



Palabra Clave

ISSN: 0122-8285

palabra.clave@unisabana.edu.co

Universidad de La Sabana

Colombia

Shahin, Saif

A Critical Axiology for Big Data Studies

Palabra Clave, vol. 19, núm. 4, diciembre, 2016, pp. 972-996

Universidad de La Sabana

Bogotá, Colombia

Available in: <http://www.redalyc.org/articulo.oa?id=64948471002>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System

Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal

Non-profit academic project, developed under the open access initiative

A Critical Axiology for Big Data Studies

Saif Shahin¹

Recibido: 2016-09-12
Enviado a pares: 2016-09-12

Aprobado por pares: 2016-09-30
Aceptado: 2016-10-02

DOI: 10.5294/pacla.2016.19.4.2

Para citar este artículo / to reference this article / para citar este artigo

Shahin, S. (2016). A critical axiology for Big Data studies. *Palabra Clave*, 19(4), 972-996.
DOI: 10.5294/pacla.2016.19.4.2

Abstract

Big Data is having a huge impact on journalism and communication studies. At the same time, it has raised a plethora of social concerns ranging from mass surveillance to the legitimization of prejudices such as racism. This article develops an agenda for critical Big Data research. It discusses what the purpose of such research should be, what pitfalls it should guard against, and the possibility of adapting Big Data methods to conduct empirical research from a critical standpoint. Such a research program will not only enable critical scholarship to meaningfully challenge Big Data as a hegemonic tool, but will also make it possible for scholars to draw upon Big Data resources to address a range of social issues in previously impossible ways. The article calls for methodological innovation in combining emerging Big Data techniques with critical/qualitative methods of research, such as ethnography and discourse analysis, in ways that allow them to complement each other.

Keywords

Big data; technology; social media; critical research; surveillance (Source: Unesco Thesaurus).

¹ Bowling Green State University. Estados Unidos. sshahin@bgsu.edu

Una axiología crítica para los estudios de *Big Data*

Resumen

Los datos masivos (*Big Data*) han tenido un gran impacto en el periodismo y los estudios de comunicación, a la vez que han generado un gran número de preocupaciones sociales que van desde la vigilancia masiva hasta la legitimación de prejuicios, como el racismo. En este artículo, se desarrolla una agenda para la investigación crítica de Big Data y se discute cuál debería ser el propósito de dicha investigación, de qué obstáculos protegerse y la posibilidad de adaptar los métodos de Big Data para llevar a cabo la investigación empírica desde un punto de vista crítico. Dicho programa de investigación no solo permitirá que la erudición crítica desafíe significativamente a Big Data como una herramienta hegemónica, sino que también permitirá que los académicos usen los recursos de Big Data para abordar una serie de problemas sociales de formas previamente imposibles. El artículo llama a la innovación metodológica para combinar las técnicas emergentes de Big Data y los métodos críticos y cualitativos de investigación, como la etnografía y el análisis del discurso, de tal manera que se puedan complementar.

Palabras clave

Big Data; tecnología; medios de comunicación sociales; investigación crítica; vigilancia (Fuente: Tesauro de la Unesco).

Uma axiologia crítica para os estudos de *Big Data*

Resumo

Os megadados (*Big Data*) têm tido um grande impacto sobre o jornalismo e os estudos de comunicação, e têm gerado um grande número de preocupações sociais, desde a vigilância em massa até a legitimação de preconceitos, como o racismo. Neste artigo se desenvolve uma agenda para a investigação crítica do *Big Data* e se discute qual deveria ser o propósito dessa investigação, de quais obstáculos se protegerem e a possibilidade de adaptar os métodos de *Big Data* para realizar a pesquisa empírica a partir de um ponto de vista crítico. Esse programa de pesquisa não apenas permite que a erudição crítica desafie significativamente os megadados como uma ferramenta hegemônica, também permite que os acadêmicos usem os recursos de *Big Data* para abordar uma série de problemas sociais de formas antes impossíveis. O artigo pede uma inovação metodológica para combinar técnicas emergentes de *Big Data* e os métodos críticos e qualitativos de pesquisa, tais como a etnografia e a análise do discurso, para que possam se complementar.

Palavras-chave

Big Data, tecnologia, mídias sociais, pesquisa crítica, monitoramento (Fonte: Tesouro da Unesco).

