

DOSSIER

TWITTER AS A NEWS SOURCE IN DATA JOURNALISM



MARÍLIA GEHRKE

Universidade Federal do Rio Grande do Sul, Porto Alegre – Rio Grande do Sul – Brazil

ORCID: 0000-0003-3538-1600

MARCIA BENETTI

Universidade Federal do Rio Grande do Sul, Porto Alegre – Rio Grande do Sul – Brazil

ORCID: 0000-0002-1621-8421

DOI: 10.25200/BJR.V16N3.2020.1277

Received in: March 09th, 2020

Desk Reviewed: May 03rd, 2020

Desk Review Editors: André Lemos and Lia Seixas

Approved on: September 20th, 2020

ABSTRACT – This paper discusses the use of *Twitter* as a news source, especially in data journalism. This practice uses public databases as sources for its investigations, but has now started to make use of less conventional sources, such as *Twitter*, which politicians now use as an official communication channel and thus avoid giving interviews to the press. In this study, we analyze news articles published over a period of a little less than two months by the data journalism team at *Vortex Media*. Our focus was on political coverage. *Twitter* may be classified as a reproductive or statistical documentary news sources, depending on the case. We argue that, by using this social networking site as a news source, data journalism could be able to provide new knowledge. Its limitations lie within the working order of the social network itself, which is highly mediated by algorithms and conditioned to spread disinformation.

Key words: News source. *Twitter*. Data journalism. Social networking site. Disinformation.

O TWITTER COMO FONTE NO JORNALISMO GUIADO POR DADOS

RESUMO – Este artigo discute o uso do *Twitter* como fonte, em especial no jornalismo guiado por dados. Conhecida por investigar a partir de bancos de dados públicos, esta prática tem acessado fontes menos convencionais, a exemplo da plataforma, à medida que os políticos a tratam como canal oficial de comunicação e evitam conceder entrevistas à imprensa. Neste estudo, analisamos notícias publicadas pela equipe de dados do site

jornalístico *Vortex Media* em pouco menos de dois meses, observando que a cobertura política se destaca nesse contexto. O *Twitter* pode ser classificado como fonte documental do tipo reprodução ou estatística, dependendo do caso. Defendemos que o jornalismo guiado por dados tem potencial para promover conhecimento inédito a partir do uso desse site de rede social como fonte; suas limitações estão na própria condição da rede, mediada por algoritmos e propícia para o espalhamento de desinformação.

Palavras-chave: Fonte. *Twitter*. Jornalismo guiado por dados. Site de rede social. Desinformação.

TWITTER COMO FUENTE DE PERIODISMO DE DATOS

RESUMEN – Este artículo analiza el uso de *Twitter* como fuente, especialmente en el periodismo de datos. Conocido por investigar desde bases de datos públicas, esta práctica ha accedido a fuentes menos convencionales, como la plataforma, ya que los políticos la tratan como un canal oficial de comunicación y evitan dar entrevistas a la prensa. En este estudio, analizamos las noticias publicadas por el equipo de datos del sitio periodístico *Vortex Media* en poco menos de dos meses, señalando que la cobertura política se destaca en este contexto. *Twitter* puede clasificarse como fuente documental reproductiva o estadística, según el caso. Argumentamos que el periodismo de datos tiene el potencial de promover el conocimiento sin precedentes del uso de este sitio de redes sociales como fuente; sus limitaciones están en la condición de esta red, mediada por algoritmos y propicia para la difusión de la desinformación.

Palabras clave: Fuente. *Twitter*. Periodismo de datos. Sitio de red social. Desinformación.

1 Introduction¹

Information obtained from sources (in this paper these sources are people and documents) are the raw material of journalism. In a democratic society, which in principle should be governed by a policy of transparency and accountability (Bennett, 2010), official sources need to provide clear information to society. In order to fulfill its role as a mediator, journalism consults databases and obtains statements to extract information and create knowledge. Through the disintermediation of media, social networking sites such as *Twitter* are being used as communication channels between governments and parts of society. Journalists use this platform to indirectly obtain statements and data.

Previous studies have discussed the use of *Twitter* as a source for journalism (Broersma & Graham, 2013; Recuero, 2009; Zago, 2011) and have shown that journalists prefer this social networking site to obtain their information because it is a space in which the public are present and interact (Nordheim et al., 2018). However, obtaining discourses and data from this platform is not completely reliable because they are difficult to verify (Heravi & Harrower, 2016)

Data journalism, which involves everything from collecting to viewing information, is important in network journalism (Heinrich, 2011) and uses data as the protagonists of the narrative.

This article discusses data journalism as a source and argues that *Twitter* can be classified as a reproductive or statistical documentary news source (Gehrke, 2018), depending on the case and the type of analysis conducted. In a post-truth scenario where platforms are highly mediated by algorithms, data journalism tends not only to use traditional public databases but also alternative tools for extracting information, like *Twitter*. In order to address this topic, we analyzed publications produced by the data journalism team at the digital native *Vortex Media*, an online journalism site specializing in political coverage. About one-fourth of the 19 news articles which we analyzed use *Twitter* as a source – the remaining articles use public repositories.

The results of our analyses showed us that, for the most part, *Twitter* is a versatile source in data journalism. It allows for complex and inferential analyses to be conducted, going beyond simply reproducing the discourse of a particular source. We agree with Diakopoulos (2019) when he says that algorithms and automation processes already exist in the news production chain. This can even be seen in the dependence that some news outlets have on platforms, particularly in terms of content distribution. Similar to other relatively new practices, using *Twitter* as a source has its limitations, especially regarding the disinformation scenario in which it operates, but it also has great potential, such as generating new products.

