Revista de Administração Contemporânea
Journal of Contemporary Administration

ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO

e-ISSN: 1982-7849

Technological Article

Like, Share, and React: Twitter Capture for Research and Corporate Decisions

Curtir, Compartilhar e Reagir: Captura de tweets para Pesquisas e Decisões Corporativas

Djeison Siedschlag*[1]
Jeferson Lana[2]
Roberto Gonçalves Augusto Junior[2]
Rosilene Marcon[1]

ABSTRACT

Context: social media have an immense amount of information, being a space for its dissemination. Individuals, online connections, are able to filter or give visibility to certain information, to the detriment of others. The central problem lies in monitoring posts and reactions aimed at corporate actions and strategies. In addition to this monitoring, companies can make decisions based on the data collected. **Objective:** to develop and structure a social media management tool. **Methods:** to achieve the general objective, the article was developed in three main steps. The first was to suggest a free software script for capturing and initial analysis of Twitter posts. The second step was to categorize this analysis and identify resources and competencies needed by companies. Finally, actions to be taken by companies for social media management were suggested. **Results:** the developed script enabled the automated extraction of data, which were stored in a database for analysis and management of online interactions. The actions were proposed based on the case study developed. **Conclusions:** in the practical field, this study contributes to the process of extracting data from Twitter by proposing a new script for capturing data, identifying the main categories of influence of digital activists and monitoring social media through strategic actions. By demonstrating that the script is effective in extracting data, it is possible to carry out further studies and implement the social media management monitoring process.

Keywords: on-line social activism; social media monitoring; Twitter data extraction; secondary data.

RESUMO

Contexto: as mídias sociais dispõem de uma imensa quantidade de informações, sendo um espaço para a sua difusão. Indivíduos, conexões on-line, são capazes de filtrar ou dar visibilidade a determinadas informações, em detrimento de outras. O problema central reside no monitoramento de postagens e reações voltadas a ações e estratégias corporativas. Além desse monitoramento, as empresas podem tomar decisões a partir dos dados coletados. **Objetivo:** desenvolver e estruturar uma ferramenta de gestão em mídia social. **Métodos:** para alcançar o objetivo geral, o artigo foi desenvolvido em três etapas principais. A primeira foi sugerir um script em software livre para captura e análise inicial das postagens no Twitter. A segunda etapa foi categorizar essa análise e identificar recursos e competências necessários às empresas. Por fim, foram sugeridas ações a serem tomadas pelas empresas para a gestão da mídia social. **Resultados:** o script desenvolvido possibilitou a extração dos dados de forma automatizada, os quais foram armazenados em banco de dados para análise e gerenciamento das interações on-line. As ações foram propostas com base no estudo de caso desenvolvido. **Conclusões:** no campo prático, este estudo contribui para o processo de extração de dados do Twitter com a proposição de um novo script para captura de dados, na identificação das principais categorias de influência de ativistas digitais e no monitoramento de mídia social por meio de ações estratégicas. Ao demonstrar que o script é eficaz na extração de dados, é possível realizar novos estudos e implementar o processo de monitoramento gerencial de mídia social.

Palavras-chave: ativismo social on-line; monitoramento de mídias sociais; extração de dados do Twitter; dados secundários.

* Corresponding Author.
1. Universidade do Vale do Itajaí, Programa de Pós-graduação em Administração, Biguaçu, SC, Brazil.
2. Universidade do Vale do Itajaí, Itajaí, SC, Brazil.

Cite as: Siedschlag, D., Lana, J., Augusto, R. G., Junior, & Marcon, R. (2023). Like, share, and react: Twitter capture for research and corporate decisions. *Revista de Administração Contemporânea, 27*(2), e220008. https://doi.org/10.1590/1982-7849rac2023220008.en

Published as Early Access: September 13, 2022.
Assigned to this issue: January 11, 2023.

JEL Code: M1.

Editor-in-chief: Marcelo de Souza Bispo (Universidade Federal da Paraíba, PPGA, Brazil)
Associate Editor: Gustavo da Silva Motta (Universidade Federal Fluminense, PPGA, Brazil)
Reviewers: Márcio Abdalla (Universidade Federal Fluminense, Brazil)
Two reviewers did not authorize the disclosure of their identities.
Peer Review Report: The disclosure of the Peer Review Report was not authorized by its reviewers.

Received: September 13, 2021
Last version received: May 30, 2022
Accepted: June 11, 2022

Note: This text is translated from the original Portuguese version, which can be accessed here.

# of invited reviewers until the decision:

	1	2	3	4	5	6	7	8	9
1st round	✓	✓	X	✓					
2nd round	✓	X	✓						

INTRODUCTION — FROM THEORY TO PRACTICE

Monitoring is a fundamental step in the administrative routine. There are many examples that show the importance of monitoring, in the most diverse areas of knowledge: routine physiological examinations help identify health changes that can lead to more serious complications; climate monitoring enables us to anticipate environmental disasters; and monthly assessments allow the teacher to assess whether the student has absorbed the class contents. While planning charts the course, monitoring helps managers correct the course along the way.

With the development of digital technology, monitoring has advanced at a rapid pace. The computational power used to capture and analyze data allows companies to assess, in real time, the volatilities of market preferences. Although there are discussions about how companies should use this data strategically, the capture and consequent analysis of data have become indispensable, as the widespread use of the internet, social media, and free file and video sharing sites have facilitated the transmission of persuasive information on diverse issues, as well as information involving organizations or their leaders in relation to social issues (Briscoe & Gupta, 2016).

Faced with the possibility of expressing themselves with just a few clicks, citizens, social movements, diverse groups, and even political parties have been using this type of internet platform as a tool to disseminate different types of information. The purpose of these actions is also quite diverse, and can serve both democratic interests and the right to information and freedom of expression, as well as the propagation of fake news and manipulation of public opinion (Fundação Instituto de Administração [FIA], 2021).

In the area of research on social activism, there are methodological choices and challenges. In particular, information from the internet can be used to analyze both social and organizational interactions and corporate responses to online activism. In this sense, it is likely that electronic traces will be identified that reveal various intentional and unintentional impacts of activism against organizations. These residuals can be used, for example, to map the dynamic emergence of a network of internal (and external) activists over time, or for an event over history (Briscoe & Gupta, 2016).

Given the multifaceted nature of activism around organizations, studies have approached the phenomenon through a variety of theoretical, empirical, and epistemological lenses. Much research has developed from traditional sociology, which has expanded, from a focus on the role of the state to the study of protests in organizational fields involving companies, universities, hospitals, and other types of market actors (Davis & Thompson, 1994; Etter & Albu, 2021; Rao et al., 2000; Van Dyke et al., 2004).

There is a consensus in the literature that the way activists operate, and influence organizations, is evolving over the years with information and communication technologies (ICTs). Earl and Kimport (2011) argue that advances in ICTs have dramatically increased instances of online activism, reducing the costs of participation for activists and the need for social movements to mobilize infrastructure. Along these lines, a study by Zhang and Luo (2013) showed that online activist campaigns using popular social media platforms can generate significant concessions and increase corporate philanthropy.

The importance of social media for contemporary protest movements has been theorized from different perspectives. Overall, the functionality of social media has been emphasized as a means of sharing information and a tool for organizing protests (Bimber, Flanagin, & Stohl, 2012). Less attention has been paid to the potential impact of internet users using social media to effect social change and how companies will respond to these demands. Digital artifacts linked to protests and other forms of online activism offer valuable information, but these are rarely analyzed by researchers (Jenzen et al., 2021).

In the organizational research field, this information could be analyzed, for example, through data mining and applying statistical methods or qualitative analysis (Mercado & Silva, 2013; Ross & Cruz, 2021). However, in the practical field, the entire data generated on social media needs to be analyzed in a process carefully, closely monitored, and measured. That is done by monitoring social media, in which companies observe how the public interacts and reacts to specific topics on their digital channels.

While, on the one hand, online activism is a phenomenon of relevance in academic and organizational environments, on the other, drastic changes from the year 2018 onward have negatively marked the history of social media and its application programming interfaces (API), which is a series of routines and standards established by a software program for the use of its features (Meirelles, 2019). Thus, where previously researchers were able to collect and monitor all publications posted on a users' wall/feed, as long as they were in public (exhibition) mode, after the discontinuity and restrictions of APIs, this practice was no longer possible. This had a significant impact on the process of monitoring social media. Based on this problem, the main objective of this study is to develop and structure a social media management tool.

