

Articulation between metrics and textual data as methodological experimentation for circulation studies

Articulación entre métricas y datos textuales como experimentación metodológica para los estudios en circulación

Articulação entre métricas e dados textuais como experimentação metodológica para os estudos em circulação

DOI: <https://doi.org/10.1590/1809-58442024114en>

Luan Moraes Romero
<https://orcid.org/0000-0003-4495-6672>

Viviane Borelli
<https://orcid.org/0000-0003-0643-2173>
Federal University of Santa Maria, Postgraduate Program in Communication. Santa Maria - RS, Brazil

Abstract

This work's objective was to experiment the possibilities of articulation between metrics and textual data in researches focused in circulation of meanings. This is an excerpt from a broader research on the circuits formed around the first three reports of the Vaza Jato series published by The Intercept Brasil, on June 9, 2020. To guide the discussion, we briefly mention some studies on circulation processes and we discuss Braga's (2017) theoretical approach on the notion of communicational episodes. To show the possibilities of integration between metrics and the inferential perspective, as referred by the author, we present the episode named "Shameless". It was through the observation of clues left by the digital tracks that we managed to make a connection between the quantitative data generated by software and the texts.

Keywords: Circulation; Metrics; Textual data; Communication episodes; Vaza-Jato.

Introduction

In March 2021, the Minister of the Federal Supreme Court (STF) Edson Fachin annulled all the convictions imposed on former president Luis Inácio Lula da Silva by the Federal Court of Paraná that were related to the investigations into Operation Lava Jato¹. The Intercept Brasil (TIB) points out in a newsletter, sent to its readers on March 8, 2021, that the judge recognizes something that prosecutor Deltan Dallagnol knew, and that the news agency revealed in 2019. Part of the text states: "Curitiba was not the competent court for the triplex process, as this court only judged complaints related to Petrobras. Just the fact that this process is being processed in Paraná constitutes an obvious procedural fraud."

This finding is part of one of the first three reports in the Vaza Jato series published by TIB, on June 9, 2020, which were the subject of investigation in a master's thesis that analyzed how the circulation of meanings occurred on digital platforms (VAN DIJCK, POELL & DE WAAL, 2018). This larger scope research identified and described the constitution of a communication circuit (BRAGA, 2017) around these reports, which integrate different platforms used by TIB, such as the website, fan page and YouTube channel. From the perspective developed by Braga (2017), each of these nodal points highlights communicational episodes that were analyzed in more detail in master's research (ROMERO, 2021).

Starting from this broader context, we chose as the focus for this article² the integration between methodological perspectives that consider metrics generated with the help of software, such as Iramuteq, with an inferential perspective on communication (BRAGA, 2017). Thus, the proposal is based on analytical tactics, which

1 Information at: <<https://g1.globo.com/politica/noticia/2021/03/08/entenda-a-decisao-de-fachin-que-anulou-as-condenacoes-de-lula-e-o-que-acontece-agora.ghml>> . Accessed on March, 21, 2021. On March, 23, 2020, the Federal Supreme Court also considered that the then judge Sergio Moro had been partial in the trials of former president Luiz Inacio Lula da Silva in Operation Lava Jato. In: <https://noticias.uol.com.br/ultimas-noticias/bbc/2021/03/23/com-voto-de-kassio-nunes-stf-forma-maioria-para-rejeitar-recurso-de-lula-contra-moro.htm> . Accessed on March, 30, 2021.

2 We consider that everyone who signs up on these different platforms is a participant in these communication circuits. As the focus of the article is not a conceptual discussion of terms used in the area of Communication such as "actor, subject, reader, producer, co-producer, commentator, interactant, prosumer, receiver, user", among others, we will call them social actors or participants, based on the theoretical proposition of Braga (2017) and Verón (2004, 2013) and also to have a more fluid reading.

involve revisiting the texts and considering the signals revealed by the metrics, with the aim of making inferences that allow the extraction of relevant knowledge to address the research question. In our study, we consider the term “texts” to refer to both what was produced by the reports published by The Intercept Brasil and what was constructed by different social actors in the comments spaces of the website, the Fanpage and the YouTube channel³.

To do this, we will start with a brief reflection on some perspectives related to studies in circulation in the field of Communication, which trigger different methodological approaches, so that we can then move forward in outlining a conception of integration of different forms of analysis. Through this methodological experimentation, we intend to contribute to the theoretical and methodological enrichment of current research.