Introduction

The techno-euphoria spurred by the advent of Big Data (e.g. Anderson, 2008) is slowly giving way to uneasiness about the social effects of enormous datasets and the algorithms used to compile and analyze them (boyd & Crawford, 2012; Crawford, Miltner, & Gray, 2014; Mahrt & Scharkow, 2013; Manovich, 2012; Shahin, 2016a). Reports of malpractices by major Big Data-enabled enterprises such as Facebook and Google that compromise user privacy (Dwyer, 2011; Rubenstein & Good, 2012), along with Edward Snowden's revelation that the U.S. government was running surveillance programs on a global scale in collusion with technology companies (Bauman et al., 2014; Lyon, 2014), have made it plain that Big Data is not the panacea for all human problems that it is sometimes made out to be. Instead, Big Data may be reinforcing social divides and exacerbating a variety of social concerns.

A *ProPublica* investigation revealed that a criminal risk assessment algorithm developed by a commercial enterprise, widely used by courts and law enforcement officials across the United States, “was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants” (Angwin et al., 2016, para. 16). A *New York Times* article highlighted a series of “mistakes” committed by commonly used Big Data technologies, including Google Photos tagging black people as “gorillas,” Nikon cameras asking Asians – who often have small eyes compared with Caucasians – if they were “blinking” (Crawford, 2016). Meanwhile, reports continue to emerge about social media companies becoming ever more intrusive, collecting increasing amounts of users' personal data to serve advertisers and even running experiments manipulating user sentiments (Dewey, 2016).

What do these concerns mean for journalism and communication research, a field in which Big Data is having a huge impact? Scholars in our field quickly took to Big Data studies: partly because much of Big Data is generated by media and communication technologies – mobile telephones, social media, and so on – and partly because Big Data started altering the economic and operational dynamics of established media institutions,

especially news organizations. The surge of interest in Big Data research, and awareness of its game-changing potential, is evident in the deluge of Big Data articles being published in communication journals; special issues on Big Data that several journals of note have come up with, including the *Journal of Communication*; *Journalism & Mass Communication Quarterly*; *Journal of Broadcasting and Electronic Media*; *International Journal of Communication*; and *Media, Culture & Society*; and the emergence of new journals devoted to Big Data research, such as *Big Data & Society* and *Social Media + Society*.

This article provides an assessment of what Big Data research has come to mean in journalism and communication studies, identifying two expansive categories: research *with* Big Data and research *on* Big Data. Then, drawing on Gitlin's (1978) well-known critique of Katz and Lazarsfeld's (1955) two-step flow theory as the "dominant paradigm" in media studies, the article examines the ideological underpinnings of Big Data research – now regarded as a "paradigm" in its own right (Burgess, Bruns, & Hjorth, 2013). Building on this critique, the article charts an agenda for critical Big Data research, discussing what the purpose of such research should be, what pitfalls it should guard against, and the possibility of adapting Big Data methods themselves to conduct critical research. It argues that a critical approach to Big Data is necessary not only because the problems posed by Big Data need to be explicitly examined in line with critical theory and methods, but also because developing such a research agenda can help critical scholarship in journalism and communication studies draw upon Big Data resources to address a broad range of social concerns in previously impossible ways.

What is Big Data Research

Big Data research is commonly understood to be research that uses massive datasets. But attempts to forge a formal definition of Big Data aren't always consistent with each other. For instance, data is deemed to be Big only when "the current techniques and technologies may not be able to handle [its] storage and processing" (Suthaharan, 2014, p. 70). But Big Data is also defined as "a capacity to search, aggregate, and cross-reference

large data sets” (boyd & Crawford, 2012, p. 663). These definitions contradict each other: Big Data must be processible, otherwise it ceases to be useful no matter how Big it might be, but if data *can be* processed, then by Suthaharan’s definition it is no longer Big. To sidestep this paradox, some scholars have defined Big Data in terms of data volumes that only supercomputers – as opposed to personal computers – can process. But this distinction between personal and supercomputers is also problematic: after all, processing capacities once limited to supercomputers are now common for personal computers as well (Manovich, 2012; boyd & Crawford, 2012).

