Mixing SNA and Semiotics: a Methodological Framework for Analyzing Online Collective Subjectivities

Mr. LUCAS CALIL Guimarães Silva, MSc in Linguistics, Universidade Federal Fluminense
Ms. TATIANA Terra RUEDIGER, MSc in Sociology, Universidade do Estado do Rio de Janeiro

Fundação Getulio Vargas, Department of Public Policy Analysis www.dapp.fgv.br

Methodology paper presented at the 2018 IPSA World Congress of Political Science, which took place at Brisbane, QLD, Australia, from July 21-25, 2018

ABSTRACT
This paper proposes a combination of linguistics and sociology in order to develop an empirical methodological approach for analyzing different interest groups on social media. From sociology, we articulate the concept of collective subjectivities (CS) for the virtual environment and propose to call its virtual layer online collective subjectivities (OCS). The causality of a CS is different of the weberian active causality and of the durkheimian conditional causality. The collective causality has to be understood as a property of the social systems, which refers to their interactions. For the OCS analysis, we consider a specific social media debate as a proxy of a social system. Thus, OCSs and their members interact on a social network in such a way that they have a causal effect over the world that may be intentional or not. The line of argumentation developed here is based on empirical finding that clustering algorithms divide a retweet network in groups that have distinct worldviews – in this specific case, political views.

This sociological approach is integrated with a linguistic structure of search rules, which defines the selection of themes for analysis and, under a French Semiotics-Greimasian theoretical standpoint, organizes the recovery of the public debate, from social media publications, in accordance with the selected topics of study. We understand the gathering of semantic data not as a combination of keywords and hashtags, but contemplating, at the center of rule-construction, the different topics and thematic associations inherent to the subject, from abstract (as public health and corruption) to concrete (an authority, a news event, a brand) concepts. To do a comprehensive search
rule, therefore, it is necessary to do a quantitative and qualitative definition of the semantic corpus to collect the data from social media – and identify the OCSs, that can vary, for example, by the opinion concerning a single topic or by the debate about two different topics related to the same theme.

As an illustrative case study for this paper, we propose to analyze the Twitter debate regarding the first month of federal intervention on public security at the state of Rio de Janeiro, by the Armed Forces, as sanctioned by president Michel Temer on February 16th, 2018.

Keywords: semiotics; collective subjectivities; social network analysis; public policy analysis

1. Introduction

Facing the contemporary relevance of social media as a public platform for heterogeneous debate among regular citizens, influencers, government and private institutions, at the first years of 21st century it has emerged, and not only in developed nations, a still-growing academic and commercial market dedicated to social media analysis. Platforms like Twitter, Facebook and Instagram have been understood as new “public spheres”, in accordance with the concept initially forwarded by Jürgen Habermas, into which information and opinions can easily be shared, accepted or rejected, even without the intervention of large, unified media groups and networks. The traditional conglomerates from the press (in special newspapers), equally facing the wide open public sphere of each social media channel, still struggle to adapt themselves to the competition of celebrities, politicians and other relevant players, using those platforms to stay relevant and bolster dissemination of content while sustaining profits. Until now, it has been not so simple a job to accomplish.

As different social media companies have developed into international, multicultural virtual spaces of discussion about any topic of relevance for its users, governments, think tanks, academic institutions and the business sector got interested on monitoring the public debate to evaluate sentiment and recognize matters of prominent importance. To obtain policy feedback and better understand groups’ opinions, political positions and tendencies. And, from marketing to polling, social media, as one of the elements of Big Data Universe, have become more than a source of information, but a research object as well, allowing the advancement of innumerable methodological approaches to study it.
2. Opposing perspectives for social media analysis

Most methodological approaches (and the most popular and widely applied) on social media have opposite perspectives, and with different purposes. Studies and theories from the social sciences (c.f. CASTELLS, 2009; CASTELLS, 2012; MARGETTS et. al, 2015; MURTHY, 2012; STROMER-GALLEY, 2002; CONOVER et al., 2011; GERBAUDO, 2012; GROSHEK and RAWI, 2013; BARBERÁ et al., 2015; BOCZKOWSKI and PAPACHARISSI, 2018) that use social media interactions and analyze them through Social Network Analysis — from now onwards abbreviated as SNA — devote attention over the cultural, economical and political impact of social media in relation with external phenomena and for human societies. Public policies, political engagement, government accountability, consumption of news and overall information etc. Even when studies, instead of investigating social networks by a qualitative-only focus, delve into large databases of publications collected from social media platforms, it is seldom presented a methodological framework about the process of defining the textual corpus, under a semantic or linguistic theoretical context. Regularly, either the data gathering is restricted to hashtags from specific events or the search rule (stringline) contains only a combination of isolated keywords and expressions comprising two or three words in sequence. Recent examples of papers which abided by such procedure of filtering investigate very diverse subjects: the spreading of fake news and disinformation (HOWARD et al., 2017); the influence of bots on political propaganda (HOWARD & KOLLANYI, 2016); communication between victims of a natural disaster (Takahashi et al., 2015); use of marijuana and alcohol (KRAUSS et al., 2016).

In a more complex scope of filtering, compared to the studies cited above, Majo-Vázquez et al. (2017) aggregated, on its database selection, hundreds of names of candidates, political parties, hashtags, news media outlets and proprietary accounts to analyze the impact of online and legacy news media on 2017 German Federal Election. On detailing the collection of data (2017, p. 2), the authors explain every criteria followed to gather the tweets pertaining to the problem proposed by the paper, whose final sample was filtered on two levels (p. 10). However, without a consideration about the semantic aspects of the subject selected for analysis (debate about the German elections on Twitter) or about the long-established linguistic reflections on meaning, textuality or the social component of natural languages – or the relation between words, syntax and morphology and the cultural representation of concepts by verbal systems.

Apart the social sciences perspective, less attached to the textual aspects of data filtering and with the primordial intention of apprehending the results derived from the datasets, the other
prevalent approach occupies itself with the improvement of ontologies, lexicons and easy-to-use tools to handily categorize and structure large amounts of unprocessed (raw) data. Scholars from fields like statistics, mathematics, computational linguistics and computer sciences contemplate building efficient algorithms and automatic filters, enhanced by machine learning, neural networks and other technical resources to process big data adequately. By this apprehension, perfecting automatic structures to process language intends to simulate the properties of human communication — patterns, concepts and phrasal structures, identifying the underlying organization of every linguistic system; even images, music, gestures.

In opposition to the social sciences perspective, the pattern-recognition approach lays focus over the data collection stage of the analysis. It proposes — functioning almost as a necessary counterpart to the other approach — to ease human work over sets of information impossible to study by analogic eyes, such as reading sentence after sentence of millions of public documents or tweets. “How to make feasible the analysis of so much data in an organized, comprehensible and relevant way” is the central problem conducting the quantitative-based side of data mining and SNA. Therefore, the pattern-recognition approach builds models of Natural Language Processing (NLP) and breaks down verbal speech and textual structures to find out equivalences and contrasts, associations and variations, most of all interested in finding statistical recurrences and predictable sets of rules to comprise multiple language possibilities.

Many fields of modern linguistics — cognitive linguistics, psycholinguistics, language acquisition and formal semantics, to name a few — dedicate heavy attention to that question, and hundreds of freely available, open-source coded bases of data mining have been available to researchers and companies, in dozens of languages, and an even greater number of paid and free softwares for SNA which apply internal and algorithmic processing of data to provide information to consumers and researchers. Every social media platform as well — Twitter, for instance, provides access to its data by a public API, with limited available features, or by a subscription service (Gnip) with greater resources to gather public posts, interactions and other sorts of information (demographics, gender, impressions, geolocation etc.).

In general, studies apply NLP techniques to previously-selected databases or after a initial filtering of topics or words — once the data (corpus) is selected, the mining process structures the raw, unclassified information, based on the parameters by which it operates. Another rich field for NLP and machine learning projects has been sentiment analysis, with the systematization of automatic operations to classify emotions over a certain topic; recently (POZZI et al. 2017; LIU, 2012), the extraction of textual subjectivities and opinions has got less universal — language-based
lexicons and ontologies made to be applied onto every database —, instead concentrating on comprehend specific aspects of textuality that differ from one topic/subject to another: context, irony and ambiguity, for example (c.f. BARBIERI & SAGGION, 2014). And, while automated structures have been essential to analyze big data and represent an important advancement to improve the ability, by business and academic professionals, to extract valuable insights from substantial volumes of information, the pattern-recognition approach remains, overall, a perspective which considers text as an end in itself, with the identification of recurrences and associations obeying syntactic properties of languages, probabilities and word and phrase-based categories. Akin to the prevalent social sciences approach, it doesn’t consider, specifically on the previous stage of research, before selecting the corpus, the social construction of meaning on verbal systems, and how textual representation articulates the organization of contextual and semantic values, concrete and abstract notions. Aware of these challenges, this paper proposes a different perspective.

2.1. A new methodological framework

To implement an unified structure of social media analysis, comprising both a linguistic outlook — on the introductory stage of definition of a textual database — and a sociological procedure of analysis, resorting to SNA techniques and methods currently available, we articulate the concept of collective subjectivities (CS) (DOMINGUES, 1995; 1999; 2004) for the virtual environment, into which social media influx of information emerges, and propose to call the virtual layer an online collective subjectivity (OCS). As such, we therefore understand a specific social media debate as a proxy of a social system, with its members interacting on a social network in a causal effect over the world. And, to establish that, we go back to the empirical reflection over how clustering algorithms — on a SNA software overlay of the interactions and textual contents filtered by a linguistic subset of search rules — divide a retweet network in groups with distinct worldviews, be it political, social, cultural or economical.