CONCEPTUAL ASPECTS

The term 'social media' refers to the practice of using any specific medium for social purposes. It is also commonly used to describe web applications, such as websites and user-generated content sharing platforms (Kaplan & Haenlein, 2010). The authors highlight that social media refers to dynamic internet platforms and their ability to support interactivity and real-time communication.

The adoption of social media by society has led organizations to adopt technological tools to disseminate important organizational information to their stakeholders (Kim & Youm, 2017; Smits & Mogos, 2013). According to Lee, Hutton and Shu (2015), social media have become a viable channel for disseminating strategic information, when compared to traditional information channels. They enable greater reach and allow companies to directly and quickly send the intended message.

However, social media can be viewed from two contradictory perspectives. The first, a positive perspective, is that it reduces the cost of communication, increases the speed of information, and eliminates physical distance. However, the negative perspective is that social media impoverish political debate through empty discussions (predominantly simplification), the use of fake news as a strategy to reiterate ideologies, and the loss of privacy in cyberspace (Cavalcanti et al., 2019).

Initially, social media were used by managers for marketing purposes (Arnaboldi et al., 2017). Later, organizations began to use these tools mainly to convey their idealized image (Harquail, 2011). This means of disclosure has been gaining popularity, and its use for regular operations is increasing among companies (Smits & Mogos, 2013).

Within this dynamic, those that use the latest social media technologies seem to outperform their competitors and report some benefits, such as lower costs and better efficiency (Harris & Rea, 2009). Therefore, researchers have reported the importance of understanding the relationship between social media use and organizational performance (Smits & Mogos, 2013; Ghardallou, 2021; Wetzstein et al., 2011).

Compared to traditional media, the internet is much more difficult to regulate. In addition, social media and digital communication significantly expand the participation of various actors in collective action. In this context, social media has been highlighted by researchers and managers, in particular, for imposing pressure for quick responses and its impact on the performance of organizations (Cavalcanti et al., 2019; Ghardallou, 2021; George & Leidner, 2019; Gomez-Carrasco & Michelon, 2017; Lewis, 2005; Luo et al., 2016; Van Dijck & Poell, 2013).

Given the relevance of social media for research in the organizational field, its understanding becomes crucial. Anecdotal evidence demonstrates the importance of social media. In 2021, the Twitter application base in Brazil, for example, comprised approximately 17.46 million users. The number of users in Brazil is forecast to reach 18.39 million by 2025 (Statista, 2021).

In the academic/practical field, Twitter has significant relevance for understanding people's actions and reactions on a given topic, or even for measuring reputation and organizational strategy. In a recent survey, Abdulaziz, Alotaibi, Alsolamy and Alabbas (2021) studied, through Twitter, the main topics and changes in people's concerns about the COVID-19 pandemic.

A study by Shin and Ki (2022), explored, through tweets, how for-profit and non-profit organizations position themselves in relation to the environment and public response. The analysis showed that for-profit companies tend to discuss their green products, while non-profit organizations are more inclined to describe a degraded environment. Furthermore, the study revealed that tweets generate a high number of likes and responses when organizations are for-profit and the messages emphasize green products.

Vogler (2020) evaluated the reputation of Swiss universities using Twitter. According to the author, reactions to tweets in the form of retweets and likes indicate how much attention the tweets received. Reactions to social media content depend on how the content is presented, with emotions being an important predictor of so-called virality.

Regardless of the research topic, Twitter is an important source of information for organizational studies. For Jung, Naughton, Tahoun and Wang (2018), Twitter is one of the most used social media platforms for corporate purposes. However, for data collection, the current API limits the number of tweets collected, making its use impossible for studies that demand timely data. Furthermore, it is not clear, in the current literature, which managerial actions at the strategic level companies should develop for monitoring and managing social media, which is the central problem to be discussed in this study.

Application programming interfaces

To assist in the integration process between systems, application programming interfaces (APIs) act as bridges, transporting data between a client and a server.

Without this process being even noticed by the user, APIs are present in the functioning of several programs and applications. They consist of a series of commands that allow

users and applications to communicate with websites and request data hosted on their servers (Silva & Stabile, 2016).

The operation of APIs generally occurs when the customer requests access to certain data made available by an information bank, service, or device, for example. To access this specific information, the client makes a request to the API. The interface is able to search the server and return the requested responses in data format, which are delivered in their purest state.

D'Andréa (2021) highlights that through APIs, online platforms offer access to part of the data generated and/or collected, such as musical genres assigned to an artist (Spotify), posts linked to a term or a hashtag (Twitter) or the total views, likes and comments, up to a given particular moment, of the videos posted by a channel (YouTube).

Regarding security, through an API, it is possible to have greater control over access permissions to a company's software and hardware. It identifies who has attempted to access the system, and locates where the request came from. In addition, it is possible to configure which information will be available to the client at the time of integration.

According to Meirelles (2019), 2018 negatively marked the history of social media APIs. After so many (political) scandals involving the irresponsible use of data, the Facebook conglomerate was forced to act publicly in an attempt to argue that — at least from that moment on — it would be more committed to the privacy and data of its users. According to the author, this fact directly impacted the social media monitoring market, which saw its data source decrease even further.

From a broader perspective, it can be said that anyone — independent researcher or academic — who worked with social media data ended up suffering from the company's latest decisions. This reality, however, is no longer new: in recent years, drastic changes in APIs have already pointed to more severe restrictions (Meirelles, 2019).

As per the timeline (Figure 1), the Facebook and Twitter APIs were officially released in 2006. In April 2010, Facebook released the first version of the Graph API.

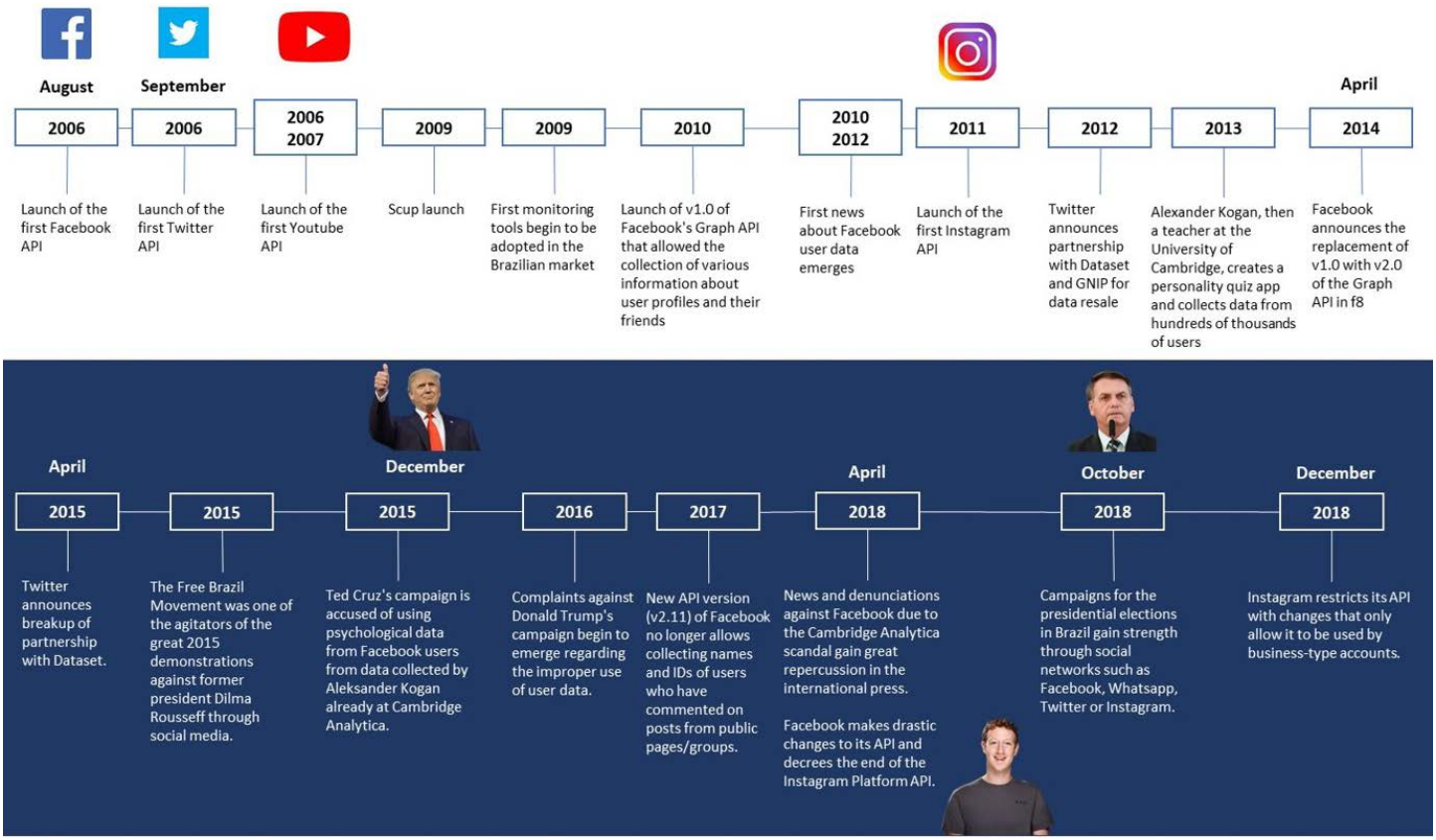

Figure 1. API timeline.