Research *with* Big Data

Instead of hampering it, this definitional ambiguity may have helped Big Data find its way into a variety of academic spaces and quickly become the zeitgeist of social science research, including and especially journalism and communication studies. Large numbers of research projects are being envisaged and carried out using previously unheard of data volumes. The very size of the dataset is often their biggest – if not only – selling point. Discourses native to Web 2.0, including social media such as Twitter, Facebook, and YouTube and sites such as Wikipedia, often provide the “Big” data for these projects. “Older” forms of discourse – news articles, political speeches, etc. – that are available in digital formats are also used.

Research with Big Data has sparked innovative methodological thinking to handle new forms of data and new levels of data volume. Techniques such as network analysis have found fresh relevance for social media research using Big Data (Guo, 2012; Kitts, 2014). In addition, scholars are coming up with ever newer methods of collecting and analyzing data from different kinds of digital platforms. Algorithmic techniques are being borrowed from computer science and computational linguistics, especially for automated content analysis, semantic analysis, and sentiment analysis (van Atteveldt, 2008; DiMaggio, Nag, & Blei, 2013; Su et al., 2016).

Parks, therefore, proffered a methodological definition of Big Data research as “the analysis of large social networks (including online networks such as Twitter), automated data aggregation and mining, web and mobile

analytics, visualization of large datasets, sentiment analysis/opinion mining, machine learning, natural language processing, and computer-assisted content analysis of very large datasets” (2014, p. 355). As the field evolves, the limits of these Big Data methodologies are also being recognized and addressed – often by combining multiple techniques that offset each other’s shortcomings (Lewis, Zamith, & Hermida, 2013; Shahin, 2016a, 2016b).

Research *on* Big Data

As several scholars acknowledge, the idea of Big Data as a social phenomenon goes beyond issues of data volumes and processing speeds (boyd & Crawford, 2012; Crawford, Miltner, & Gray, 2014; Mahrt & Scharkow, 2013; Manovich, 2012). Big Data has enabled and empowered a range of institutions and practices that are changing the world as we know it (see also, Shah, Cappella, & Neuman, 2015). Understanding them and their impact constitutes research *on* Big Data.

Studies about major internet and social media corporations, focusing on how they make their products and services work online to how they operate offline and what kinds of effects they have, are examples of research on Big Data. For instance, scholars are trying to understand the process by which search engine companies write their algorithms and how these algorithms promote their business models (Introna & Nissenbaum, 2000; Mager, 2012; Rohle, 2009). Others are focusing on the ways in which social media are having an impact on both participatory and contentious politics (Bennett & Segerberg, 2012; Gil de Zúñiga, Molyneux, & Zheng, 2014). Studies looking at the impact of Big Data on social phenomena and issues that have themselves emerged in the digital age – digital communities, digital labor, digital divide and so on – are also examples of such research (Andrejevic, 2014; Graham, Straumann, & Hogan, 2015; McChesney, 2013).

The emergence of Big Data has raised or reframed a number of ethical questions and legal challenges. Exploring these also constitutes research on Big Data. Some of these challenges are technological – the issue of internet governance, for instance, especially its contentious aspects such as net neutrality (Quail & Larabie, 2010; van Eeten & Mueller, 2012). Perhaps more

significantly, mass supervision and the threat to personal privacy have become two of the biggest human concerns of the so-called Petabyte Age. Research on Big Data, therefore, includes how governments and corporations compile, store, and use personal data, and the effects of these practices on citizens (Stoycheff, 2016; Tene & Polonetsky, 2012).

Big Data is not only enabling new types of institutions and practices but also altering previous ones, sometimes quite dramatically. News organizations, for instance, are witnessing changes at multiple levels. The news they produce is becoming increasingly data-driven and techniques such as data visualization are gaining in importance (Coddington, 2014). The kind of people working in news organizations is also evolving (Lewis & Usher, 2014). While reporters and editors are expected to develop their technological savvy, there is also an influx of technologists “to identify and appropriate suitable technological systems and solutions from external providers, or develop and reconfigure such systems and solutions themselves” (Lewis & Westlund, 2015, p. 450).

News organizations will change even further as they experiment with the possibilities of “immersive” and “robotic” journalism (Carlson, 2015; de la Peña et al., 2010). Meanwhile, the marketing of news and the way news organizations think about their business are also changing. Cumulatively, these shifts are not only transforming news organizations internally but will potentially also change them as social institutions – altering their relationships with other social institutions such as advertisers, political parties, and various levels of government, which, in turn, are undergoing similar transformations enabled by Big Data.