Initially, we present the theoretical standpoint of the linguistic approach over how the object of analysis and the corpus of research are selected, under a French Semiotics-Greimiasian methodology of comprehension of the universal structure inherent to textual narratives and the emergency of meaning. As a theory developed since the 1960s, predominantly on France, Italy, Belgium and Brazil, by scholars of the movement baptized as the School of Paris, Greimiasian Semiotics differs from other Semiotic appreciations about the relations between social constructions
and semantic representation (e.g. the Semiotic model forwarded by Charles Sanders Pierce’s theory of signs), proposing a generative process of analysis in layers of observation of abstract concepts and semantic traces identified on textual units.

We consider the collection of social media data not as a combination of keywords and hashtags, without a fundamental step-back to recognize the thematic and semantic elements the study intends to analyze, but the elaboration of search rules (stringlines) as a posterior process of selecting linguistic units that comprise discursive figures understood as semantic traces manifested on texts by a collectivity — in this case, on social media. After the corpus selection, it is possible to identify the OCSs regulated by influencers, popular tweets and opinions regarding one or several subjects of public discussion online.

Finally, to illustrate how the proposed methodology can be applied on social media, we analyze the Twitter debate, on Brazilian Portuguese, about the first month of federal intervention on public security at the Brazilian state of Rio de Janeiro, by the Armed Forces, following a decree signed by president Michel Temer on February 16th, 2018.

3. On French Semiotics

Social media analyses invariably require: a) access to database of publicly available information, as an API connection or a programmed, independent system to crawl (gather) engagements, posts and metadata related to the object of study; b) a processing software to convert the raw database, be it on a JSON-formatted or downloadable at .csv files, on a categorized and visually comprehensible structure; c) a platform to apply boolean (logical) operators to the stream of data (real-time streaming of historical) and collect the information.

To a quantitative-only proposal of analysis, any of the conditions above can be ignored — it is not necessary to have a system of automatic data collection if the object consists of a few Facebook public pages or the tweets made by some authors for a short period of time. But, for analysis requiring large sets of data, in some cases dozens of millions of posts and interactions, the conditions above must be observed. Daas and Puts (2014, p. 5), looking for correspondence between public confidence on the economy and social media sentiment, acquired more than 3 billion posts from a private company database, without a previous rule filtering — 80.1 percent Twitter publications, 10.6% Facebook posts and the remaining from news websites, blogs, Linkedin and other social media applications —, and applied a sentence level-based sentiment algorithm to measure positive or negative opinions. Further on, evaluated the entire corpus over
how metrics surveyed by Statistics Netherlands about confidence on Dutch economy can be correlated to the chronological variation on sentiment from social media publications, splitting the database by aggregation of words (p. 15) related to economic topics (consumer, confidence, spending etc.) and the most written words on Dutch language, as pronouns and articles — which have a syntactical, not semantic, predominant function on verbal textuality.

The findings obtained by Daas and Puts suggest an identifiable association between overall perception by citizens and changes on economic sentiment on social media, with the intention of develop a predictive model, cheaper and faster than regular polling and surveys used to identify public perceptions. However, at the stage of gathering a textual corpus, either a sampled one (normally when the only stream of data available is a public API) or an integral one, as exemplified by the study about Dutch consumer confidence, some questions emerge. First and foremost, to the purposes of this paper, the understanding of the specific subjects analyzed and measured, and how to obtain an accurate database of posts related to them. By choosing not to provide a previous stringline to filter only mentions about economic topics, Daas and Puts present different selections based on isolated words identified on the entire corpus and tested many options of keyword-filtering (p. 12) to observe which ones better correlate to the research from Statistics Netherlands. Combined “words related to consumer confidence” with pronouns expressing “personal or group experiences”, such as “I, you, me, we”, without detailing how, on a sentence level, the pronouns articulate syntactically with the words equated to consumer confidence or if the morphological variations of the keywords chosen were considered. “Spending”, for example, can operate as part of a nominal construction (e.g. public spending, deficit spending, money spending) or on a verbal capacity, expressing an action (e.g. “I’ve been spending a lot of money”; “I was spending a lot of time on this”). Moreover, not necessarily the presence of a pronoun and a substantive on the same text determines a relation between subject and object connecting both. A Twitter post is limited to 280 (previously 140) characters, and normally only a few phrasal structures can be combined to produce a single text, making easier to recognize subject/object associations, but on a Facebook publication a substantive can appear on a paragraph, while a pronoun can be identified many paragraphs later, without any semantic or syntactic liaison between them.

At another sphere, the study lacks a detailed information about what, thematically, represents an economic subtopic. Several words and actions could be understood as pertaining to the conceptual idea of a textual representation about a country’s economy: unemployment (and the act of finding a job); exchange and interest rates; inflation; taxes; public investment and budget; industrial and agricultural sectors and products; exportation and importation; mentions to federal institutions and authorities, as a Central Bank president or Treasury secretary. All the terms above
could be relevant to the debate on consumer confidence and how it relates to confidence on the economy. And confidence, as an expression of a personal intention/opinion, is frequently replaced at regular speech by the use of the verb *to trust/, not present on the keyword combinations specified at the research (in Dutch or in English).

After the consideration of the thematic framework by which the economic debate is referred at a manifested level, a second stage of semantic reflection about the corpus is therefore relevant: what subparts of the debate on economy express perceptions of confidence? Spending money and job-related issues, as described by the paper, but what defines a negative or positive opinion from a personal perspective concerning consumer confidence? And how, based on the single identification of substantives — “consumer”, “confidence”, “job”, “economy”, “finance” etc. —, it is possible to obtain both a semantic identity with an economic theme and a expression of confidence about the economic theme itself? A person may have confidence on innumerable affairs, from a loving partner to a football team, and the public display of confidence, on a post, can replace a broad concept, like “economy”, for an economic decision of everyday practices, as withdraw money from a savings account or not to travel on the next holiday. The statistic correlation between traditional polling and social media sentiment is empirically demonstrated by Daas and Puts, but such questions present a legitimate problem about the causal inference extracted from the corpus analyzed.

### 3.1. Theoretical standpoints and practical challenges

Research on social media covering complex, wide-ranging subjects as economy, security, public health, education and politics demands an equally complex identification of the object of study in relation to how this object is textually filtered from the public debate online. When the object is solely a search for a brand, a name or a hashtag, the data filtering follows a strict limitation on the covered subjects. For KRAUSS et al. (2016), it is easier to collect posts only about marijuana and alcohol than to identify mentions about any kind of drug use, for example. Yet an equivalent approach for something as broad as economy does not operate in a language level at the same way, as much for the difficulty to embrace the multiple words, expressions and topics internal to the abstract understanding of “what is economy” as for making sure the database analyzed has a valid correspondence — regarding accuracy and volume — with the hypothetical amount of posts, commentaries and engagements published about the theme on the internet.

This question, in regards to the academic and commercial interest on the sizable amount of information available to analysis on social media, is particularly concerning when a company...
of researcher intends to develop a large structure of real-time monitoring about many different topics and themes simultaneously. Such demand requires regular updates and corrections to the stringlines collecting data and a system which integrates search rules in layers of detail — broad themes divided into categories internal to the overall themes, stringlines that can be combined to evaluate specific debates and subjects. Trees of semantic categorization, with multiple levels articulated, by a linguistic reasoning, to generate distinct (and accurate) databases to study.

In contemplation of these pragmatic and theoretical issues, the semiotic methodology proposed here (HJELMSLEV, 1963; GREIMAS, 2011, 2012; FONTANILLE, 2008; GREIMAS & COURTÉS, 2011) acknowledges the concept of meaning as a result from semantic articulations between linguistic elements (signs, in a minimal apprehension), as organized internally within any object identified as a text. This follows the pioneering reflection launched by Ferdinand de Saussure (2011) about the central object of Linguistics as an independent field of academic study. Therefore, the elaboration of search rules, instead of identifying words and syntactic constructions, is designed to identify, with precision, the varied textual forms of expression about a selected subject, with a specific manifestation of meanings — either focused on a restricted topic, like a person or a brand, or about a wide-ranging, abstract theme.

The Greimasian-French Semiotics, incorporating many of Saussure’s initial linguistic principles, is based on the premise that the semantic values inherent to any text have to be studied by the structural analysis of the semantic relations restricted by the limits of the textual corpus, and not as if the meaning (the multiple, possible meanings materialized from a text) has an external dependence upon a context to be analyzed. Under the Greimasian approach, a text is any selected object of study with an internal coherence of linguistic articulations, such as a poem, a painting, an excerpt of a novel or a tweet. This assumption does not exclude cultural, social and historical contexts, but determines that what is external to the selected corpus must be semantically apprehended from inside — as the text recovers, at the level of enunciation¹, what is memory, history and culture (BARROS, 2005).