Source: Adapted from Meirelles, P. (2019). Histórico das APIs no monitoramento e pesquisa em mídias sociais. Brasília, DF: Instituto Brasileiro de Pesquisa e Análise de Dados.

The first monitoring tools were adopted by the Brazilian market at the end of this period, with names such as Scup, Radian6, Trendrr, and Socialmention.

According to Meirelles (2019), at this first moment, the infinite amount of data made available by Facebook and Twitter was already seen with great optimism by different actors: companies themselves, marketing professionals, researchers (academics), etc.

However, after years of exponential growth and great enthusiasm, the first problems began to emerge, especially for Facebook. Between 2010 and 2013, some articles 'denounced' the absurd amount of data and information that Facebook held about its users, and that it shared too readily with third parties.

After four years of practically unlimited access to the infinity of user data, v1.0 of Facebook's Graph API was finally discontinued in April 2014. Previously, the tools had been able to collect all the posts on users' walls/feeds, as long as they were in public (exposure) mode, but from the discontinuation of the API this would no longer be possible.

After the 2015 crash, the social media monitoring and research market was forced to reformulate the way it worked with data in practice. The biggest change in this regard was the redirection of the focus on Facebook data: once it was no longer possible to follow what people shared on their public profiles, the tools — and the market — turned mainly to other social media.

In late 2017, Twitter announced the release of premium APIs. The proposal was to find a middle ground between the limits of the free (and public) API, most commonly used by monitoring tools and independent researchers, and the whirlwind of data from the enterprise API, the so-called firehose available through GNIP (which gave access to all platform tweets since 2006). According to the Brazilian Institute of Research and Data Analysis (Instituto Brasileiro de Pesquisas e Análise de Dados (IBPAD, 2021), Twitter has always maintained a favorable stance toward requesting data via API. Over the years, it has developed different API 'options,' mainly to meet market demand: while, on the one hand, the free version has always been easily accessible (with a limit of seven days backdated and a request every 15 minutes), at the other end, the so-called firehose (or enterprise) was very costly for companies interested in unlimited access to the platform's publications (backdated to 2006, when it was founded). To fill this gap (with a possibility of backdating for up to one month and greater volume of data in the request), the company launched the premium APIs.

Thus, the limitation of platforms in accessing data requires of researchers and managers alternative ways to process, systematize, and manage these data. It is not enough simply to extract the data; it is also necessary to develop an analysis capable of helping companies in strategic decision-making.

ANALYSIS OF THE SITUATION/PROBLEM AND THE PROPOSED SOLUTION

This section presents the steps developed for extracting data from the web, the script (set of commands), the segmentation of the processing for extracting the posts, the organizational actions, and the flowchart for the corporate response to online activism.

Data extraction from the web

The theme proposed for this study is part of an academic research project, which analyzed the tweets of the hundred largest higher education institutions (HEIs) in Brazil in terms of the number of students. To this end, it was observed whether the HEIs used social media, then we sought to evaluate the posts of their users on Twitter from 2013 to 2019. Due to the limited access to historical data on the platform, which would make retroactive analysis of the posts unfeasible, it was decided to develop a script (a set of commands) capable of recording data from web pages when accessing the site.

For the development of this study, the web bot algorithm and the web scraping technique were used to extract data from Twitter. The web bot is an algorithm used to analyze and extract information from websites in a systematic and automated way (Omari et al., 2016). These bots capture information from web pages and register the links identified, so that they can be used later. Bots crawl and scrape data using two techniques: web crawling and web scraping. These techniques can be used simultaneously or as two separate tasks (Khalil & Fakir, 2017).

Web crawling is the process responsible for performing searches for web pages and indexing them. This technique captures information from websites and registers the links found, so that they can be located in the future (D'Haen et al., 2016). The process starts from a seed, which is a starting point for finding new addresses to visit. As the bot visits these addresses, hyperlinks are identified and added to the list of addresses to be visited (Khalil & Fakir, 2017), providing the locations of other pages, which keeps the database up to date.

Web scraping is the extraction process used to automatically collect different data from websites, converting unstructured into structured information, for subsequent analysis (Zhao, 2017), in a procedure known as data scraping. The bot, which uses the web scraping technique,

is programmed to make requests to a web server from a predefined list of URLs. After the request has been made, the necessary data is extracted. The data obtained is copied and can be exported in files in JSON (acronym for JavaScript object notation) and comma separated value (CSV) formats, among others. Normally, this process simulates human navigation when using a website. However, the bot can make more requests than those made by a person.

It should be noted that Twitter — the social media that is the focus of this study — has adopted a privacy policy (Twitter, 2022a) that makes all tweets registered by its user's public, i.e., they become visible and searchable by anyone. As per Twitter's privacy police, whenever content is publicly shared through a tweet, the user agrees that this information can be widely disseminated. Thus, the user is responsible for his/her own tweets, and for other information provided through the services made available.

The construction of the script for this article followed the Twitter guidelines, as can be seen at the link https://twitter.com/robots.txt. Figure 2 shows the parameters used to develop the bot.

```
# Every bot that might possibly read and respect this file
# ======================================================
User-agent: *
Allow: /*?lang=
Allow: /hashtag/*?src=
Allow: /search?q=%23
Allow: /i/api/
Disallow: /search/realtime
Disallow: /search/users
Disallow: /search/*/grid

Disallow: /*?
Disallow: /*/followers
Disallow: /*/following

Disallow: /account/deactivated
Disallow: /settings/deactivated

Disallow: /oauth
Disallow: /1/oauth

Disallow: /i/streams
Disallow: /i/hello
```

Figure 2. Parameters used for the Twitter bot development.
Source: Twitter (2022b). Robots. Retrieved from https://twitter.com/robots.txt

The tool was developed following the guidelines of the platform, as well as best practices in software development, in accordance with the provisions of Law No. 13,709 of August 14, 2018 (Lei n.º 13.709 [LGPD], 2018), by a natural person or by a legal entity governed by public or private law, with the objective of protecting the fundamental rights of freedom and privacy and the free development of the personality of the natural person.

Script (set of commands) for data extraction

In the task of extracting data from Twitter, the most objective way is to use the tools provided by the social network itself. Twitter, as mentioned, offers a data extraction API (Meirelles, 2019). However, this API limits the user to extracting recent tweets, published no more than seven days prior to the post extraction date, making it impossible to extract data published before that period.

Historical posts are relevant to a significant amount of analysis, including informal reading of the social network by users, who tend to look for old posts. The establishment of a historical monitoring strategy for social networks can also bring a series of benefits to the company, such as understanding the behavior of the public by observing their preferences, as well as understanding the brand perception of the public that uses these media. The relevance of historical data, in addition to that provided by the API, is significant. Twitter offers advanced searching, which allows users to filter keywords and date ranges, as shown in Figure 3.

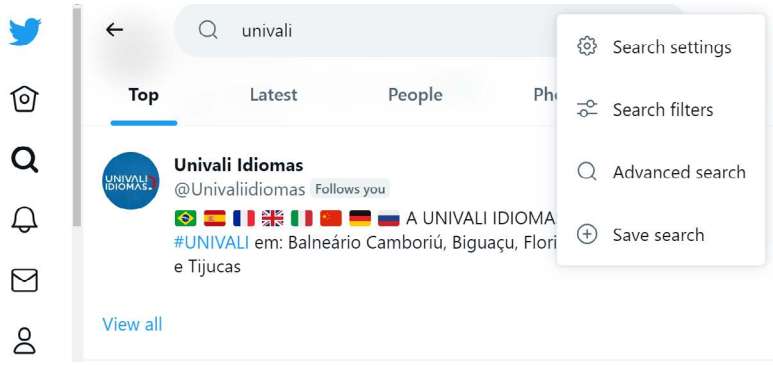

Figure 3. Twitter advanced search.
Source: Survey data — Twitter.

