

## **EPISTEMOLOGY, METHODS AND THEORIES OF COMMUNICATION IN THE BIG DATA ERA: A CRITICAL PANORAMA OF SOCIAL MEDIA RESEARCH**

**Ana Thereza Nogueira Soares**

---

### **ABSTRACT**

This paper proposes a critical reflection on the epistemological, methodological and theoretical implications of the researches based on Big Data – especially on social media data – for the scientific field of communication. From an epistemological point of view, it reveals the unsustainability of analytical models based on static frameworks of communication, claiming that the social processes that emerge with the influence of the internet are unequivocally presented in fluid and contingent formats. In this context, it highlights that the evolution of technology itself has the potential to boost the construction of data collection and analysis tools capable of grasping the communication movements, justifying the need for alignment between ontology, epistemology and methodology in scientific research. The text, also, poses questions about communication theory and its concepts. It is believed that the relevance acquired by data in recent years should not point to a domain of the empirical over the theoretical. Effectively, the strengthening of the communication science demands precision and care with the use of terms, models and theoretical references historically consolidated in the problematization and explanation of the contemporary.

### **KEYWORDS**

Big Data; communication theory; epistemology; social media

---

### **RESUMO**

O artigo propõe uma reflexão crítica sobre as implicações epistemológicas, metodológicas e teórico-conceituais das pesquisas baseadas no Big Data – especialmente em dados de mídias sociais – para o campo científico da comunicação. Do ponto de vista epistemológico, revela a insustentabilidade de modelos de análise baseados em enquadramentos estáticos da comunicação, alegando que os processos sociais que emergem com a influência da internet apresentam-se, inequivocamente, em formatos fluidos e contingentes. Nesse âmbito, destaca que a própria evolução da tecnologia tem condições de impulsionar a construção de ferramentas de coleta e análise de dados capazes de apreender os movimentos comunicacionais, fazendo jus à necessidade de alinhamento entre ontologia, epistemologia e metodologia na pesquisa científica. O texto, ainda, problematiza questões relativas à construção teórico-conceitual da comunicação. Acredita-se que o relevo adquirido pelos dados, nos últimos anos, não deve apontar para um domínio do empírico sobre o teórico. Efetivamente, o fortalecimento da ciência da comunicação passa pela precisão e pelo cuidado com o uso de termos, modelos e referências teóricas historicamente consolidados na problematização e explicação do contemporâneo.

### **PALAVRAS-CHAVE**

Big Data; epistemologia; mídias sociais; teoria da comunicação

---

## INTRODUCTION

Can Big Data, upon the disruptive potential it presents to the organization forms of both society and science, generate inputs to the re-discussion of communication concepts and theories? This seemingly rhetorical question may be essential for a reflection on the future of communication science. About a century after the first theoretical elaborations on the communicational phenomenon and its effects on the psychic and social configurations of individuals in mass society, the ubiquity of messages and media, materialized by the appropriation we make of technology, generates intense debates that reaches all spheres of society.

Cukier and Myer-Schonberger (2013), Boyd and Crawford (2012) Coté, Gerbaudo and Pybus (2016) and Gil de Zúñiga (2015) are just some of the authors dedicated to the reflection on the strength of Big Data and algorithms in the conformation and reproduction of sociocultural practices. According to Kitchin (2017), there is increasing pressure to analyze critically how algorithms have interfered in the production of scientific knowledge. It is difficult, in fact, to identify a field that is not directly or indirectly affected by the speed, variety and volume of data production and its conversion into information with high diffusional power. From science to public health, from the banking system to the production of news: the amount of data spontaneously or compulsorily produced and then recorded by society grows in an unimagined fashion, generating new political regimes that have the potential to challenge existing social conventions. Something quite extraordinary is printed in social reality when, for example, a company like Google becomes a trustworthy base of knowledge to prevent and control a flu epidemic that had the potential to spread itself around the world (Cukier & Mayer-Schönberger, 2013; Helbing, 2015).

Thus, the objective of this text is to point out, in a preliminary way, critical points and paths for the communication researcher's practice today, given the vast possibilities of data collection and analysis enabled by internet platforms, tools and applications. The focus is on the impact of Big Data on epistemologies, theories and methods on which the researches that translate the reality of communication are based, through data coming from the diverse media and digital applications available to society today.

## **BIG DATA: CIRCUMSCRIBING THE MEANINGS OF THE PHENOMENON TO SCIENCE**

The exponential growth of research and articles that involve the measurement of social life through digital environments, such as social media platforms and open government data systems, reveals the pressing need for debate on epistemological and theoretical-methodological issues that data has helped to forge.