Traditionally, Greimasian methodology has been used in analyses with restricted corpora of objects. However, instead of representing a methodological problem to big data studies, this

¹ Emile Benveniste, in Problèmes de Linguistique Générale, II, first describes enunciation as the process by which the language is put into operation with an individual act of use (1974, p. 86). In other words, enunciation is, semiotically, a non manifested, but inherent, stage of every linguistic process. The individual — not necessarily a “concrete” person — apprehends the linguistic system and organizes a semantic object, while regulated by choices, restrictions and rules. This happens, primarily, with projections of the individual at the manifested level, over categories of space, person and time — such as the use of a first person approach to write an opinion column or the verbal choices to describe a scene from the narrative past, for example.
opens up different possibilities to the field of applied linguistics. By the semiotic theoretical position, as already explained, social media posts and interactions can be selected as textual objects in no different perspective than a novel or a poem, and the increasing relevance of the public debate on web platforms – digital public spheres – has stimulated and compelled the demand for practical efforts to consolidate a method of semiotic analysis on a large scale, with a comprehensive corpus of, sometimes, millions of texts.

The internet as a virtual support for linguistic objects, nonetheless, has raised conceptual problems, by changing the physical and material relation inherent to every textual object. This process, defined by the digitization of photographs, books, songs and paintings, represents an aesthetic change by the perspective of the person who interacts with those objects, replacing the concrete space of the world and altering the sensible engagement with them (c.f. ZILBERBERG, 2006; FECHINE, 2018). This, by no means, destitutes digital objects of being texts – especially verbal texts, which keep, online, the same rules, patterns and systemic regularity of the offline counterparts, with distinctions derived from the new possibilities of expression made possible by the internet (emojis, memes, hashtags etc.). The communication process, however, retains the same system: the natural languages.

As first described by Hjelmslev (1963), the semiosis – semiotic manifestation of meaning – is, essentially, a product of the junction of two plans (expression and content) derived directly from the elementary definition of sign, by Saussure, as the junction of a signifier (expression) and a signified (content) on natural languages. A verbal text articulates a combination of signs (minimal units able to produce meaning for every language), and the meaning emerges from the association of each sign to the other signs from the same textual object — any word is either a sign or is comprised of signs manifested in linearity, and, although every system of a natural language has a limited set of combinatory signs, a new one can be integrated to a language. When a neologism is incorporated, for example.

When the object of analysis is a collection of online posts, the semiosis — concept that expands Saussure’s initial focus only on verbal systems to any linguistic manifestation by human societies, be it verbal or non-verbal — is dependent on the properties of social media. The limit of characters, the ability to retweet and have an instantaneous interaction with other profiles, the articulation of photos and text onto a single post. To better understand how public debate happens online, it is important to investigate the process of construction of verbal objects, with one of the most prominent developments given to this question, in Semiotics, coming from Jacques Fontanille (2015) and Fontanille and Zilberberg (2001), whose reflections upon the concept of “forms of life”, extracted from Wittgenstein, aim to expand the semiotic notion of “object”. By Fontanille’s
argument (2015, p. 14), the semiotic object must be understood by a social perspective, and the regular, stable and accepted practices of a society interfere in the construction of meaning — and, as such, certain strategies, already consolidated by repetition, recognition and efficacy, have a central positioning inside the social discussions of a group (canonical works of art, for example), while others, innovative and disruptive, have a smaller audience and can be surprising at first.

This understanding of a Semiotics of Cultures, into which the dominant forms of life (being natural languages the most important of all) regulate the public debate, is at the heart of a wider consideration of the semiotic corpus, once the monitoring of verbal discussions on social media as Twitter or Facebook emerges from social collectivities who share a natural language and use regular expressions — with the intent of being comprehended by other members of society. It is the regularity of verbal manifestations that allows the elaboration of stringlines to predict and collect the overall debate about any given topic on social media. As natural languages have limited graphic forms of expression, and words and semantic constructions have a collective nature, a specific rule of search can recover a very broad sample of the public debate about any semantic subject.

3.2. Building a stringline

Calil and Ruediger (2017, p. 7), describing the semiotic approach, explain that “the referential for coherent identity is the presence, in each post that integrates a corpus, of discursive figures belonging to the semantic field that one wishes to study, which is determined a priori.” A hypothesis about the semantic fields of choice (not merely words nor expressions) must be established first, be those restricted or broad, so adequate texts can be recovered from social media. Considering, for example, the federal intervention on the state of Rio de Janeiro — the case study selected in this paper —, the previous stage of defining how it may be verbally represented, on Brazilian Portuguese, presents a laborious task. To formulate a hypothesis and construct stringlines about such a topic, it is important to establish the boundaries of what comprises public security in Brazil, and, after that, to identify the particular debate about one of the subtopics related to the overall conception of public security in a country (Brazil) in special.

At this stage, we carried out a overview of policy papers, state and federal government publications and news articles about the federal intervention and public security in Brazil. So, as a preliminary hypothesis, we defined as object of study a combination of actions, institutional references and semantic figures: mentions to president Temer’s decree; references to all main actors involved in the intervention (Rio de Janeiro’s State Department for Public Security, the generals
named by Temer to conduct the operations at Rio; and the federal ministries devoted to the initiative, the Ministry of Justice and the Ministry for Public Security, the last one specially devised to oversee the intervention). The search rule focused on identify, at Twitter, how a selected assortment of events can be manifested, semantically, in public posts: a presidential decree; the foundation of a “Cabinet of Intervention” commanded by an Army general; the impact of the intervention as seen by regular citizens; the use of military personnel, instead of police forces, in operations at slums (“favelas”) and strategic regions of the state, like internal borders and cities from Rio’s metropolitan region with high crime rates.

A stringline has to condensate, in a limited number of linguistic elements and logical articulations, the multiple ways people can speak about the selected object on social media. This method pretends an iconization (GREIMAS and COURTÉS, 2011, p. 250-251; BARROS, 2005, p. 69) of the semantic values collectively related to the topic, and, once the initial hypothesis is established, with the definition of the semantic figures central to the case study, a succession of quantitative and qualitative verifications is applied on a software with access to the database — Twitter. Based on a preliminary stringline, starting from the words and relations more closely associated with the subject — “intervention”; “Armed Forces” and “Rio” simultaneously at the same tweet; the name of general Walter Braga Netto, head of the Cabinet of Intervention etc. —, we evaluate the first results. The preliminary tests ran on the database get to verify the volume of mentions found and categorized and the assertiveness of the stringline, making possible to remove incorrect expressions, filter relevant synonyms and identify ironies and metaphorical constructions widely used on social media, but not correlated to the topic studied. For instance, it was necessary the exclusion of publications about military intervention at the Brazilian federal government — nowadays a fervid subject of debate on Brazilian social media platforms — that did not mention the federal interference at Rio.

If the stringline is very strict, it may lack many accurate posts that use abbreviations, slang, regional variations and might be included in the final database, potentially causing an error of corpus representation. If it is very open, it may gather many incorrect posts, potentially causing the same problem regarding corpus representation. The recurrent verifications intend to minimize such issues, and the data collection only starts when the stringline presents both a high percentage of accuracy and a slight variation on the proportional volume of texts searched, given a posterior addition or exclusion of a linguistic element to the rule.

At the final stage of building a stringline, it must contain textual attributes — structured by the boolean operators used to write the rule — to identify any semantic traces pertaining to the
chosen theme. To Greimasian Semiotics, a semantic recurrence within a text is defined as an *isotopy* (GREIMAS and COURTÉS, 2011, p. 275-278), a combination of semantic values that, in articulation, reproduce *meaning effects* on a certain semantic direction. These values can be figurative or conceptual: mentions about the federal intervention combine abstract notions from political, military and police spheres and figures that manifest them, on the text surface. The discursive figure of a soldier, a rifle, a military apparel or a tank; descriptions of operations against drug-dealers; the act of sanctioning a presidential decree about public security. Words do not have an inherent semantic value determined only by graphic expressions; the semantic value of each figure and theme emerges from internal relations between opposing elements (e.g. Saussure’s fundamental principle of *linguistic value*) at the textual object. That is why a stringline built only with hashtags, isolated words and expressions, without considering how verbal languages articulate linguistic units in mutual relation to generate meaning, may not get an adequate corpus on a text-based social media research.

To understand the restrictions provided by the timeframe of the search is as well fundamental, as a pivotal procedure to correctly gather the debate for a specific period of time, with its influencers, actors, emerging events and particular discussions — a journalistic coverage of a previously undiscovered matter influences enormously on the influx of publications about a given subject. This is a necessary, from a theoretic perspective, return to the central dichotomy defined by Saussure between the synchronic and diachronic approaches (2011, p. 133), by which every subtle change to the dynamic of values inside the linguistic system constitutes a specific ‘linguistic state’, able to be analyzed as an independent object. Because of it, new textual figures from outside the semantic fields regularly recognized as pertinent to the search of a subject may start to appear in relation to the topic, as the nomination of a new minister to a public office or the sarcastic use of a word to describe a concept from a distinct semantic field — in Brazil, it is very frequent to replace, in usual discussions on Twitter, the nation’s monetary unit (real) for a product heavily affected by an inflation or for the name of a president. For example: “this ticket is going to cost me a thousand Temers” or “gasoline’s average price is four potatoes”.

4. Collective subjectivities

In this section, we explain the research method to discover the different collectivities that were formed online during the first month after the announcement of the federal intervention in Río de Janeiro. By doing so, we wish to undercover how citizens might have reacted to the news online and how their interactions online have formed different collectivities. Furthermore, we wish to
understand the differences between the interpretations that may have risen regarding the issue and what were the main explanations endorsed by each group. First, we shall introduce the concept of Collective Subjectivity, central to the proposed theoretical approach, and then link it to their online mode of existence in which we will elucidate how the networked framing processes occur. After that, we will explain how we think SNA can undercover what we describe analytically as Online Collective Subjectivities. Finally, we will do the analysis of our chosen case study in order to show how a qualitative analysis of he most relevant data uncovered by both the linguistic data aggregation method and the SNA help us find the main interpretations arising from the debate.