In a second moment, to corroborate this perception, we present the procedurality of a communicational episode, called by us as “Cara de Pau”, which is materialized in the video published on the YouTube channel entitled “Sérgio Moro diz que não é um judge investigador”⁴ on June 9, 2019. Based on the election, description and analysis of this communicational episode, we will highlight the importance of integrating analytics based on metrics applied to textual data with the methodological tactic of returning to texts.

Circulation studies and methodological approach

When beginning the theoretical reflection, it is important to briefly mention some contributions on how the instance of circulation has emerged as an epistemological approach for Communication studies. The investigations that are based on the theoretical articulation related to circulation are guided by a diversity of theoretical approaches. According to Grohmann (2019), it would be possible to tentatively encompass perspectives that are affiliated with at least three theoretical horizons: those that dialogue with socio-discursive analyses, others that are based on reflection from a culturalist perspective in sync with cultural studies, and also an approach that seeks to align reflections on the circulation of capital with the circulation of meanings.

Among them, we are affiliated with semiodiscursive investigations that seek to develop a critical perspective on the appropriation of collected data, with a close eye on the traces left by different social actors on the most diverse digital platforms. We thus engage in dialogue with the propositions of GINDIN and BUSSO (2018) about semiodata and the need for criticism and an interpretative and critical approach. We share such concerns, that there should be no “enchantment of data” or mere description, since such signs must be analyzed, as well as it is necessary to critically reflect on such processes of production of meanings (Borelli, in press).

Thus, we conceive, like Fausto Neto (2018), that investigations in circulation through a semiodiscursive approach go through different moments and possible angles, when we consider the Latin American context of research in Communication. For the author, such questions emerge in the work of Eliseo Verón at different moments and address circulation as a space of discursive lag (VERÓN, 2004). As Fausto Neto (2018, p. 15) highlights, circulation goes beyond a “transition zone” and represents much more than an “intermediate link”.

Throughout his academic production, the Argentine semiologist updated his concepts in light of the diversity of communicational realities he observed. In his last published book, Verón (2013) states that it would no longer be possible to consider circulation only as a place of time lag, a place of passage between production and recognition, since he observed logics of structural couplings¹ that complexify these relationships.

We rely on Braga’s (2017) conception, which elaborates an intermediate theory about episodes, interaction devices and communication circuits. We consider that one of the contributions of this theoretical articulation is the possibility of using methodologies based on an evidentiary paradigm, proposed by Ginzburg (1989) and discussed by Sebeok and Umiker-Sebeok (2004)⁵. Both problematize that for the construction of knowledge it is necessary to follow traces, signs and revelations that often manifest themselves in details and that end up going unnoticed. In the case of our proposal, the metrics generated by Iramuteq required a closer look at the textual data, as it was necessary to tension what had been said by the different social actors in relation to the quantitative data. In this way, it was from the observation of clues and tracks that we were able to make inferences.

3 Such reflections were based on the investigations of the German sociologist Niklas Luhmann, especially in the work LUHMANN, Niklas. *Social Systems*. Stanford. Stanford University Press, [1984] 1995.

4 There is no scope for a deeper discussion of the authors’ works here, but in summary, we highlight that Sebeok and Umiker-Sebeok (2004) problematize the method adopted by Charles Peirce and Sherlock Holmes, which are based on looking at small facts and specificities. For them, only after collecting a series of clues will it be possible to make inferences. Ginzburg (1989, p. 177) understands that the evidentiary or semiotic paradigm ended up penetrating various areas of knowledge “deeply shaping the human sciences. Tiny paleographic particularities were used as clues that made it possible to reconstruct cultural exchanges and transformations.”

5 To highlight the texts, we use the double resource of bold and underlined letters for the selected words, and only bold to guide the synthesis of the relevance of the highlight.

The authors inspire Braga (2008, 2017) to reflect on Communication studies through an inferential perspective. For the author, it is inference that “allows the updating of the code, exercising it beyond a mechanical process of coding/decoding, which would do nothing more than transmit information” (BRAGA, 2017, p.31). Based on this perspective, Braga (2017) conceives that communication occurs through inferential processes between social actors who share common codes. Thus, for the author, the formation of interactional devices occurs from a strategic reiteration of attempts at mutual understanding in the social fabric. Communicational episodes would be the order of materiality and singularity of exchanges. Braga et al (2017) use classes as an example: if they were about a specific subject, they would be something of the dimension of the device, while the materiality of the encounter between the teacher and his students in a single class would be its episode. In this way, for him, the communication circuit would be constituted from the articulation between different devices that intertwine.

