*PA*rtecipazione e *CO*nflitto
** The Open Journal of Sociopolitical Studies*
http://siba-ese.unisalento.it/index.php/paco
ISSN: 1972-7623 (print version)
ISSN: 2035-6609 (electronic version)
***PACO*, Issue 11(2) 2018: 332-360**
DOI:10.1285/i20356609v11i2p332

Published in July 15, 2018

RESEARCH ARTICLE

SOFTWARE POWER AS SOFT POWER
A literature review on computational propaganda effects in public opinion and political process

Rose Marie Santini, **Larissa Agostini**, **Carlos Eduardo Barros**, **Danilo Carvalho**, **Rafael Centeno de Rezende**, **Debora G. Salles**, **Kenzo Seto**, **Camyla Terra, and Giulia Tucci**
Federal University of Rio de Janeiro, Brazil

ABSTRACT: This article draws on a systematic literature review of recently-published material to investigate the state of the art of research into how social network profiles are being used to manipulate public opinion on political issues. The aim of this review is to discuss how the use of these profiles impacts on constructing consensus, authority, legitimacy and representativeness of politicians, parties, hegemonic and minority groups. In total, 369 articles were obtained following a bibliographic search on four different scientific citation indexing databases. Of those, 16 were considered relevant to the research question. In order to identify general trends and specific results, the pertinent papers were subjected to descriptive and critical analysis. Five main approaches were observed: A) Usage policies and platform neutrality; B) Definition, detection and description of the manipulation agents; C) Political role of agents on social networks; D) Attempted electoral predictions via social networks; and E) Network dynamics and strategies for disseminating information. In the conclusion, the results of the literature review are discussed.

KEYWORDS: Bots, manipulation, politics, public opinion, social media, systematic literature review

CORRESPONDING AUTHORS: Rose Marie Santini: marie.santini@eco.ufrj.br

1. Introduction

Discussions about the strategies, methods and degrees of public opinion manipulation on social media began to gain attention from the press following the US presidential campaign of 2016 and the United Kingdom's Brexit referendum. With Donald Trump's victory, the concept of computational propaganda was reinforced with substantial evidence: bots generated 20% of political messages posted on Twitter during the US presidential campaign. On Facebook, Russian-controlled entities bought advertisements and posted controversial content with polarizing opinions that went viral with the use of bots. Facebook estimates that the content produced by the Russians, including posts and paid ads, reached 126 million Americans, around 40% of the nation's population (The Economist 2017).

According to Woolley and Howard (2017), computational propaganda refers to the use of algorithms, automation and human curation to intentionally distribute misleading information over social media networks. According to Bradshaw and Howard (2017), the use of social media to manipulate online public opinion has been found in more than 28 countries since 2010. Among the different strategies, the authors highlight cyber troops as a global phenomenon: groups of bots created by governments, military organisations or political parties, and funded by public money, to engage in social media networks and manipulate local or foreign publics.

Despite the presence of online bots since the Internet became public in the 1990s (Woolley 2018), social bots are a phenomenon that has emerged with the advent of social media. Social bots are algorithms that can automatically produce and/or disseminate content and interact with humans (Ferrara *et al.* 2016), creating a kind of artificial public opinion on social media networks. The algorithms are able to imitate human communication, including time patterns of content production and diffusion and the expression of feelings. According to Ferrara et al. (2016) social bots have already been used to infiltrate political debates, manipulate the stock market, steal personal information, mirror fake news, generate disinformation and noise.

However, bots have also been used in acts of resistance. Hacktivists like Anonymous participated intensely in the Arab Spring, supporting the popular movements in Tunisia and Egypt through bots and fake profiles. The members admit to having produced content and posted on Twitter and Facebook as ghost writers "in the name of the people". They also gave up their own personal accounts on social networks for posting content in favour of the protests, seeing as the Internet was being censored and even switched off in the Middle East. The aim was to enable the local population to expose the political happenings of the region to the rest of the world (Knappenberger 2012).

What is of particular interest, according to Bradshaw and Howard (2017), is that the preferred method of organizing cyber troops has been changing. It has evolved from engaging military units into the development of online strategies and experiments to attempt to directly manipulate public opinion. Bolsover and Howard (2017, 273) maintain that, "the anonymity of the Internet has allowed state-produced propaganda to be presented as if it were not produced by state actors." Strategic communication firms are now hired by all sorts of organisations, but mainly by governments, to develop and execute online manipulation campaigns on social networks.

Despite media scrutiny on this topic, incipient scholarship approaches still face several controversies and challenges. We propose the first Systematic Literature Review (SLR) about the effects of computational propaganda in public opinion and political process. The SLR is a widely used method in Health Sciences, and recently, some scholars have attempted to apply it in Social Sciences (Dacombe 2017). SLR allows the researchers to exhaustively analyse the publications on a theme, pointing out theoretical and methodological tendencies.

This method can contribute to the emergent computational propaganda research area by systematising results, consensuses, disagreements, limitations and field paradigms. Some of the SLR advantages are the transparency and reproducibility of the method, enabling the comparison of results and meta-analysis (Dacombe 2017). It is important to understand the general impacts of computational propaganda on the construction of public opinion, authority, legitimacy and representativeness of different agents in order to discuss its effects on the contemporary political processes.

In this paper, we map the existing research into the use of social media profiles for manipulating public opinion on political matters. In the next section, we discuss the scientific challenges of studying the effects of algorithmic manipulation on the global political context. In the third section, we present the research protocol and describe how it was carried out. In the fourth section, we perform a descriptive analysis of the selected articles. In the fifth section, a critical analysis of the primary findings is presented. Lastly, we conclude with a discussion on the ethical, social and political challenges in this field.

2. From Cold War to Cyber War

During the cold war, the Soviet Union and the USA disputed the technological domain without direct military action against each other, which led to the space and nuclear-arms races. The internet itself, as a global network, was initially created to be a

military war-machine in the context of the Cold War (Barbrook 2009). Today, it seems the world is once again experiencing an escalating dispute for the technological domain, now characterised by an algorithmic turn (Uricchio 2011) and ideology (Mager 2012), involving the private sector and government agents.

Information warfare is not new in the history of East and West battlespace, but the use of social media leads the contemporary power dispute to a new dimension (Bradshaw and Howard 2017). The private sector, including Internet giants like Facebook, Twitter and Google, has been acting not only as a technology developer and supplier, but also as a strategic operator in the competition for information and power (Sanger 2018).

We are currently experiencing a cyber warfare, characterised by an algorithmic race, that has the potential to cause devastation: an evident example of the combined action of governments, political parties, civil organizations and technology companies is the Russian manipulation on the US 2016 presidential election (Shane and Mazzetti 2018). The algorithmic race involves the combination of big data, data mining and artificial intelligence that can be applied to store and target information on users for ideological influence and emotional contagion.

The development of political bots arises within this computational dispute to influence, manipulate and modify public opinion and behaviour by dissimulated and indirect means (Bradshaw and Howard 2017), representing a new kind of soft power. For Joseph Nye (2012), within the concept of soft power, the best propaganda does not look like propaganda. Bots are easily taken for ordinary users and, according to The Economist (2017), they are more effective at disseminating content than real people and media companies.

