance of online services like Twitter and Facebook. Dimensions such as affordances, datafication/algorithms, governance, business models, and infrastructures evidently interfere with the possibilities and constraints imposed on empirical data-based research.

The first aspect to be revisited in these final considerations are the relationships between the constant changes in how APIs work and the role played by these "border resources" from the beginning of the platformization of the web. In "(Politics of) APIs: First Initiatives," we discussed issues surrounding the development of protocols geared towards a controlled data exchange, the legal pressures over the increasing concentration of power by Internet-related companies (telecoms and Microsoft), and the euphoric adherence to the "collaborative" proposition of web 2.0. Using examples from Twitter and Facebook, we sought to synthesize how both the economic and governance strategies of platforms have been progressively changing in a trend that complexifies and generally worsens the conditions for conducting empirical data-based research.

"Affordances, governance, and algorithmic mediations" detailed these three dimensions to, once again, point out how the socio-technical logics and corporative choices decisively shape data supply and possibilities. The platforms' efforts to standardize user activities and the implications of uncritical adherence to the logistics of popularity or engagement are issues to be observed by researchers seeking to question and not only follow API affordances. The significant changes in data governance after the Cambridge Analytica scandal suggest a crisis of the public API-based research model and an increased asymmetry between studies conducted in countries and institutions with

precarious conditions of infrastructure and political articulation with the platforms. Finally, we sought to highlight how the centrality and diversity of algorithmic mediations applied to data establish diversified regimes of visibility and hierarchization for each user or calculated public, which exemplifies the accentuated experimental dimension of platforms, mainly by consolidating machine learning as a priority model of data processing.

After all these considerations, a question one might ask is whether our argumentation in this article implies an exhaustion of the public API model (still) adopted by some social media platforms. More interested in raising the question rather than answering it, we seek here to alert that API data are indeed rich and powerful empirical material when they are understood as “clues” that signal how social issues, methods, and platforms that co-produce themselves are interdependent. To avoid the instrumental and positivist uses of traces provided by Twitter, Facebook, or YouTube is, ultimately, to refute the “ideology of dataism” (van Dijck, 2017), that is, to go beyond the enthusiastic belief in the supposed possibility of revealing *the reality* through data overriding a critical reading that seeks to denaturalize and, why not, repropose the dynamics of the platforms. ■

REFERENCES

- Ahmed, W. (2019, June 18). Using Twitter as a data source: An overview of social media research tools. *LSE Impact Blog*. <http://bit.ly/3r14GWP>
- Ahmed, W., Bath, P. A., & Demartini, G. (2017). Using Twitter as a data source: An overview of ethical, legal, and methodological challenges. In K. Woodfield (Ed.), *The ethics of online research* (Vol. 2, pp. 79-107). Emerald. <https://doi.org/10.1108/S2398-601820180000002004>
- Bodle, R. (2011). Regimes of sharing. *Information, Communication & Society*, 14(3), 320-337. <https://doi.org/10.1080/1369118X.2010.542825>
- Borra, E., & Rieder, B. (2014). Programmed method: Developing a toolset for capturing and analyzing tweets. *Aslib Journal of Information Management*, 66(3), 262-278. <https://doi.org/10.1108/AJIM-09-2013-0094>
- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662-679. <https://doi.org/10.1080/1369118X.2012.678878>
- Bruns, A. (2019). After the ‘APIcalypse’: Social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, 1-23. <https://doi.org/10.1080/1369118X.2019.1637447>