4.1. The concept of collective subjectivities

According to Domingues (1995), the concept of collective subjectivity (hereinafter CS) is an attempt to solve the classical sociological dilemma between action and structure. In his description of the concept, he decenters not only the individual, but also the collectivity that is not seen as an homogeneous unity. His emphasis is on the interaction between individuals and collectivities (DOMINGUES, 1995, 2004). Thus, any social system could be seen as a CS, varying in causal effect and centring level – that is, of internal homogeneity. Corporations, social classes, friend groups, states, social movements and even families can be understood as CSs (DOMINGUES, 1995, 1998). The concept takes into account four different dimensions of social relations: the hermeneutical, the material, the power, and the space-time (DOMINGUES, 1998, 2002). All these dimensions affect and help delineate the CSs, but at the same time they are also being shaped and reshaped by the interaction of the CSs with each other (1995).

CSs’ main characteristic is their causal effect – or collective causality, as Domingues puts it (1995). This means that the CSs and their members interact in such a way that they have a causal effect over the world, that may be intentional or not. If we assume that social systems do indeed exist, we have to classify the causality of a CS as being different of the active causality, typical from the weberian tradition, that focuses on individual activity; and of the conditional causality, that is derived from the durkheimian tradition and focuses on the effects of society over itself (DOMINGUES, 1995, 1999). The collective causality has to be understood as a property of the social systems, which refers to their interactions.

This causality may be manifest (intentional) or latent. Many times the impacts of interaction – i.e. their causality – on a specific system may not be intended or felt by their members (DOMINGUES, 2004). However, it is important to keep in mind that certain collectivities are conscious of their collective causal effect, even though this consciousness may only be practical – in
the sense that Giddens distinguishes practical and discursive consciousness (GIDDENS, 1986). This means that individuals are incapable of expressing discursively the impact they have caused and/or suffered. This process leads Domingues (2004) to develop the concept of involuntary centering to explain the causality that derives from a collectivity whose actors ignore their impact as a group. The centering depends on the intrinsic potential of each social system and does not necessarily take place.

Two central aspects are responsible both for the changes in the social systems as well as their reproduction and “inertia”: social memory and social creativity. While both are tied to the individual subjectivity and are established by interactive processes, memory is composed of mnemonic elements that result from interactive processes of the collective subjectivities themselves, which crystallize over time in a variable way, depending on the collective subjectivity in question, both materially in symbols and objects as well as in the individual body, brain and subconscious themselves (DOMINGUES, 1998).

These elements, allied with the unintentional consequences of action, power and the interactive inclinations of individuals and collectivities, help organize social systems, sedimenting rules, behaviors and patterns of interactions between individuals and collectivities. However, they can also be transformed. The primary source of transformation lies in the unconscious (DOMINGUES, 1995, 1998) and is due to what Freud called the primary process – that is, the incessant slide of meaning typical of the unconscious. Later, this has to be translated to the social reality of the individual through the secondary process, in which he tries to match the mental image created by the id’s primary process with an object of the real world (FREUD 1900).

Thus, according to Domingues, social life is interwoven and constituted as an interactive, multidimensional network in which individual actors and collectivities exert causal influence on each other (DOMINGUES, 1995, 2002, 2004). Independent of whether interactions happen face-to-face or on a virtual space, from the moment actors find themselves in a specific social situation, they will try to interpret the situation in which they are inserted in. We can draw a parallel with Goffman’s idea of definition of the situation (GOFFMAN, 1974). The interpretation of a social setting, however, depends on the way actors perceive themselves and each other, and the interpretations arising delineate the common ground of interaction for individuals and collectivities (MEAD, 1934; DEWEY, 1927; STONE, 1962; SNOW, 2001; CICOUREL, 1993).
4.2. Framing Theory and Sensemaking

One theory that conceptualizes how different interpretations of reality arise is framing theory. In *Framing analysis* (1974), Erving Goffman explains which circumstances make individuals accept a given situation as real and how individuals select specific elements in order to be able to define the situation. Goffman argues that individuals utilize frames in order to highlight specific aspects of the situation in question. The different forms to delimit and describe the situation are capable of altering the perception people have of this situation and therefore have a causality over the world.

For example, Tversky and Kahneman (1981) show how presenting a situation differently might alter individual decision and, thus, actions, which directly influence reality. If we enhance a positive outcome of a policy instead of a negative one, even though both are going to take place, people are more likely to be in favor of the policy. According to the authors, small, apparently non consequential changes in the way problems are formulated and presented might entail significant changes in individuals preferences (1981, p.457).

Similarly, following the sensemaking literature (MORGAN et al., 1983; WEICK, 1995, 1999) might be useful for framing theory in order to understand, on one hand, how individuals give meaning to reality by articulating and drawing from different pre-existing framing possibilities and cues from the environment, and on the other how this meaning is contested (FISS & HIRSCH, 2005). Each culture, according to Modigliani and Gamson (1987), would encompass a series of interpretative packages that articulates the culture’s symbols and ideas. Framing theories also highlight the individual’s capacity to manipulate expressions in order to persuade others of a specific framing (KLANDERMANS, 1992; SNOW & BENFORD, 1988, 1992; SNOW et al., 1986).

Drawing a parallel with the concept of collective subjectivities, we could say that each collective subjectivity will build collectively through their internal and external (with other CSs) interactions a discursive field, from which they can draw certain symbols whose constant use leads to the crystallization of different interpretative packages that are to a bigger or smaller degree used and shared by its members. These interpretative packages could also be described as framings for real-world situations and subjects, narratives built collectively by each collective subjectivity to make sense of reality.

According to Snow and Benford (1988), not all frames are able to have resonance, which means they might not find support in the public. At the same time, since the identities of individuals are part of the universe of discourse that is built taking into account the norms of groups and continuously transformed by its interactions, as explained by MEAD (1932); the framings that find more resonance in the audience will have more power in the sense that they have had more
impact in the universe discourse and thus, in individuals and groups themselves. This process, in which identities are shaped following group-specific guidelines (SNOW & MACHALEK, 1984) is called identity talks (HUNT & BENFORD, 1994). Both personal and collective, identities are products of social interaction, shaping and being shaped by collective actions — which is one of the possible collective causalities CSs have in the world. Any kind of mobilization might produce meanings that are used collectively and personally in building identity, and these change and are also reproduced through talking (HUNT & BENFORD, 1994). In our case study, whenever we limit the object’s scope, while what have is a frozen picture of reality, analytically we always keep in mind that these are fruit on an ongoing process, of which we are just looking at a temporal and context specific dimension.

While at the same time each individual and collective are simultaneously trying to understand the world deriving from previous framings/interpretative packages or adapting those to create new ones that better explain the situation, there may be fights to try to impose framings as hegemonic interpretations of the same situation both among sub collectivities inside a CS and also between different CSs. The symbols that become crystallized are the ones that are used the most, up to the point they become internalized by CSs – which is the process through which new symbolic meanings are created, passing from the individual dimension to the social one. This means that the longer time and the more people utilize a new symbolic articulation, the more it sediments its social meaning. In this sense, there is a constant fight for meaning fixation, and each different individual and CS will struggle to impose their meaning on others.

These linguistic social constructions – be them words, meanings, narratives – are a collective creation that are altered and dislocated, on the one hand, through the power battles (as before described) to establish meaning and truth, and on the other through the processes of sensemaking of reality on both the individual and the collectivity. In practice, this means that we borrow elements from each others’ narratives, and after successive reconstructions and reinterpretations of symbols through metaphorical and metonymic processes we may even collectively create new meanings and symbolic constructions. Just as nations share certain patterns of cultural, linguistic and narrative crystallizations of their symbols, so do smaller sub collective subjectivities build their own collective narrative as their individuals try to make sense of reality. These are inevitably related: a) to the group’s own narrative, by selecting symbols and utilizing or rearranging existing frames of interpretation; and b) to the other group’s narratives, since CSs exert influence on each other and react to each other’s’ movements.
4.3. Collective Subjectivities’ Communication in the Digital Era

Communication theory traditionally conceived the social influence process as having two steps in which news flows from opinion leaders to followers (BERELSON et al., 1954; KATZ & LAZARSFELD, 1955; LAZARSFELD et al., 1948; MERTON, 1949). However, according to Gamson (1992), opinion formation is not as simple as people mimicking opinion leaders or mediatic discourse. In the author’s book Talking Politics, he shows how the everyday citizen is capable of framing issues by connecting the information coming from the media with a reflection based on their own worldviews and experiences, as well as their peers’. Furthermore, Gamson concludes that even though many people are not likely to engage into political action a priori, they can be stimulated to do so, activating a sort of political consciousness if there is what he calls “action context”. Thus, with the advancement of informational technology and communications technologies, it has become increasingly easy to find people that share your grievances (GAMSON, 1992), and even to build common grievances collectively and turn them into action, as we have seen happen ever since events like the Arab Spring and Occupy Wall Street.

Social network theory allows us to study this process taking into account the growing importance of steps that take place horizontally, between citizens (FREEMAN, 2000; NOOY et al., 2005; SCOTT, 2000; WASSERMAN and FAUST, 1994). Today, because of the internet, it is easier to diffuse information and researchers are growingly emphasized the importance of the role normal citizens have had in creating content and influencing major events online (CASTELLS, 2012; MERAZ & PAPACHARISSI, 2013). What is interesting in the internet as a potential public sphere of debate is that, although opinion leaders exists — the so-called digital influencers, which range from regular internet celebrities to even journalists and celebrities of other venues, that have huge amounts of followers and thus have an advantage position to make their messages diffuse through the network — the average citizen still has potential to influence the discussion and even become an influencer himself.