As stated in the 2016 Incapsula report (Zeifman 2017), 48.2% of Internet traffic is human activity, and 51.8% is bots, classified in eight types, which can be split into good bots and bad bots. Bearing this in mind, the following questions arise: how are we to identify and track the bots? What are the differences and similarities between fake profiles, whether human-controlled or automated? How do these different agents operate? Who builds the bots, who sells them, and for whom do they work? When do they generate content and how do they act? What sources do they use? Who are their target audience? What do we know about their effects on public opinion formation and on the political process? Many of these questions remain unanswered. After the SLR, we observed that there are some robust studies on this theme, but they are few and far between. This is not only due to it being a new subject matter, but also due to the conceptual and methodological problems typical of this field of study.

At a conceptual level, there are difficulties in defining and distinguishing agents and actions of manipulation on social media. The boundaries that divide bots from human behaviour are becoming more and more tenuous (Ferrara et al. 2016). Among all these agents, there are fake profiles that are bots behaving like an army of humans, and others that are people behaving like an army of bots (such as "click farms"). There are trollers and fanatical users who produce and diffuse messages mechanically, performing the role of buzzers, with the aim of viralizing all that they publish.

The bots can also infiltrate networks of real connections, produce credible content and adopt human-like time patterns to hinder their detection. They can automatically gather content from several predetermined sources, post material in real time, establish conversation, comment on posts and answer questions. Moreover, there is a growing trend of combining human action, big data and automation to further refine their application (Ferrara *et al.* 2016), bringing the concept of bots close to that of the cyborg.

In relation to methodological aspects, the research challenges are huge, especially as the activity is questionable on both legal and moral grounds. On the one hand, there are no official data about malicious profiles available from social media companies, which suggest that they benefit from the presence of bots, boosting their audiences (Dayen 2017). On the other hand, these platforms are proprietary, making it difficult to access data. Twitter data, i.e., can be extracted free of cost via Twitter API, but the content retrieved is limited, approximately 1% of all tweets, with no assurance of a random sample (Kim *et al*. 2013).

Another important fact to consider is transparency in relation to the number of bots active on each platform and how their characteristics might compromise the reputation and credibility of these social media companies in the eyes of their users, and hence their turnover and market value.

Furthermore, the bots are manufactured behind a veil of secrecy, hindering the identification of the manufacturers and the collection of information. In the academic arena, researchers face ethical dilemmas to conduct non-obstructive experiments. Other qualitative methods of investigation (such as interviews, participant observation and focus groups) present challenges, as the bot operators and the companies that sell and hire them do not want to be identified with such a suspicious activity.

Quantitative methods are also insufficient for the automatic identification of bots, and especially of cyborgs. Identification methods based on graph analysis tend to fail because the bots easily connect with real people for camouflage. Human identification tends to be very precise and effective (Ferrara *et al.* 2016), however it is unfeasible for

shortage of time and economic reasons by virtue of the amount of social network user data.

A Systematic Literature Review (SLR) was conducted to develop a diagnosis of the scientific research in this emerging field. We indicated the types of case studies, the approaches and methods employed, and the practical and theoretical conclusions drawn to date. A systematised overview of the studies can increase scientific understanding of the phenomenon and its geopolitical dimensions. The SLR contributes to the field so that specific generalizations can be avoided (Dacombe 2017) and biased conclusions diminished, according to the centrality of one particular case or research context.

3. Method

A systematic literature review should comprise three main stages: planning based on a research protocol, execution and reporting of results (Kitchenham *et al.* 2009). The protocol consists of a research question, search strategy, and selection criteria. The execution phase involves the research itself and, finally, the findings reporting stage, which encompasses the analysis, discussion, and interpretation of the results. Based on the discussion presented earlier, two research questions were defined:

RQ1: How social media profiles are used as intervention agents for the manipulation of public opinion in politics?

RQ2: What are the social and political consequences of these intervention agents on social media?

Considering the research questions, four conceptual axes were established in order to develop the search expression. These axes are (1) intervention agents; (2) social media; (3) politics; and (4) consequences. For each axis, a list of keywords and their respective synonyms was created in a multi-step process to enable inquire of the databases. Initially, brainstorming sessions were carried out based on previous literature to define the list of concepts and synonyms. Several databases search pre-tests were carried out in order to refine results. Each keyword was tested individually and combined with the other axes. The final search expression was based on Boolean logic: "OR" for the synonyms, and "AND" to associate the four conceptual axes, as shown below:

TITLE (troll OR bots OR computational propaganda OR fake OR fraud OR hack OR spam* OR hate) AND TOPIC (facebook OR twitter OR 4chan OR reddit OR app OR osn OR online social network OR social network service OR digital media OR forum OR internet OR social media) AND TOPIC (politic* OR elect* OR candidat* OR democra* OR vote OR govern* OR state OR presiden* OR polls OR activis*) AND TOPIC (crisis OR manipulat* OR monitoring OR influen* OR public opinion OR journalis* OR audience OR misinformation OR consen* OR popular*)*

We conducted our search in four main indexed scientific literature databases, covering the areas of Politics, Information Science, Social Sciences and Computer Science: Scopus, EBSCO, IEEE XPlore, and Web of Science[1].

3.1. Exclusion Criteria

When defining the exclusion criteria, we focused on the design, purposes and outcomes of the primary studies. It is important to acknowledge that exclusion decisions remain relatively subjective (Gimenez and Tachizawa 2012) and may have biased our results, but this limitation is already envisioned by the SLR approach (Véras *et al.* 2015).

Initially, some steps were carried out in order to organize our results, either through the databases affordances or through automated evaluation of our results spreadsheet. In order to capture the international debate (Petticrew and Roberts 2006), only scientific articles, conference proceedings, books or book chapters written in English were considered. Duplicated results were also excluded.

The exclusion criteria were established according to the research question and were drawn up based on thematic relevance, as recommend by Petticrew and Roberts 2006). Since we are interested in a qualitative discussion, the systematic processes for assessing weight of evidence did not unduly restrict our findings (Davies *et al.* 2013). Articles regarding the production of yellow journalism news (fake and junk news) were not among our interest. Therefore, if the answer to the question (a) 'Is the research focused on the production of Fake News?' is positive, the article is excluded from our selection. Likewise, articles on technical issues, i.e. source code review or development, should be excluded. So, if the answer to the question (b) 'Is the research

[1] For each database, adjustments were made to the expression search according to the platform's technical operation.

focused on computer programming issues?' is positive, the article is excluded from our selection.

Furthermore, the research must be focused on manipulating agents, and therefore the article must be excluded only if the answer in all of the following questions are negative: (c) focused on bots, (d) focused on fake profiles, (e) focused on trolling, (f) focused on information cascade campaigns, and (g) focused on the mechanisation of real profiles (click-farm or fake like).

Finally, to evaluate the pertinence regarding the political context, at least one of the answers to the following questions must be positive, otherwise the article must be excluded: (h) Is the research about State Policies (government image building)? (i) Is the research about elected officials or political figures? (j) Is the research about political parties? (k) Is the research about electoral dispute? (l) Is the research about disputes for meaning or an ideological dispute over minority groups (feminism, xenophobia etc.)?