2 *Twitter* and disinformation

Network journalism (Heinrich, 2011) involves thinking beyond the traditional relationships between journalists and personal news sources to include other types of sources triggered and mediated by software (Manovich, 2013). Platforms have an influence on journalism as they determine how they will produce and distribute content (Bell & Owen, 2017).

Minimal time for consulting sources and the limitations of newsrooms both make it so that, in order to optimize their search for information, reporters need to use quick tools to investigate and verify facts. Thus, it is not unusual for them to use materials available on the internet or press releases (Broersma & Graham, 2013),

resources that require less time to verify. This is on top of the need to create innovative agendas and the public figures's dislike of the press. Given this context, using social networking sites as a source seems unavoidable. *Twitter*, for example, is used by politicians and celebrities to communicate directly with the public (Nordheim et al., 2018) and to avoid uncomfortable questions from journalists. This strategy, although it allows one to bypass traditional media, is subject to other conditions such as algorithmic instructions and the structure of the network itself (size of publication, modes of interaction, etc.).

In a post-truth era where opinions seem to be worth more than facts, populist politicians create discourses of an 'us versus them' mentality where the press is presented as an enemy. "Post-truth communication is exactly where populism wants politics to be – the realm of divided truth, binary thinking, and broken-up communication" (Waisbord, 2018, p.14). As a result we have presidents like Jair Bolsonaro in Brazil and Donald Trump in the United States who attack journalism in an attempt to undermine its credibility by claiming that the verified content is actually "fake news". The term is an oxymoron as the concept of news presupposes truth and verified facts. It makes reference to disinformation, that is, content created with the intent to damage an individual's or an institution's reputation (Wardle & Derakhshan, 2017).

The aversion politicians have for contacting journalists leads reporters to use other sources. Known for its limited number of characters, *Twitter* is a social network site founded on July 15, 2006, which at the time was called *Twttr* (Bell & Owen, 2017). It is among the top ten in user count in the world – in 2019 it reached 330 million users. In the beginning, *Twitter* had a character count of up to 140 characters, doubling to 280 characters in 2017 without the use of photos or videos. The use of other media in publications began in 2011 and, since then, more than 50% of the platform's content has featured images and videos (Ortiz-Ospina, 2019).

Ellison and Boyd classify (2013) *Twitter* as a social network site (SNS) and as a microblogging server. It has several interactional features such as replying to a post, republishing and sharing a post, tagging other users, and generating threads (a series of connected texts) to create longer publication sequences. The platform also allows content indexing through the use of hashtags (#) which allow a topic to be accessible in trending topics to find out, in real time, what the latest topics on *Twitter* are. Although the default configuration for a profile is set to public, it is possible to create a private user profile.

With a private profile other users need authorization by that profile's user in order to interact, but publications cannot be shared. Basically, a post from a private user profile cannot be spread organically across the network, although it can be spread by using the print screen feature. In March 2020, the company launched a new sharing format (called Fleets) where texts last up to 24 hours and cannot receive likes, public replies, or retweets.

There are three points that allow a social network site to be defined as a network communication platform, according to Ellison and Boyd (2013): 1) it presents identifiable profiles which are provided by the user; 2) it shows public connections; 3) it consumes, produces, and/or interacts with content generated by users and their connections. For these authors, an SNS is more than just a social network, which is a popular term used to define social relationships (but not necessarily on the internet). These authors believe an SNS has different uses. *Twitter* has a prevalence of asymmetric relationships; it is more common to follow the profiles of people you do not personally know. This is one explanation for why a large number of politicians and the public use this platform. Users with a large number of followers have their authenticity verified by the platform, which in turn increases a profile's level of reliability.

Spaces such as *Twitter* are favorable environments for the dissemination of disinformation and robots, or bots; an automated software application programmed to do tasks, like posting content. This content is published without the use of any filter and does so at an alarming speed (Heravi & Harrower, 2016). Social network sites are part of the logic of software culture (Manovich, 2013), which includes the production, distribution, and reception of most of the content and experiences. Thus, as pointed out in a previous study (Brenol & Gehrke, 2017), one can imagine there is a Software Factor in news content production. As a source, *Twitter* publications suffer from interference that ends up being reflected in its use as a raw material for journalism.

3 *Twitter* as a documentary news source for journalists

Journalists use sources – people or documents that present information on a given topic – in order to produce content. While personal news sources are interviewed, documentary news sources

have to be approached differently. Broersma and Graham (2013) claim that journalists use *Twitter* for four reasons: finding new stories, helping them find sources and obtain information, extracting statements, and verifying information – often using the public as a kind of collective verification technique (crowdsourcing).

Recuero (2009) is one of the Brazilian researchers who has studied *Twitter* as a journalism source. In her study she identifies social network sites as: a) sources that produce information; b) information filters; or c) spaces for repeating this information. In addition to using the platform to search for qualified sources and unpublished information, she points out that the content circulating on *Twitter* can shape the media's agenda. Zago (2011) also discusses the use of *Twitter* as a source, analyzing the websites for the *Folha de S.Paulo*² and *Zero Hora*³ newspapers. This author shows how the news on the platform deals with a range of topics, but most of it focused on celebrities and politics.