If a person is retweeted enough, he can either slowly or abruptly grow his audience and thus his influence. People with more followers have a direct channel of communication and idea exposition to a larger audience than those with less followers, who depend on the agreeableness of their own followers to the point that they retweet the message in order for it to have a chance at influencing a large number of people. Still, if the followers of the person who retweeted you do not believe in your message enough to share it, your tweet will die and probably will exert little causal influence to the collective subjectivity as a whole (MERAZ & PAPACHARISSI, 2013).
Everyone has the potential to become an influencer himself by having his statements amplified by the connectedness of online networks (CASTELLS, 2012; MERAZ & PAPACHARISSI, 2013). In fact, it has already been shown that this was what took place in events such as the Egyptian Revolutions during the Arab Spring, in which a collective process of networked framing (MERAZ & PAPACHARISSI, 2013) took place. Networked framing is defined by the authors as “process through which particular problem definitions, causal interpretations, moral evaluations and/or treatment recommendations attain prominence through crowdsourcing practices” (2013, p.22).

This means, even though internet celebrities have a vantage starting point when compared to the average citizen user because of the direct audience they have, the average citizen still has potential to influence the discussion both microscopically (endorsing whomever framing of the event he agrees with by retweeting the message (HANSEN et al. 2011; BOYD, GOLDER & LOTAN, 2008) and macroscopically thanks to the logic of connective action. According to Bennet and Segerberg (2013), this logic explains how large-scale processes like Occupy Wall Street and the Arab Spring emerged in a more or less organic form, with no specific social movement organization playing a central role. The importance of the retweeting process is that this repetition process helps the movement grows (MERAZ & PAPACHARISSI, 2013). The more someone is exposed to a frame, the more readily it becomes available in the individual’s mind (IYENGAR, 1990). Thus, the more people repeat the same framing/retweet, the more likely it will be of influencing public opinion (HEMPHILL, CULOTTA & HESTON, 2013). By repeating the same tweet, just as with a refrain, the tweet’s disruption augments and might even overthrow the entire order of discourse (BERTELESEN & MURPHIE, 2010; DELEUZE, 1995).

Social media platforms propitiate more information diffusion, on the one hand because of the fast speed with no a priori geographical constraints, but also because of the weak ties that facilitate information propagation on social media platforms (BARBERÁ, 2014). A study by Stroemer-Galley (2002) has concluded that many people that are discussing issues online are doing it also with people they would not discuss with face-to-face, which leads the authors to conclude that the internet may be seen as a public sphere in the sense that it offers a space for political conversation among a broader spectrum of the population, connecting people and groups that would otherwise not interact. Hence, the internet seems to be a fruitful way to observe how groups fight for these processes of opinion formation online.

In terms of information diversity, Granovetter (1983) concluded that weaker ties might even be more important than stronger ties in the sense weaker ties are responsible for spreading of diverse information to groups, in a sense, heterogeneity of thought in a sense. If we consider some authors even hypothesize this process creates “contention in framing by the actors in these
fields” (FISS & HIRSCH, 2005), weaker ties become even more important in facilitating framing dispute inside and between CSs. Since they help disseminate information to heterogeneous groups, there will be more communities talking about the same events and thus mobilizing common symbolic articulations into different framings that might vary, depending on the CS.

For the scope of this research, we will be looking at traces of interactions online — specifically, as retweets. Following the line of thought elucidated so far, if a person X sees Y’s tweet online and agrees to the point he identifies with it and wants people to know it or he wants to disseminate a specific message such as a news event, X will retweet Y’s tweet.

### 4.3.1. SNA as a method to undercover Online Collective Subjectivities

We argue in this paper that groups formed when people are discussing a specific issue online can be understood as a virtual layer of the groups that Domingues (1995) calls collective subjectivities and explained previously here. However, since we work with real world-data, these could never comprehend the full extent of a CS, for we will analyze interaction only on one layer — that is, online, and specifically on Twitter. However, they could be composed of other social interactions. Thus, we will call these groups Online Collective Subjectivities (hereinafter OCS).

The line of argumentation we develop here is based on the empirical finding that clustering algorithms applied on sharing networks, regarding posts that show polarization (GUERRA et al., 2013) and are seen as controversial, divide a retweet network of profiles into groups that have distinct worldviews — as shown to be the case in political topics (Conover et al., 2011). Thus, based on the previous explanation of CSs, we will move to elucidate how we define this concept analytically in order to look at the data and then the methodology used to identify these groups in the retweet network.

An OCS is formed through interactions online. They are a way to group profiles online by their messages sharing patterns, in which the symbolic articulation proximity of user defines their position in the Network. What arises from the phenomenon is a division that is defined by us analytically as OCS, united in a way that, aware of it or not, individuals’ profiles online form a group that publicizes collectively a narrative by sharing intensely the same set of messages (in this case study, tweets) which end up thus, collectively yet individually, amplifying certain viewpoints that are hidden in these messages. Tweets, in this context, also due to their size, are seen as circumscript framings of situation that arises from the interactions between individuals, that draw from their discursive and symbolic fields to form these tweets, and OCSs. However, it is important to remember that individuals symbolic fields also change as they are affected by other individuals’
and OCSs’ discursive and symbolic fields. These are formed in its more obvious form by retweeting someone, but could be analyzed more precisely by looking at expression uses and new linguistic articulations that are incorporated and spread on the network in a new written tweet from another actor that has come in contact with these elements through his network.

Much like the emergence of a social movement that depends on where individuals focus their attention (HUNT & BENFORD, 1994), OCSs’ existence depend on individuals’ decision to focus their attention collectively on an issue and writing and sharing content about it. This means that there must be produced content online that finds frame resonance — which means, like in social movement theory (NOAKES & JOHNSTON, 2005), having the power to motivate a part of individuals — in this case, among the online interacting profiles on Twitter. Logically, if there was no one talking about the federal intervention, there would be no data for us to analyze. This awareness in part of the population that generates debate online is many times induced by a feeling of indignation or injustice triggered by a real world event that becomes popular either through the efforts of an intentional real-world social movement, or through the network’s own natural activity — i.e., by an individual decision that makes a topic go viral. But the main difference from OCSs to social movements is that OCSs’ members form a group without necessarily having the awareness that they belong to this group, and whose limits are, on the one hand, not well delimited, sometimes being blurry where one OCS end and another begins and, on the other hand, very fluid, in which alliances may have big oscillations from time to time.

Having this in mind, when looking at how the most popular tweets frame the issue being analyzed, we are able to see what found more frame resonance, being brought to the status of importance by the number of retweets that show support for these framings. These can thus be analyzed in order for us to understand the big narrative supported by the OCS, without this implying that this is a vision shared homogeneously on the individual level by all the OCS’s members. This narrative, subjectively analysed by connecting the most popular frames, is fruit of the processes and symbolic negotiations between an OCS’ different framings of the situation. In these negotiations, tweets with more reach (in this case, retweets) are the winners because they get amplified and thus has a higher probability to exert more symbolic and persuasive causality on the network, being thus potentially more powerful. This intuition is supported by (MERAZ & PAPACHARISSI, 2013, p. 18), who state that different than framing negotiation processes in other mediums: “the practices of repetition on Twitter, embedded into the ambient and affective nature of the medium, were essential to crowdsourcing a frame to prominence”.
We know activists act online, and people also use online platforms to show their personas, and we know that identity formation at the individual level might be also a part of world sense-making and that this process takes into account the groups’ worldviews, identities and opinions. By allying the notion of collective subjectivities from Domingues to the processes of identity talks of (HUNT & BENFORD, 1994) among individuals and CSs, and having in mind that there may be militants of specific social movements acting, we can define the Online Collective Subjectivities that arise online as the framing process that was indeed formed between social movements and individuals that are interacting with social movements, forming a sort of organic social movement which frames collectively, even if they are aware of this of not (like a CS), and thus exerts collective causality on the world because their framings of the situation will impact the way other OCSs will frame the situation too. Of course in real life these analytical distinctions between CSs and OCSs do not exist, for in life all spheres of the social are interwoven, and when interacting online individuals’ mental processes also depend on: conversations they have in real life, in other communication mediums as well as on the informations they received and the order in which they were received — Lau et al. (1991) showed that users will accept the first point of view to which they are exposed as long as it’s not too far from their current view of their matter.

OCS, differently than social movements, have no awareness about their belonging to a group and its limits, as a network can be at times unclear and fuzzy. New connections of support may grow at incredible speeds, just to disappear days later again. OCSs can be unstable, as they are formed organically around world events and topics, in which 1) people may tweet about current events sharing opinions about it, that may even be borrowed from someone else (in the form of a retweet), trying to position themselves towards a situation in order to, on the one hand, make sense of reality and, on the other, present an identity of themselves to the world; and 2) real world social movements find opportunity to mobilize people to their cause and build framing bridges with their own social causes, trying to gain more supporters.