The exclusion of primary studies was based on an analysis of titles, keywords, and abstracts. When information provided in the above-mentioned sections was absent, the introduction, method and conclusion sections were examined. Two authors evaluated each article and in the event of discrepancy between their appraisal, a third author would make a definitive assessment.

After applying the selection criteria, we evaluated the articles in terms of quality, in order to reduce bias and indicate the strength of evidence provided by the review (Petticrew and Roberts 2006). Based on a list of criteria adapted from Najafabadi and Mahrin (2016) and Kitchenham *et al.* (2009), four questions were defined for checking the methodological and scientific rigor that relate to the clarity of the objectives, methods, results, and limitations of each study. These issues have three possible answers, to which different scores are given: Yes = 1; Partially = 0.5; and No = 0. The questions and their respective responses are presented in Appendix 1. We defined that a total score of less than or equal to two (score ≤ 2) represented low methodological quality and these articles were excluded from the systematic review.

3.2. Execution of the review

The database search was conducted in November 2017. Following the procedure described above resulted in the return of 467 documents and, with the exclusion of duplicate items, a total of 369 texts. These documents were then assessed based on the exclusion criteria. After the abstracts had been read by the pairs of authors, 45

articles were considered relevant. With the validation of a third author, 34 of those articles remained in the body of analysis.

These texts were subsequently quality assessed, which resulted in a further five articles being reassessed as irrelevant to the scope, five articles being excluded due to low methodological quality (scoring 2 or less) and another two could not be assessed as the authors were unable to access the entire text. Once the complete texts had been read another six articles were excluded on the grounds of being inapplicable to the scope of the review. Thus, 16 pertinent articles were left.

4. Results and descriptive analysis

With the 16 relevant articles selected, we examined the following metadata of the works: year, source and field of knowledge of the publication, quantity of authors and in which country they were based. In each study we identified the authors' genders, the method used, the social media studied, the country where the intervention was conducted, the nationality of the journal and the scope of analysis.

Of a total of 50 authors, 31 were men (62%) and 19, women (38%). As regards where the authors work, most of them are at US institutions (19 authors) and the rest work in: Germany (7), United Kingdom (6), Brazil (4), Indonesia (3), Ecuador (4), Colombia (2), Spain (2), Israel (1), Australia (1) and Canada (1). There is a concentration of authors engaged at US institutions (38%).

The first two works on the theme of this review were published in 2015. In 2016, eight were published and in 2017, six. The search does not reflect the complete result of publications in 2017, as it was conducted in November and some articles may have not been indexed by that month even if they had already been published.

Two major areas of knowledge are producing works on this theme: communication (five works) and computer science (eleven works). Eleven studies are articles published in periodicals and five in conference annals, all presented in the area of computer science. The five works from the area of communication were published in scientific periodicals, four of which were gathered in a single special edition of the International Journal of Communication. Of the journals that have already published pieces on the theme, nine are US publications, one German and one Spanish.

Among the selected studies, four were conducted with a focus on the USA (Bessi and Ferrara 2016; Mustafaraj and Metaxas 2017; Sadiq, Yan, Taylor, Shyu, Chen, and Feaster 2017; Stieglitz *et al.* 2017), two on Spain (Abril 2016; Ben-David and Matamoros-Fernandéz 2016) and just one on each of the following countries: Brazil

(Oliveira, França, Goya, and Penteado 2016), Canada (Ford, Dubois, and Puschmann 2016), Colombia (Cerón-Guzmán and León 2015), Ecuador (Recalde *et al.* 2017), France (Ferrara *et al.* 2017), Indonesia (Ibrahim *et al.* 2015) and United Kingdom (Murthy *et al.* 2016). Three articles did not specify any single location (Woolley 2016; Ferrara *et al.* 2016; Marechal 2016).

Of the sixteen works reviewed, six are primary mixed-method studies (Sadiq *et al.* 2017; Murthy *et al.* 2016; Ibrahim *et al.* 2015; Cerón-Guzmán and León 2015; Stieglitz *et al.* 2017; Abril 2016), five are primary qualitative studies (Ferrara *et al.* 2017; Ben-David and Matamoros-Fernandéz 2016; Ford *et al.* 2016; Marechal 2016; Recalde *et al.* 2017), two are primary quantitative studies (Bessi and Ferrara 2016; Oliveira *et al.* 2016) and three are literature reviews, however none of them systematic (Woolley 2016; Mustafaraj and Metaxas 2017; Ferrara *et al.* 2016). As no systematic literature review was identified in the search, we can consider that this is the first on the theme.

Regarding which social media are analysed, twelve studies investigate data extracted from Twitter, one extracts information from Facebook, two address data from more than one social network and one does not specify which platform is studied. The research protocol included discussion forums in the searches, and the only one mentioned was Reddit.

One of the limitations of empirical studies in this field is the availability of data. Most the research projects work with data from Twitter because this platform allows broader access than the others. There are a few issues related to this concentration of research on a single platform, namely, (i) inability to draw generalizations, (ii) priority on specific user groups, (iii) potential to turn obsolete, and (iv) limitations of the analyses according to the specific characteristics and uses of the platform (Rains and Brunner 2015). Furthermore, when compared to Facebook Inc., Twitter has a lot lower penetration and user numbers, thus compromising the representativeness of the data (Molina 2017).

In relation to the research scope, eleven articles focus on bots, nine on trolling, three on fake profiles, three on spammers and three on information cascade campaigns. No study was found that investigated the mechanisation of real profiles (i.e., Click Farms and Fake Likes). Regarding the political contexts addressed, ten studies look at the electoral dispute; eight discuss elected offices or political personalities; five involved political parties; three address State politics (construction of the government image) and three analyse discussions involving minority groups (feminism, xenophobia etc.). Some articles analysed more than one object of study and considered a variety of political contexts.

5. Critical analysis

This section proposes a critical analysis of the pertinent articles and introduces a discussion regarding the relevant points of each work. This interpretative evaluation enabled us to identify the objectives, trends and limitations of the studies. We found five primary, non-exclusive approaches in the articles, presented in Figure 1: (A) Usage policies and platform neutrality; (B) Definition, detection and description of the manipulation agents; (C) Political role of agents on social networks; (D) Attempted electoral predictions via social networks; and (E) Network dynamics and strategies for disseminating information.

As demonstrated by the intersections between approaches (B) and (C) and between (B) and (E), observed in four and three articles respectively, (B) is the most commonly addressed theme among the selected articles and also the one that converges most with the other topics. This shows that there remains some theoretical-methodological discrepancy regarding the definition, detection and characterization of the manipulation agents. Hence, when discussing the social impacts of the agents and strategies of information propaganda, it would seem that studies are still prioritising a technical and conceptual discussion about the agents.

In the following sections we group the articles by thematic focus and propose a discussion based on their similarities and differences.