Following these understandings, Borelli & Kroth (2020) developed research in which they discuss the reconfigurations of the radio device in light of the emergence of circulation and the complexification of the process of mediatization of society. Frigo, Romero and Borelli (2019) articulated the notion of interactional device (BRAGA et al, 2017) with that of platform (VAN DIJCK, POELL & DE WAAL, 2018), highlighting the need for codes and inferences to constitute a dispositional arrangement of interaction (BRAGA, 2017).

It is important to note that Van Dijck, Poell & De Waal (2018) conceive that the platformization of society has undergone three processes: a) datafication, b) commodification and c) selection. For the authors, such processes affect the news sector, which becomes dependent on infrastructural technology companies, such as Facebook and Google. In addition, editorial changes occur, as many journalistic organizations and their editors begin to make important decisions and review distribution strategies based on data produced by digital platforms.

Therefore, it is important to understand how such processes work in the context of the development of digital platforms, since in scientific investigations that focus on digital traces it is necessary to problematize the contextualization of the data obtained (BREITER and HEPP, 2018), such as the process of capturing, processing and analyzing data.

However, given the impossibility of expanding on such questions about how platforms work, platformization and the different repercussions on society and modes of interaction and communication (we recognize their importance, but they are not central to this reflection), we will now specify how we constructed our methodological path based on the graphs and metrics generated by the free lexicometric analysis software Iramuteq.

It is important to consider that the application of quantitative methods can help identify patterns and other observable forms that are not possible using other methodologies. Thus, we understand that the use of software for data collection can be a powerful tool for problematizing such processes of production of meanings in circulation. However, the quantitative perspective is tensioned with a qualitative view, without the intention of placing these two approaches in opposition, but rather, reflecting on their specificities and transversalities.

We integrated statistical analysis, word clouds and maximum tree graphs from similarity analysis (DEGENE and VERGES, 1973) that takes into account the betweenness centrality metric (NEWMAN and GIRVAN, 2004; BRANDES, 2001) applied to textual data. Each of them contributed as a way of looking at our empirical data and, together, in an integrated way, they gave us support to return to the texts produced and observe the meanings in context, as well as the processuality of the communication episode.

The data extraction and mining process is complex and involves several stages. The statistical analysis of the words used in a specific text was performed both to perform an initial purification of the textual data, conducted manually, aiming to mitigate disparities between common expressions and proper nouns. This procedure is crucial to improve visualizations, as has been problematized in other reflections on the potential, limitations and possibilities in the use of software for data analysis and visualization in the area of Communication (WOBETO, ROMERO AND BORELLI, 2023).

Additionally, we developed metrics such as the percentage of occurrence of analyzed words in relation to the total occurrences, the percentage of frequency of a specific word in relation to the total occurrences, and the average occurrence per corpus. By creating word lists, it is possible to return to the original texts based on a word chosen by the researcher, as well as by selecting a group of words. This approach has proven to be significant for returning to the original texts based on the selection of words, providing an idea of the relevance of a given word or group in relation to the others used in a given context.

Furthermore, when a word cloud is created, it is based on a frequency notion of the uses of a given term in relation to the total set of words in a text. It is a visual graph that can help in certain cases to understand which themes are most frequently used in a text, or set of texts. In our research, the inferences made from it are made in conjunction with the maximum tree graphs.

Similarity analysis is the result of the incorporation of studies on graph theory into mathematics by linguistic studies (MARCHAND; RATINAUD, 2012). When performing this type of analysis, Iramuteq generates a graph of a network formed, with the words that make up the text as nodes, and assigns weights to the edges that connect them by the co-occurrence measure. Thus, according to the proximity and frequency of a word with others in certain texts, it organizes the edges. In other words, with greater frequency and greater proximity, the connection between them is greater.

There are several ways to observe and build this graph, however, we use algorithms that form the maximum tree of these relationships between words in a given text that take into account the betweenness centrality metric⁶ (NEWMAN and GIRVAN, 2004; BRANDES, 2001). Before delving deeper into the notions surrounding this metric, it is important to understand that a maximum tree is a form of extreme simplification, making the nodes that represent maximums emerge in a visually larger way, that is, that are articulated more frequently and in greater proximity with other nodes (DEGENE and VERGÈS, 1973).