Coté, Gerbaudo and Pybus (2016), for example, place Big Data on the political spectrum that guides the manipulation of algorithmic data by social agents. From the more general question, about which power relations influence the knowledge generated by the analysis of digital data, the authors reflect critically on the meanings of Big Data appropriation by scientists. To them, there is no convincing arguments that favors an exempt knowledge and universal validities, guaranteed by the supposed neutrality and technical

objectivity of massive data collection from the internet. The Silicon Valley market power and its preponderant role in this new economic chain of production, the structural inequalities between data generation by citizens and their control by corporations, as well as the algorithm opacity that decides network configurations and user perceptions (Coté, Gerbaudo & Pybus, 2016) are just some of the signs that we have already entered a new interface in society's relationship with technology.

There is in the spirit of our time an urgency in understanding how the dyad of algorithms and Big Data can enable new social and cultural forms. One must understand how they are employed and develop a critical sense regarding their limits, capabilities, implications and possibilities. These technological arrangements provide a wide range of possibilities for knowledge, however, they also need to be understood facing the human aspects that guide formal procedures for pattern recognition, set programming parameters and generate data that eventually entails semantic gaps and prejudices (Uricchio, 2017; Kitchin, 2017).

At the same time, the domination mechanisms exercised by the capitalist monetization dynamics of digital data have been constantly stressed by the social agents themselves. That is, the uses and appropriations of data by economic and political spheres, because they have acquired enormous dimensions and influence, also become inputs to the constitution of knowledge, in the inevitable reflexive movement that governs individual and collective consciousness. Milan and van der Velden (2016) propose the concept of "data activism", the glimpse of a knowledge that escapes the reification of the future, based on a critical development of science and technology, inscribed by the researchers themselves. In this context, the increasing availability of data is seen as a powerful and unprecedented opportunity to bring about social change. "Data activism" supports the emergence of new epistemic cultures that "challenge the mainstream readings of reality" and "shape the way we relate to knowledge and its validation" (Milan & van der Velden, 2016, p. 63).

From the point of view of research on political communication in social media or digital social media platforms, Gil de Zúñiga (2015) also introduces a set of works that offer unique ways of combining methods and techniques related to Big Data that, in addition, take into account the ethical component of personal data use in the shaping of research processes. The author acknowledges, however, that knowing how to interpret the immensity of data provided by our digital footprints "remains an enormous challenge" (Gil de Zúñiga, 2015, p. 2). In our present context, are at stake notions such as the validity of scientific constructs, tests of hypotheses and generalization, that is, some of the paradigmatic canons of modern science. The mentioned authors are concerned with issues such as the elimination of theories as starting point for knowledge production, the loss of explanatory power of the structural dimension of society (Coté, Gerbaudo & Pybus, 2016), as well as models of data collection potentially capable of gathering endless digital databases (Gil de Zúñiga, 2015).

The reversal of methodological procedures that place collections and correlations of extensive data upon the social as maximized and improved expressions of knowledge is taken as an indicator of a macro-scientific revolution, in the sense Kuhn (1998) implied

to the term. Kitchin (2014) explores the extent to which Big Data leads to alternative epistemologies in the social sciences, based on a critical assessment of occurring changes in research practices. The author challenges the idea that emerges in this context of a purely inductive *modus operandi* in research, arguing that, rather than fostering empiricism over theoretical construction, data-driven science is a reconfigured version of the traditional scientific method. For this, however, data must be collected and interpreted in the light of relevant issues to existing knowledge. The disaggregated sets of “photographs”, relating to specific circumstances and contexts of the digital reality, illustrating some kind of correlation on social traces, arranged on the internet and visualized by the researcher, do not necessarily answer theoretical-scientific questions.

The seductive aspect of Big Data-based research also asserts the construction of a mythology, according to the term used by Boyd and Crawford, which is based on the “widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights there were previously impossible, with the aura of truth, objectivity and accuracy “ (2012, p. 663). Such a symbolic dimension of data acts as a reinforcement to the illusion of informational comprehensiveness, representativeness and trustworthiness and tend to be held as ontological foundation, guiding and shaping scientific knowledge and its public diffusion. There is a guarantee of verisimilitude that underlies this new epistemology of highly empiricist bias, grounded in a profusion of exploratory dives into dataset mining on the internet. In this movement, however, Kitchin (2014, p. 5) warns that “data are examined through a particular lens that influence how they are interpreted” and, furthermore, that “correlations between variables within a data set can be random in nature and have no or little causal association”.

The naturalization of ecological fallacies, pointed out by Kitchin (2014) as a potential risk to the epistemology of science that emerges within Big Data, is not an isolated issue for reflection. Among other observed challenges to traditional scientific field practices is the excessive individualization of search and data collection methods. This leads, in the extreme, to the incommensurability of the various analyzes and, therefore, to the impediment of critical procedures that form the basis upon which theories are tested and improved.