5.1 (A) Usage policies and platform neutrality

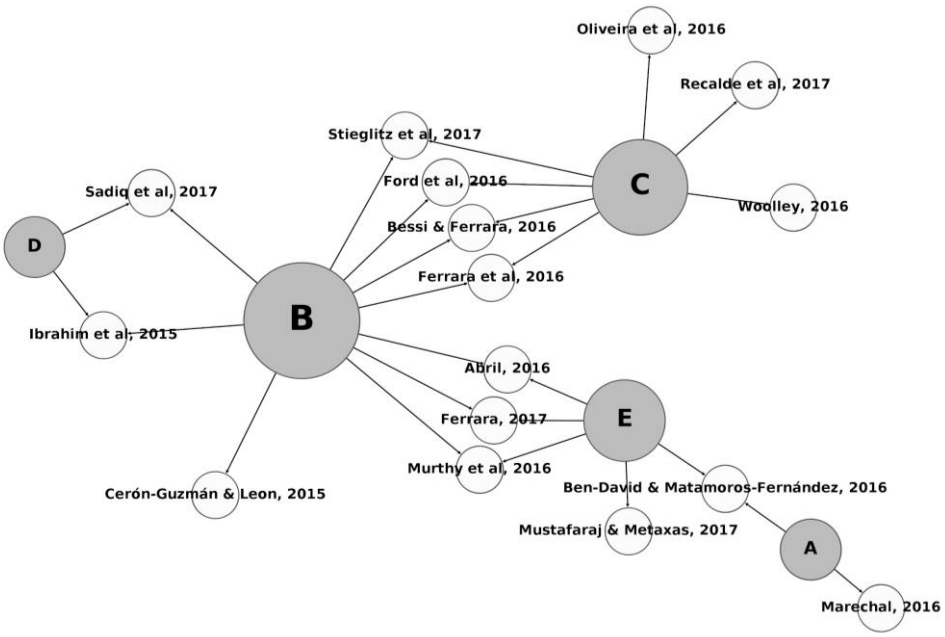
In this section, we discuss the usage policies and responses by social media platforms to online public opinion manipulation initiatives. The studies conducted by Ben-David and Matamoros-Fernandéz (2016) and by Marechal (2016) describe two major regulatory gaps: (1) the circulation of hate speech and (2) the non-existence or lack of regulation governing the political use of bots.

Based on the analysis of pages published by Spanish far-right groups on Facebook, Ben-David and Matamoros-Fernandéz (2016) describe how covered discrimination and hate speech circulate on the platform, favoured due to its technical characteristics and limitations on monitoring in its usage policies. The authors indicate that most of the hate speech is circulated through the publication of links to external sites, shared content or comments on the page. These means are not subject to algorithmic filters, monitoring by the platform editors or subject to being reported as improper by users.

Hence, the accountability is transferred through a chain: from governments to corporations, from these to users, and from these to external sources through links.

Marechal (2016) adds that this transfer of accountability leads to a culture of mutual surveillance, not far removed from totalitarian states, and reflects the power imbalances on the web. As well as proposing a universal regulatory standard in relation to hate speech, the author also defends general regulation for the use of bots on online platforms. Analysing the terms of use of Facebook, Twitter, Reddit and Telegram, the researcher argues that none of the companies fully guarantees the users' rights. She therefore suggests three parameters for regulating bots on social networks: (1) transparency: bots should be presented as such, without trying to resemble real people; (2) consent: contact with and sending messages to users should only occur against their authorization, and (3) limitations on the secondary use of data: data captured by bots should only be used for purposes as explicitly consented to by the users.

Figure 1 - Graph of connections between research approaches



In conclusion, both Ben-David and Matamoros-Fernandéz (2016) and Marechal (2016) agree that social media companies, due to the priority they give to guaranteeing revenues over user rights, have not yet taken such measures. The approach for the regulation of the platforms may, in the future, be supported by broader legal developments (Ben-David, Matamoros-Fernandéz, 2016) or by competition among companies in the market (Marechal, 2016).

5.2 (B) Definition, detection and description of the manipulation agents

In this section we have grouped the articles that, through a variety of methods and forms of analysis, seek to detect and identify intervention and manipulation agents, as well as to describe their activities on social media.

Each work uses its own definition of bot, however there are considerable similarities between the definitions. Stiegliz *et al.* (2017) use the term bot to describe a software designed to automate a task in a computational environment. For Murthy *et al.* (2016), bots are social media accounts that are entirely or partially controlled by software. Ferrara *et al.* (2017) refer to bots as automated accounts on social media, which have the purpose of deceiving and manipulating.

According to the definition proposed by Bessi and Ferrara (2016), bots are accounts controlled by algorithms that emulate human user behaviour, but operate at far greater speed. Ferrara *et al.* (2016) define the term social bot: a computer algorithm that automatically produces content and interacts with humans on social media, trying to emulate, and possibly influence, their behaviour.

Sadiq *et al.* (2017) define the term chatbot as a rudimentary software with minimal automation and basic conversation capabilities, that directs its script to social networks with the aim of attacking, manipulating facts, or simply causing noise. Ford *et al.* (2016) examine the relationships of the WikiEdit bots with journalists, politicians and the government. The WikiEdit bots are transparent bots that announce on Twitter when certain institutions and organisations, especially of the government, edit or publish articles on Wikipedia.

Bots are the most widely investigated intervention and manipulation agents, but other characters also appear as the object of study, such as trolls, buzzers and spammers. All these agents can also be automated and can be used with the purpose of causing noise. The work by Abril (2016) evaluates the role of trolls in the context of the 2012 parliamentary elections in Catalonia. The author indicates that trolls are profiles that generate racist, xenophobic, homophobic, misogynous, class-

discriminatory or similar content (*apud* Tabachnik 2012) with the intention of disturbing conversation flows.

Ibrahim *et al.* (2015), trying to predict the result of the 2014 presidential elections in Indonesia, define buzzers as accounts that support one of the candidates running for election while defaming the other. These can be divided into two categories: bots and accounts of either paid or fanatical users.

Cerón-Guzmán and León (2015) define spammers as accounts designed to infiltrate social media to be detected by security systems, imitate human behaviour and gain the trust of real people. The authors describe two main objectives of spammers: (i) to create the appearance of popularity as followers of one user and (ii) to defame a profile, indirectly benefitting the other.

Automatic bot detection has become more complex with their increased sophistication. Ferrara *et al.* (2016) suggest a taxonomy for bot detection: (i) Bot detection system based on the network of friends; (ii) Detection system based on crowdsourcing, in which a single profile is analysed by several people; (iii) Machine learning methods based on the identification of features that discriminate bot from humans; and (iv) Mixed methods: a combination of methods that analyses multiple dimension of user behaviour.

Cerón-Guzmán and León (2015) develop a study against the backdrop of the 2014 Colombian elections, and using Twitter data, create a machine learning technique to identify spammers according to the following account characteristics: number of followers, number of tweets, age of the account, rate of user mentions, rate of hashtag usage. They identified that 22% of all the accounts were spammers.

In a study conducted during the 2016 French presidential elections, Ferrara *et al.* (2017) also use machine learning techniques to filter data sets and detect bots, considering three criteria: (i) default or customized profile (customization requires human action), (ii) lack of geographical metadata (the use of Twitter app through smartphones records real users physical location) and (iii) activity statistics like total number of tweets and frequency of posts, proportion of retweets to tweets, proportion of followers to accounts followed. Hence, Ferrara *et al.* (2017) conclude that, based on usage patterns of bot accounts, there might be a black market for bots reutilization for political disinformation.