When they need a systematic analysis of historical publications, supported by statistical methods and computational tools, those interested in analysis must carry out the collection and cataloging manually, so that the data can be effectively analyzed. This demands significant effort on the part of researchers and managers. It is in this first stage of collecting historical publications that this work is important, as the new form of consultation solves the limitation of the API offered by the social network, and significantly reduces the data collection time.

The tool developed consists of a script (set of commands) in the Python programming language, the Selenium browser, and the BeautifulSoup library, to implement a robot that simulates the use of the social network by a human performing manual searches, as per the schema shown in the Figure 4.

The script was developed based on the work of Choudhary (2021), available on github.com. At the time of writing this article, the last update of that tool was on 10/03/2017, and we were not able to extract data from the current version of the Twitter website.

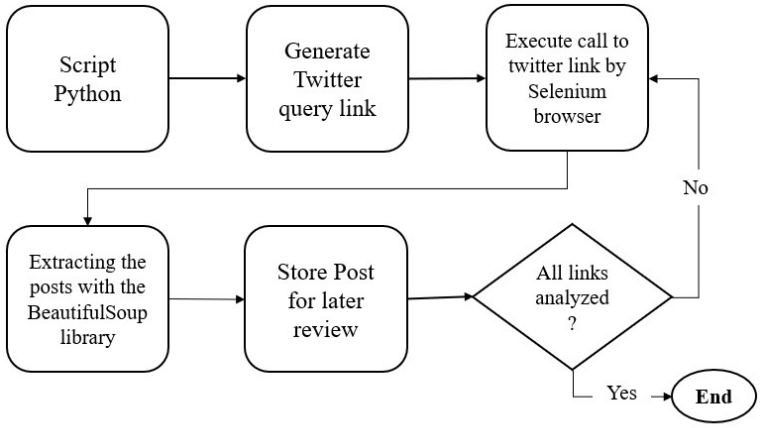

Figure 4. Robot schema for data extraction.
Source: Prepared by the authors.

In the first step, the web crawling technique was used to track the links generated by the advanced searching, offered by the social media platform Twitter. For the set of test keywords, we used terms that identify the HEIs studied. It should be noted, as described in the introduction, that the proposed topic for testing the robot is part of an academic research project analyzing tweets on HEIs in Brazil. However, other keywords could also be used, depending on the desired research objectives.

Each keyword was crossed with each of the days in the period 01/01/2013 to 12/31/2019, generating Table 1, which has three columns: one for the keyword searched on

Twitter, one for the reference dates, and one for the reference dates + 1 day. The higher education institution Universidade do Vale do Itajaí (Univali) is used here as an example.

In the advanced search on Twitter, it is possible to inform a wider range of dates (more than one day). However, we chose to analyze posts day by day in order to minimize the possibility of a large volume of data on the same page.

For each row of Table 1, a link was created, which can be generalized by the following model: <https://twitter.com/search?q=[**KEY_WORD**]%20until%3A[**DATA_REFERENCE+**]%20since%3A[**DATA_REFERENCE**]&src=typed_query>

Using the first row of Table 1, the following substitutions were made, as shown in Table 2.

Table 1. Data to generate queries on Twitter.

Keyword	Reference date	Reference date + 1
univali	01/01/2013	01/02/2013
univali	02/01/2013	03/01/2013
univali	01/03/2013	01/04/2013
…		
univali	12/31/2019	01/01/2020

Note. The date range used in this query is for illustrative purposes only, and any desired range can be used.

Table 2. Keyword and reference date.

Items	Replaced by
[KEY_WORD]	replaced by 'univali' — example
[DATA_REFERENCE]	replaced by '2013-01-01' due to the year-month-day format required by Twitter
[DATA_REFERENCE]	replaced by '2013-01-02' due to the year-month-day format required by Twitter

Note. The keyword 'univali,' used in this query, refers to the research object; any term can be used here, according to the needs of each study.

After replacing the data from Tables 1 and 2 in the model link, the following link was generated:

<https://twitter.com/search?q=univali%20until%3A2013-01-02%20since%3A2013-01-01src=typed_query>.

This link returned the posts in which the keyword 'univali' was mentioned in posts dated 01/01/2013. As there were no posts that met this query, a page with no posts was returned, usually with some advertising, as in the example of Figure 5, presented in the browser, when the link is typed in the address bar.

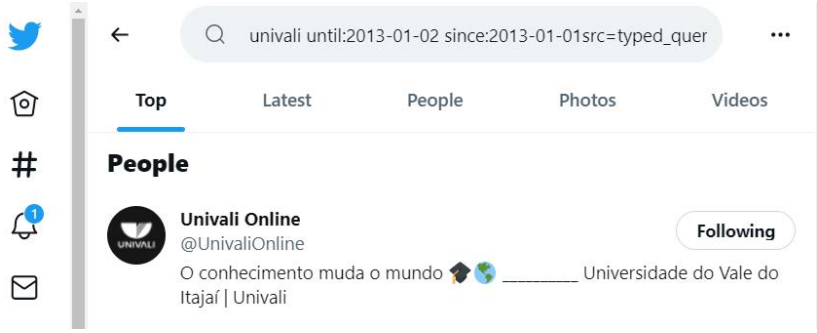

Figure 5. Page returned by the link.
Source: Research data — Twitter.

In total, 265,811 links were created, which answered queries for 91 keywords, multiplied by 2,921 days.

In the second step, the robot executed the link call and waited for the results page to be made available by the social media servers, simulating a query performed by a human user. In this access, the Selenium tool was used, which is nothing more than a web browser. Selenium can perform tasks such as clicking on a link, using a Python script (Selenium, 2021).

Once access to historical posts was made possible, web scraping techniques were used (Zhao, 2017), to enable the collection of the text of the posts. Despite the apparent complexity of the commands used, on each page presented in the queries, Twitter developers chose to use tags with semantics that are easily interpreted by most English-speaking readers (W3schools, 2021).

It is worth mentioning that this analysis described refers to the version of the Twitter pages available in January/2021, and the company responsible for the social media may change the way the website is built at any time. In the analyzed version of the website, each of the tweets is initially delimited by the HTML tags <article> and </article>, described in detail at <https://www.w3schools.com/tags/tag_article.asp>.

The Python programming language offers some libraries that facilitate access to these tags. In this work, we chose to use the BeautifulSoup library.

In order to access all <article> tags and their content, the find_all ('article') method is used to keep all the posts separate. After this initial collection, the CSS and JavaScript tags of the page can simply be ignored, so that the post text can be collected. An example of this relationship between image and code is presented in Figures 6 and 7.

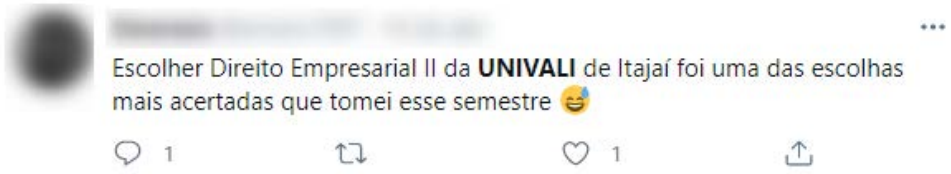

Figure 6. Tweet as presented to the user by the browser.
Source: Research data — Twitter.

```
<article role="article" tabindex="0" class="css-1dbjc4n">
  <div lang="pt" dir="auto" class="css-901oao ">
    <span class="css-901oao">Choose Business Law II from </span>
        <span class="css-901oao css-16my406 ">UNIVALI</span>
        <span class="css-901oao">
            from Itajaí was one of the best choices I made this semester
        </span>
        <span>
       <img alt="Smiling face with open mouth and cold sweat" draggable="false"
         src="https://abs-0.twimg.com/emoji/v2/svg/1f605.svg" class="css-9pa8cd" />
    </span>
  </div>
</article>
```

Figure 7. Tweet <article> tag from Figure 5 (summary).
Code segment that builds the image in Figure 6. Source: Prepared by the authors.

Social media posts, however, are not limited to alphanumeric symbols, such as the letters A to Z and the numbers 0 to 9. A very valuable semantic content for any analysis can be inferred by analyzing emoticons, which are small images that are added to the text of the posts.

In this work, extraction of emoticons was treated separately. The social media Twitter has chosen to use custom images to present its emoticons, many of them non-standard Unicode commands (Unicode®, 2021). These images are shown through the HTML tag <img>, used to load the images from the server to the web browser.

Even with this particularity, Twitter gives the Unicode character (Unicode®, 2021) or the description of the image in the alternative text attribute to the image. This extraction was performed from this attribute.