The autonomy given to researchers regarding the elaboration of their own tools for tracking, mining, collecting and analyzing data has the potential to become, paradoxically, what depletes modern science in one of its main forces – the possibility to evolve in truth searching through the comparison of results of investigative tests and hypothesis refutation. Because it emphasizes the empirical disposition on digital platforms and establishes it, a priori, as the truth to be unveiled, the customization of Big Data-based research individualizes scientific practice, transforming multiple reality frameworks into undisputed knowledge. The research results, particularized by algorithmic instruments that collect and systematize data, tend thus to lose the explanatory potential that is the hallmark of generalization and theory formulation.

Today, numerous laboratories and research groups around the world, dedicated to internet research and interested in exploring digitally stored data, have their own models

and digital data collection tools<sup>1</sup>. There is a profusion of algorithms created for selecting and combining social media data, and not all are open source. The growing interest in research on social networking on the internet (which has gained prominence in the field of communication's studies over the last decade) is reflected on the wide range of tools and software available exclusively for the treatment of this type of data, each with its functionality, its path of algorithmic coding and aesthetic formulation<sup>2</sup>.

Although this diversity of approaches represents an overcoming of models and techniques that limit the understanding of social reality, precisely because they are constructed from a perspective of control and fixity of that reality, in its static interpretative frameworks (and therefore incongruent with movement inherent to the communication), one must look for some characteristics that distinguish these research tools. The construction of technical devices for data capture and analysis is done by means of algorithms, whose heterogeneous and opaque form, possibly programmed to perform the function of artificial intelligence, has the potential to extend the beaconing intentions of its design. As Kitchin stresses it, algorithms work in multivariate ways, “need to be recognized as being ontogenetic, performative, and contingent: that is, they are never fixed in nature, but are emergent” (2017, p. 21). “Many algorithms are designed to be reactive and mutable to inputs” (Kitchin, 2017, p. 21), continually unfolding, through their artificial intelligence, new data and information to be interpreted.

Little is known, however, about the effects to science on Big Data manipulation, let alone our multiple forms of contemporary communication and the metadata mined from them by communication scientists. Because the algorithmic coding that boosts Big Data tends to be obscured by the aesthetics of their applications, software, and other technical devices, it is laborious to uncover the original intents that lead to the diffusion and sharing of data by various agents. Such directions depend after all on reverse translations of previously programmed codes. Kitchin (2017) also refers to the ‘translation issue’ in his critical review of algorithms. The programming of digital cyber spaces depends on two translations, mainly:

the first is translating a task or problem into a structured formula with an appropriate rule set (pseudo-code). The second is translating this pseudo-code into source code that when compiled will perform the task or solve the problem. Both translations can be challenging, requiring the precise

<sup>1</sup> Among the many research centers operating today are The Citizen Lab, a laboratory at Munk School of Global Affairs, University of Toronto; The Social Media Lab of Ryerson University, also in Toronto, the Center for Digital Culture in Mexico; the Nordic Center for Internet and Society, University of Siegen, Germany; the Singapore Internet Research Center and Media Lab Research of the Massachusetts Institute of Technology (MIT) in the United States. In Brazil, counting only research centers focused on social media data, we can cite at least three, the Laboratory of Image and Cyberculture Studies (Labic), Universidade Federal do Espírito Santo; the Interdisciplinary Research Group on Social Network Analysis, Universidade Federal de Minas Gerais and the Laboratory of Media, Speech and Analysis of Social Networks (MIDIARS), Pelotas Catholic University. This list is not exhaustive, however.

<sup>2</sup> By way of illustration, the following tools, among many others, make up the list of alternatives for collecting, processing and analyzing internet data: NodeXL, iGraph, Netlytic, Uberlink, Issue Crawler, Gephi, Cosmos, Chorus, VT Tracker and Netvizz, the latter is specific for Facebook data collection. The resources and applications allowed are numerous, generating the most diverse analytical specificities.

definition of what a task/problem is (logic), then breaking that down into a precise set of instructions, factoring in any contingencies such as how the algorithm should perform under different conditions (control). (Kitchin, 2017, p. 17)

Although it seems as an automatic construction process, effectively controllable, the challenge posed by algorithms and the profusion of data they drive is precisely that society could effectively commune with the translations carried out by programmers. This possible communication between the meanings of algorithms and the implications they bring to the organizational dynamics of society is one of the most underestimated needs of contemporary politics.