Bessi and Ferrara (2016) analyse bot content from the period of the 2016 US presidential elections and use BotOrNot, a tool developed with the use of machine learning, to detect bots. The authors analyse how the conversations timing on social network were related both to external events and to behaviours originating within the network.

Sadiq *et al.* (2017) use machine learning techniques to analyse the stance of real Twitter users through their posts. To detect bots, they develop an Associative Affinity Factor Analysis (AAFA) framework, which when compared to more commonly used tools for semantic analysis, demonstrates more effective performance, with a 96% precision rate.

Evaluating data from the 2016 US elections, Stiegliz *et al.* (2017) opted not to use machine learning. They base their assessment on literature and define a series of hypotheses in an attempt to differentiate suspicious behaviour. The following criteria were evaluated: number of followers, retweets, mentions of other users, links posted and quantity of posts. Murthy *et al.* (2016) do not use any detection method at all. Their strategy was to purchase existing bots on the market, connect them to Twitter accounts created by voluntary users to then, through observation, try to understand the behaviour of these bots.

After identifying intervention agents, some articles then focus on characterising the activities of those agents on social media. Murthy *et al.* (2016) analyse the ways in which bots are annexed to social media accounts. Their possible functions are: injecting content, disseminating spam, guiding or hindering discussions. Bessi and Ferrara (2016) also identify social bots as a manipulation tool in political debate.

5.3 (C) Political role of agents on social networks

Another approach that appears in seven articles is how technologies are used to intervene in automated fashion in online political debate. Woolley (2016) highlights the effectiveness of bots in attacking, hijacking and altering lines of argument on social networks. His work draws on a selection of 41 news items and articles between 2011 and 2014 that reports of political use of bots in 18 countries. As well as simulating human interaction, social bots can mine user data and manipulate the perception of public opinion. Three forms of activity were identified: (i) demobilizing adversary initiatives; (ii) supporting partners; (iii) driving the number of followers of certain profiles. The first two forms tend to be more commonly used in more authoritarian-leaning countries. The third strategy tends to be more frequent in more democratic countries. The circumstances of these interventions were also categorised as follows: elections, protests, political support and safety issues.

Ferrara *et al.* (2016) conduct a literature review about the influences of these agents in online and offline environments. The authors stress that the presence of bots on the internet is generally inoffensive and even positive for the user. However, both these

bots and those produced with malicious intentions can create information cascades based on false content and rumours, causing network noise and artificially inflating the follower bases of certain profiles.

Both reviews identify governmental sectors (chiefs of state, parties and military entities) as the main beneficiaries, and likely employers, of these bots. There are two common characteristics found in the remaining articles grouped in this category: (i) they focus particularly on case studies related to governmental sectors, such as electoral disputes and protests, and (ii) they only analyse data from Twitter.

Bessi and Ferrara (2016) observe social bot activity and their influence on online political discussions during the 2016 US presidential election. After carrying out a sentiment analysis of tweets posted by humans and bots, the authors identified that human participate more than bots during debate peaks, concomitant with network-external events. This occurs because bots are programmed to divulge information in a systematic and continuous manner, whereas humans are more susceptible to engaging in political debate.

Moreover, the pro-Trump tweets were significantly more positive that the pro-Clinton tweets. The fact that bots produce more positive content to support one candidate can influence the perception among users on the network, suggesting the existence of an organic support for a specific candidate. Hence, the presence of social bots can raise three central issues: (i) influence is redistributed through suspect accounts that might be operated with malicious intentions, (ii) the political debate can become even more polarised, and (iii) the dissemination of disinformation can increase. One interesting point of this analysis is the fact that it is impossible to identify those responsible for automated accounts (Bessi and Ferrara 2016).

Many bots are programmed to simulate human user behaviour with the objective of influencing public opinion. However, Stieglitz *et al.* (2017) identify behavioural characteristics that differentiate them from non-automated profiles. The authors analyse 1,464 active accounts in the debate about the same 2016 US election. However, these programs can be just as effective as humans in communicating with real users. Indeed, the emotional response people display to content published by bots is comparable to the responses to content published by other people. Consequently, bots are capable of distorting perceptions regarding content and disputes between different positions on networks.

Another type of situation involving the political use of bots and affecting relations between governments and citizens is that of public protests. Oliveira *et al.* (2016) investigated the dynamics of message dissemination on Twitter during two opposing protests that occurred in Brazil in 2015, one against the government and one in favour.

The article focuses on the diffusion of retweets of nine posts for five days (before, during and after the demonstrations). The authors argue that a large proportion of the retweets are bot-published content. Therefore, when analysing just the raw numbers of a sharing of content on the network, without considering the role played by bots, there is a real risk of overestimating impact.

Recalde *et al.* (2017) discuss the profiles classed as hashtag hijackers, defined as accounts that empty out the original meaning of their political opponents' hashtags to create noise on the networks. Based on a body of Twitter publications about the 2015 and 2016 demonstrations in Ecuador, a semantic analysis of messages was performed, in which 17% of the users were classed as hashtag hijackers. The authors allege that it is essential that Twitter include some identification of these hijackers in their algorithm that recommends which people to follow, since this distorts the political perspective of its users.

Opposing this phenomenon is the rise of "transparency bots". Ford *et al.* (2016) analyse the WikiEdits bots, which are programmed to monitor and report on Twitter any edits made on Wikipedia by IPs associated to the government. But the surveillance is not neutral, but rather directed and limited by the users of these tools. Therefore, the authors argue that the WikiEdits do not guarantee positive results for democracy. However, the role of these bots opens up the debate about values such as transparency and confidentiality in the dispute of a broader political field.

To sum up, the potential number of bot publications is therefore disproportional to that of human users, which increases the bots' capacity to generate empathy, engagement and social impact. We also observed typical characteristics of an emerging field of research, such as methodological difficulties in identifying and classifying the automation strategies used on social networks for political manipulation.

5.4 (D) Attempted electoral prediction via social networks

The articles analysed in this section discuss the possibility of predicting election results based on data extracted from social networks. Hence, in online electoral analyses, it is necessary to distinguish between the influence of automated profiles and the exclusively human opinions. The works of Ibrahim *et al.* (2015) and Sadiq et. al. (2017) seek not only to measure the weight of this automated influence, but also to propose correlations between the scale of real support for each candidate on the networks and the result in number of votes at the end of the election.

Ibrahim *et al.* (2015) proposed a prediction method based on Twitter for the Indonesian presidential election in 2014, with a margin of error of 0.61%, which is less than that of the traditional opinion polls. Based on a sentiment analysis of the content of tweets and their meanings, the messages were classified as positive or negative in relation to the mentioned candidates. This score was used to define the voting intention of the users.

To study the 2016 US election, Sadiq *et al.* (2017) collected Twitter data from the accounts of influential figures, both Clinton and Trump supporters. The analysis also included 3,000 bot accounts hired by the researchers and programmed to emulate the action of supporters. The proposed method attributed a positive, negative or neutral sentiment in relation to each candidate and demonstrated an 80% effectiveness rate in detecting the electoral affinities of the human accounts.