One of the studies that has contributed the most to understanding this theme is that of Nordheim et al. (2018) which offers a longitudinal analysis on the use of *Twitter* and *Facebook* as sources for journalists over a ten-year period. They analyzed the following newspapers: *The New York Times* (United States), *The Guardian* (United Kingdom) and *Süddeutsche Zeitung* (Germany). Their analysis started when platforms first launched and concluded in 2016. In general, their study pointed out that *Twitter* is one of the newspapers' choices for obtaining information. The authors attribute this choice to the fact that "elite sources", such as politicians and celebrities, also use the site. Since they are unable, or reluctant, to interview these sources, journalists use the statements they post on *Twitter*. Although there are similarities between the topics explored in the three countries mentioned above (the use occurs in subjects such as crime, media, technology, and international politics, among others) German journalists are the ones who use social network sites as a source the least.

There is, however, a reluctance to trust the disseminated content. A study conducted by Heravi and Harrower (2016) on Irish journalists showed that *Twitter* is used in several stages of news production, such as searching for sources and events, but it is done so with caution, especially when it comes to verifying information. When asked how they go about identifying reliable users they intend to use as a source, the respondents in the study said they look for institutional evidence (for example, the name of the company they

work for) in profiles, as well as validation outside the platform. The journalists in the study also reported that they base their reliability on the quantity and quality of posts made on a *Twitter* account, as well as the number of followers that account has. Lastly, they look at how long the profile has been active for and the user's profile photo.

What researchers are doing here is creating alternative ways of assessing the credibility of profiles on social network sites. One example can be seen with Fletcher et al. (2017), who developed Truthmeter, an automated tool for classifying the credibility level of *Twitter* profiles – on a grading system of 0 to 9 – and their potential to be used as a source. Designed to aid journalists when verifying sources, this resource is part of the SocialSensor project, a collaborative project of ten European institutions for developing software to optimize how social network sites are used in professional news routines. A number of tests and calibrations were performed, and a final list of criteria for establishing what makes a credible *Twitter* source was defined as follows: number of tweets, number of retweets, number of followers and proportion of followers in relation to the people you follow, verification, frequency, and popularity.

When accessed without interviewing the authors, *Twitter* posts become a documentary news source. In data journalism, social network sites can be seen as an alternative or complement to public databases. By working with complex issues that explain the facts, we see data journalism as a digital form that produces news about relationships (Nielsen, 2017), and seeks to contextualize events.

In general, documentary news sources are databases, laws, memos, publications on social network sites, scientific studies, and other documents that cannot be consulted directly through interviews, but have the potential for information to be extracted and used in news articles. *Twitter* is essentially a reproductive documentary news source that:

Gathers information or statements previously issued by people and organizations and registers it, and does not occur through direct interviews [...] This category also includes publications from other media and posts on social network sites that, without any kind of interview being conducted, receive document status when accessed by journalists. (Gehrke, 2018, p.94).

It can also be understood as a documentary news source on statistics, since the mass amount of data extracted from the platform serves as raw material for analysis. It covers, in general, numbers used to describe an event.

As with any source, *Twitter* must be critically analyzed – even documents considered to be trustworthy and objective could be subject to some type of influence or bias (Steenen, 2019). Official sources, which journalists generally use, can act as “information subsidiaries” (Gandy, 1982, p.62); people who release small amounts of information that journalists can use and which do not require too much verifying, thus, fitting easily and working well in the busy routine of these professionals.

For Diakopoulos (2019), platforms like *Twitter* are also important to help journalists expand on their sources. What’s more, using this social network site as a raw material, in the case of data journalism, contributes to the exclusive character of the content of news articles through the type of data collected and the analysis employed. Structured information in public databases is openly available and generally ready for use (in the case of simple extractions and analyses). This is not the case with *Twitter* data, which requires further collection and analysis efforts.

By incorporating unusual sources into their work, journalists transform seemingly disordered information into knowledge. This is because the content that circulates on social network sites does not necessarily have to be in the public’s interest or have any kind of social commitment (Recuero, 2009). As Diakopoulos (2019, p.24) points out, journalists mainly make rough observations of the world (which includes talking to personal news sources and examining documents) to communicate content that the audience can then make sense of: “As data becomes information and then knowledge, value is added. This is exactly what journalists do: increase the value of information for news consumers and for society”.

Broersma and Graham (2013) describe a “marriage of convenience” between journalism and social network sites, as through them it is possible to both obtain information and disseminate news and reach readers. However, we need to recognize that this relationship has ethical implications. According to Christofoletti (2019), journalism is expected to address issues of public interest while respecting the privacy of people, groups, and organizations. In order to understand how Brazilian journalists relate to issues of privacy in their daily lives, the author conducted an online survey of 120 professionals in 2018.

When asked about how concerned they are about the privacy of their sources, 65% of respondents believed there should be less privacy for celebrities than for average people; a further 19.2% said there should be no privacy for public figures. In terms of how these journalists regard social network sites being used as a source, most consider these spaces

to be public, free, and independent of the user's consent to use. For example, 48.3% of journalists stated that it is completely acceptable to use photographs from social network sites, 40% said it is acceptable only if they have permission from the profile owner, and only 5.8% consider that this is an invasion of privacy; a further 4.2% state that this type of conduct comes with the territory. These numbers led Christofolotti to believe that most journalists consider social network sites to be free repositories.

While people do make content available about themselves in these spaces – which Martins (2019, p.15) refers to as “offered privacy” – it must be taken into account that some of this content, when shared only between friends, may be of a private nature, despite the fact that one could use the screen capture feature and later circulate said content between people who would not have initially received it. Using this content involves ethical implications, especially from non-public figures. “[...] journalists cannot disregard their ethical obligation to assess, in each situation, whether they have a right to that information or not, an obligation they are incumbent on upholding. Or to assess if there are personal rights (privacy, honor, reputation, good name) potentially present” (Martins, 2019, p.18).