Given the presented panorama, the field of communication's research faces at least two broad categories of issues regarding its current and future developments. The first refers to a necessary alignment between ontology and research methodology, in the sense that Hall (2003) employed to discuss the foundations of comparative politics research. It is a question of reflecting on what seems to be the emergence of an epistemological orientation in communications sciences, whose maxim, despite having its focus on the technical development of the media, does not necessarily contemplate its implicit presuppositions and implications of use to the theoretical-methodological organization of the researches.

The second category, inevitably associated with the first, encompasses the fear of theory decline, favoring the datafication of reality in scientific research, as described above. In this context, it is pertinent to rethink the tendency towards the indiscriminate use of theoretical conceptions and concepts displaced from their original production circumstances, which describe and anachronistically design data sets and their communicational meanings. The argument is that this is one of the guarantees for the evolution of scientific knowledge in communication.

The experiences of social immersion in the internet have shown that we are facing a new context, with new demands and processes, permeated at all levels by the production, diffusion and seizure of data, in short by communications between these spheres. Web 3.0 configures itself as a kind of stage for these connections between data, a new dimension of space in which relationships happen in a diverse, dispersed and multiple way (Amaral, 2012). Arguably, there are not many points in common between communication as undertaken in the field of mass communications and that which emerges on the digital pathways. Given this perspective, it would be erroneous to interpret the latter reality based on theoretical-conceptual artifacts constituted to explain another communicational pattern of society.

About content generated by social agents in digital networks, for example, it is argued that "their potential classification, even with the formal variety and diversity of the platforms, goes beyond standardized categorizations" (Amaral, 2012, p. 142). If Big Data and the metadata that allows its configuration are the new support of social ties, as Amaral (2012) also argues, it is necessary to move swiftly in the comprehension about its semantics, about the senses imbedded in its conception and evolution.



This concern also refers to that one expressed by Shaldbolt, Hall and Berners-Lee (2006), risen from the increasing need to integrate digital data into common denominators of meaning. The semantic web advocated by the authors is perceived beyond the so-called “folksonomies”, languages shared by a considerable number of people interested in replicating certain data through information, to reverberate it.

In information science, this organization of the information flow and shared construction of web semantics is characterized by the decentralization of access, and the ease through which data are spread through social media. The icons of these “folksonomies”, however, are several. From hashtags to memes, various forms of aesthetic representation can represent the relational dynamics of data on the internet. However, it remains to be seen which are the similarities and parallels, the most appropriate comparisons to be made between these semantic sets, which enables their indexing to ontologies, as defined by Shaldbolt, Hall and Berners-Lee (2006). Ontologies, in this sense, are capable of removing the ambiguities between the countless paths through which data travel and are diffused in the digital space. These are more perennial semantics, upon which taxonomies are based, enabling the comprehension of the social in the context of the internet.

Through these intricacies, Big Data seems to be a promising phenomenon to understand the reality of contemporary social communication that, in addition, activates communication and its informational processes as a relevant theoretical and interpretative axis for the human and social sciences. The emergence of data as a dominant methodological paradigm reveals the convergence between the hard and soft sciences. In this scope of research that confronts the scientific practices, there is no way to ignore communication as a vital explanatory component of the connections enabled today by the interrelations between individuals and data. However, it is necessary to inquire how the communicational phenomenon is conceived and diffused by the researchers in their constructions on what is outlined in and by the digital media and apparatuses.

#### **ONTOLOGIES OF COMMUNICATION AND METHODOLOGICAL RESEARCH DESIGNS – SOME REMARKS**

What communicational reality are the researchers talking about? What questions lead scientific investigations that are based on internet data?<sup>3</sup> Which ontology supports the methodological choices of the researcher? The argument here is that the angle from which the works based on Big Data are oriented helps to configure the methodological and theoretical paths of the investigations. For example, the prevalence of a communication’s view as a static phenomenon, composed by fixed portraits of reality, located in predetermined and controlled plans of direct and impartial observation, induces a methodological research design that favors linear mechanisms of information circulation.

<sup>3</sup> It is important to reiterate that data from social networks platforms, or social media (according to the concept popularized in English), are just one kind of Big Data. Its growth as research topic has not been the sole task of communication scientists, but also of political scientists, anthropologists, psychologists, geographers, computer scientists, and academics from various fields of knowledge.