Both the aforementioned studies test the efficiency of emotional analysis models applied both to bots and human accounts. They attempt to calculate the likelihood of each account support a candidate and, consequently, to measure total endorsement on the network for each of the presidential candidates (Ibrahim *et al.* 2015; Sadiq *et al.* 2017). Excluding the support by bots from the sample, the proportion of positive and negative support for each candidate in the expected total number of votes is projected, thus giving a prediction of vote distribution and of the election results (Ibrahim *et al.* 2015). It is important to question the allegedly superior efficiency of electoral predicting by analysis of social networks, since it opens up an epistemological discussion within political science regarding the methods of opinion polling and demographic vote projections.

5.5 (E) Network dynamics and strategies for disseminating information

Of the pertinent articles, five indicate a research trend related to network dynamics and possible strategies of information propagation. In the study conducted during the 2012 parliamentary elections in Catalonia, Abril (2016) investigated the existence and behaviour of trolls on Twitter using two central questions: (i) to what extent was trolling successful in Twitter discussions about independence in Catalonia, and (ii) what were the characteristics of the successful trolls? The work identified a minimum amount of successful trolling on Twitter during those elections. However, the author recognised the importance of alerting any sign of trolling whatsoever, as a single successful action can generate the artificial diffusion of information and, consequently, a manipulative impact on public opinion.

Murthy *et al.* (2016) conducted an experiment on Twitter whereby 12 volunteers created accounts on the platform and interacted with media events related to the UK general election of 2015, by means of a set of hashtags. To half of these accounts the authors attached bots that automated the follow and retweet functions, and replicated only the content created by the volunteers, with the aim of increasing the chance of followers seeing the posts. The bots failed in trying to influence the logics of information diffusion on the network because the created accounts had none of their content retweeted and were unable to gain any central importance. This is due to the fact that the created accounts remained distant from previously existing influencers on the network, since the stance strongly depends on the social capital of the profiles.

Ben-David and Matamoros-Fernandéz (2016) focus their study on the propagation of hate speech on social networks. The research monitors the Facebook pages of parties from the Spanish far-right, considering the hypothesis that the network dynamics maintain algorithms that interfere in sociability, thus favouring the diffusion of hate speech manifested by the users. The logic of this algorithm creates filter bubbles, in reference to a personalised use of the network (*apud* Pariser, 2011). The most obvious consequence of this filtering is that users with racist behaviour on Facebook are recommended more similar content within the platform. Therefore, the study concludes that the network dynamics of Facebook contribute to the dissimulated propagation of hate speech.

The network dynamics are also the focus of a study conducted during the 2017 French presidential election. Ferrara (2017) investigates disinformation campaigns operated by social bots on Twitter, among other objectives, in order to understand how the functional characteristics of the platform might contribute to information propagation. The study analyses bot and human behaviour on Twitter and also the interactions between these profiles and concludes that bots engaged in disinformation propagation in the 2017 French presidential election, albeit to no great effect.

In the study conducted by Mustafaraj and Metaxtas (2017) during the 2016 US presidential elections, the objective was to evaluate organised efforts to disseminate disinformation in recent online political campaigns. The results reinforce the importance of the network dynamics, signalling that the most important step for the propagation of disinformation is the infiltration of fake profiles in real user networks so that the real ones pass on the content within their own networks.

6. Conclusion

Based on the findings of the systematic literature review presented in this paper, it can be concluded that computational propaganda is a two-pronged phenomenon – technological and social. In the academic arena, it is an emerging theme in the area of computer science, yet still incipient in communication studies, which indicates an important research agenda for the social sciences. On the one hand, broad technical knowledge is required to extract information, deal with big data, create algorithms, produce sophisticated bots and assess them. On the other, to consider it from a technical perspective alone would hinder us from understanding the possible social effects, such as, its impact on democracy, on the dissemination of fake news, on the contagion of hate behaviours and on people's perception of reality.

The objective of this SLR was to reflect on the state-of-the-art and draw a diagnosis of the field, including the identification of possible gaps and new directions to explore (Kitchenham *et al.* 2009). However, the limitations of a SLR applied to a new and interdisciplinary topic must be considered. There are still no specialised journals, the articles are dispersed and indexed in different databases, and there is no controlled, standardized vocabulary. In these cases, a SLR is incapable of covering all the existing literature, but can rather indicate a likely map of the studies.

The researchers are still attempting to develop precise bot detection tools, but little is known about how they act and what their consequences will be. It can be noted that two thematic gaps are conspicuous by their absence. We failed to find a single study that addresses the mechanization of real profiles through "click farms" and "fake likes", used as a tactic to increase the impact of posts and the popularity of certain profiles on social media. We also found no studies that present strategies and methods to identify who creates, who funds and who sells these bots or "cyborgs", and for whom these bots work for. There is a latent demand for effective fieldwork methods to gain a deep understanding of the actors and intentions behind digital astroturfing.

One of the solutions raised in the articles to avoid the anonymity of the bots and their "master of puppets" on social networks is the regulation of the platforms, and this argument is reinforced in sub-section (A). However, regulation is a complex issue met with strong resistance by social media companies who do not want their business models to be subject to legal restrictions and control. Presently, there is not enough transparency about the mechanism of their algorithms, nor about the behaviour of malicious agents to develop an effective regulation.

It is also immensely difficult to define a conceptual and methodological framework that is capable of differentiating between artificial behaviours and ill-intentioned

online actions, for example: bots; fanatical users; personal accounts administered by companies; trollers; buzzers and spammers. To what extent do these intervention agents differ as regards their impact on public opinion? Or do they just represent different means to the same end? Might we consider that they play mutually complementary roles in relation to computational propaganda? These questions remain open and require the development of rigorous concepts and critical stances to be adopted by researchers with respect to the technological, political and communicational phenomena.

Another gap demonstrated in the results is the relationship between computational propaganda and the dissemination of hate speech. Many articles were excluded due to their discussing the hermeneutics of online hate speech, but building on a questionable premise: that the behaviour of the actors was always spontaneous. It is necessary to consider the hypothesis that, in some cases, the production and dissemination of hate speech can be an orchestrated action. Therefore, the opposite premise should be tested through critical studies based on robust and consistent empirical data.

An additional problem raised from the literature search is the concentration of empirical works based on data extracted from Twitter. There are very few studies about the presence of social bots on other social media platforms. In addition to the reasons and limitations presented in section (4), there are other specific characteristics of Twitter that exacerbate the bias in the results of these studies. For example, among the most assiduous and influential content producers on Twitter, we see brands, celebrities, politicians and famous figures of various kinds and niches (Jin and Phua 2014). And its user base, totalling roughly 330 million (Molina 2017), corresponds to only 16.5% of the Facebook user base, with over 2 billion (Facebook Newsroom, 2017), which reflects a low level of representativeness in relation to online social media audiences.

Therefore, the role played by Twitter in public opinion formation is something that should be questioned in more depth. Could it be that the Twittersphere has a two-step flow function, of emphasising the role of opinion leaders in influencing common users, but in a distributed manner, that is, on a micro scale? Perhaps, despite its low penetration, the platform holds some agenda-setting power over the traditional media? Or is it that Twitter functions primarily as a second screen, like in live broadcasts, complementing the consumption of the mass media? Against this backdrop, there is the need to discuss how traditional communications theories might be revisited and updated.