Thus, the metric used is based on the intensity with which a given node (in our context, words) is located between two other nodes (NEWMAN and GIRVAN, 2004; BRANDES, 2001). Contextually speaking, in our investigation, the more a word is used in a given text together with other words, the more it assumes local importance, even if in terms of frequency in relation to the entire network, it has little intensity. Based on this metric, it is possible to highlight groups of words that have affinity in their recurring uses in a given text, organizing them into classes.

Thus, the maximum tree graphs were generated by Iramuteq and exported to be processed in the Gephi data visualization software. This processing was done to highlight the betweenness centrality metric, presenting the intensity in a gradually increasing manner - from blue to red - in the graphs that make up this investigation. In addition, the modularity of each of the trees was calculated to see if the formation of the word clusters is minimally reliable, as highlighted by Newman and Girvan (2004).

Below, we present the application of the proposed analysis and highlight some of the procedures carried out. Thus, it is necessary to say that we have highlighted only the morphological classes: adjectives, nouns, complementary nouns, verbs, supplementary verbs and unrecognized forms. We consider that our choice is based on our research objectives, which seek to understand how meanings circulate, just as Camargo and Justo (2013) corroborate the notion that such classes show words with full meanings. We believe that even if a given word is used in its polysemy, by moving back to the texts we will be able to evaluate this issue.

The maximum tree graphs were created using Iramuteq software and then transferred to Gephi, where we applied the “non-overlapping” visualization algorithms. We calculated modularity and highlighted the size of the words according to their frequency. The colors of the edges show the intensity of the betweenness centrality metric, gradually going from blue to red. The intensity of co-occurrence between words is shown according to the size of the edges, i.e., thicker wires show high intensity.

In addition, we present the relative frequency percentages of the terms in relation to the total number of occurrences in the corpus when presenting the word cloud data. These percentages were calculated manually based on the absolute frequency table exported from the software. When we present the trees, we refer to the betweenness centrality metric.

Regarding the presentation of word groups, we will expand with examples those that have a noun or adjective as the first word, in accordance with our methodological option of seeking those that have full meaning. We display the groups going from the one that presents the word with the greatest intermediation centrality to the smallest.

Also for the purpose of systematizing the data, some comments will be chosen as examples and labeled according to the selected episode: (“**example** of selected comments” with the acronym C 1 and so on). In addition, with the processing of the textual data, the punctuation marks were lost and the spelling presented in such comments was maintained. In this process of mining and processing the data, it was also necessary to manually change terms that contained characters not recognized by Iramuteq, such as the hyphen, and it was necessary to use terms in lower case, such as proper nouns (Brazil, Lava Jato, Lula, Moro, Glenn Greenwald, among others). In this way, the writing that makes up the corpus was maintained, since it is necessary to know how the software language operates.

Based on the construction of these graphs and the analyses generated, it is possible to make inferences within the scope of the research about the processes observed in the communication episode chosen here for analysis. To do so, it is necessary to return to the texts originally produced (by TIB and by the readers) so that we can understand the meanings evoked by the words in their specific contexts. Below we detail the process involving the data collected,

⁶ A first version of the article was presented to the Working Group on Reception, Circulation and Social Uses of Media at the XXX Annual Meeting of Compós, Pontifical Catholic University of São Paulo, São Paulo - SP, July 27 to 30, 2021. After discussion and review theoretical-methodological, the case to be detailed in this reflection has changed.

the statistical analysis, the generation of the word cloud and the similarity analysis for the set of comments in the communication episode.

The communication episode “Cara de pau” (Shameless)

The episode “Cara de pau” (Shameless) is related to the video posted on the TIB YouTube channel, entitled “Sérgio Moro says he is not an investigating judge”. The video is 58 seconds long and shows a public speech by former judge Sérgio Moro when he declares: “I have no investigation strategy whatsoever”. The speech is contrasted with the text that follows in the video description, in which the news agency states: “we showed that he suggested to Deltan Dallagnol that he change the order of the Lava Jato phases”. Inspired by Braga (2017), we consider that the notion of communication episode encompasses the combination of what was exposed by the TIB together with the comments found.

On June 6, 2020, when the videos were collected, all comments on the video post were collected, thus, our corpus of analysis consists of 1,072 texts⁷, from which 21,015 occurrences emerged (words, agglutinated forms or words) with 3,311 distinct words and 1,873 appearing only once. The average number of occurrences in this corpus is 19.60. With the methodological choice of excluding certain grammatical classes, such as pronouns and adverbs, we were left with 11,578 occurrences, representing 55.09% of the total. Of these, 9,911 (47.16%) are active and 1,667 (7.93%) supplementary. The maximum tree graph was generated with all words that appear more than ten times in the corpus, and had a modularity of 0.763 and presented 12 classes.