A tweets’ size limit makes it impossible to express the whole ideology of a social movement or a manifesto. It is more likely, if one disconsiders the text that may be in the an URL attached to the tweet, that it has to be limited to a couple of main ideas. This makes it necessary, just like a newspaper headline, to have to summarize a position — in this study, a frame — within Twitter’s textual limits. This might at first seem as an disadvantage, but it is easier for specific, limited diagnostic or prognostic framings in the format of tweets to resonate with the general population, because its size demands succinct and more general constructions. Thus, people can pick and assort online their own identity by stitching together these different framings that are being offered
in the forms of tweets and present on their profile page a more general narrative that shows an individual self, albeit influenced by the OCSs.

### 4.3.2. Retweet Network: Map of Online Interactions

A retweet is a way of sharing a message or image on Twitter. As one looks at a retweet network, one manages to understand the relationship created between different user profiles, which can be seen as a trace of their virtual interaction. On more practical terms, whenever someone shares someone else’s post, this link can be understood as the propagation of information from user A to user B (CONOVER, 2011; BOYD, GOLDER & LOTAN, 2008). This is due to the fact that when I share someone else’s post, it is just as if I am, on the one hand, agreeing with the statement and, on the other, repeating it and propagating this idea for my whole network (MACSKASSY, 2011; BOYD, GOLDER & LOTAN, 2008).

In this work, we support the idea that framing processes take place in Twitter. We use a tweet as a unit of analysis because, thanks to the linguistic string method described in this paper, the tweets in our datasets are all selected because they present a textual correspondence which ensures that it is making a reference to the analyzed theme. In this reference to the object, which generally comes in form of text but can also include pictures, memes and videos, the author can be addressing the theme with neutrality, but many times he is also manifesting his opinion to a specific theme - even if this opinion is given in the form of a retweet. Thus, we analyze a tweet trying to understand the framing supported by the author of the tweet/retweet. In this sense, we consider a tweet is a text excerpt in which the author manifests his support to a specific framing of the situation, topic or subject related to the theme being discussed.

Moreover, as said before, as I share a tweet, thus disseminating its framing, words and theme, I raise the chances of any of these elements to be utilized by another user. Still, it is important to keep in mind that people, discourses and collectivities are in constant movement, for interactions are dynamic and not fixed, being developed as a process. Whenever we look at a graph we are looking at a static, circumscribed space-time picture of a virtual reality.

In order to build a retweet network, it is possible to extract from Twitter’s API a table containing all the tweets that match the search stringline. From that, we built a script in order to create another table with three columns: one is the source, another is the target, and the last one is the weight. Whenever user A retweets user B, the script fills “A” on the source column, “B” on the target column, and one is added to the weight column. At the end of this process, we have a
table that informs us how each node interacts with other nodes. This table can be inserted on the software Gephi to create visualizations that undercover the dynamics between different users.

4.3.3. Modularity and Clusterization

After inserting the data into Gephi, we run the software’s modularity algorithm. Modularity maps different groups inside a network (NEWMAN, 2004, 2006). The algorithm separates the network in different groups (GIRVAN & NEWMAN, 2002), each of them having more links between the nodes inside of its cluster than between the nodes of other clusters – which coincides with Newman’s intuition that individuals will establish more relationships (which, in a network, are represented by edges) inside their groups than with other groups.

Once modularity has been calculated, the clusters are defined. We will consider these clusters are representations of the OCSs. The connections between the nodes – or profiles – should be understood as traces of the OCS members’ social interactions. We used Gephi free software to colour each OCS in order to differentiate them visually. Afterwards, based on the retweet number, we manage to establish the most important nodes of each cluster – i.e. their authorities. This will be useful for the qualitative analysis, since the authorities indicate which messages resonated the most within each OCS. Thereby, we will be able to discover what framings managed to aggregate people in specific clusters and what general narratives we can derive for the OCS with these frame analyses.

Finally, we implement the ForceAtlas2 algorithm. This attracts nodes whose edges are thicker, while it also repels nodes that have thinner edges between each other (Jacomy et al. 2014). This means that nodes that establish more relationships with each other will be attracted and will be shown spatially close to each other on the graph. This method allows us to visually understand how different individuals and their groups – here defined as OCSs – establish networked relationships (NOACK, 2009, JACOMY et al. 2014). According to Conover (2011), clusterization algorithms like the ForceAtlas2 divides retweet networks in profile groups that express different political points of view. This has also been verified in our case study when analyzing the debate about the federal intervention in Rio de Janeiro.

This high level of sharing inside of clusters leads us to assume we can see these groups as OCSs, because this indicates a high level of interaction among cluster members (NEWMAN, 2004, 2006). In other words, if I am a member of cluster x, I will probably share the same tweets or news other members of my cluster share. This behavior indicates that there is something similar in the train of thought of these individuals, that there is a logic that is common to them and makes a certain content appears sufficiently relevant to the point that individuals of the same cluster need
to share and publicize it (BOYD, GOLDER & LOTAN, 2008). It also indicates that there is common
ground of understanding of symbols and even values. Since people interact in a more or less or-
ganical way online, the emergence of OCSs takes place through involuntary centering, as

The visual representation we obtain from the graph is informational in a few aspects: links
between the nodes, that represent relationships between users’ profiles; the position of the node
in relation to others; the size of nodes’ labels, that are proportional to their importance; and their
colours, that indicate the groups to which the nodes belong. Along with the visual interpretation
that comes from these characteristics of the graph, we will do a qualitative analysis of the most
retweeted tweets per cluster. We will look at how each OCS mobilizes different framings for the
same event by using different words, meanings and narratives, which will help us determine what
are the groups’ worldviews and what unites individuals to form a specific OCS. Moreover, we will
be able to understand if a group’s framing of the situation is disputed by another group’s framing.
This would help us understand the interactive inclinations of OCSs, which refer to the way CSs
interact with one another, and would also make us understand who is trying to influence the de-
bate and public opinion and what are the arguments that find most adherence among other citi-
zens that are using social platforms. This will vary according to the centering level of each collect-
tivity which, in turn, depends on the level of internal identity and organization.

5. The debate about federal intervention at Rio de Janeiro

This paper integrates two fields of research to conduct its methodology and conceptual under-
standing of social media analysis, namely semiotics and sociology. And, as already described, we
will be working with one dataset created by gathering posts about the federal intervention, by the
Armed Forces, in the state of Rio de Janeiro, following president Michel Temer’s decree on Feb-
uary 16, 2018.

The dataset comprises Twitter publications (original, replies and retweets) filtered with
the application of six different stringlines using a Gnip API subscription (access to all public posts
available) after a default common filtering for Portuguese-only mentions. This language filter is
provided by Twitter, and we do not have the algorithmic set of parameters the platform uses to
recognize a post as written in Portuguese or in other language\(^2\). We have tested the same string-

\(^2\) For more information on Gnip PowerTrack (real time and historical) operators, see: http://support.gnip.com/apis/power-
track2.0/rules.html.
lines used to collect the data with and without using the language operator “lang:pt”, and the percentage difference between the quoted estimation of mentions downloaded before and after using the operator was smaller than 3%. Language filters have been susceptible to errors, but the absence of a language operator to filter publications may result in an expressive volume of data wrongly collected, as natural languages from the same historical origin — derived from Latin, for example, like Portuguese, Spanish, Italian and French — share similar syntactic structures, words, morphological variations and patterns. Is it relevant to point out, as well, that Twitter does not have a filter for the Brazilian Portuguese variation; however, as the difference between applying or not a language operator was quantitatively very low, it did not substantially affect the corpus selection, although it can be problematic for other thematic objects of study.

Each stringline was applied to a delimited time period, with modifications inserted to the search rules in order to filter new terms and expressions not relevant and observing the changes on public debate regarding the impact of federal intervention. As the weeks passed after Temer’s decree in February, the overall debate about the topic diminished; it was necessary, to ascertain the accuracy of the stringline, to allocate in proximity relation many linguistic attributes that, in the first days, could be filtered in isolation, such as the name of Eduardo Villas Boas, head of Brazilian Armed Forces — on the first days after the presidential order, every mention to general Villas Boas was pertinent to the heated discussion about the subject on Twitter, but, after a few weeks, he became greatly associated with other military topics not directly related to the object of study. Further rule modifications important to the data collection were linked to external events and news of great and timely impact for Twitter users, prompting extensive volume of posts associated to the intervention — most of all, the murder of Rio de Janeiro councillor Marielle Franco, on March 14, 2018.

The first applied stringline gathered 676,438 tweets from February, 16, to February, 19; the second one, with slight modifications compared to the starting search rule, gathered 592,122 posts only from February, 20, to February, 27; the third one, to complete the database from day 27, another 26,784; the fourth one, adding relevant hashtags created by users favorable to the federal intervention, collected 335,649 tweets from February, 28 to March, 05; the fifth one, with several editions and much more strict, correlated with the overall contraction of the debate, got additional 119,304 mentions from March, 06 to March, 12; the last stringline, very similar in semantic scope to the fourth one, closed the data collection with 258,621 tweets selected from March, 13 to March, 19. The graph below has, in total, 1,405,834 retweets from a complete database of 2,008,918 publications.
6. Findings and Analysis

6.1. Visual Analysis and Findings

The modularity algorithm helped uncover the different groups arising from the retweet network about the federal intervention in Rio de Janeiro. Each one of these groups are to be understood analytically as OCSs, and are colored differently in order to be visually distinct. We analyzed the four biggest groups because the other ones were small and unrepresentative, comprising less than 1% of the profiles interacting online on the debate. Thus, these smaller groups were excluded from the visualization and analysis due to their small role in the debate.
The biggest OCS found is the red one, that mobilized about 34.14% of the profiles discussing the issue. The second one was colored in light blue, and comprises about 22.09% of the profiles. Following, we have dark blue OCS with 19.18% of the profiles and pink OCS, that represents 18.65% of the individuals debating. On the retweet graph, we can see these four bigger groups comprise a vast majority (about 94.06%) of the profiles talking about the issue. This indicates that, by analyzing separately each groups’ more relevant content, we will be able to understand what were the most important interpretations arising from the announcement of the federal intervention and how the different actors positioned themselves regarding the issue.