Regarding the methodological aspects, it is common for analyses of social networks sites to be restricted to the number of friends, followers, posts or shares, and the most

sophisticated studies are able to semantically assess published content. Twitter does not provide much data about the socio-demographic profile of its users, which imposes restrictions on the scope of the researches. Geographical location, religious affiliation, political preferences, gender, level of education and other variables related to social behaviour are data that are almost impossible to extract automatically from social media. Little is known about how these variables might affect the circulation of computational propaganda and how they might influence online public opinion. Qualitative and mixed methods are recommended in order to develop these sociological aspects, which can make studies slower and more costly.

Sub-section (D) summarises the articles that present models for predicting election results. However, it is necessary to assess the academic contributions and the social impacts of the predictive models. The prognosis of phenomena like elections, protests, referenda etc. can give rise to distortions and abusive interpretations, precipitated results and false-positives, or they can become self-fulfilling prophecies. Biased social predictions can distort the contextual perception needed for decision-making, essential to a healthy democratic process (Ascher 1982). We observed that there is greater concern about defending the assertive capacity of the prediction methods than there is about discussing the subjective and political consequences that the disseminated narratives may cause in the long term.

This SLR presents the current discussion on online political manipulation agents and identifies future research opportunities in this field, producing insights, comparing evidence and discussing methodological practices for empirical studies. Future investigations should build upon this existing knowledge, and, as Dacombe (2017) argues, the SLR could also prevent unnecessary replications in future works.

SLRs have become highly valued by policy makers (Dacombe 2017), offering a synthesis for legislators and civil society organizations to understand and act on a particular social issue. Given the current importance of social media for contemporary democracies, this SLR can substantiate debates on the regulation of these platforms. The broader discussion on the phenomenon of computational propaganda is as urgent as it is emergent. It is important to the academic community to share data and knowledge in the face of the ethical, social and political challenges, in order to preserve democracy in the era of Cyber Warfare.

References

Abril E. P. (2016), "Unmasking Trolls: Political Discussion on Twitter during the Parliamentary Election in Catalonia", *Trípodos*, 39: 53-69.

Ascher W. (1982), "Political forecasting: The missing link", *Journal of Forecasting*, 1(3): 227-239.

Barbrook R. (2009), *Futuros Imaginários: das Máquinas Pensantes à Aldeia Global*. São Paulo: Peiropolis.

Ben-David A., A. Matamoros-Fernandéz (2016), "Hate speech and covert discrimination on social media: Monitoring the Facebook pages of extreme-right political parties in Spain", *International Journal of Communication*, 10: 1167-1193.

Bessi A., E. Ferrara (2016), "Social bots distort the 2016 U.S. Presidential election online discussion" *First Monday*, 21(11).

Bolsover G., P. Howard (2017), "Computational Propaganda and Political Big Data: Moving Toward a More Critical Research Agenda", *Big Data*, 5 (4): 273-276.

Bradshaw S., P. N. Howard (2017), "Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation". In: S. Woolley and P. N. Howard (eds). Working Paper 2017.12. Oxford, UK: Project on Computational Propaganda. comprop.oii.ox.ac.uk<http://comprop.oii.ox.ac.uk/>. 37pp.

Cerón-Guzmán J. A., E. León (2015), "Detecting Social Spammers in Colombia 2014 Presidential Election". In: Lagunas O, Figueroa G. and O. Alcántara (eds.) *Advances in Artificial Intelligence and Its Applications*, 14th Mexican International Conference on Artificial Intelligence, MICAI 2015: Cuernavaca, Morelos, Mexico, October 25–31, 2015 Proceedings, Part II, Springer, Cham, pp. 121-141.

Dacombe R. (2017), "Systematic Reviews in Political Science: What Can the Approach Contribute to Political Research?", *Political Studies Review*, 16(2): 148-157.

Davies D., Jindal-Snape D., Collier C. Digby R., Hay P., A. Howe (2013), "Creative learning environments in education—A systematic literature review", *Thinking Skills and Creativity*, 8: 80-91.

Dayen D. (2017), "How twitter secretly benefits from bots and fake accounts". T*he Intercept* [online]. November 6th. Retrieved March 1st , 2018 (https://theintercept.com/2017/11/06/how-twitter-secretly-benefits-from-bots-and-fake-accounts/).

Facebook Newsroom (2017), *Company info*. Retrieved March 1st , 2018 (https://newsroom.fb.com/company-info/).

Ferrara E., Varol O., Davis C., Menczer F., A. Flammini (2016), "The rise of social bots", *Communications of the ACM*, 59 (7): 96-104.

Ferrara E., Varol O., Davis C., Menczer F., A. Flammini (2017), "Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election". *First Monday*, 22 (8).

Ford H., Dubois E., C. Puschmann (2016), "Keeping Ottawa honest - One tweet at a time? Politicians, Journalists, Wikipedians and Their Twitter Bots"**.** *International Journal of Communication,* 10: 4891–4914.

Gimenez C., E. M. Tachizawa (2012), "Extending sustainability to suppliers: a systematic literature review", *Supply Chain Management: An International Journal*, 17(5): 531-543.

Ibrahim M., Abdillah O., Wicaksono A. F., M. Adriani (2015), "Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Results in A Twitter Nation". In: *IEEE 15th International Conference on Data Mining Workshops*, Atlantic City, pp.1348-1353.

Jin S. A., J. Phua (2014), "Following Celebrities' Tweets About Brands: The Impact of Twitter-Based Electronic Word-of-Mouth on Consumers' Source Credibility Perception, Buying Intention, and Social Identification With Celebrities", *Journal of Advertising*, 43(2), 181–195

Knappenberger B. (2012), "We Are Legion: The Story of the Hacktivists". Documentary Film, 93 min. Retrieved March 1st , 2018 (https://www.youtube.com/watch?v=-zwDhoXpk90&t=99s).

Kim A. E., Hansen H. M., Murphy J., Richards A. K., Duke J., J. A. Allen (2013), "Methodological Considerations in Analyzing Twitter Data", *JNCI Monographs*, 2013 (47): 140–146.

Kitchenham B., Brereton P., Budgen D., Turner M., Bailey J., S. Linkman (2009), "Systematic literature reviews in software engineering—a systematic literature review", *Inf Softw Technol*, 51(1): 7–15.

Mager A. (2012), "Algorithmic Ideology", *Information, Communication & Society,* 15(5): 769–87.

Marechal N. (2016), "When Bots Tweet: Toward a Normative Framework for Bots on Social Networking Sites", *International Journal of Communication*, 10: 522-531.

Molina B. (2017), "Twitter overcounted active users since 2014, shares surge on profit hopes". *USA Today* [online]. Retrieved March 1st , 2018 (https://www.usatoday.com/story/tech/news/2017/10/26/twitter-overcounted-active-users-since-2014-shares-surge/801968001/).

Murthy D., Powell A. B., Tinati R., Anstead N., Carr L., S. Halford et al. (2016), "Bots and Political Influence: A Sociotechnical Investigation of Social Network Capital", *International Journal of Communication , 10*, 4952-4971.