Of the possible inferences about the word cloud, we have at the center: “moro” (0,795%); “lula” (0,533%); “juiz” (judge) (0,504%); “cara” (face) (0,443%); “brasil” (0,443%); “TIB” (0,414%); “risos_k” (0,409%); “bandido” (bandit) (0,357%); “glenn_greenwald” (0,343%); “só” (0,328%). The centrality is occupied, again, by the mention of Moro and Lula, as well as laughter and the word “bandit” appearing with high frequency.

GRAPHIC 1: Wordcloud



SOURCE: Authors

In relation to the maximum tree graph generated, 12 classes or communities of words emerge that are close to each other. The classes are as follows: class 0 { “TIB” (3043), “glenn_greenwald”, (940) “the” (188) }; class 1 { “brasileiro” (brasilian) (929), “boca” (mouth) (188), “deixar” (to leave) (0) }; class 2 { “lula” (5141), “preso” (arrested) (561), “apoiar” (to support) (374) }; class 3 { “bandido” (bandit) (2006), “ministro” (minister) (375), “defender” (to defend) (375) }; class 4 { “to fall” (561), “casa” (house) (0), “conje” (spouse) (0) }; class 5 { “ver” (to see) (1824), “querer” (to want) (745), “gente” (people) (188) }; class 6 { “brasil” (8121), “lava_jato” (375), “esquerdo” (left) (374) }; class 7 { “juiz” (judge) (4875), “público” (public) (375), mão (hand) (374) }; class 8 { “cara” (face) (927), “falar” (to

⁷ With changes in the privacy policies of several platforms, in line with discussions emerging after the Cambridge Analytica case in 2015, and with the implementation of the European General Data Protection Regulation in 2016, programming application interfaces (in English, API) were changed, which makes it difficult to collect comments automatically, as well as not fully showing interactions on public pages.

In class 0 { “TIB” (3043), “glenn_greenwald”, (940) “the” (188) }, with 77 comments, there are congratulations with greater intensity to the work done by the news agency (“ao glenn_greenwald and all the team -TIB congratulations for the work follow firm” C 10), but also criticism appears (“-TIB is a joke -laughs- are against the punishment and fight against corruption in Brazil -laughs- you will sink” C 11).

In class 3 { “bandido” (bandit) (2006), “ministro” (minister) (375), “defender” (to defend) (375) }, with 69 comments, there are curses to Sergio Moro (“this country is a joke the garbage of the justice minister is a bandit of the most unscrupulous” C 12), as there is also a notion that even before the presented by the series, Lula continues to have guilt (“your god Lulinha was condemned in all legal instances _Moro was only the first until the STF ministers appointed by him himself condemned him to defend a bandit boy” C 13). As also arise comments with marks of legal argumentation (“in the criminal process of the current constitution the judge can not have any active role in the process defend one bandit against another but the first is still banished these citizens of good are ridiculous” C 14).

In class 1 { “brasileiro” (Brazilian), (929), “boca” (mouth) (188), “sair” (to leave) (0) }, with 45 comments, there are comments that mention the journalist Glenn Greenwald, one of the authors of the series, in a disrespectful way⁹ (“this greenwald has to suck a [...] big enough to occupy the beak with another [...] and not that worm of his husband nor that besteriol constituted family with his Brazilian gay male to simply do weakleaks in brazil” C 15), others that ask him to disclose the material received as is customary from the WikiLeaks site (“glenn_greenwald should do as WikiLeaks put all the audios in the media for evaluation and expertise including glenn_greenwald should be fair and clip the companions who stole everything from all Brazilian people since 2003 including their children be fair glenn_greenwald” C 16). In addition, there is a request for specialized legal information to better understand the situation (“someone can tell me what is the system of the Brazilian criminal procedure code and if I could explain this system in a few words based on some article obs I am not understanding anything of what is being said” C 17).

In class 8 { “cara” (face) (927), “falar” (to talk) (375), “peixe” (fish) (0) }, with 81 comments, there are curses to Sérgio Moro (“cara de pau” C 18), as well as the expression “cara” is used for interaction with other users (“is to talk about party investigation has already forgotten that the PF was in the party headquarters of its president and the PSL is being investigated too -laughs- “cara” you really believe in what you are writing but it’s just a fanatic who finds himself aroused” C 19).