Similarly to what Hall (2003) has observed in international comparative policy research, we can say that in the field of communication, too, the value of the methods employed depends on its congruence with the way in which the researcher understands the causalities in the processes that are taken for analyzes. Thus, the capture of digital data at a given moment of time and in a specific context of performance represents exactly what, within the universe of possible communicational explanations? If the perspective of a linear reality in communication and the control of its results is what prevails ontologically in research, Big Data will have little to offer in terms of knowledge construction. This way to represent reality, as composed of fixed entities and organized a priori in linear forms of causes and consequences, is limited in its explanatory capacity. By denoting a way of thinking about how social reality works, such a world view incorporates linear models as representative patterns of reality (Abbott, 1998). Thus, in this perspective, epistemological advances will hardly be possible in terms of expanding the explanatory possibilities of the communicational phenomenon. Research that sees communication as a punctual phenomenon, situated in time, and more, sees its purpose as instrumental and controllable by the intentions of the issuer, creates a kind of paradox in the way they interpret digital data.

On internet, data and information seem to be able to be identified as units of a social complexity cut out by the most diverse variables and is therefore multilinear. Insofar as they are translated through textual, graphic, imaginary, auditory expressions and also through structures and architectures of relationship between nodes, the data collected in digital environments allow us to glimpse - and even represent - the power of human communication.

It is not a case of arguing that what happens on the internet in terms of social action and expression is fundamentally different from the processes already experienced in the “analog society”. However, if we take the digital dimension of social existence as a starting point to explain it, there must be a careful and adequate adjustment between our interpretative inventories, hitherto accumulated as a scientific field, and this reflexive and constant reality, woven by individuals and organizations. The argument here is, therefore, toward the impossibility in sustaining the same conceptions about a communicational reality that prevailed in the twentieth century at this moment in history, in which the internet composes a significant part of the functioning of social life.

An epistemology based on complexity is much more coherent with the dynamics of social media, based on the nonlinear, multi-causal, and perpetual movements observed through the expression of tensions and connections in dispute. Relationship networks, such as Facebook and LinkedIn, combine with information platforms such as Twitter. Social networking blogs, such as Medium, have replaced the role of traditional websites as place for public and identitarian presentations by individuals and organizations, fostering the emergence and construction of a new media ecology, resulting from the growing proliferation of applications as identified by Amaral (2012) in her analysis on Web 3.0. Audiovisual content gains through Vimeo and YouTube meanings that vary depending on the perceptions and appropriations that the users make of them. Social media makes



more evident, in short, a whole range of variables that challenge the processes of communication, denoting more clearly the insufficiency of world representations imprisoned exclusively in punctual and self-centered manifestations of informative broadcasts. In case of platforms and mobile applications such as Snapchat, or in features in the type of Instagram and WhatsApp's stories, messages become effectively a literal representation of everyday life, in streamings edited and customized according to production contexts suitable for the issuer's world vision, as they continually transform by relationships they establish with external partners.

Circumscribing social media as an object of communications studies becomes, therefore, the great challenge of refining the available methods and tools in order to assure the possibility to apprehend the most varied dynamics of expression, the appropriations of technologies and their meanings by emitters and receptors, and also the communication improbabilities (Luhmann, 2006) that permeate digital reality. In this sense, initiatives such as that of Procter, Vis and Voss (2013) are inspiring, since they propose methods capable of incorporating in a more explicit way the interpretation and representation of virtual reality, the movement inherent to the disputes and confluences of meaning in the processes of social communication. The authors succeeded to elaborate a collection tool of Twitter posts (tweets and retweets), in which the algorithmic codification is capable to track the dynamics of disturbances, rumors and gossips in the platform. Moreover, the representation of the analytical results, in a timeline of messages emitted and passed on in Twitter, addresses the fluidity of information sharing actions (which comprise a portion of Big Data disposed in social media) with its capacity of diffusion and, in short, the disputes about its truthfulness and meanings.

In this way, communication scientists can find a privileged space for the application of innovative technological tools, capable of apprehending the paths of the emitted messages and especially their appropriation and resignification by the receiving instances. The 'interactive media critique devices' analyzed by Braga (2006) may finally, according to the methodological scopes, tools and techniques of data collection and analysis, improved with Big Data, become visible in their different layers of collective construction.

According to Martino (2009), the media news, constituted as revealing objects of communication or themes of public interest, crosses, with speed and sometimes unexpected directions, the daily life of the internet. To understand the mechanisms by which the news emerges in digital environments, whether as an expression of individual and organizational volitions, or as spontaneous events, intrinsic to the fluid dynamics of social relations, becomes a relevant research goal in the field of communication. The continued technological enhancement of digital data mining and analyses has the potential to create mapping instruments that cover more complete media communication frames, rendering their trajectories and spanning information emissions with their subsequent directions through social media. The paths opened and traced by the messages themselves would represent, in this sense, a digital architecture of the communicational space-time. Thus, the messages, with all their interpretive nuances, naturally emerging and changeable, would complete the complex analytical framework of communication in social media.