Both red and pink OCSs present arguments that indicate implicitly or explicitly that they position themselves against the federal intervention, which explains their proximity in the graph figure. The blue OCSs (dark and light), on the other hand, are in favor or the federal intervention, which makes sense given their visual proximity in relation to each other in the graph. We can notice from the figure that the red and the dark blue OCSs have a more dense structure in which its nodes are being pulled with a stronger force towards opposing sides, which indicates the groups must have opposing views regarding the military interpretation. Looking closely at the image, one can see that the red OCS has nodes concentrations on the left side of the graph, while the dark blue OCS has a higher agglomeration of nodes and edges on the right side of the graph. This means that there are few nodes that articulate tweets arising from both of these OCSs, which indicates these OCSs might have opposing interpretations of reality. Still, we can see some bridges, probably from users mobilizing each others’ tweets to criticize the framing of the opposing OCS while defending their own. In fact, as said before, when we look at the most retweeted messages from the groups, we can see this visual interpretation is also found on our qualitative analysis, which we will develop on the next section.

The red OCS was composed of 119024 different users and 773390 total tweets, which means the tweeting average per user was about 6.5 tweets. The pink OCS, also against federal intervention, had a much lower tweeting average: only 1.9 tweets per user. Similarly, although light blue OCS had the second highest number of unique users (76982), it also holds a low tweeting average: 1.8 tweet per user. Dark blue OCS, on the other hand, had the highest tweeting average: 11.11 tweets per user. So, although dark blue OCS was only the third biggest OCS in terms of unique users, when we take a look into the total number of tweets, it has the second highest number of tweets (742690), which represents 36.97% of the total number of tweets. Still, the most important OCS in terms of debate contribution was the red one, being responsible for 38.50% of the conversation. This high number of tweets from the dark blue and red OCSs explain the more opaque visualization generated by these groups – because there are more retweets among them, there are
more edges between the nodes and thus their representation becomes more opaque. This high level of tweeting also indicates that these groups might be more concerned and engaged in this debate than the pink and the light blue OCSs, which only generated 6.18% and 6.9% respectively and whose average tweets indicate a smaller effort/concern in publicizing the matter.

6.2. Content Analysis: OCSs and their frames

If we take a closer look at the data of the light blue and pink OCSs, we notice that these groups influencers’ have a disproportionate weight on the debate of their clusters. They do not have a really high followers count, but have written tweets that have found high resonance on Twitter during this debate. When looking at the most popular tweets from the different OCSs, we notice that red OCS shares many of pink OCS’s top tweets, just as dark blue shares many of light blue OCS’ top tweets. This indicates that both pink and light blue OCSs are composed majorly of individuals that participated in the discussion by engaging primordially with the highly viral content, contributing to the dissemination of the subject outside of the groups that usually discuss politics online. As we look at the influencers of the red and dark blue OCSs, on the other hand, we find that there are many politicians, journalists and generals of the army among the most influencing authors. The discussion on these groups is more intense, as indicated by the average number of tweets per user, and more ideologically based. The influencers of these groups appear to have a clearer political orientation than the ones from light blue and pink OCSs. Albeit the four OCSs have a clear opinion regarding the federal intervention, red and dark blue OSCs have stricter views than pink and light blue OSC.

The members of the red OCS retweet messages that mention police abuses, identifying the military police as a problem and using it as an argument against the federal intervention. Red OCS was also highly mobilized by the assassination of councillor Marielle Franco, with many of the top tweets talking about the event. It seems that either the assassination awoke sentiments of indignation and solidarity regarding violence, and/or the people opposing the intervention thought it was a good strategy to manipulate these sentiments to gather more people to their side of the discussion. Either way, the assassination of the city councillor indeed made conversation spike again a month after the federal intervention’s announcement, which led to conversation in the group about the intervention to rise from 2814 tweets on March, 14th, to 62750 on March 15th. This can be seen as a political opportunity structure (BENFORD & SNOW, 2000) that facilitated collective action framing processes that opposed federal intervention.
Indeed, of the 20 most retweeted messages of the group, 10 mentioned the assassination of the city councillor in some way. A common association was a temporal one: many tweets brought attention to the coincidence of Marielle's execution and her appointment as referendary of the Commission that would supervision the federal intervention. Furthermore, other tweets criticized those who, like president Michel Temer, tried to use her murder as an argument for the federal intervention even though she was openly against it. The president was also criticized for associating Marielle Franco’s assassination to organized crime without proof and for using her death as an argument for the federal intervention, even though she opposed militarization.

Another framing that appeared after the assassination of Marielle Franco mobilized a pop culture reference – the movie “Elite Troop”. One of the popular tweets talks about how the traffic and the militia are involved with the military police and the government. The group criticizes the ones that are in favor of the federal intervention because they only remember the parts of the movie in which the military police tortures drug dealers, and not the structure of the violence that is behind traffic. Another tweet that also alludes to the movie reminds the audience that the movie was more than fiction and that the director even had to leave the country because he was threatened for the movie’s death. This framing induces the audience to assume the police is involved with Marielle’s execution. In this framing, the police is the one portrayed as the enemy and the reason to be fearful. Another tweet reminds the audience that the weapons used by the drug dealers come from the police, also strengthening the narrative that the police plays a role in traffic violence.

One other common framing of the situation that resonated in the red OSC was the inequality of rights between citizens and an emphasis on the lack of human rights by certain section of the population - in this case, people that live in slum. We can see this, for example, on the third most popular tweet of the group. The author employs prognostic framing by implying that the existence of groups in Brazil who do not understand structural oppression and thus negate the existence of racism, sexism, homophobia, etc. are a problem for the country. In this framing, supporting a federal intervention would be as bad as advocating that feminicide is not an important issue.

The fourth most supported tweet by the OSC was a post which says there has always been violence in poor parts of Rio de Janeiro. However, the author evaluates that when robbery got to the rich zones, the elite pressured for a federal intervention. The author evaluates the problem is that the government only cares about one social class - the elite. This tweet was the most popular one after the intervention announcement, but lost its place after the death of Marielle Franco. Another popular case brought to prominence by the OSC that showed indignation was the inspection of children’s backpacks in slums. This military action was framed as “absurd” by the author.
of the tweet. The tweet then poses the question of what would people think if children of rich neighborhood would also have their backpacks inspected, trying to make people put themselves in the slums inhabitants’ shoes.

Following on the human rights and inclusive agenda, the group also renders popularity to a tweet that refers to the case of Rafaela Silva, an olympic champion who was inspected at random for no reason in a major avenue of Rio de Janeiro. The olympic athlete stated that this was due to prejudice on the police officer’s side, since he told her taxi driver that was pulled over that he “thought he had picked up the athlete at a slum”. Therefore, while not referring directly to the federal intervention, by portraying the police as racist, the tweet is supporting a framing normally mobilized to oppose the intervention, which evaluates that more police action in slums will not only not solve violence, but will also worsen the quality of life of poor and black people, that become targets of the institution’s prejudice. The OSC also used the inspection of children’s backpacks to awaken indignation, characterizing the military action as “absurd”.

Other framings of the situation emphasized how other states have worse violence statistics, and questions why Rio de Janeiro, which holds the tenth position, is the only State that is receiving federal intervention. The tweet then emphasizes the action is political and aims to attract the media’s attention. The OCS also emphasizes how normal military actions can result in the death of innocent civilians by army men that cannot even be held accountable by their actions. A few of the most popular tweets emphasize this framing of reality: some by remembering situations in which this happened in the past, and another by criticizing a tweet from a rightwinged, conservative politician, Eduardo Bolsonaro, that tweeted it was absurd that a military man could be brought to prison if he shot a civilian in the midst of a military action.

The pink OSC also mobilizes framings that oppose the federal intervention in Rio de Janeiro. The most important influencer has contributed with some of the top tweets of the group, which had also found resonance among the red OSC due to its framing devices that go against the federal intervention. This influencer was the one who articulated the pop culture reference about Elite Troop movie. The group also mobilizes injustice framings, remembering past cases in which the military police would have allegedly killed people who were denouncing malpractices inside the organization. Just like in the red OSC, the enemy, according to pink OSCs narrative, is the police and any efforts that involve more federal intervention would thus fail due to the organizations’ corruption schemes. The injustice framing is also mobilized in order to pressure for equal rights and human rights for all Brazilians. Marielle’s assassination is also mobilized by the group, although to a lesser extent than in the red OSC.
The light Blue OCS, on the other hand, sheds good light on the military police. Most re-tweets depict a soldier handing a flower to an old lady on a bus, giving a sense of serenity, civility, humanity and community for the Army. It tries to deconstruct the image of fear portrayed by groups red and pink. The second most shared tweet shows a picture of a kid with drugs and arms, and says “I think it’s absurd for the military to inspect a child’s backpack”, clearly confronting with irony red OCS’s critique towards the military, with a visual argument that seeks to legitimate this military action. This framing wants to bring the audience to think that it is necessary to do that to children, because they help spreading violence and illegality and thus have to be inspected for a matter of public security. Another popular tweet in the group says that he would rather have his child’s backpack inspected than have him being used by the traffic. This framing also tries to present argumentation in favor of the militarization and legitimation of their actions, and this time the framing of the issue is a matter of public security for the child himself. This tweet is also present as one of the most popular in the dark blue OSC.