Mustafaraj E., P.T. Metaxas (2017), "The fake news spreading plague: was it preventable?", *Proceedings of the 2017 ACM on Web Science Conference* (pp. 235-239). ACM.

Najafabadi M. K., M.N. Mahrin (2016), "A systematic literature review on the state of research and practice of collaborative filtering technique and implicit feedback", *Artificial Intelligence Review*, 45(2): 167-201.

Nye J. (2012), "China's Soft Power Deficit". *The Wall Street Journal* [online]. May 8. Retrieved March 1st , 2018 (https://www.wsj.com/articles/SB10001424052702304451104577389923098678842).

Oliveira E. C., França F., Goya D., C. Penteado (2016), "The Influence of Retweeting Robots During Brazilian Protests" In *49th Hawaii International Conference on System Sciences (HICSS),* Koloa, pp. 2068-2076.

Petticrew M., H. Roberts (2006), *Systematic reviews in the social sciences: a practical guide*. Hoboken: Blackwell Publishing.

Rains S. A., S. R. Brunner (2015), "What can we learn about social network sites by studying Facebook? A call and recommendations for research on social network sites", *New Media & Society,* 17(1): 114–131.

Recalde L., Mendieta J., Boratto L., Terán L., Vaca C., G. Baquerizo (2017), "Who You Should Not Follow: Extracting Word Embeddings from Tweets to Identify Groups of Interest and Hijackers in Demonstrations", *IEEE Transactions On Emerging Topics In Computing*, XX(XX).

Sadiq S., Yan Y., Taylor A., Shyu M. L., Chen S. C., D. Feaster (2017), "AAFA: Associative Affinity Factor Analysis for Bot Detection and Stance Classification in Twitter", *IEEE International Conference on Information Reuse and Integration (IRI)*, San Diego, pp.356-265.

Sanger D. E., (2018), *The Perfect Weapon: War, Sabotage, and Fear in the Cyber Age*. New York: Crown Publishers.

Shane S., M. Mazzetti (2018), "Inside a 3-Year Russian Campaign to Influence U.S. Voters", *The New York Times* [online]. February 16th 2018. Retrieved July 4th 2018. (https://www.nytimes.com/2018/02/16/us/politics/russia-mueller-election.html).

Stieglitz S., Brachten F., Berthelé D., Schlaus M., Venetopoulou C., D. Veutgen (2017), "Do Social Bots (Still) Act Different to Humans? – Comparing Metrics of Social Bots with Those of Humans", In: Meiselwitz G. (ed), *Social Computing and Social Media. Human Behavior*. SCSM 2017. Lecture Notes in Computer Science (LNCS), vol 10282, Springer, Cham, pp. 379–395.

The Economist (2017), *Social Media's threat to democracy*. November, 4th-10th, 2017, pp. 20.

Uricchio W. (2011), "The algorithmic turn: Photosynth, augmented reality and the changing implications of the image". *Visual Studies*, 26(1): 25–35.

Véras D., Prota T., Bispo A., Prudêncio R., C. Ferraz (2015), "A literature review of recommender systems in the television domain", *Expert Systems With Applications*, 42 (22): 9046–9076.

Woolley S. C., P. N. Howard (2016), "Automating power: Social bot interference in global politics", *First Monday*, 21 (4).

Woolley S. C., P. N. Howard (2017), "Computational Propaganda Worldwide: Executive Summary" In: "S. Woolley and P. N. Howard (eds.) Working Paper 2017.11. Oxford, UK: Project on Computational Propaganda. comprop.oii.ox.ac.uk. 14 pp.

Woolley S. C., P. N. Howard (2018), "The Political Economy of Bots: Theory and Method in the Study of Social Automation" In: R. Kiggins (ed.), *The Political Economy of Robots: Prospects for Prosperity and Peace in the Automated 21st Century*, International Political Economy Series. Cham, Switzerland: Palgrave Macmillan, pp.127-155.

Zeifman I. (2017), "Bot Trafic Report 2016: Bots & DDOS, Performance, Security". Imperva Incapsula. January 24, 2017. Retrieved March 1st , 2018 (https://www.incapsula.com/blog/bot-traffic-report-2016.html).

Appendix – Criteria for quality assessment

Question	*Answer*
Are the objectives of the study clearly indicated?	Yes: The aim is described clearly and explicitly. Partially: The essential purpose of the research is not mentioned clearly. No: No phrase is mentioned about the aim of the research.
Are the methods used in the study adequate to the purpose and well executed?	Yes: The search method is adequate and well executed. Partly: The method is appropriate or well executed. No: The method is inadequate and poorly executed.
Are the conclusions or expected results relevant and achieved?	Yes: The article fulfilled the objective and presented relevant results. Partially: The article has achieved the expected results or has relevant results for the theme. No: Expected results are irrelevant and have not been achieved.
Are the limitations of the work clearly documented?	Yes: The text clearly explains the limitations of the study. Partially: The article mentions the limitations, but does not explain the reasons why they exist. No: The limitations of the study are not mentioned.

Authors' information

Rose Marie Santini is a Professor at the School of Communication of the Federal University of Rio de Janeiro (UFRJ), as well as Professor of the Graduate Program in Information Science of UFRJ and of the Graduate Program in Communication Technologies and Languages of UFRJ. Founding member and Research Director of NetLab / UFRJ. Her research interests involve algorithmic culture and sociology applied to network studies.

Larissa Agostini graduated in Social Communication at the Federal University of Rio de Janeiro (UFRJ) and is interested in microssociology, network studies and politics. Larissa is currently a fellow researcher of NetLab.

Carlos Eduardo Barros is pursuing a graduation in Social Communication at the Federal University of Rio de Janeiro (UFRJ). His research interests involve the issues about freedom of communication and collective action on digital media era. Carlos is currently a fellow researcher of NetLab.

Danilo Carvalho studies Social Communication at the Federal University of Rio de Janeiro (UFRJ) and is a developer at Fundação Getúlio Vargas (FGV). His research interests include microsociology, social media and politics, and computational social science. Danilo is currently a fellow researcher of NetLab.

Rafael Rezende graduated in Social Communication at the Federal University of Rio de Janeiro (UFRJ) and his research interest involve the relation between social movements and social networks. Rafael is currently a fellow researcher of NetLab.

Débora G. Salles is a PhD student in the Information Sciences Graduate Program at the UFRJ, acting as a fellow researcher in the NetLab research group. She holds a master's degree of Communication and Culture also from UFRJ. Her research focuses on digital media, cultural consumption, social networks and collaborative production.

Kenzo Seto graduated in Social Communication at the Federal University of Rio de Janeiro (UFRJ) and is pursuing a master degree in Communication also at UFRJ. His research interests involve digital populism and the political impacts of algorithms and social media. Kenzo is currently a fellow researcher of NetLab.

Camyla Terra graduated in Social Communication at the Federal University of Rio de Janeiro (UFRJ) and is pursuing a master degree in Information Science also at UFRJ. Camyla is currently a fellow researcher of NetLab. Her research interests involve sociology applied to politics, digital media and cultural studies.

Giulia Tucci is a Chemical Engineer, graduated at PUC-Rio and holds a Master degree in Biomedical Engineering at UFRJ. Her research interests involve coding and politics studies. Giulia is currently a fellow researcher of NetLab.