Discussions on using SNS information are complex, and a number of issues need to be addressed by journalists. The fact that platforms such as *Twitter* and *Facebook* are used as official channels of communication by governments and politicians intensifies the debate even further since these publications are indeed of a public nature. What is most important is that discussions are held prior to extracting information on how personal data is used, including by large technology companies. To get an idea of the complexity of this topic, the EU's General Data Protection Regulation (GDPR), in force since May 2018, is a regulation with 99 chapters. Law No. 13,709 in Brazil, known as the General Personal Data Protection Law (LGPD) of August 2018, is a law to protect the rights of freedom and privacy of citizens, including digital media, when it comes to processing personal data.

4 *Twitter* as a final source for the public

Twitter can also be a final source for the public – who not only receive their news on the social network site, but can also follow the profiles of public figures and institutions. “*Twitter* becomes a system where news is reported, disseminated and shared online in short, fast, and frequent

messages” (Hermida, 2010, p.310). The definition of what a source is has several dimensions to it: the posts serve as raw material for journalists; the content published by journalists, institutions, and public figures is a source of information for audiences; and in some circumstances involving crowdsourcing, *Twitter* users themselves become a source for journalists.

When searching for information on social network sites, the public often comes across a mix of content. Not only do they find news media committed to producing and distributing truthful and accurate content, but they will also find unverified information, which may be spread by friends or even robots, the algorithms of which respond to certain types of triggers. There are a variety of events, sources, and reputations (Sterrett et al., 2019) that make it difficult to trace the origin of the information.

In this space of interactions in which official and unofficial sources of instant communication and fragmented information mix, Hermida (2010, p.298) defines *Twitter* as an integral part of “systems of consciousness” that characterize what he calls “environmental journalism”. If a journalist who is technically qualified finds it difficult to check for publications of this nature and distinguish reliable content from questionable content (Martins, 2019), it may be even more challenging for a reader to do so.

The different levels of information sources can be observed in opinion polls. In Brazil, a survey conducted by the DataSenado Research Institute⁴ (2019) asked 2,400 people to name which source they thought was the most used for news consumption. The questions were multiple choice and the results showed that 79% of the interviewees constantly used the *WhatsApp* messaging application as a source, followed by television (50%), and *YouTube* (49%). *Twitter* is used on a recurring basis as a source by only 7% of respondents. In all, 98% of respondents reported accessing the internet by cell phone, and 61% claim to consume content through computers or laptops. Mobile device applications are used heavily to *Twitter* search for news on social network sites.

This same survey also showed that 73% of the interviewees said that the most important thing in determining whether an article or report is reliable is to look at the source that published it; another 24% said it depends on who sent the news. A survey conducted between November and December 2016 in the United States contained similar results: almost 1,500 adults were included in this survey, their responses demonstrated that the public tends to trust content shared by public figures (Sterrett et al., 2019). According to the authors, the survey indicates there are two factors that influence opinions about what is shared on social network

sites: 1) trust in the person who shares a particular story; and 2) the credibility of the news site that reports the story. This reinforces that the power lies with those who do the sharing.

In general, the discourse of public figures that use social network sites is that of disintermediation; they speak directly to the final public without the need for any press to mediate it. Digital platforms are no exception. Algorithms determine the rules of the game and even how the speech will be delivered – a post of up to 280 characters, for example, can produce simplistic and volatile responses. And it gets updated constantly, which makes it challenging to keep up with a story's developments.

In view of the high consumption of information through social network sites, communication vehicles create profiles on a number of different platforms to distribute their content. Bell and Owen (2017, p.49) raise questions about the impact of these sites on the financial sustainability of companies. They need to produce attractive content which can then be circulated among the platforms' algorithms; however, there is little financial return for the news outlets. "Nowadays, platforms no longer just play the role of a distribution channel – they determine what the public sees, who gets paid for that audience, and even what format and genre of journalism they exhibit".

Twitter, as we have seen, is a source that journalists and the public both use. It is a social network site which people access in order to search for news despite being inserted in an environment which is conducive to the spread of disinformation and heavily mediated by algorithms. In the next section, we shall discuss the use of *Twitter* as a specific source for data journalism. Despite the problems that come with this environment, the platform seems to be making unprecedented contributions to this type of journalism, which usually uses public databases for content production. We argue that there is a potential here to expand political coverage since government repositories alone are not adequate enough to create new guidelines.

5 The *Vortex* case: analysis and discussion

Numerical information that quantifies and translates a certain event is the starting point for news articles that present data journalism techniques. Although statements from personal news sources and including individuals in its narratives are not excluded

from this type of journalism, it is public databases, documents, and statistics that are used as its main sources. If these sources are not openly available on transparent portals, they can be obtained through the Citizen Information Service (SIC), a channel that receives public requests via the Access to Information Law (LAI), which came into force in Brazil in May 2012. Clearly, this access to information hinges on the political decisions of government officials, as government agencies can ignore requests or directly refuse them on the grounds that they deem the information to be confidential. Data journalism involves everything from collecting to visualizing information. It derived from precision journalism and Computer-Assisted Reporting (CAR), and comprises the application of computer and social science knowledge in its processes, in addition to expanding the role of the press as a defender of the public interest (Träsel, 2014).

Let's now revisit the discussion of *Twitter* as a source for data journalism. We believe that this practice expands on the variety of sources, especially when it comes to political coverage. In times of post-truth and criticism of government officials in the press, statements are being obtained more through structured databases and less through interviews. Unlike traditional journalism, data journalism does not only use the statements that circulate on the platform. In general, data is extracted from *Twitter* for a more complex analysis. As a starting point, this article uses the tweets extracted from the platform as a reproduction documentary news source, which may be able to fit the statistical type depending on its use (Gehrke, 2018).