In class 11 { “país” (country) (930), “colocar” (to put) (0), “mundo” (world) (0) }, with 55 comments, there emerge comments that speak about how the country seems to be (“well passive that release of audio to the JN of an acting president with a former president appointed to the ministry after the very determination of the end of the listening and released in record time in the serious country this would have shocking consequences” C 20) and (“in a serious country this Green, whatever, was already seeing the sunrise square that yes” C 21) or be (“so this country is ruined people who defend bandits total inversion of values the PT did its job well and destroyed the morale of a part of the population” C 22).

Congratulations do not appear prominently when we look at the metrics used in our research, but they appear in the corpus in 43 comments. There are congratulations to the work of TIB (“the mask of this impostor fell congratulations for the work glenn_greenwald and team the _tib” C 23) and congratulations to former judge Sergio Moro (“congratulations _Moro put in jail corrupt and powerful politicians and businessmen in a country that only arrested poor” C 24).

Thus, after the presentation of emerging classes, when resuming the discussion on the procedural of the interactional episode we can see as already presented in the previous episode, that there is a sharing of common codes between TIB and the commentators, as there is an inferential process of the interactional exchanges (BRAGA, 2017). It is possible to note that the posting of the video on the YouTube channel aims to present another evidence for the series of reports because it appears in the body of the third report published on the site.

Thus, taking the notion of communication flows, one of the orientation TIB-readers and another readers-readers, it is possible to note that in the first flow, there is a highlight to the confluence of positions concerning the criticism made to the ex-judge Sergio Moro, the only character present in the video, who is accused of treason and cursed in many of the comments.

Regarding the other axis, we note that there are signs of negotiation of meanings concerning what was exposed by the news agency (“if the **_Moro** were talking to Lula you would come full of whining and you get what to be livestock of the myth” C 25). As also appear to mark the establishment of an open conflict, when they oppose the notions of who are the bandits (“the Brazil gang of vagabond that defends Bolsonaro and _Moro they stole thousands of reais and you are licking the ass of these **bandits**” C 26) (“laughs, look who talks the guy who wants the **bandit** released and the judge arrested you is pathetic guy” C 27).

⁹ We choose to omit obscene expressions.

In addition, we note that the exchanges with legal argument marks were those that led to a search for negotiating the meanings in forward flows, as the comments that emerged in class 7. It is also possible to find marks that highlight the space of the comments being considered an ambiance. (“I went by and decided to give a little stop and leave a note the nine-finger **bandit** is in cana there in Curitiba” C 28). Thus, taking into account such aspects it is possible to realize that the logic of the platform can be considered important for understanding the procedurality of the interactional episode, as well as the political context.

We infer that this episode is marked by distinct conflicts, both in the dispute for meanings between what the reports reveal and what readers say in the comments. Such findings point to complex dynamics of circulation, in which participants produce multiple meanings through diffuse processes and in forward flows (BRAGA, 2017).

After detailing the theoretical and methodological path, in which we proposed the articulation between the metrics generated through the use of the Iramuteq software and the graphs constructed with the help of Gephi and the tension with the textual data, we proceed to list some notes as a closing of our proposal.

Final considerations

We understand that the inferential approach integrated with the use of graphics and metrics contributes to deepening the reflection around the proceduralities that affect the circulation of meanings. Thus, taking into account that the methodology to be used depends on several factors, such as the problem involved, the researcher’s journey and their theoretical affiliation, we consider that the directions adopted in our investigation can contribute to other research on the circulation of meanings. It is not uncommon to see research in Communication that limits itself to presenting quantitative data, without problematizing it. Our experimentation detailed here was precisely to make a theoretical and methodological effort so that the data would not remain loose, but could give us clues for a closer look.

Regarding the chain of our investigation, we consider that starting from a perspective based on the graphic and numerical use of the percentages of relative frequency of word clouds, for the maximum trees, with the betweenness centralities (NEWMAN and GIRVAN, 2004) and the modularity, may be another methodological approach to be undertaken in ongoing investigations. In our research, we considered that word clouds contribute to having a visual idea of what the main subjects or themes present in the comments would be. However, the integration with the visualization of maximum trees and their metrics proved to be important for us to be able to delve deeper into the meanings present in the communication episode. Here it is important to highlight the transition made from Iramuteq graphs to Gephi, as it is in the latter that you can increase or decrease the proportion of both trees and nodes, in addition to calculating metrics.