Finally, dark blue OCS mainly disseminated tweets that frame the federal intervention in a good way, showing support for it. Its most popular tweet is a video of an actor saying that those who are against the federal intervention should stop buying drugs because these drugs are what gives money that allows drug dealers to buy bullets that end up hurting innocent people. Another popular framing criticized Brazilians for felling indignation towards the police officer that found drugs and weapons on a kid’s backpack rather than feeling indignation towards the drug dealer that was using this child to practice illicit activities. Jair Bolsonaro, a right-winged, conservative federal congressman, posted a tweet that framed the federal intervention as a small solution to the problem. According to him, until Brazil changes its legislation regarding criminal activity, an intervention will only offer a mending that does not solve the issue. For him, it is also necessary to ensure that officers will not be punished if they kill an innocent man by accident during the military actions. Other tweets of the group not only support the federal intervention in Rio de Janeiro, but even defend the military dictatorship that took place in Brazil and ask for its return.

The Rafaela Silva case also appears in the dark blue OSC, but being framed in an entirely different manner. For the OCS, there was no authority abuse. According to the tweet mentioning the case in the dark blue OCS, the military police’s action is just a part of their jobs, justified by the intervention and the high violence level in Rio de Janeiro. For the author of the tweet, and opposed to the framing we found in the red OCS, the current problem of Brazil is seeing everything as a result of racism. Another of the most popular tweets of the group is a counterframing that arises in order to oppose another group’s framing of the situation regarding the military inspection of children’s backpacks. On this tweet, the dark blue OCS brings attention to the fact that it is odd
for Brazilians to start questioning the military’s action as they saw the photo of the child that was caught carrying drugs and arms, instead of focusing on blaming the drug dealing for using children for their profit.

To show support for the federal intervention, the group popularizes tweets that shed good light on the institution of the army and the police, probably as a reaction to the bad framings they also received online. For example, a tweet questions how can the military be the one to be responsible for traumatizing the children living in the slums when they also see drug dealers carrying rifles around their necks. The group also wishes to raise empathy towards the institutions by saying that no one cares when a military police officer dies. On another popular tweet, they criticize the UN for saying nothing about the police officer casualties’ numbers or the number of murders in a year, but when they hear about the federal intervention which “may get in the way of the drug dealing”, the UN expresses preoccupation regarding the use of armed forces on police-oriented activities.

Keeping the critical tone towards human rights movements as the UN, the dark blue OCS also confronts part of individuals that follow left-wing ideology that support the concept of human rights. After councillor Marielle Franco’s murder, the left is accused of being condescending with criminality by the dark blue OSC, which also asks these people what they propose we do since they do not support militarization or repression in any form. In another tweet, a military man says it is impossible to win crime while these “human rights organizations and defenders do not let the army and the police do their jobs”. The OSC also tries to delegitimize the opposers of the federal intervention on a personal level, calling them “little shits” at times or saying the only people that would oppose the intervention are people that “smell coke” and are “potheads” (slang for weed smokers). The group also react to the insinuations of the pink and the red OSCs that the police and/or the military might have been involved in Marielle Franco’s assassination, reaffirming that criminals killed the city councillor.

7. Conclusion

According to Benford and Snow (2000), “Controversies regarding whom or what to blame frequently erupt between the various SMOs comprising a social movement as well as within movement organizations”. Indeed, when it comes to controversial issues like the federal intervention in Rio de Janeiro, different groups had different worldviews and interpretations regarding the matter. By allying a network analysis with a qualitative analysis of Twitter data referring to the event,
we were able to understand how each different OCS framed the issue, and which controversies existed between the different groups.

We can see that some frames of situation that become popular in a specific OSC may have found opposing framings of the same situation in another OCSs. By doing a qualitative analysis of the most important groups, tweets and authors involved in the discussion online, we were able to uncover the counterframes that were articulated by people online in order to make sense of reality and try to offer a solution to the problems of Rio de Janeiro. A big mobilizer of speech during the time period associated with the topic of federal intervention was the murder of city councillor Marielle Franco, a black woman member of a left-winged party that came from the slums and defended a lot of minorities’ rights. Marielle’s assassination brought a new emergence of frames about the subject of federal intervention, since she so vehemently opposed it.

These frames, contained on tweets, were created and retweeted in such a manner that, after using SNA processing, one can uncover groups what we described analytically as OSCs, an online layer of interaction of the CSs (DOMINGUES, 1995). Taking a look at the tweets that were repeated the most and thus were more likely to have had higher affective intensity (MERAZ & PAPACHARISSI, 2013), we could see the dominant frames that emerged on the different OCSs and understand how one OCS was dialoguing with another, trying to win a broader group of supporters on the online sphere and that hopefully and possibly would translate to the real sphere too. This allowed us to understand how the dominant OSCs made sense of the federal intervention.

The red OSC and the pink OSC brought to prominence framings that accentuated the problems arising from the federal intervention. In general, the tweets offer diagnostic framings, which associate the event as a cause of social problems. In these framings, the police, especially the military police, but also the army, are seen as the enemy, and their actions towards the population are seen as responsible for causing of violence. Meanwhile, on the other side of the political spectrum, the dark and the light blue OSCs supported the federal intervention, framing the issue prognostically as a solution to the violence of Rio de Janeiro. When mobilizing diagnostic framing, it was always directed at the people who had different views on the matter than them, that are portrayed as being responsible for the malfunctioning of military activity. According to this framing of the situation, the people who criticize the police’s actions are actually making their job harder and thus are indirectly helping the drug dealers and criminals, who are the enemy that must be fought. The group is so concerned about the violence that they defend a series of harsh measures to fight criminality, like inspecting children for security issues. These measures, however, are seen by the pink and red OSCs as a violation of human rights and fruit of an oppressive State that ends up hurting poor people the most.
Under the linguistic-semiotic perspective, we proposed to demonstrate how a semantic approach over the construction of the corpus for social media studies can present a new position regarding the preliminary stage of data collection, with the elaboration of a semantic selection of themes and discursive figures that compose, socially, the verbal manifestation about an object of study. Instead of analyzing words and textual expressions as isolated, independent entities, or applying mathematical-only models of proportional association — which rarely present textual reflections about how language and meaning emerge and how words have a system-dependent representation —, we have elaborated specific search rules (stringlines) to different time slots, obeying the frequent changes on the public debate about the federal intervention, using multiple lexical associations and morphological constructions (with the use of Gnip’s available operators) to be able to accurately capture a reasonable sample of the many positions, opinions, language uses and actors interacting on Twitter, in several OCSs, about the subject.

It would be impossible to measure the discussions about such a topic only with isolated words, hashtags and expressions, as it was necessary, first, to delineate what actions, narratives, actors and semantic elements — president Temer’s decree, the impact of Marielle Franco’s murder, the operations conducted by the Army at different locations, the generals in charge of the task, the mandatory articulation between figures related to the Armed Forces (tanks, soldiers etc.) and the geographical restriction of the intervention scope. A search for “intervention” only would gather thousands of references to unrelated debates, and a more strict search for “federal intervention” would miss valuable posts about the object and the variable uses people use to talk about the intervention on social media.

Adhering to a linguistic-semiotic methodology, this paper does not exclude the necessary use of algorithms and automatic models to conduct research on social media, but sees the use of both procedures in conjunction as a step forward. On studies that analyze textual expressions and need to survey multiple texts to compose an object, a previous reflection about the corpus has central importance, as the product of the data collection — either social media posts or other big data applications — depends on the verbal manifestations by social collectivities, dependent upon semantic relations and language properties by which meaning emerges.
References


Barberá, Pablo. (2014) How social media reduces mass political polarization. Evidence from Germany, Spain, and the US. Job Market Paper, New York University, nº 46


Conover, Michael; Ratkiewicz, Jacob; Gonçalves, Bruno; Menczer, Filippo; Flammini, Alessandro; Francisco, Matthew. (2011). Political polarization on Twitter. Icwsm 2011


Guerra, Pedro; Meira Jr, Wagner; Cardie, Claire; and Kleinberg, Robert. (2013). *A Measure of Polarization on Social Media Networks Based on Community Boundaries*. ICWSM 2013


Howard, Philip N.; Kollanyi, Bence. (2016). *Bots, #StrongerIn, and #Brexit: Computational Propaganda during the UK-EU Referendum*. Cornell University Press

Howard, Philip N.; Kollanyi, Bence; Bradshaw, Samantha; and Neudert, Lisa-Maria. (2017). *Social Media, News and Political Information during the US Election: Was Polarizing Content Concentrated in Swing States?*. Oxford: Project on Computational Propaganda


Jacomy, Mathieu; Venturini, Tommaso; Heymann, Sebastian; and Bastian, Mathieu. (2014). *ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software*. PloS One, v. 9, n° 6


Krauss, M.; Cabrera-Nguyen, Peter E.; Cavazos-Rehg, P.; Bierut, Laura J.; Moreno, Megan A. (2016). *Young Adults’ Exposure to Alcohol- and Marijuana-Related Content on Twitter*. J Stud Alcohol Drugs, v. 77(2), pp. 349–353


Merton, Robert K. (1949). On sociological theories of the middle range

Morgan, Gareth; Frost, Peter J.; and Pondy, Louis R. (1983). Organizational symbolism