In order to understand how a digital native news outlet uses *Twitter* as a source in data journalism, we analyzed the publications of the data team at *Vortex Media*⁵, a site which specializes in politics, one of the areas probably most affected by the public figures's constant interactions on social network sites. Founded and directed by Diego Escosteguy⁶ in October 2019, *Vortex* was born from the idea of creating "journalism on the move" claiming "radical transparency of methods and intentions". However, the website's main source of funding has never been revealed. Escosteguy (2019) publicly stated that *Vortex* would be funded by subscriptions to ensure the outlet's independence. He also said that ads from private companies would be welcome, but ads from public agencies would not be accepted in order to avoid any conflict of interest.

Vortex built newsrooms in Brasília and in São Paulo; the data team's headquarters and whose work we will analyze here. However,

due to problems with the site's unclear form of financing, the data team was short-lived: in less than two months, journalists and other professionals who were part of this team were dismissed and the São Paulo headquarters was shut down. We chose to analyze this as the news outlet used data analysis as a major differential in attracting subscribers and, in fact, made significant strides in this area.

From October 8 to December 3, 2019, the *Vortex* Data section released 27 publications. Eight of these were not included in this analysis as they correspond to the "Data of the Week" section; a compilation of topics which is structured differently from the news. Thus, the initial *corpus* contains 19 news items. All were read, entered into a spreadsheet, and classified according to the main source. Four (21.05%) of these 19 news items use *Twitter* as a source; another 15 (78.95%) use public data available on the transparent portals of the Executive, Legislative, and Judiciary branches, or from requests via the Access to Information Law. We were able to identify, in this section, that more than three quarters of the 19 news articles analyzed correspond to politics; 15 (78.95%) on publications, two (10.55%) on the environment, one (5.25%) on the economy, and one (5.25%) on the judicial system.

After this first screening process, we selected only the following four news items that use *Twitter* as a main source: "As interações da bancada do PSL⁷ on Twitter" (Hirota et al., 2019); "No *Twitter*, comentários negativos dispararam após reação de Bolsonaro sobre caso Marielle"⁸ (Spagnuolo et al., 2019b); "Trump, Bonoro Pistola e Jacaré de Tanga⁹: quem Bolsonaro segue no Twitter" (Almeida, 2019) and "Campanha no *Twitter* por impeachment de Gilmar Mendes¹⁰ tem marcas de ação coordenada" (Spagnuolo et al., 2019a).

In general, these news articles either focus on President Jair Bolsonaro¹¹ and his interactions on *Twitter* or on the current Brazilian political environment. There are even discussions about who actually writes the tweets for the president (Santi, 2019). The *Vortex* data group tried to make journalism more transparent to readers. One way it did this was to use a box called "why does it matter?", a section which explains to the public why a particular agenda had been developed. The explanation is the same for the three texts that address politics: guidelines of this nature are important because the Brazilian president uses *Twitter* as one of his main channels to communicate with the public and as a space to interact with other public figures.

As Zago points out (2011), *Twitter* is generally used in conventional journalism when it is necessary to extract statements from the

public which might otherwise be difficult to obtain through an interview. This platform is used as a source in data journalism not just for qualitative analysis or extraction of testimonials; through programming codes and statistical models, data journalists are able to make quantitative analyses that help to give a broader understanding of a scenario.

One of the main strategies used by the *Vortex* data team for working on content extracted from *Twitter* was sentiment analysis (Hirota et al., 2019; Spagnuolo et al., 2019b). Through algorithms, it is possible to detect the meaning attributed to the collected words, generating a unique analysis on the content chosen by the team. When it comes to profile interactions on *Twitter*, network analysis techniques are generally used to understand which profiles have the most or least interactions within a selected spectrum. To do this one needs to access the Application Programming Interface (API) of this social network site, which requires a technical knowledge which conventional journalism professionals do not always possess.

The news article about interactions between PSL politicians (Hirota et al., 2019) – then former president's political party – sought to reveal internal disputes within the party through public speeches and even name-calling on *Twitter* profiles. In this case, it is possible to quantify and systematize information through these interactions on the platform, which often only occur behind the scenes in politics. Analyzing the positive or negative aspect of publications, one can infer, for example, which politicians are the closest and most engaged, and what this means in the Brazilian political scene. According to Bolsonaro's guidebook, several members of the PSL broadcast on social network sites in order to speak directly to the public. It is necessary to remember, however (and perhaps this is the biggest criticism of using tweets as a source), that the platform stimulates dramatic and volatile speeches in hopes of drawing more attention to the platform itself.

In data journalism, identifying patterns is one of the main resources to help with understanding large amounts of data. For example, *Vortex* analyzed the user profiles that the president follows on *Twitter*, as well as the profiles that had retweeted messages (Almeida, 2019).

In addition to official organs, whose feeds are commonly followed by presidents of the Republic, profiles associated with political militancy (such as right-wing and conservative groups) are among some of Bolsonaro's favorites to follow. In this case, the use of data journalism helps to quantify and standardize information that is sometimes presented as perception. The existence of numbers helps

to contextualize an event, going beyond declaratory journalism. In addition, data can either confirm or debunk a hypothesis, which is usually a starting point for most data journalism productions.