Figure 8 shows the alternative text to the emoticons presented in Figure 5, highlighted here with dark borders. In the highlighted segment, it is possible to observe the description of the emoticon presented: 'Smiling face with open mouth and cold sweat'.

```
<img alt="Smiling face with open mouth and cold sweat" draggable="false"
  src=https://abs-0.twimg.com/emoji/v2/svg/1f605.svg class="css-9pa8cd" />
```

Figure 8. Alternative emoticon text, image tag.
Twitter page source code segment. Source: Prepared by the authors, based on survey data.

Segmentation of processing for extracting posts

Data extraction robots can execute a far greater number of requests to the site than a human browsing the same site in the normal way. Often, for security reasons, sites fail to respond to a large volume of requests from the same computer to avoid overloading the servers. When developing the data extraction script, it was decided to make accesses with intervals of about 20 seconds between each request. As shown in Table 3, the estimated time needed for all the requests to be carried out was around 61 days.

Table 3. Calculation of requests.

Links	Seconds between requests	Time in seconds	Time in minutes	Time in hours	Days to request links
265,811.00	20.00	5,316,220.00	88,603.67	1,476.73	61.53

Note. The number of links refers to the search terms used. Source: The authors, based on survey data.

Given the initial estimate of about 61 days to request the links, added to the page processing time, the time for data storage and the time of inoperability, whether due to internet failure, computer crash, or error and correction time of the script, we opted to segment the analysis of the links in simultaneous execution, using 20 computers.

As shown in Figure 9, the computers were used simultaneously, dividing the work of extracting the data between them. This work was orchestrated manually, via a remote computer. Based on this scheme, it was possible to increase the processing capacity in the data extraction.

Using the scheme illustrated in Figure 6, it was possible to extract about 1.8 million tweets between 12/10 and 12/20/2020. All data were kept in archives and are available for subsequent reading and analysis.

From the database, it is possible to perform various technical procedures, according to the purpose of the analysis. This database can also be easily read by other software, for the presentation of management information, such as sentiment analysis, tweets polarization, descriptive analysis, data mining approach with exploratory graph analysis, among other quantitative or qualitative techniques. Table 4 shows the fields extracted for each tweet.

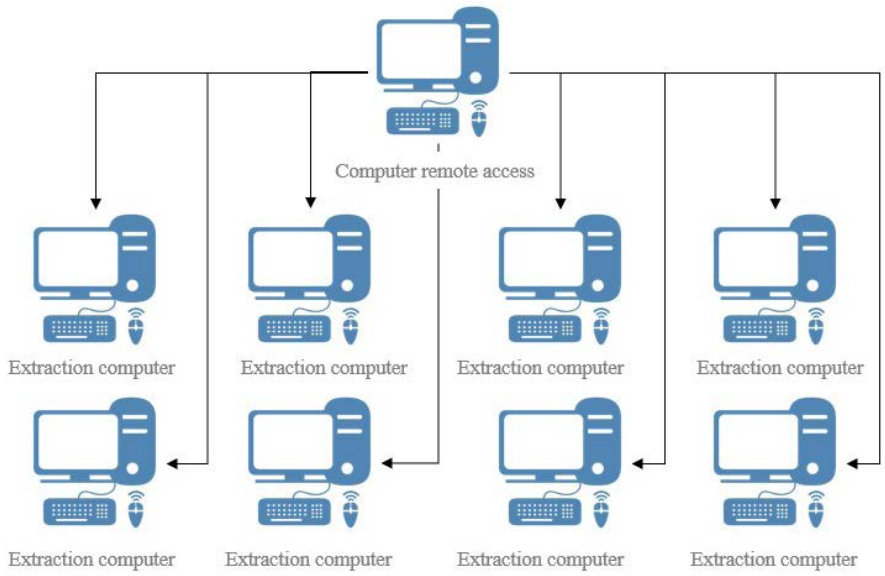

Figure 9. Processing segmentation scheme.
Source: Prepared by the authors.

Table 4. Example data extracted by tweet.

Fields	Extracted data
search term	univali
url_day	https://twitter.com/search?q=univali%20since%3A2019-12-18%20until%3A2019-12-19&src=typd
url_tweet	/hanna_coisinha/status/1207147824615493632
date	2019-12-18T03:57:23.000Z
date_descr	Dec 18, 2019
name	@hanna_coisinha
tweet_text	What will I miss most about Univali???? The teachers 42
hashtags	
respost	4
retweet	0
like	2
emoticon	|| Guy crying a lot|Red heart

Note. A URL refers to the network address where some computing resource is located. Source: The authors, based on survey data.

Through the search terms, it was possible to carry out queries that met the desired data extraction, from which the tweet URL, date, name, tweet text, hashtags, replies, retweets, likes, and emoticons were extracted. Using this data extraction process, it is possible for researchers to advance in the field of organizational research, whether by influencing performance, the corporate response to social activism, the impact on reputation, or institutional image driven by social media. The developed script is presented in the shared data in the end of this article.

Corporate response to online activism

From the script developed for the data extraction, it is possible for organizations to monitor the online environment in a systematic way, focusing on responding to demands that arise instantly and that may impact organizational management, whether through new strategies of online social activism, leading groups of activists to empowerment by expanding the reach of claims assumed as collective rights, or on isolated issues, in defense of humanitarian, political,

cultural, and economic causes, but which, at the same time, demand attention from organizations due to their potential impact.

In this study, from the collection and analysis of tweets from higher education institutions (HEIs) in Brazil, it was possible to highlight the main categories of influence of activists and stakeholders in the online environment and, from that, to infer strategic actions in response to the identified demands. Finally, the implementation of the social media monitoring process for companies is proposed.

Practical example (case study)

For the purposes of practical demonstration, from section Script (set of commands) for data extraction, tweets from an HEI in the state of Santa Catarina were extracted. The choice of HEI was due to its relevance in national and international rankings. Data extraction resulted in a total of 24,787 tweets, and the random sample extracted by the software comprised 5,984 tweets from the IES, in the period from 2013 to 2019. It is worth noting that the choice of the period sought the highest number of interactions with the IES at different times. For the data analysis, the Orange Data Mining tool was used. First, the categories were identified through cluster analysis.

Based on the frequency of terms and through cluster analysis, three categories of interaction with the HEI were identified: financial, pedagogical, and social. When a user makes a comment about the HEI on Twitter, whether positive or negative, it relates to financial, pedagogical, or social issues. Table 5 presents the keywords with the highest frequency, by category.

The word 'university' was used most frequently (610 times) and is related to the expression 'I'm at UNIVERSIDADE' or 'I am at the University', being classified in the social category.

Figure 10 shows the word cloud without distinction of categories, covering the entire database.

Table 5. Keywords by category.

Keywords	Categories	*No. tweets*	%
Monthly payment; pay; money; traffic ticket; credits; ticket; real; Fies; handbag; demand; to spend.	Financial	56	0.9%
Classroom; course; studying; lecture; test; grades; teachers; search; entrance exam; registration; education; graduation.	Pedagogical	1,154	19.3%
Campus; university; college; crowd; people; left; tomorrow; early; I arrived; I passed; to wake up; to sleep.	Social	4,033	67.4%
Out of context		741	12.4%
Grand total		5,984	100%

Note. Source: Prepared by the authors, based on survey data generated by the Orange Data Mining tool.

Figure 10. Word cloud.

Source: The authors, based on survey data.

From the identification of the most frequently used terms, it is possible to identify the main categories of online interaction with the organization and then carry out the sentiment analysis.

For the sentiment analysis and classification of the 5,984 tweets from the IES, the Orange Data Mining system (Python programming) was used. First, the file was read. Next, the text processing was performed, transforming the letters into lowercase, removing accents, URLs, and stop words, then the word clouds were generated to verify and clean the database. The results are shown in Table 6.

Table 6. Sentiment analysis by category.

	1. positive	2. negative	3. neutral	Grand total	%
F — financial		3	53	56	1%
P — pedagogical	69	127	958	1,154	19%
S — social	177	338	3,518	4,033	67%
O — out of context	38	91	612	741	12%
Grand total	**284**	**559**	**5,141**	**5,984**	**100%**
%	5%	9%	86%	100%	

Note. Source: The authors, based on survey data generated by the Orange Data Mining tool. The O — out of context category refers to tweets not classified in the financial, educational, or social categories.

From the collection and classification of the HEI's tweets, it was found that the highest frequency of interactions was related to the social category, and, in relation to the sentiment scale, 86% were classified as 'neutral.'

Among the classifications of positive and negative scales, the negative one stands out, with 559 tweets. These are concentrated in the social and pedagogical categories, indicating questions to be answered and reassessed by the HEI.