Related, but Different

Research *with* Big Data and research *on* Big Data are closely interrelated. Studies that use massive datasets or computational techniques also often investigate social institutions and practices that have been enabled by voluminous datasets and algorithms. Research on social media effects using large volumes of social media data is an example. A number of scholars are extending the agenda-setting theory by investigating the effects of social media conversations on public opinion – even using social network analy-

sis to do so (Neuman et al., 2015; Vargo et al., 2015). Other scholars are examining emerging practices of media consumption, such as second screening (Giglietto & Selva, 2015), through large-scale social media analyses.

But research *with* Big Data need not always be research *on* Big Data. Scholars may use Big Data to investigate issues that have little to do with Big Data as a social phenomenon. Westwood et al. (2013) examined 3.2 million articles to identify which foreign countries and regions receive most coverage in U.S. newspapers. Sjøvaag et al. (2012) used computer-assisted data gathering and structuring to study the online news content of the Norwegian public service broadcaster. Even social media studies need not be about social media as a social phenomenon. Park et al. (2015), for instance, used 1.7 billion tweets to examine how individualist and collectivist cultures differ in their use of emoticons. Emery et al. (2015) studied the effectiveness of a health campaign through responses on social media. Guo et al. (2016) examined 77 million tweets to identify the key topics being discussed during the 2012 U.S. presidential election campaign, while McGregor and Mourão (2016) also used Twitter data to explore the gendered distribution of relational power.

Similarly, research *on* Big Data is not always conducted *with* huge datasets or computational techniques. The consumption practices and behavioral effects of social media are also being investigated using traditional survey methods and samples of a few thousand to even a few hundred respondents (Gil de Zúñiga, Garcia-Perdomo, & McGregor, 2015). Stoycheff (2016) conducted an experimental study, with 255 participants, on the effects of social media surveillance on democratic discourse. Clerwall (2014) and Carlson (2015) studied “automated/algorithmic journalism” using small-scale experiments and textual analyses. And through 17 expert interviews, Mager (2012) shed light on how Google’s search engine feeds its business model.

Why do Big Data Research?

Research is always rooted in certain values and beliefs – its axiology – which serve certain purposes. These values are not always acknowledged, or even

realized – especially by social scientists who believe their scholarship to be “objective” and “impartial” (Schutt, 2009). That, indeed, is one important reason why Big Data has found such a ready audience among scientifically-minded scholars: it promises access to a pristine, out-there “truth” unhindered by human subjectivity. And yet, even the most positivist of research has an axiology – the inability or unwillingness of social scientists to recognize it only indicates that their axiology is hegemonic and has assumed the status of a Kuhnian “paradigm” (Kuhn, 2012).

Administrative Axiology

In his well-known critique of Katz and Lazarsfeld’s (1955) two-step flow theory as the “dominant paradigm” of media research, Gitlin observed that the theory was “consonant with an *administrative* point of view, with which centrally located administrators who possess adequate information can make decisions that affect their entire domain with a good idea of the consequences of their choices” (1978, p. 211; my emphasis). In other words, the purpose of research conducted from the two-step flow perspective is to provide administrators with the information they need to come up with policies that would have the desired effects. Gitlin further located this administrative point of view in “academic sociology’s ideological assimilation into modern capitalism and its institutional rapprochement with major foundations and corporations in an oligopolistic high-consumption society; ... a concordant marketing orientation, in which the emphasis on commercially useful audience research flourishes; and ... a justifying social democratic ideology” rooted in consumerism (p. 224).

Much the same could be said about a great deal of Big Data research. To begin with, the very label of “Big Data” is oriented toward administrative control and consumer marketing (Lewis & Westlund, 2015; Puschmann & Burgess, 2014). It is meant to indicate a paradigmatic shift from previous forms of data, invoke “newness” and thereby enhance marketability. The mythology of Big Data, Puschmann and Burgess have argued, frames it in two interrelated ways: “as a natural force to be controlled and as a resource to be consumed” (2014, p. 1690). Talking of Big Data as a natural force detracts from the constructed nature of datasets, ascribing greater authenticity to products

and services associated with Big Data. Simultaneously, this mythology allocates power to those who *can* control this natural force.