One of the advantages of using *Twitter* as a source is the ability to measure and assign meaning to subjects that are otherwise difficult to quantify, not to mention being able to gauge public perception on certain matters. One example of this can be seen in a news article on the tone in which Jair Bolsonaro spoke about the press and its coverage of the murder of councilwoman Marielle Franco in Rio de Janeiro – the president's name was mentioned in the inquiry into the death of the former councilwoman (Spagnuolo et al., 2019b). While extracting data from *Twitter*, journalists noticed that interest in the subject had grown during the *Jornal Nacional* broadcast, the television news program with large audience which originally broadcast the news. Discussions reached their peak after the president's live broadcast across his social networks. The *Vortex* data journalism team, in just one minute, was able to quantify a total of 1,767 tweets containing the expression "Bolsonaro", which showed that the president's statement generated more negative backlash than the original news article itself. Journalists were also able to measure the total number of tweets in which the expression was mentioned: about 850 thousand tweets in a span of 19. Using graphics, the team managed to show that there was only minimal support for the president during this period, which occurred around the time his live speech was broadcast.

The news on the backlash from the Marielle case shows that data journalism does not always make use of public repositories; it is open to employing other types of quantifications. More than simply mentioning the case, journalists were able to reveal mobilization numbers on *Twitter*, bringing together elements of social relevance and accuracy. By offering the programming code adopted in the collection and analysis of data, the team also submits its work to public scrutiny, looking to establish an honest relationship with the reader. Furthermore, journalists offer a skeptical view of the content circulating on *Twitter*, admitting that the pro-president publications may have been orchestrated. As we mentioned earlier, the platform has characteristics that make it a breeding ground for robot activity. It is well known that these automated profiles are used to divert attention from relevant issues, stimulate political events, and feed populist politics, which holds the premise that the true voice of the people must be heard (Frost, 2019). In Brazil, social network sites are

the barometer for measuring actions of the Jair Bolsonaro government.

Something similar happened in another news article published by *Vortex*. Using *Twitter* as a source, the team once again recognized the presence of orchestrated actions. Out of all the *Vortex* texts we analyzed, the most significant was probably the one on the campaign to impeach a Supreme Federal Court (STF) minister. The team did more than just collect the tweets from certain hashtags; it went further and suggested that this campaign had been orchestrated by a small group of *Twitter* users (Spagnuolo et al., 2019a).

This was not just an article, it was an investigative report, using *Twitter* as a source and accessing its API. The data journalism team also used statistical resources in their analysis in addition to exploring the texts from users' profiles, thus revealing some points in common (like adopting a conservative posture while supporting the president) among those who reproduced a large number of messages asking for Gilmar Mendes to be impeached.

The type of material taken from the platform for *Vortex's* analysis is still categorized as a reproductive news documentary source as the analysis obtained its information through documented records and not interviews. In addition, this reproductive type of news source presupposes more mediation (Gehrke, 2018) since its content is submitted to other platforms it hosts. For news articles presenting a network or sentiment analysis, they use a statistical news documentary source which includes numbers to describe events. The statistical documentary news source is the one that differs the most from 'declaratory' journalism. Very few personal news sources were used in the data journalism conducted by *Vortex* which we analyzed. The texts have qualitative and quantitative features, both of which help generate new knowledge.

Based on both the bibliography consulted and the analysis proposed in this study, we were able to infer that the use of *Twitter* as a reproductive documentary news source in data journalism is both limited and has potential. One of its limitations is the world of disinformation in which the platform is inserted, as well as less personal interactions between personal news sources and journalists, and the reliance on data journalism from official sources (even though *Twitter* is not a public repository, it does contain profiles from the public and is used as an official channel of communication by some governments). Its potential is in the transformative power of data journalism analyses, which provide readers with unique guidelines and even provide raw material for new

analyses, as the programming codes are shared with the public.

Generating new analyses leads to advancements in knowledge, especially in the field of politics. With new, digitized data made available (Diakopoulos, 2019), other journalists have the opportunity to make new inferences and create unique agendas. The use of documentary news sources in data journalism also forces one to think about the changes in the dynamics involved between reporters and personal news sources (Broersma & Graham, 2013). By accessing published content without having to interview a source, journalists risk losing subtle nuances and distinctions that can only be observed through personal contact or journalistic investigation. Almost one quarter of the published content we analyzed from *Vortex Media* used *Twitter* as the main source in its political coverage.

6 Final considerations

This article discussed the use of *Twitter* as a source. In data journalism, this platform is used as either a reproductive or statistical documentary news source (Gehrke, 2018), depending on the event being reported on. At a time when journalism is being discredited, particularly by populist politicians on social network sites (Waisbord, 2018), it is important to study how sources are appropriated by journalism. Concluding this work helps us to infer that, unlike traditional practices which normally reproduce the statements propagated on *Twitter*, data journalism attempts to generate knowledge (Nielsen, 2017) by using strategies of network analysis and analysis of meaning in content production.

The object of analysis in this article shows that *Twitter* is a productive source for data journalism on politics primarily because government officials who could be described as anti-media use social network sites as a form of communication they believe is free from mediation. For journalism, the challenge of using *Twitter* as a source seems to rest on issues of privacy and ethics when extracting information, as well as trying to verify facts in an environment rife with robots and orchestrated actions. This is added to the fact that data journalism is highly dependent on public databases. Although the analyses conducted on the platform shows it can create new content, we are still talking about a source that is originally official – after all, it deals with how governments communicate with the public. In this sense, what researchers criticize is that journalists are dependent on official communications accessed from a distance, thus losing the subtle nuances and distinctions of dialogues

between journalists and personal news sources.