In this context, communication's research can finally overcome its instrumental bias, insofar as it incorporates in its descriptions and analyzes not only the intentionality of the emission, but also the vanishing points that escape attempts to regulate its meanings, as analyzed by Fausto Neto (2011). The instrumentalisation of media – which shaped the orthodoxy in the field of communication and still dominates its spheres of knowledge production directed by or for market and state institutions – is not congruent with the spontaneous processes of communication clearly observable throughout the internet.

### **THEORIES AND CONCEPTS OF COMMUNICATION FROM A SOCIAL MEDIA ANGLE**

In this section of the text, the idea of Kitchin (2014) is retaken, for whom the paradigmatic revolution provided by Big Data cannot justify the absence of theoretical reflection. On the contrary, it is argued that the rise of a datafied society generates a propitious moment for the deepening and re-discussion about the validity of concepts and their epistemic assumptions. Problems and research questions grounded in theories should thus prevail as compasses guiding the empirical gaze, and not be discarded in the name of sovereigns “data that speak for themselves”.

The continuous revisional exercise and debate about the theoretical-conceptual basis of the communication field is a guarantee of scientific knowledge validity, after all. It is assumed that this depends on the interpretational accuracy of the data, expressing feelings, worldviews, assertions, deliberate attempts at persuasion, as well as individual and collective actions. Ultimately, social media and the information they have on reality relate to a wide range of social expressions.

Thus, it is appropriate and urgent to discuss how classical conceptual and theoretical frameworks of communication can contribute to thinking about research issues involving Big Data, or how, in an inverse relation, the same theoretical repertoires can be rethought from social reality presented online. One of the starting points for reflection, according to previous arguments, is the very volatile, ephemeral and contingent nature of data production that permeates research endeavors. The numerous platforms and devices in which social agents can project their voices, images and concerted actions are the base of the data we manipulate. They are crossed and constructed by most diverse motivations and interests, that appear and disappear, as analyzable information, in uninterrupted and interconnected movements of individual and organizational expressions.

As outlined in the previous section, a most diverse array of social media compete for the attention of users, in a universe full of communicational possibilities, each with its materiality and symbolism. This scenario shows that lifestyles, priority interests in social communication, spatial location, economic conditions, among other factors, make up the list of variables that influence the dynamics of digital society. Cultural and national contexts can, for example, induce the use of one platform to the detriment of another. Twitter is a very interesting case in this sense. While in countries like the United States the platform is regarded as a thermometer of politics in the public sphere, in Brazil its

impact is not so comprehensive. On the other hand, studying the theme of contemporary politics in Brazil without including at least one mention of the prominent role of Facebook as a platform for public debate will fall short in giving the complete picture.

Therefore, in this multifaceted context, working in a tight manner with the concepts of emitter, medium, message and receiver, among others, disregarding the inherent movements of meaningful connections in digital media, seems to be counterproductive. A conceptual appropriation that shows well the dilemmas faced by theory is the idea of broadcasting, today applied to understand data generated in a digital, centralized and unidirectional informational network. It is believed that such a concept gives little value to the understanding of internet communication, since the data it normally designates can not necessarily reveal the seizure of information by an alleged audience, much less what is done out of this information. The notion of 'audience' as referred to groups of followers or friends, similar to a broadcasting type of emission, also put the receiving instance in a condition of passivity that goes back to theories such as the hypodermic needle. But can we really interpret the internet's configurations based on conceptual mechanisms that go back to the historical context of 20th century communication, and thus use primal concepts of the field to explain current reality?

It is appropriate to recall that from information, in its pure state, one cannot deduce communication, whose existence is tied to shared connections of meanings, mediated by technique and / or by individuals. Followers of a Facebook page should not be equated, therefore, with audiences of mass society, although the performance of companies such as Cambridge Analytica, based on psychometrics, takes on a great deal of importance in relation to traditional mass media, when it comes to the process of achieving attention and persuasion of individuals through personal data. As noted above, a linear, mechanical and emission-centered view of communication does not seem compatible with fugitive and impermanent processes that are understandable out of Big Data.

Thus, it is argued that the predominant theoretical paradigm for conducting social media research in this complex scenario that links the traditional contexts of mass communication to the digital contexts of reality production is that proposed by Luhmann (1995) (Figure 1). Communication, from this perspective, is only complete when issued information [Mitteilung] is systematically selected and understood [Verstehen], being able henceforth to generate new information and, potentially, new communication cycles.

It is precisely this idea of communication as a central element of social coordination – as unity materialized by uninterrupted cycles of production and deconstruction of meanings – that is best paired with the instantaneous exchange of data and information in social media or with the concomitance of emission and reception spheres, if we stay at traditional concepts of the field.