The purpose of Big Data research thus becomes *how* to control this “natural force.” Methodological research enables administrators – governmental and corporate – to figure out new sources of data, new ways of mining it, and new techniques of analyzing it. That is why techniques such as opinion mining and sentiment analysis are becoming so popular, because they make administrators better understand how their consumers are feeling about particular products and customize product placement more efficiently. The same techniques also allow governments to discern how the public is thinking or feeling. Indeed, research has gone beyond analyzing to manipulating sentiment. In 2014, Facebook infamously tinkered with the news feeds of more than half a million users to test how positive and negative posts affect consumers’ emotions on social media – so that it doesn’t simply have to react to sentiments but can even shape sentiments to benefit advertisers (Kramer, Guillory, & Hancock, 2014; see also Panger, 2016).

This administrative axiology extends into political communication research too. Studies focusing on how particular aspects of social media and particular ways of using them shape political behavior allow political parties to run their campaigns more effectively on social media, and even come to regard social media as an increasingly important site of political campaigning. In this orientation, the voter is the consumer while political parties are no different from corporations selling consumer products – even as social media themselves become the all-encompassing environment within which the buying and selling of everything from fast-moving consumer goods to political parties takes place. Not surprisingly, all this research is typically carried out in the name of social democracy, which as Gitlin (1978) noted, forms the ideological justification for the administrative point of view.

Critical Axiology

As opposed to the administrative axiology, which helps produce, sustain, and normalize structures of power, a critical axiology of research questions the legitimacy of such power structures and uncovers the process by which

they come to be powerful. Big Data has empowered governments and corporations by giving them greater control over our lives. Critical Big Data research is aimed at (1) unearthing the ideological underpinnings of Big Data-enabled institutions and services; (2) investigating the norms and practices through which they exercise power; and (3) examining the effects that such power may have on people's lives.

Critical Research *on* Big Data

As critical Big Data research focuses on institutions and practices enabled by Big Data, it would typically constitute research *on* Big Data. There are several important studies in this domain, even though their authors do not always refer to them explicitly as Big Data research. As a general survey of such scholarship is not possible here, I discuss a few crucial examples.

Mager's (2012, 2014) research on "algorithmic ideology" exposes how the logic of revenue generation and profit maximization dictates the functioning of search algorithms. Through interviews with computer scientists and programmers, journalists, net activists, and jurists, she shows that "corporate search engines and their capitalist ideology are solidified in a socio-political context characterized by a techno-euphoric climate of innovation and a politics of privatization" created by mass media (2012, p. 774). Everyone from website builders to individual web users are embedded in this hegemonic structure, and that is what allows the business model of search engines such as Google to function: "If website providers or users broke out of the core network dynamic, the power of search engines and their schemes of exploitation would fall apart" (p. 782).

Andrejevic's (2007, 2009) critique of interactivity, a cornerstone of what has come to be known as Web 2.0, reveals how seemingly democratizing practices actually provide administrators greater control over people's lives and undermine social justice. He observes that "whenever we are told that interactivity is a way to express ourselves, to rebel against control, to subvert power, we need to be wary of power's ruse: the incitation to provide information about ourselves, to participate in our self-classification, to complete the cybernetic loop" (2009, p. 41). It is the "active audience's"

ability to provide “feedback” that has allowed marketers to “envision a world in which it becomes increasingly possible to subject the public to a series of controlled experiments to determine how best to influence them” (p. 42). The 2014 Facebook study (Kramer, Guillory, & Hancock, 2014) is one example of such mass experimentation.

Experimental research can also be informed by a critical axiology. A study by Stoycheff (2016) indicates that the U.S. government’s mass surveillance of internet users, exposed in 2013 by Edward Snowden, has had a “chilling effect” on public discourse online. It has especially undermined the expression of opinions that people consider to be unpopular. The government’s justification of its surveillance program has also affected online behavior: “when individuals think they are being monitored and disapprove of such surveillance practices, they are equally as unlikely to voice opinions in friendly opinion climates as they are in hostile ones” (p. 305).

As these studies demonstrate, a critical approach to Big Data research questions many of the assumptions upon which the administrative approach is based. It challenges the climate of techno-utopia that has been spawned by and is constantly revitalized in conventional Big Data discourses. It questions the “normalcy” of the neoliberal worldview, in which big corporations and their pursuit of profit are seen as the natural path of human progress. It also disputes the capitalist appropriation of human agency and social democracy, and exposes the nexus of Big Data, Big Business, and Big Government that makes such appropriation possible. And it often does so without working with Big Data.