- Bucher, T. (2013). Objects of intense feeling: The case of the Twitter APIs. *Computational Culture*, 3. <http://bit.ly/3c0jWPJ>
- Bucher, T. (2018). *If... then: Algorithmic power and politics*. Oxford University Press.
- Bucher, T., & Helmond, A. (2018). The affordances of social media platforms. In J. Burgess, A. Marwick, & T. Poell (Eds.), *The SAGE handbook of social media* (pp. 233-253). Sage. <http://dx.doi.org/10.4135/9781473984066.n14>
- Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. <http://bit.ly/3vyYXLo>
- d'Andréa, C. F. de B. (2018). Cartografando controvérsias com as plataformas digitais: Apontamentos teórico-metodológicos. *Galáxia*, (38), 28-39. <https://doi.org/10.1590/1982-2554234208>
- d'Andréa, C. F. de B. (2020). *Pesquisando plataformas online: Conceitos e métodos*. Edufba. <http://repositorio.ufba.br/ri/handle/ri/32043>
- d'Andréa, C., & Melgaço, L. (2019, June 11-14). *Tecnologias emergentes: "Questões problemáticas" do VAR tuitadas durante a Copa 2018* [Presented article]. XVIII Encontro Anual da Compós, Porto Alegre, RS, Brasil. <https://bit.ly/3c5mHz9>
- Davies, H. (2015, December 11). Ted Cruz campaign using firm that harvested data on millions of unwitting Facebook users. *The Guardian*. <http://bit.ly/2OOxmVV>
- Doctorow, C. (2006, July 3). T-shirt design: "The Internet, a Series of Tubes". *Boing Boing*. <http://bit.ly/3eUgs2N>
- Freelon, D. (2018). Computational research in the post-APIs age. *Political Communication*, 35(4), 665-668. <https://doi.org/10.1080/10584609.2018.1477506>
- Gadde, V., & Gasca, D. (2018). Measuring healthy conversation. *Blog Twitter*. <http://bit.ly/3r05Q53>
- Gerlitz, C., & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society*, 15(8), 1348-1365. <https://doi.org/10.1177/1461444812472322>
- Gerlitz, C., & Rieder, B. (2018). Tweets are not created equal. A platform perspective on social media metrics. *International Journal of Communication*, 12, 528-547. <https://bit.ly/3lty2fm>
- Gillespie, T. (2010). The politics of "platforms". *New Media & Society*, 12(3), 347-364. <https://doi.org/10.1177/1461444809342738>

- Gillespie, T. (2018). A relevância dos algoritmos. *Parágrafo*, 6(1), 95-121. <http://bit.ly/38VfMq5>
- Gorwa, R. (2019). What is platform governance? *Information, Communication & Society*, 22(6), 854-871. <https://doi.org/10.1080/1369118X.2019.1573914>
- Helmond, A. (2015). The platformization of the web: Making web data platform ready. *Social Media + Society*, 1(2), 1-11. <https://doi.org/10.1177/2056305115603080>
- Helmond, A., Nieborg, D. B., & van der Vlist, F. N. (2019). Facebook's evolution: Development of a platform-as-infrastructure. *Internet Histories*, 3(2), 123-146. <https://doi.org/10.1080/24701475.2019.1593667>
- Histórico das APIs no monitoramento e pesquisa em mídias sociais. (s.d.). IBPAD. <http://bit.ly/2NxpDor>
- Introna, L. D. (2016). Algorithms, governance, and governmentality: On governing academic writing. *Science, Technology, & Human Values*, 41(1), 17-49. <https://doi.org/10.1177/0162243915587360>
- King, G., & Persily, N. (2019, April 28). First grants announced for independent research on social media's impact on democracy using Facebook data. *Social Science One*. <http://bit.ly/3r42Y74>
- King, G., & Persily, N. (2020). A new model for industry-academic partnerships. *PS: Political Science & Politics*, 53(4), 703-709. <https://doi.org/10.1017/S1049096519001021>
- Law, J. (2017). STS as method. In U. Felt, R. Fouché, C. A. Miller, & L. Smith-Doerr (Eds.), *The handbook of science and technology studies* (pp. 31-57). MIT Press.
- Mackenzie, A. (2018). From APIs to AI: Platforms and their opacities. *Information, Communication & Society*, 22(13), 1989-2006. <https://doi.org/10.1080/1369118X.2018.1476569>
- Marres, N. (2015). Why map issues? On controversy analysis as a digital method. *Science, Technology & Human Values*, 40(5), 655-686. <https://doi.org/10.1177/0162243915574602>
- Marres, N., & Gerlitz, C. (2016). Interface methods: Renegotiating relations between digital social research, STS and sociology. *The Sociological Review*, 64(1), 21-46. <https://doi.org/10.1111/1467-954X.12314>
- Marres, N., & Weltevrede, E. (2013). Scraping the Social? Issues in live social research. *Journal of Cultural Economy*, 6(3), 313-335. <https://doi.org/10.1080/17530350.2013.772070>
- Mintz, A. G. (2019). *Visualidades computacionais e a imagem-rede: Reapropriações do aprendizado de máquina para o estudo de imagens*