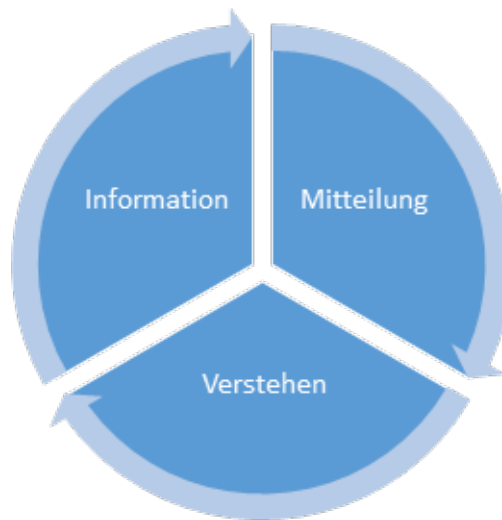


Figure 1: Communicational unity in the theory of Niklas Luhmann (1995)

Sbardelotto (2017), by other theoretical paths, presents a convergent idea to the Luhmannian perspective when he advocates the production-reception commutability in the processes of network circulation and “that such poles exist only reciprocally, constituting themselves mutually through the dynamics of circulation” (p .4). There would be no empirical support in this context for the reiteration of theoretical communication models in which instances of production and assimilation of information are isolated from each other. “In the circulatory flow of communication networks, specifically, the poles of production and reception do not disappear, but it is no longer possible to fix them in a specific social subject, such as media corporations or traditional social institutions” (Sbardelotto, 2017, p.19). The reality of digital life today is that it is possible for an agent to perform transmitter and receiver functions simultaneously. In addition, this agent can be false, artificially controlled by algorithms, or be representative of an organization, an ideological collective, etc. There are many facets of who produces and consumes information.

Without to exhaust the reflection on how communication theories and concepts can guide research in the context of increasing complexity revealed by the internet, it is worth a final reflection on the conceptual meaning of messages. Are we talking about messages in their conventional sense when we opt for research questions that focus on hashtags, conversations, shares, likes, photographs, audiovisual records or retweets in social media? This issue resumes the above argument about semantic web ontologies (Shaldbolt, Hall & Berners-Lee, 2006). Without determining the parameters on which digital spaces are constructed, it is difficult to understand how they can effectively favor social communication. The hypothesis is that the choice of the communication platform in which a message is expressed predisposes specific types of content and aesthetics, also consistent with its social use. This has consequences not only when formulating broader categorizations about the meanings of messages and their contents, along with their coupling to media (applications would now be only one of the media types), as well

as the articulation between this variable that “carries the meaning” and the social dimensions of its circulation.

It is fundamental to note, moreover, that meanings collectively constructed around messages can be interpreted in the light of different angles, starting from the original messages that reach the world, to the type of digital profile that diffused it initially and arriving at the configuration of the social network that sustained it for a while in evidence. These choices undoubtedly assume epistemological implications of more individualistic or holistic views of internet reality.

### CONCLUDING REMARKS

Last but not least, there is a reflexive note about the role of artificial intelligence, as well as the presence of so-called bots, in the production and circulation of information, and therefore in the communicational power of digital life. The issue emerges mainly through the evidence of influence that these systems have on political-electoral processes (Bessi & Ferrara, 2016), but whose implications have not yet been sufficiently debated on the scientific arena. Following Dusek’s (2006) philosophical postulate, according to which all technological means of communication shape social experiences in some measure, it is here suggested the necessity to radically rethink the classic view of an exclusivity of human agency in communicational processes. An objectified view of the technology has to be questioned, based on the understanding that bots have acquired, in current contexts, capacity to express and to process information, since they are constituted as algorithms. As suggested above, out of the need to translate algorithmic codes for different contingencies, “developments in machine learning have enabled algorithms to self-optimize and generate their own improvements. They can now self-author and self-create” (Uricchio, 2017, p. 127). It is worth stressing that authorship and agency have become relative concepts. The heterodoxy of a theoretical conception that attributes to technique a potentially active role in the communicational processes is necessary, therefore, in the transcendent movement of scientific knowledge marked by the digital social life, but the validity of this assumption can only be measured in the course of procedures in collecting, systematization and analysis of Big Data. The analytical work on digital traces has hardly begun. If data science wants to live up to be the basic status of democracy, as Berners-Lee advocates in a recent interview (Baker, 2018), it will have to run preferably in the speed of the internet. ✍

Translated by Leon Rabelo

### BIBLIOGRAPHIC REFERENCES

- Abbott, A. (1998). Transcending general linear reality. *Sociological Theory*, 6, 169-186.
- Amaral, I. (2012). Participação em rede: do utilizador ao “consumer 2.0” e ao “prosumer”. *Comunicação e Sociedade*, 22, 131-147.