In addition to inferences about the visual form, the numerical exploration of the graphics can guide the analytical direction, which in our case, we preferred to expand the look at the groupings found through the word with greater intermediation centrality (NEWMAN and GIRVAN, 2004). However, it is possible, if it is of interest to other researchers, to take other routes, depending on the guiding research question and research interests. We chose to highlight what emerged in excess, that is, we sought to look at recurrences with greater intensity.

Finally, our research adds to others in the area that explore the complex universe of circulation and its interpenetrations (FAUSTO NETO, 2018) and continuous flows and beyond (BRAGA, 2017). Still, tentatively, we seek to look at the communicational episode instigated by investigations into the circulation of meanings through a critical approach (GROHMANN, 2019). We also intend to contribute to discussions initiated by Borelli and Dias (2018) with the aim of indicating possible methodological advances and limitations in current research.

Our methodological experimentation that seeks to integrate different analytical tactics requires care in preparing the argument, as there are signs that reinforce each other, as well as others that seem to indicate divergent understandings. In the communicational episode presented, the discussion emerges when we look at the centrality of the mention in the maximum tree with two antagonistic branches linked to Sérgio Moro, who is sometimes mixed with the notion of a criminal, traitor to the country, and sometimes with the image of an exemplary judge.

Regarding the challenges faced during the research, it is important to highlight that in the data mining process, some subtleties of the comments are lost, such as the use of emojis and other images. Another challenging point is related to the cleaning of textual data, which was carried out in an artisanal way. We know that it is possible to develop algorithms to automate this issue, so it is important to approach researchers from other areas of knowledge so that this can become viable through partnerships.

We seek to integrate quantitative notions with qualitative ones. In other words, detailing the context for textual data from an inferential perspective was the way found in this research to address this problem. This

approach proved to be promising, especially when looking at the mentions of the character in the video, showing both discontent and idolatry, revealing a wealth of details and nuances, which integrated with statistical analysis gives greater robustness to the research.