Critical research with Big Data

But critical questions – relating to Big Data, digital technology, or social phenomena in general – may also be explored *with* Big Data, that is, with the help of enormous datasets and emerging computational techniques that facilitate their analysis. Such research would be motivated by a spirit of social justice – as opposed to advancing the interests of governments and businesses. Equally importantly, it would pay heed to the epistemological, methodological, and ethical/normative concerns that have been raised vis-à-vis conventional Big Data research (see also Shahin, 2016a).

The biggest such concern, of course, is the “rhetoric of objectivity” surrounding Big Data – the notion that Big Data somehow provides access to a pristine, “out-there” reality, an access untainted by fallacious human beliefs, emotions, attitudes, or values (Crawford, Miltner, & Gray, 2014). Critical research would instead view datasets as constructs that are shaped by how human beings perceive the world, and how datasets, in turn, represent the world in ideologically motivated ways (Gitelman, 2013; Helles & Jensen, 2013; Puschmann & Burgess, 2014). Respecting people’s privacy concerns is another important issue for critical research, especially in the context of social media. While it is impossible for a scholar to get permission from every social media user whose posts are part of a massive dataset, the scholar would take care to ensure that the data being collected is at least in the public domain.

Another problem is the superficiality of conventional Big Data research. Mahrt and Scharkow called “comparatively shallow measures” and “lack of context awareness” as two of the most frequently discussed issues with Big Data studies (2013, p. 26). Talking specifically about textual data, Lewis, Zamith and Hermida observed that “when turning to computerized forms of content analysis, many scholars have found them to yield satisfactory results only for surface-level analyses, thus sacrificing more nuanced meanings present in the analyzed texts” (2013, p. 38). That is mainly because “the computer is simply unable to understand human language in all its richness, complexity, and subtlety as can a human coder” (Simon, 2001; cited in Lewis, Zamith, & Hermida, 2013, p. 38). In contrast, critical Big Data studies would attempt to be more contextually sensitive and fine-grained. A final problem is *apophenia*, or “seeing patterns where none actually exist, simply because enormous quantities of data can offer connections that radiate in all directions” (boyd & Crawford, 2012, p. 668). Humongous datasets can readily yield “statistically significant” relationships among variables, and post-hoc theorization makes these “findings” even more problematic (Mahrt & Scharkow, 2013). A critical approach to Big Data research would avoid research designs that rely on such findings.

Superficiality and apophenia, in particular, are functions of the enormity of datasets. But as Mahrt and Scharkow suggested, “Big Data can

safely be reduced to medium-size data and still yield valid and reliable results” (2013, p. 28). One way to deal with these problems, therefore, is to reduce the volume of data used for analysis through randomized or purposive sampling. Computational methods can help sample data in theoretically meaningful ways, reducing Big Data to more manageable sizes. Once sampled, the data may be analyzed in a nuanced, contextually sensitive manner.

Murthy and colleagues have published multiple articles on how to conduct research with Big Data on smaller scales. Their work is aimed at helping scholars short on financial and technical resources – in other words, scholars who are not affiliated with businesses and governments – access, store, and analyze Big Data, especially social media data. For instance, Murthy and Bowman (2014) discuss a cost-effective mechanism to collect, store, and study nearly 150 million tweets a month. They compare some easy-to-use databases in terms of their value for social researchers, explain the hardware requirements and technical details of setting up a collection and storage system, and provide an experimental case study that takes readers through every step of the process all the way to the analysis. Murthy (2013) explains how to conduct ethnographic research through Facebook and how to use iPhones as data-gathering devices for such research. He argues that digital ethnography is not just feasible but necessary because “our respondents now spend significant portions of their occupational and social lives online... If we do not keep pace in our research methods, we risk not collecting data from spaces which are important to the daily lives of many of our respondents (e.g. Facebook).”