It would also be necessary to evaluate the tweets with the highest engagement (number of likes and retweets), with a view to identifying activist movements, whether for individual or collective reasons. In addition, through the sentiment scale, it is possible to carry out a qualitative analysis, highlighting the main motivators for positive or negative issues of interactions with the organization.

Implementation of the social media monitoring process

The social media monitoring and management process needs to be implemented in companies not only to evaluate institutional reputation, or for communication and marketing purposes, but in a broader way, covering an intersectoral perspective. It is at this point that the data extraction script developed in this study can help with the extraction, categorization, and distribution of posts, and with the classification of positive, negative, and neutral tweets, as presented in the practical example of subsection 3.4.1.

In general, the social media monitoring process is the continuous monitoring, structured by metrics, of all interactions, reactions (e.g., likes, content, and shares), and comments from the public, showing what has been said about the company in the social media. The intersectoral view proposed in this article requires a broad analysis of all interactions that occur in social media. These interactions can be captured from the following points:

- the brand name and its most common variations, if any;
- the names of the company's main products and services;
- mentions, with social media tags and hashtags; these mentions can have several search terms, such as the name of the brand, product, service, campaign slogan launched, among others;
- keywords that are close to the subject of the brand, product, or service;
- mentions filtered by location, when applicable;
- additionally, search for mentions of the brand's main competitor, to understand the public's perception of companies in the same segment.

Through the social media monitoring process, the company will be able to understand who are the supporters and detractors, social activists, the posts with the most engagement, the words that are most associated with the company, and the reaction of the general public, over time.

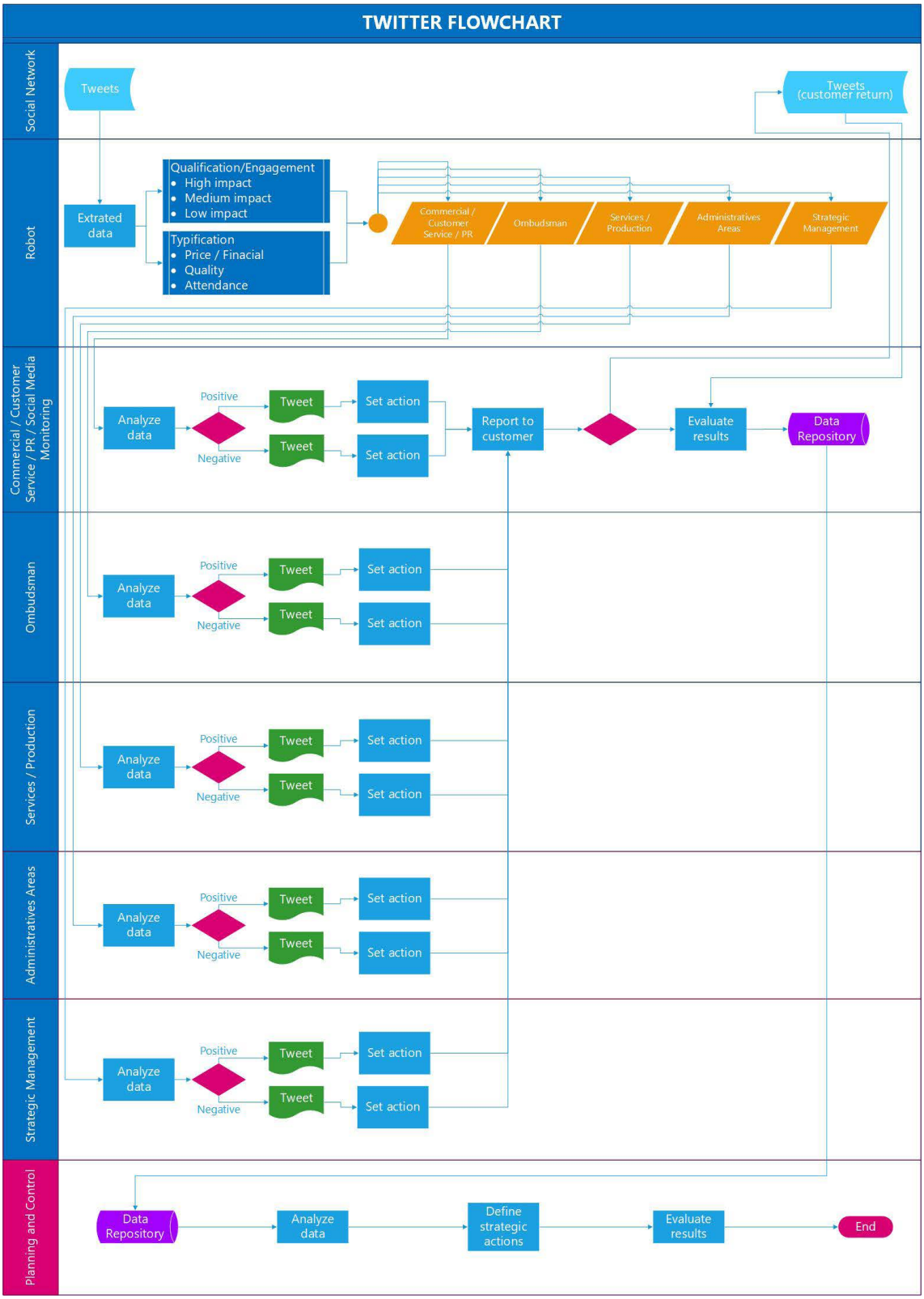

Figure 11. Social media monitoring process flowchart.
Source: The authors, based on survey data.

Figure 11 shows the flowchart of the monitoring process. This process starts with posts on social media. The robot — script (set of commands) for data extraction developed in this study — performs the capture and storage of posts in a structured database, containing some important fields such as: search term; url_day; url_tweet; date; date_descr; name; tweet_text; hashtags; answer; retweet; like; emoticon.

In the database, it is possible to qualify the posts on a scale: high, medium, or low impact. The metric to measure the impact can be based on likes and retweets, that is, when a post with a significant number of likes and/or retweets occurs, it is classified as having a high impact, whether positive, negative, or neutral.

In addition to the impact classification, the categorization of tweets is necessary for the distribution of messages to the respective sectors of the company. In the practical example presented in this article (in a higher education institution), the classification was given by three main categories: financial, pedagogical, and social. However, these categories may vary depending on the type of company, segment, and customers, among others. Thus, the flowchart presents three essential generic categories: price/financial, quality, and service.

The price/financial category is associated with all mentions of the price of products and/or services, issues associated with the company's financial aspects. The quality category is related to the level at which the product or service meets the previously established requirements. Finally, the service category concerns the level of satisfaction of customers and other company stakeholders. Although this category is considered subjective, each person has a different perception of quality, and this is a strategic tool used by companies in their fundraising.

After classifying and categorizing the tweets, the next step is to distribute the information for the analysis and response of the responsible sectors. Based on the case study developed in this research, distribution to five fundamental areas is proposed: (1) commercial, customer service — SAC, and public relations — RP, (2) ombudsman, (3) services/production, (4) administrative areas, and (5) strategic management, which may vary according to the organizational structure of each company or HEI. For the implementation of the process, the importance of structuring the social media monitoring area is highlighted. Although this activity may be linked to sectors such as commercial, customer service, PR, it is essential that the company directs resources for this purpose.

Each sector analyzes the data and develops specific actions, according to the demand presented through the classification and categorization of tweets. If positive, the sector evaluates the points mentioned and sends the response to the social media. If negative, the sector takes action to correct or improve it, according to the published feedback, and communicate it to the commercial, customer services, public relations (PR), and/or social media monitoring sectors, which send the response to the social media.

The commercial, customer service, PR, and social media monitoring sectors evaluate the results, which are stored in a data repository. From this repository and with the metrics of classification and categorization of tweets, the strategic area can analyze the data and define strategic actions. This last stage of the process is fundamental for the identification of necessary resources and competences and for the strategic positioning of the company.

Strategic organizational actions

The process of extracting, monitoring, and managing social media requires a set of proactive organizational actions that, if implemented, can contribute to the management and corporate response to the demands of the online environment. From the extraction of data from the web, through the script (set of commands) presented in section 3.2, and the case study carried out, it was possible to infer a set of improvement actions for the process of data extraction, monitoring, and management of Twitter user interactions with the organization. Table 7 shows these actions.

With the implementation of the actions described in this study, it is possible to improve the process of monitoring and managing social media, having as premises the improvement of stakeholder engagement, the identification of activist movements and dissatisfied followers, the detection of false information, the monitoring of actions competition, and the identification of the organization's strengths and weaknesses. Therefore, it is necessary to implement the social media monitoring process in companies. This process requires integration between areas/sectors, so that the company can respond to demands arising from the online environment.