References

- BORELLI, Viviane V. (2023no prelo). A problematização teórica-metodológica nas pesquisas de circulação e plataforma. In: J. Ferreira, Jairo,; A. M. C. M. Silveira, V. Borelli ; Dalmolin, Aline ; Rosa, Ana Paula. Löfgren, Isabel (Orgs.) . **IA, algoritmos e plataformas: Questões e hipóteses na perspectiva da midiatização**. 1. ed. SANTA MARIA: FACOS-UFSM. v. 1.
- BORELLI, V; DIAS, M. S. M. Circulação discursiva: desafios metodológicos para compreender as interações entre jornais e leitores. **Rizoma**, [S.L.], v. 6, n. 2, p. 95-109, 7 jul. 2018. APESC - Associação Pro-Ensino em Santa Cruz do Sul. <http://dx.doi.org/10.17058/rzm.v6i2.12602>.
- BORELLI, VIVIANE; KROTH, M. E. . Circulação e construção de circuitos: a dinâmica do rádio. **E-COMPÓS (BRASÍLIA)**, v. 23, p. 1-25, 2020.
- BRAGA, José. Luiz. Dispositivos Interacionais; Circuitos de Comunicação. In: BRAGA, José Luiz; CALAZANS, Regina; RABELO, Leon et al. **Matrizes interacionais: a comunicação constrói a sociedade**. Campina Grande: EDUEPB, 2017. P.71-64.
- _____. Comunicação, disciplina indiciária. **Matrizes**, v. 1, n. 2, p. 73-88, abr. 2008. Disponível em: <www.matrizes.usp.br/ojs/index.php/matrizes/article/download/46/28>. Acesso em: 6 mar. 2016.
- BRANDES, U. A faster algorithm for betweenness centrality. **The Journal Of Mathematical Sociology**, [S.L.], v. 25, n. 2, p. 163-177, jun. 2001. DOI: <https://doi.org/10.1080/0022250X.2001.9990249>
- BREITER, A.; HEPP, A. The Complexity of Datafication: putting digital traces in context. In: HEPP, A.; BREITER, A.; HASEBRINK, U. (org.). **Communicative Figurations**. Londres: Palgrave, 2018. p. 387-405.
- CAMARGO, B. V.; JUSTO, A. M.. IRAMUTEQ: um software gratuito para análise de dados textuais. **Temas em Psicologia**, Ribeirão Preto, v. 21, n. 2, p. 513-518, dez. 2013. Disponível em <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-389X2013000200016&lng=pt&nrm=iso>. Acesso em: 11 nov 2020.
- DEGENNE, A. e VERGÈS, P. Introduction à l'analyse de similitude. **Revue Française de Sociologie**, v. 14, n. 4, p. 471, Out 1973. Disponível em: <<https://www.jstor.org/stable/3320247?origin=crossref>>. Acesso em: 24 maio 2020.
- FAUSTO NETO, A. Circulação: trajetos conceituais. **Rizoma**, [s.l.], v. 6, n. 2, p.08-40, 7 jul. 2018. APESC - Associação Pró-Ensino em Santa Cruz do Sul. <http://dx.doi.org/10.17058/rzm.v6i2.13004>.
- FRIGO, D.; ROMERO, L. M.; BORELLI, V. Plataformas, dispositivos interacionais e circulação: mapeamento do episódio “Vaza Jato”. **Anais de VIII Colóquio Semiótica das Mídias**, p. 17, 2019.
- GINDIN, Irene Lis; BUSO, Mariana Patricia. Investigaciones en comunicación en tiempos de big data: sobre metodologías y temporalidades en el abordage de redes sociais. In: **adComunica**. Revista de Estrategias, Tendencias e Innovación en Comunicación, n. 15. Castellón: Asociación para el Desarrollo de la Comunicación adComunica y Universitat Jaume I, 2018, p. 25 - 43.
- GROHMANN, Rafael. Os rastros digitais na circulação de sentidos: pela desnaturalização e contextualização de dados na pesquisa em comunicação. **Galáxia** (São Paulo), São Paulo , n. 42, p. 150-163, Dec. 2019. Disponível em <<http://dx.doi.org/10.1590/1982-25532019340183>>. Acesso em 20 Jan. 2021.
- GINZBURG, Carlo. **Mitos, emblemas, sinais**. São Paulo: Companhia das Letras, 1989.
- MARCHAND, P.; P. RATINAUD. (2012). L'analyse de similitude appliqueé aux corpus textuelles: les primaires socialistes pour l'élection présidentielle française. Em: Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles. JADT 2012. (687–699). Apresentado em 11eme Journées internationales d'Analyse statistique des Données Textuelles. **JADT 2012.**, Liège, Belgique
- NEWMAN, M. E. J.; GIRVAN, M.. Finding and evaluating community structure in networks. **Physical Review e**, [S.L.], v. 69, n. 2, p. 69-74, 26 fev. 2004. Mensal. American Physical Society (APS). DOI: <http://dx.doi.org/10.1103/physreve.69.026113>.
- ROMERO, L. M. Série **Vaza Jato: Análise do circuito comunicacional formado em torno das três primeiras reportagens**. 2021. 102 f. Dissertação (Mestrado) - Programa de Pós-graduação em Comunicação, Universidade Federal de Santa Maria, Santa Maria, 2021.

WOBETO, Samara Letícia; ROMERO; Luan Moraes e BORELLI, Viviane. A utilização de softwares para análise e visualização de dados nas pesquisas em Comunicação. **Anais da ABCIBER XVI - SIMPÓSIO NACIONAL DA ABCIBER 2023**. ISSN : 2175-2389. In: <https://abciber.org.br/simposios/index.php/abciber/abciber16/paper/view/2071>. Consulta em 3 abr 24.

VAN DIJCK, J; POELL, T.; DE WAAL, M. **The Platform Society. Public Values in a Connective World**. Nova York: Oxford University Press 2018.

VERÓN, E. **Fragmentos de um tecido**. São Leopoldo: Unisinos, 2004.

____. **La semiosis social**, 2. Ideas, momentos, interpretantes. Buenos Aires, Paidós Planeta, 2013.

About the authors

Luan Moares Romero

luan_155@hotmail.com

Journalist from the Federal University of Santa Maria (UFSM) Master and PhD student in the Postgraduate Program in Communication.

Viviane Borelli

viviane.borelli@ufsm.br

Professor of the Postgraduate Program in Communication and the Department of Communication at the Federal University of Santa Maria (UFSM), Santa Maria, RS, Brazil. Leader of the research group “Media Circulation and Communication Strategies” (Cimid). PhD in Communication Sciences from Unisinos.

Received on: 11/26/2023

Accepted on: 24/08/2024

Responsible editors:

Marialva Barbosa and Sonia Virgínia Moreira



This article is published in Open Access under the **Creative Commons Attribution 4.0 International License (CC-BY)**. The authors retain all copyrights, transferring to Intercom: Revista Brasileira de Ciências da Comunicação the right to carry out the original publication and keep it updated