In my own research (Shahin, 2016a), I have used a methodological approach that combines natural language processing with Python and interpretive analysis to study large-volume textual data in a theoretically grounded and contextually sensitive manner – illustrating it with two case studies. The first case study examines the Inaugural Address Database, a collection of the inaugural addresses of all U.S. presidents from George Washington to Barack Obama. Using Python, I extract two purposive samples from this database: each sample includes all occurrences of a theoretically significant keyword (“constitution” and “public”) along with a certain

number of characters on either side that provide the contexts in which the keywords were used. Next, these samples are studied using the interpretive technique of cluster criticism, in which the words being used in the vicinity of the keyword are coded into semantic categories that, in turn, suggest how the presidents interpret and relate to the two keywords. In the second case study – examining year-long news coverage of two separate shootings at a U.S. army camp – I use Python to extract all paragraphs in which the word “terror” in all its forms (terrorism, terrorist, terrorists) was used. These paragraphs are then analyzed using ideological criticism to show that a shooting is considered a “terrorist attack” when the shooter is a Muslim, but not otherwise.

Conclusion

Adopting a critical axiology is never an easy task in any field of scholarship. Critical scholars, by definition, go against the norms of their field and find fault where others see merit. That makes critical research not just intellectually but also professionally challenging. And yet, a critical axiology is necessary if research has to serve the public instead of being a means of administrative control, intentionally or otherwise.

Defining the public interest is a tricky question: as we have seen, the powerful themselves justify their control over the public through ideologies such as social democracy, which are meant to empower the public. So the more pertinent question is why should *any* set of institutions or individuals – including (critical) scholars – have the capacity to define what is good for the public as a whole. Such a capacity is necessarily an exercise of power. Instead of trying to proffer a definition of public interest, the purpose of critical scholarship is to reveal the social processes by which such definitions are produced and naturalized, point out the institutions and individuals who influence or control these processes, and uncover how particular definitions serve particular ideologies and interests.

The growing influence of Big Data on human affairs and social relations necessitates a critical approach to Big Data research. Big Data is a powerful tool, and it is being used to perpetuate the ideologies and inter-

ests of governments and corporations. A critical approach is therefore required to unravel the mythology that Big Data apologists have woven around it and lay bare the ways in which it bolsters administrative control. This can, and is, being done by scholars using “small data” and traditional methods. It can also be done using Big Data itself, and the emerging computational methods needed to do research *with* Big Data – especially in conjunction with critical/qualitative methods.

Such research is still in its infancy. But that is partly because methodological Big Data research is itself developing gradually, and relies heavily on collaboration with scholars from information science, computational linguistics, and so on. As journalism and communication scholars become more adept in Big Data research techniques – and simultaneously come to recognize their limitations – the merits of combining them with more critical research methods will perhaps become apparent. In the same way, a deeper appreciation for critical Big Data studies – such as this article hopes to provide – will perhaps lead more scholars to think along these lines and develop more ways of using Big Data with a critical axiology.

References

- Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016, May 23). Machine bias. *ProPublica*. Retrieved from <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [Date accessed: May 30, 2016]
- Anderson, C. (2008, June 23). The end of theory: The data deluge makes the scientific method obsolete. *Wired*. Retrieved from <http://www.wired.com/2008/06/pb-theory/> [Date accessed: November 12, 2015]
- Andrejevic, M. (2007). *iSpy: Surveillance and power in the interactive era*. Lawrence: University Press of Kansas.
- Andrejevic, M. (2009). Critical Media Studies 2.0: An interactive upgrade. *Interactions: Studies in Communication and Culture*, 1(1), 35-51.

- Andrejevic, M. (2014). The Big Data divide. *International Journal of Communication*, 8, 1673-1689.
- Bauman, Z., Bigo, D., Esteves, P., Guild, E., Jabri, V., Lyon, D., & Walker, R. B. (2014). After Snowden: Rethinking the impact of surveillance. *International Political Sociology*, 8(2), 121-144.
- Boyd, danah, & Crawford, K. (2012). Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5): 662-679. doi :10.1080/1369118X.2012.678878.
- Bennett, W. L., & Segerberg, A. (2012). The logic of connective action: Digital media and the personalization of contentious politics. *Information, Communication & Society*, 15(5), 739-768.
- Burgess, J., Bruns, A., & Hjorth, L. (2013). Emerging methods for digital media research: An introduction. *Journal of Broadcasting & Electronic Media*, 57(1), 1-3. doi:10.1080/08838151.2012.761706
- Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, 3(3), 416-431.
- Clerwall, C. (2014). Enter the Robot Journalist: Users' perceptions of automated content. *Journalism Practice*, 8(5), 519-531.
- Coddington, M. (2015). Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3(3), 331-348. doi: 10.1080/21670811.2014.976400.
- Crawford, K. (2016, June 25). Artificial intelligence's white guy problem. *New York Times*. <http://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html> [Date accessed: May 30, 2016]