CONCLUSIONS AND TECHNOLOGICAL/ SOCIAL CONTRIBUTIONS

The main objective of this study was to develop and structure a social media management tool. In the practical field, this study helps companies systematically extract data from Twitter for the development of a robust monitoring system and process, enabling the promotion of strategic actions in the online environment.

Table 7. Strategic actions in response to online activism.

Phases	Actions
Data extraction	Automate the process of extracting Twitter data from the script (set of com-mands) presented in section 3.2 of this article.
	Decode thousands of data (texts and images, including emojis) in real time, in an organized database, for future analysis.
	Capture user mentions related to the organization, with tags and hashtags, on Twitter — hashtags are used as search engines, categorization tools, and, in particular, marketing and communication tactics.
Monitoring	Monitor with historical data organized in a chronological way and follow-up of strategic themes, through keywords.
	Set metrics for interactions, reactions (as likes, published content, and shares), and general public comments.
	Create analysis categories to identify the main drivers of stakeholder interac-tions with the organization (as shown in subsection 3.4.1 Practical example).
Management	Identify positive and negative sentiments over time and make organizational adjustments where necessary.
	Identify tweets with the highest engagement (number of likes and retweets) and which topics they are related to.
	Promote publications that generate a positive impact on the organization's image with a view to improving the organizational relationship with its stake-holders.
	Based on negative feedback on a particular strategy or action carried out by the organization, prepare an action plan to correct the aspects highlighted.
	Analyze stakeholder interactions with industry organizations, competitors, agencies, NGOs, and other institutions of interest to the organization.
	Develop intelligent reports from the monitoring of social media, via qualitative and quantitative indicators, for the strategic management of the organization.

Note. Source: Prepared by the authors.

This practice is justified by the fact that social media has become ubiquitous and is playing an increasingly critical role in today's business environments. The new informational possibilities they present affect the relationships between stakeholders and managers of organizations for the most diverse reasons and individual or collective interests.

From the case study presented, it was possible to infer a set of actions for the process of data extraction, monitoring, and management of Twitter user interactions with the HEI under study. The identification of the main categories of interaction with the HEI could help in the process of responding to the demands of the online environment.

Implementing the social media monitoring process requires cross-sector integration. Although studies in social media focus mainly on the area of marketing, this study broadens the field for the discussion on how computing techniques can serve to support administration and decision-making.

Through the systematization of the data, it is possible to carry out other studies, which may contribute, for example, to the literature that examines the strategic behavior of companies regarding the use of Twitter for the dissemination of information (Davis & Thompson, 1994; Etter & Albu, 2021; Ghardallou, 2021; Rao et al., 2000; Van Dyke et al., 2004).

Evidence that the application of the script (set of commands) is effective in extracting data may encourage other researchers to develop studies in the area. Social media monitoring allows the performance of quantitative and qualitative analysis, as well as the applicability of other methods that can help researchers in investigations into behaviors that were not previously identified in market data or prior knowledge.

Regarding the operationalization of the script for data extraction, improvements were noted that, if carried out, could automate the work even more, significantly reducing human intervention and increasing the security and reliability of the extraction process. Thus, as a suggestion for future studies and for improving the script presented in this article, we suggest the use of a centralized database, to store links and posts extracted from several machines, with simultaneous processing.

The interval time between accessing one link and another was arbitrarily defined, based on experiments carried out without methodology and cataloging. Once the cluster is built, it is possible to gradually reduce the waiting time between one link and another, until a minimum interval time between the links is defined, thus improving the data extraction process. It is also possible to virtualize computers in a cloud environment.

Finally, this study has some limitations in proposing strategic actions for the demands of the online environment. For example, tweets from only one HEI were analyzed, making it impossible to generalize the results.

REFERENCES

Abdulaziz, M., Alotaibi, A., Alsolamy, M., & Alabbas, A. (2021). Topic based sentiment analysis for COVID-19 tweets. *International Journal of Advanced Computer Science and Applications*, *12*(1), 626-636. https://doi.org/10.14569/IJACSA.2021.0120172

Arnaboldi, M., Busco, C., & Cuganesan, S. (2017). Accounting, accountability, social media and big data: revolution or hype? *Accounting, Auditing & Accountability Journal*, *30*(4), 762-776. https://doi.org/10.1108/AAAJ-03-2017-2880

Bimber, B., Flanagin, A., & Stohl, C. (2012). *Collective action in organizations*: *Interaction and engagement in an era of technological change.* Cambridge, UK: Cambridge University Press. https://doi.org/10.1080/10584609.2012.722513

Briscoe, F., & Gupta, A. (2016). Social activism in and around organizations. *Academy of Management Annals*, *10*(1), 671-727. https://doi.org/10.1080/19416520.2016.1153261

Cavalcanti, D. B., Bringel, E. P. B., Costa, F. R. J. da, Oliveira, T. M. de, & Zuccolotto, V. R. (2019). Digital activism and indignation nets in Brazil: The pressure groups. *Journal of Politics in Latin America*, *11*(1), 109–130. https://doi.org/10.1177/1866802X19840455

Choudhary, P. (2021). *TwitterAdvSearch*. Retrieved from https://github.com/prakharchoudhary/TwitterAdvSearch

Davis, G. F., & Thompson, T. A. (1994). A social movement perspective on corporate control. *Administrative Science Quarterly*, *39*(1), 141-173. https://doi.org/10.2307/2393497

D'Andréa, C. (2021). Para além dos dados coletados: Políticas das APIs nas plataformas de mídias digitais. *Matrizes*, *15*(1), 103-122. https://doi.org/10.11606/issn.1982-8160.v15i1p103-122

D'Haen, J., Van den Poel, D., Thorleuchter, D., & Benoit, D. (2016). Integrating expert knowledge and multilingual web crawling data in a lead qualification system. *Decision Support Systems*, *82*, 69-78. https://doi.org/10.1016/j.dss.2015.12.002

Earl, J., & Kimport, K. (2011). *Digitally enabled social change: Activism in the internet age.* Cambridge, MA: MIT Press.

Etter, M., & Albu, O. B. (2021). Activists in the dark: Social media algorithms and collective action in two social movement organizations. *Organization*, *28*(1), 68-91. https://doi.org/10.1177/1350508420961532

Fundação Instituto de Administração (2021). *Ativismo nas redes sociais: características, impactos e exemplos*. Retrieved from https://fia.com.br/blog/ativismo-nas-redes-sociais

George, J. J., & Leidner, D. E. (2019). From clicktivism to hacktivism: Understanding digital activism. *Information and Organization*, *29*(3), 100249. https://doi.org/10.1016/j.infoandorg.2019.04.001

Ghardallou, W. (2021). The impact of firms' and CEOs' social media usage on corporate performance. *Investment Management & Financial Innovations*, *18*(4), 21. https://doi.org/10.21511/imfi.18(4).2021.03

Gomez-Carrasco, P., & Michelon, G. (2017). The power of stakeholders' voice: The effects of social media activism on stock markets. *Business Strategy and the Environment*, *26*(6), 855–872. https://doi.org/10.1002/bse.1973

Harquail, C. (2011). Recreating reputation through authentic interaction: Using social media to connect with individual stakeholders. In R. Burke, G. Martin, & C. Cooper, (Eds.), Corporate reputation. Managing opportunities and threats (pp. 245-263). Burlington, NJ: Gower Publishing.