- Baker, D. (2018). Is the internet broken? WIRED asks Tim Berners-Lee, Jimmy Wales, James Ball, Wendy Hall and more about how we could – and should – reset the net. Start again. *Wired*. Jan., UK Edition, 120-129.
- Bessi, A. & Ferrara, E. (2016). Social bots distort the 2016 Presidential election online discussion. *First Monday*, 21(11).
- Boyd, D. & Crawford, K. (2012). Critical questions for Big Data. *Information, Communication & Society*, 15(5), 662-679. doi: 10.1080/1369118X.2012.678878
- Braga, J. L. (2006). *A sociedade enfrenta sua mídia. Dispositivos sociais de crítica midiática*. São Paulo: Paulus.
- Coté, M., Gerbaudo, P. & Pybus, J. (2016). Introduction: politics of big data. *Digital Culture & Society*, 2(2), 5-15. doi: 10.14361/dcs-2016-0202
- Cukier, K. & Mayer-Schönberger, V. (2013). *Big Data: a revolution that will transform how we live, work and think*. New York: Houghton Mifflin Harcourt.
- Dusek, V. (2006). *Philosophy of technology: an introduction*. Malden, MA: Blackwell Publishing.
- Fausto Neto, A. (2011). Comunicação das organizações: da vigilância aos pontos de fuga. In I. Oliveira & A. Soares (Eds.), *Interfaces e tendências da comunicação no contexto das organizações* (pp. 43-67). São Caetano do Sul, SP: Difusão Editora; Rio de Janeiro: Editora Senac Rio.
- Gil de Zúñiga, H. (2015). Citizenship, social media, and big data: current and future research in the social sciences. *Social Science Computer Review*, 1-7. doi: 10.1177/0894439315619589
- Hall, P. A. (2003). Aligning ontology and methodology in comparative research. In J. Mahoney & D. Rueschemeyer (Eds.), *Comparative historical analysis in the Social Sciences* (pp. 373-404). Cambridge: Cambridge University Press.
- Helbing, D. (2015). *The automation of society is next. How to survive the digital revolution*. Charleston: CreateSpace Independent Publishing Platform.
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, April-June, 1-12. doi: 10.1177/2053951714528481
- Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1), 14-29. doi: 10.1080/1369118X.2016.1154087
- Kuhn, T. (1998). *A estrutura das revoluções científicas*. São Paulo: Perspectiva.
- Luhmann, N. (1995). *Social systems*. Stanford, CA: Stanford University Press.
- Luhmann, N. (2006). *A improbabilidade da comunicação*. Lisbon: Vega.
- Martino, L. C. (2009, June). *Atualidade mediática: o conceito e suas dimensões*. Paper presented at XVIII Compós - Encontro da Associação Brasileira de Programas de Pós-graduação em Comunicação, Belo Horizonte.
- Milan, S. & van der Velden, L. (2016). The alternative epistemologies of data activism. *Digital Culture & Society*, 2(2), 57-74. doi: 10.14361/dcs-2016-0205
- Procter, R., Vis, F. & Voss, A. (2013). Reading the riots on Twitter: methodological innovation for the analysis of big data. *International Journal of Social Research Methodology*, 16(3), 197-214. doi: 10.1080/13645579.2013.774172



Shaldbolt, N., Hall, W. & Berners-Lee, T. (2006). The semantic web revisited. *IEEE Intelligent Systems*, mai/jun, 96-101.

Sbardelotto, M. (2017, june). *Circulação em rede: a comutabilidade dos polos de produção e recepção no fluxo comunicacional digital*. Paper presented at XXVI Compós - Encontro da Associação Brasileira de Programas de Pós-graduação em Comunicação, São Paulo.

Uricchio, W. (2017). Data, culture and the ambivalence of algorithms. In M. T. Schäfer & K. van Es. (Eds.), *The Datafied Society, Studying Culture Through Data* (pp. 125-137) Amsterdam, Amsterdam University Press.

#### **BIOGRAPHICAL NOTE**

Ana Thereza Nogueira Soares is PhD candidate in Communication Sciences at the University of Brasília – UnB, Brazil. She is linked to the research line of Communication Theories and Technologies and to the research groups Organizational Communication and Systemic Thinking – COMSiS and Communication, Technology and Politics – CTPol. Her research interests are focused on communication phenomena undertaken in the context of civil society organizations and movements, especially considering their forms of expression and communication in the digital space.

E-mail: [anatsoares@gmail.com](mailto:anatsoares@gmail.com)

Address: SQN 411, Bloco E, ap. 304 CEP: 70866-050. Brasília-DF. Brazil

\* Submitted: 30.11.2017

\* Accepted: 15.03.2018