- em plataformas online* [Master's Thesis, Universidade Federal de Minas Gerais]. Repositório institucional da UFMG. <http://bit.ly/2P6j3vP>
- Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, 20(1), 293-310. <https://doi.org/10.1177/1461444816661553>
- Puschmann, C., & Burgess, J. (2014). The politics of Twitter data. In K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 43-54). Peter Lang US.
- Recuero, R. (2019, July 9). Mídia social, plataforma digital, site de rede social ou rede social? Não é tudo a mesma coisa? *Raquel Recuero*. <http://bit.ly/3tHK4Vw>
- Rieder, B. (2018). Examinando uma técnica algorítmica: O classificador de Bayes como uma leitura interessada da realidade. *Parágrafo*, 6(1), 123-142. <https://bit.ly/2PagYih>
- Rieder, B., Matamoros-Fernández, A., & Coromina, O. (2018). From ranking algorithms to 'ranking cultures': Investigating the modulation of visibility in YouTube search results. *Convergence*, 24(1), 50-68. <https://doi.org/10.1177/1354856517736982>
- Rogers, R. (2013). *Digital methods*. MIT Press.
- Rogers, R. (2018a). Otherwise engaged: Social media from vanity metrics to critical analytics. *International Journal of Communication*, 12, 450-472. <https://bit.ly/3vOhqE7>
- Rogers, R. (2018b). Social media research after the fake news debacle. *Partecipazione e Conflitto*, 11(2), 557-570. <http://doi.org/10.1285/i20356609v11i2p557>
- Schneider, T. (2006, July 11). Mr. Stevens' wild ride through a "series of tubes". *Public Knowledge*. <http://bit.ly/3r04fw5>
- Shiffman, N. (s.d.). CrowdTangle for academics and researchers. *CrowdTangle*. <http://bit.ly/3r0FIab>
- Stone, B. (2006, September 20). Introducing the Twitter APIs. *Blog Twitter*. <http://bit.ly/3s5PiKm>
- van de Ven, R. (2017, January 25). Choose how you feel; you have seven options. *INC Longform*. <http://bit.ly/2Nv82nh>
- van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford University Press.
- van Dijck, J. (2017). Confiamos nos dados? As implicações da datificação para o monitoramento social. *MATRIZES*, 11(1), 39-59. <https://doi.org/10.11606/issn.1982-8160.v11i1p39-59>

- van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society*. Oxford University Press.
- Venturini, T., & Rogers, R. (2019). “APIs-Based Research” or how can digital sociology and journalism studies learn from the Facebook and Cambridge Analytica data breach. *Digital Journalism*, 7(4), 532-540. <https://doi.org/10.1080/21670811.2019.1591927>
- Vimieiro, A. C., & Bargas, J. de K. R. (2019). O uso de dados e métodos digitais nas pesquisas em comunicação. *Revista Famecos*, 26(2), e32473. <https://doi.org/10.15448/1980-3729.2019.2.32473>

Article received on October 24, 2020, and approved on March 8, 2021.