Harris, A. L., & Rea, A. (2009). Web 2.0 and virtual world technologies: A growing impact on IS education. *Journal of Information Systems Education*, *20*(2), 137. Retrieved from https://aisel.aisnet.org/jise/vol20/iss2/3

Instituto Brasileiro de Pesquisas e Análise de Dados (2021). *Twitter libera API especial para acadêmicos*: O que você precisa saber. Retrieved from https://www.ibpad.com.br/blog/twitter-libera-api-especial-para-academicos-o-que-voce-precisa-saber/

Jenzen, O., Erhart, I., Eslen-Ziya, H., Korkut, U., & McGarry, A. (2021). The symbol of social media in contemporary protest: Twitter and the Gezi Park movement. *Convergence*, *27*(2), 414-437. https://doi.org/10.1177/1354856520933747

Jung, M. J., Naughton, J. P., Tahoun, A., & Wang, C. (2018). Do firms strategically disseminate? Evidence from corporate use of social media. *The Accounting Review*, *93*(4), 225-252 https://doi.org/10.2308/accr-51906

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59-68. https://doi.org/10.1016/j.bushor.2009.09.003

Khalil, S., & Fakir, M. (2017). RCrawler: An R package for parallel web crawling and scraping. *SoftwareX*, *6*, 98-106. https://doi.org/10.1016/j.softx.2017.04.004

Kim, E. H., & Youm, Y. N. (2017). How do social media affect analyst stock recommendations? Evidence from S&P 500 electric power companies' Twitter accounts. *Strategic Management Journal*, *38*(13), 2599-2622. https://doi.org/10.1002/smj.2678

Lee, L. F., Hutton, A. P., & Shu, S. (2015). The role of social media in the capital market: Evidence from consumer product recalls. *Journal of Accounting Research, 53*(2), 367–404. https://doi.org/10.1111/1475-679X.12074

Lei N° 13.709 de 14 de agosto de 2018 (2018). Lei Geral de Proteção de Dados Pessoais (LGPD). *Diário Oficial da União de 15 de agosto de 2018*, e republicado parcialmente em 15 de agosto de 2018 - Edição extra. Retrieved from www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

Lewis, C. (2005). Unions and cyber-activism in South Africa. *Critical Perspectives on International Business*, 1(2/3), 194–208. https://doi.org/10.1108/17422040510595672

Luo, X. R., Zhang, J., & Marquis, C. (2016). Mobilization in the internet age: Internet activism and corporate response. *Academy of Management Journal*, 59(6), 2045-2068. https://doi.org/10.5465/amj.2015.0693

Meirelles, P. (2019). *Histórico das APIs no monitoramento e pesquisa em mídias sociais*. Brasília, DF: Instituto Brasileiro de Pesquisa e Análise de Dados.

Mercado, L. P. L., & da Silva, I. P. (2013). O Twitter na coleta de dados na pesquisa qualitativa. *Informática na educação: Teoria & Prática*, *16*(1). https://doi.org/10.22456/1982-1654.12378

Omari, A., Shoham, S., & Yahav, E. (2016). Cross-supervised synthesis of web-crawlers. In 38th *Proceedings of the International Conference on Software Engineering*, (pp. 368-379). New York, NY: ACM. https://doi.org/10.1145/2884781.2884842

Rao, H., Morrill, C., & Zald, M. N. (2000). Power plays: How social movements and collective action create new organizational forms. *Research in Organizational Behavior*, *22*, 237–281. https://doi.org/10.1016/S0191-3085(00)22007-8

Ross, S. D., & Cruz, B. D. P. A. (2021). Análise quantitativa de textos: Apresentação e operacionalização da técnica via Twitter. *Administração: Ensino e Pesquisa*, 22(1). https://doi.org/10.13058/raep.2021.v22n1.1859

Selenium (2021). *Selenium Automates Browsers*. Retrieved from https://www.selenium.dev/

Shin, S., & Ki, E.-J. (2022). Understanding environmental tweets of for-profits and nonprofits and their effects on user responses. *Management Decision*, *60*(1), 211-230. https://doi.org/10.1108/MD-05-2020-0679

Silva, T., & Stabile, M. (2016). *Análise de redes em Mídias Sociais*. In Monitoramento e pesquisa em mídias sociais: Metodologias, aplicações e inovações (pp. 237-260). São Paulo: Uva Limão.

Smits, M., & Mogos, S. (2013). *The impact of social media on business performance*. 21st Proceedings of the European Conference on Information Systems. p. 125. Retrieved from https://www.researchgate.net/publication/273138667_The_impact_of_social_media_on_business_performance

Statista (2021). *Forecast of the number of Twitter users in Brazil from 2017 to 2025*. Retrieved from www.statista.com/forecasts/1146589/twitter-users-in-brazil

Twitter (2022a). *Política de privacidade do Twitter*. Retrieved from https://twitter.com/pt/privacy

Twitter (2022b). *Robots*. Retrieved from https://twitter.com/robots.txt

Unicode® (2021). *Full emoji list, v14.0*. Unicode® Emoji Charts v14.0. Retrieved from https://unicode.org/emoji/charts/full-emoji-list.html

Van Dijck, J., & Poell, T. (2013). Understanding social media logic. *Media and Communication*, *1*(1), 2-14. Recuperado Retrieved from www.researchgate.net/publication/263566996_Understanding_Social_Media_Logic

Van Dyke, N., Soule, S. A., & Taylor, V. A. (2004). The targets of social movements: Beyond a focus on the state. *Research in Social Movements, Conflicts, and Change*, *25*(1), 27–51. Retrieved from www.emerald.com/insight/content/doi/10.1016/S0163-786X(04)25002-9/full/html

Vogler, D. (2020). Analyzing reputation of Swiss universities on Twitter: The role of stakeholders, content and sources. *Corporate Communications: An International Journal*, *25*(3), 429-445. Retrieved from https://www.emerald.com/insight/content/doi/10.1108/CCIJ-04-2019-0043/full/html

Wetzstein, B., Leitner, P., Rosenberg, F., Dustdar, S., & Leymann, F. (2011). Identifying influential factors of business process performance using dependency analysis. *Enterprise Information Systems, 5*(1), 79-98. https://doi.org/10.1080/17517575.2010.493956

W3schools. (2021). *On-line web tutorials*. Retrieved from https://www.w3schools.com/css/

Zhang, J., & Luo, X. R. (2013). Dared to care: Organizational vulnerability, institutional logics, and MNCs' social responsiveness in emerging markets. *Organization Science*, *24*(6), 1742-1764. https://doi.org/10.1287/orsc.1120.0813

Zhao, B. (2017). Web scraping. In L. Schintler, & C. McNeely (Eds.). *Encyclopedia of Big Data.* New York: Springer.

Authorship

Djeison Siedschlag*

Universidade do Vale do Itajaí, Programa de Pós-graduação em Administração
Rua João Coan, n. 400, Universitários, 88161-064, Biguaçu, SC, Brazil
E-mail: djeison.s@univali.br
https://orcid.org/0000-0001-6986-573X

Jeferson Lana

Universidade do Vale do Itajaí
R. Uruguai, n. 458, Centro 88302-901, Itajaí, SC, Brazil
E-mail: jlana@univali.br
https://orcid.org/0000-0002-9787-1114

Roberto Gonçalves Augusto Junior

Universidade do Vale do Itajaí
R. Uruguai, n. 458, Centro 88302-901, Itajaí, SC, Brazil
E-mail: roberto.augusto@edu.univali.br
https://orcid.org/0000-0002-8735-931X

Rosilene Marcon

Universidade do Vale do Itajaí, Programa de Pós-graduação em Administração
Rua João Coan, n. 400, Universitários, 88161-064, Biguaçu, SC, Brazil
E-mail: rmarcon@univali.br
https://orcid.org/0000-0002-0478-7715

* Corresponding Author

Copyrights

Plagiarism Check

RAC maintains the practice of submitting all documents approved for publication to the plagiarism check, using specific tools, e.g.: iThenticate.

Funding

The authors reported that there is no financial support for the research in this article.

Authors' Contributions

1st author: conceptualization (equal); project administration (equal); resources (lead); supervision (lead); validation (lead); visualization (equal); writing – original draft (lead).

2nd author: conceptualization (equal); supervision (equal); validation (equal); writing – review & editing (equal)

conceituação (igual); supervisão (igual); validação (igual); escrita - revisão e edição (igual).

3rd author: formal analysis (equal); methodology (equal); resources (supporting); software (equal); validation (equal); visualization (equal); writing – original draft (equal).

4th author: cconceptualization (equal); supervision (equal); validation (equal); writing – review & editing (equal).

Conflict of Interests

The authors have stated that there is no conflict of interest.

Peer Review Method

This content was evaluated using the double-blind peer review process. The disclosure of the reviewers' information on the first page, as well as the Peer Review Report, is made only after concluding the evaluation process, and with the voluntary consent of the respective reviewers and authors.

Data Availability

The authors claim that all data used in the research have been made publicly available through the Harvard Dataverse platform and can be accessed at:

Siedschlag, Djeison; Augusto Junior, Roberto Gonçalves; Lana, Jeferson; Marcon, Rosilene, 2022, "Replication Data for: "Like, share and react: Twitter capture for research and corporate decisions" published by RAC-Revista de Administração Contemporânea", Harvard Dataverse, V1. https://doi.org/10.7910/DVN/UZUJL3

RAC encourages data sharing but, in compliance with ethical principles, it does not demand the disclosure of any means of identifying research subjects, preserving the privacy of research subjects. The practice of open data is to enable the reproducibility of results, and to ensure the unrestricted transparency of the results of the published research, without requiring the identity of research subjects.