Embedded within a culture of software (Manovich, 2013) and algorithms, *Twitter* (a large technology company) ends up going through several stages of news production: from production, to accessing sources, to content distribution. Relying only on a platform that reveals little about its algorithms seems risky since automation limits even the type of message produced and circulated in that space. Thus, data journalism needs to maintain a variety of sources from which to extract and produce knowledge. Although there may be technical and financial limitations, the current scenario shows that newsrooms should create and maintain their own databases.

Lastly, we concluded from this study that, given the current context of press criticism, disintermediation, and disinformation, there is a greater tendency for documentary news sources (and not always conventional ones) to be used in Brazilian journalism. The current scenario requires information to be verified more closely and what is presented on social network sites needs to be questioned and looked at further. To fulfill their role as a qualified mediator, journalists need to strengthen the bonds of respect with readers and demonstrate knowledge about their own processes, which includes recognizing the increased complexity involved with channels that propagate public statements, as social network sites do. Studies like this one are conducted with the hopes of better understanding this scenario as the trend shows that *Twitter* and other similar platforms will continue to be used as sources – and to do so requires being familiar with its features and accepting its limitations, such as the propagation of purposely volatile statements and robots working among humans.

NOTES

- 1 This work was conducted with support from the Coordination for the Improvement of Higher Education Personnel (CAPES), Financing Code 001, and from the National Council for Scientific and Technological Development (CNPq).
- 2 Daily newspaper with the largest print and digital circulation in Brazil, based out of São Paulo. According to the Communication Verifier Institute (IVC), *Folha* recorded a monthly average of 328.4 thousand paid subscriptions in 2019.
- 3 Headquartered in Porto Alegre and part of the RBS Group, it is the

most widely circulated daily newspaper in the state of Rio Grande do Sul. It features both print and digital versions. It launched the GaúchaZH news portal in 2017, a combination of the newspaper *Zero Hora* and Rádio Gaúcha.

- 4 The institute was created in 2005 to monitor issues the Brazilian public has with the Federal Senate, and issues that are to be debated in this sphere.
- 5 Available at: vortex.media
- 6 Brazilian journalist who specialized in political coverage, worked at the two largest weekly information magazines in Brazil: *Veja* and *Época*.
- 7 Acronym of the Social Liberal Party, through which Jair Bolsonaro was elected President of the Republic in 2018.
- 8 Marielle Franco served as a councilwoman in the city of Rio de Janeiro for the Socialism and Liberty Party (PSOL). She was shot and killed on March 14, 2018, along with her driver, Anderson Gomes. She was known for acting in defense of communities in Rio and against the police force imposed on residents.
- 9 Self-titled humorous profiles that convey texts and images aimed at Bolsonaro's followers.
- 10 Gilmar Mendes is a minister at the Supreme Federal Court (STF), the highest court in the Brazilian legal system. The STF acts as the Supreme Court and the last instance of a constitutional court. Since 2005, 18 impeachment requests against Mendes have been filed. In 2020, he came under pressure after criticizing the presence of the military in the federal government.
- 11 Jair Bolsonaro was elected president of Brazil for a four-year term (2019–2022). He is a retired Army captain and was a federal deputy for almost 30 years, having approved only two projects in that period. He is a far-right politician who defends the military dictatorship, torture, and the right for citizens to bear arms.

REFERENCES

Almeida, R. (2019, Nov. 27). Trump, Bonoro Pistola e Jacaré de Tanga: Quem Bolsonaro segue no Twitter. *Vortex Media*. Retrieved from <https://vortex.media/dados/24355/trump-bonoro-pistola-e-jacare->

de-tanga-quem-bolsonaro-segue-no-twitter/

Bell, E., & Owen, T. (2017). A imprensa nas plataformas: Como o Vale do Silício reestruturou o jornalismo. *Revista de Jornalismo ESPM*, 6(20), 48–83. DOI: 10.7916/D8D79PWH

Bennett, W. L. (2010). The press, power and public accountability. In A. Stuart (Ed.), *The Routledge companion to news and journalism* (pp.105–115). London: Routledge.

Brenol, M., & Gehrke, M. (2017). Catraca Livre e o Fator Software para a prática jornalística. Proceedings from *8º Congresso Internacional de Ciber-jornalismo*. Campo Grande, Universidade Federal do Mato Grosso do Sul.

Broersma, M., & Graham, T. (2013). Twitter as a news source. *Journalism Practice*, 7(4), 446–464. DOI: 10.1080/17512786.2013.802481

Christofoletti, R. (2019). Percepções de jornalistas brasileiros sobre privacidade. *Matrizes*, 13 (2), 179– 197. DOI: 10.11606/issn.1982-8160.v13i2p179-197

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Cambridge: Harvard University Press.

Ellison, N. B., & Boyd, D. (2013). Sociality through social network sites. In W. Dutton (Ed.), *The Oxford handbook of internet studies* (pp.151–172). Oxford: Oxford University Press.

Escosteguy, D. (2019, n.d.). Nasce o jornalismo em movimento. *Vortex Media*. Retrieved from <https://vortex.media/vortex/>

Fletcher, R., Schifferes, S., & Thurman, N. (2017). Building the ‘Truthmeter’: Training algorithms to help journalists assess the credibility of social media sources. *Convergence: The International Journal of Research into New Media Technologies*, 26(1), 19–34. DOI: 10.1177/1354856517714955

Frost, C. (2019). The power of voice: bots, democracy and the problem of political ventriloquism. *Journal of Political Power*, online first, 1–16. DOI: 10.1080/2158379X.2019.1701831

Gandy, O. H. Jr. (1982). *Beyond agenda setting: information subsidies and public policy*. Norwood: Ablex Publishing Corporation.

Gehrke, M. (2018). *O uso de fontes documentais no jornalismo guiado por dados* [master thesis, Universidade Federal do Rio Grande do Sul]. Repositório Digital UFRGS.

Heinrich, A. (2011). *Network journalism: journalistic practice in interactive spheres*. Nova York: Routledge.

Heravi, B., & Harrower, N. (2016). Twitter journalism in Ireland: Sourcing and trust in the age of social media. *Information, Communication &*

Society, 19(9), 1194–1213. DOI: 10.1080/1369118X.2016.1187649

Hermida, A. (2010). Twittering the news. *Journalism Practice*, 4(3), 297–308. DOI: 10.1080/17512781003640703

Hirota, R., Almeida, R., & Lago, L. (2019, Oct. 23). As interações da bancada do PSL no Twitter. *Vortex Media*. Retrieved from <https://vortex.media/dados/8751/as-interacoes-da-bancada-do-psl-no-twitter/>

Instituto de Pesquisa DataSenado. (2019). *Redes sociais, notícias falsas e privacidade de dados na internet*. Brasília: Senado Federal e Câmara dos Deputados.

Lei Geral de Proteção de Dados Pessoais (LGPD), N° 13.709/2018. (2018). Diário Oficial da União. Brasília, DF: Presidência da República.

Manovich, L. (2013). *Software takes command*. Nova York: Bloomsbury Academic.

Martins, P. (2019). Redes sociais como fontes de informação jornalística: As novas fronteiras da privacidade. In R. Christofolletti (Ed.), *Privacidade, transparência y éticas renovadas* (pp.11–31). Sevilha: Ediciones Egregius.

Nielsen, R. K. (2017). Digital news as forms of knowledge: A new chapter in the Sociology of Knowledge. In P. Boczkowski, & C.W. Anderson (Eds.), *Remaking the news: essays on the future of journalism scholarship in the digital age* (pp.90–110). Cambridge: The MIT Press.

Nordheim, G. von, Boczek, K., & Koppers, L. (2018). Sourcing the sources: An analysis of the use of Twitter and Facebook as a journalistic source over 10 years in The New York Times, The Guardian, and Süddeutsche Zeitung. *Digital Journalism*, 6(7), 807–828. DOI: 10.1080/21670811.2018.1490658

Ortiz-Ospina, E. (2019, Sep. 18). The rise of social media. *Our world in data*. Retrieved from <https://ourworldindata.org/rise-of-social-media>.

Recuero, R. (2009). Redes sociais na Internet, difusão de informação e jornalismo: Elementos para discussão. In D. Soster, & F.F. Silva (Ed.), *Metamorfoses jornalísticas 2: A reconfiguração da forma* (pp.37–55). Santa Cruz do Sul: EDUNISC.

Santi, A. (2019, Feb. 23). Carlos Bolsonaro tuíta por Jair? Estes indícios mostram que o vereador usa a conta do presidente. *The Intercept Brasil*. Retrieved from <https://theintercept.com/2019/02/22/carlos-bolsonaro-twitter-jair-bolsonaro-presidente/>

Spagnuolo, S., Lago, L., & Mendes, G. (2019a, Nov. 14). Campanha no Twitter por impeachment de Gilmar Mendes tem marcas de ação coordenada. *Vortex Media*. Retrieved from <https://vortex.media/dados/19632/campanha-no-twitter-por-impeachment-de-gilmar-mendes-tem-marcas-de-acao-coordenada/>

Spagnuolo, S., Hirota, R., Lago, L., & Almeida, R. (2019b, Nov. 1). No Twitter, comentários negativos dispararam após reação de Bolsonaro sobre caso Marielle. *Vortex Media*. Retrieved from <https://vortex.media/dados/13321/vortex-data-no-twitter-comentarios-negativos-dispararam-apos-reacao-de-bolsonaro-sobre-caso-marielle/>

Steenen, S. (2019). Journalism's epistemic crisis and its solution: Disinformation, datafication and source criticism. *Journalism*, 20(1), 185–189. DOI: 10.1177/1464884918809271

Sterrett, D., Malato, D., Benz, J., Kantor, L., Tompson, T., Rosenstiel, T., Sonderman, J., & Loker, K. (2019). Who Shared It?: Deciding What News to Trust on Social Media. *Digital Journalism*, 7(6), 783–801. DOI: 10.1080/21670811.2019.1623702

Träsel, M. (2014). *Entrevistando planilhas: Estudo das crenças e do ethos de um grupo de profissionais de jornalismo guiado por dados no Brasil* [doctoral thesis, Pontifícia Universidade Católica de Rio Grande do Sul]. Biblioteca digital de teses e dissertações – PUCRS.

Waisbord, S. (2018). The elective affinity between post-truth communication and populist politics. *Communication Research and Practice*, 4 (1), 17–34. DOI: 10.1080/22041451.2018.1428928

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Estrasburgo: Conselho Europeu.

Zago, G. (2011). O Twitter como fonte para o jornalismo. *Animus – Revista Interamericana de Comunicação Midiática*, 10(20), 51–65. DOI: 10.5902/217549772786

MARÍLIA GEHRKE. Doctoral student in Communication at PPGCOM/UFRGS, scholarship holder from Capes, and a member of the Journalism Research Center (NUPEJOR – UFRGS/CNPq). Email: marilia.gehrke@ufrgs.br

MARCIA BENETTI. PhD in Communication and Semiotics from PUC-SP. Full professor for the Department of Communication at UFRGS, leader of the Journalism Research Center (NUPEJOR – UFRGS/CNPq), scholarship holder from CNPq. Email: marcia.benetti@gmail.com

TRANSLATED BY: LEE SHARP