- Crawford, K., Miltner, K., & Gray, M. L. (2014). Critiquing Big Data: Politics, ethics, epistemology. *International Journal of Communication*, 8, 1663–1672.
- De la Peña, N., Weil, P., Llobera, J., Giannopoulos, E., Pomés, A., Spanlang, B., ... & Slater, M. (2010). Immersive journalism: immersive virtual reality for the first-person experience of news. *Presence: Teleoperators and Virtual Environments*, 19(4), 291-301.
- Dewey, C. (2016, August 19). 98 personal data points that Facebook uses to target ads to you. *Washington Post*. Retrieved from <https://www.washingtonpost.com/news/the-intersect/wp/2016/08/19/98-personal-data-points-that-facebook-uses-to-target-ads-to-you/> [Date accessed: September 2, 2016]
- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, 41, 570-606.
- Dwyer, C. (2011). Privacy in the Age of Google and Facebook. *IEEE Technology and Society Magazine*, 30(3), 58-63.
- Emery, S. L., Szczypka, G., Abril, E. P., Kim, Y., & Vera, L. (2014). Are you scared yet? Evaluating fear appeal messages in tweets about the tips campaign. *Journal of Communication*, 64(2), 278-295.
- Giglietto, F., & Selva, D. (2014). Second screen and participation: A content analysis on a full season dataset of tweets. *Journal of Communication*, 64(2), 260-277.
- Gil de Zúñiga, H., Garcia-Perdomo, V., & McGregor, S. C. (2015). What is second screening? Exploring motivations of second screen use and its effect on online political participation. *Journal of Communication*, 65(5), 793-815.

- Gil de Zúñiga, H., Molyneux, L., & Zheng, P. (2014). Social media, political expression, and political participation: Panel analysis of lagged and concurrent relationships. *Journal of Communication*, 64(4), 612-634.
- Gitelman, L. (Ed.). (2013). *Raw data is an oxymoron*. Cambridge, MA: MIT Press.
- Gitlin, T. (1978). Media sociology: The dominant paradigm. *Theory and Society*, 6(2), 205-253.
- Graham, M., Straumann, R. K., & Hogan, B. (2015). Digital divisions of labor and informational magnetism: Mapping participation in Wikipedia. *Annals of the Association of American Geographers*, 105(6), 1158-1178.
- Guo, L. (2012). The application of social network analysis in agenda setting research: A methodological exploration. *Journal of Broadcasting & Electronic Media*, 4(616), 631.
- Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. (2016). Big Social Data analytics in journalism and mass communication comparing dictionary-based text analysis and unsupervised topic modeling. *Journalism & Mass Communication Quarterly*, 93(2), 332-359.
- Helles, R., & Jensen, K. B. (2013). Making data—big data and beyond: Introduction to the special issue. *First Monday*, 18(10), Retrieved from <http://firstmonday.org/article/view/4860/3748>
- Introna, L., & Nissenbaum, H. (2000). The public good vision of the internet and the politics of search engines. In R. Rogers (ed.) *Preferred Placement – Knowledge Politics on the Web* (pp. 25-47), Maastricht: Jan van Eyck Academy.
- Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence: The part played by people in the flow of mass communications*. New York: Free Press.

- Kitts, J. A. (2014). Beyond networks in structural theories of exchange: Promises from computational social science. *Advances in Group Processes*, 31, 263–298. doi:10.1108/S0882-614520140000031007
- Kramer, A., Guillory, J., & Hancock, J. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788–8790.
- Kuhn, T. S. (2012). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Lewis, S. C., & Usher, N. (2014). Code, collaboration, and the future of journalism: a case study of the Hacks/Hackers global network. *Digital Journalism*, 2(3), 383–393.
- Lewis, S. C., & Westlund, O. (2015). Actors, actants, audiences, and activities in cross-media news work: A matrix and a research agenda. *Digital Journalism*, 3(1), 19–37.
- Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of Big Data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media*, 57(1), 34–52. doi:10.1080/08838151.2012.761702
- Lyon, D. (2014). Surveillance, Snowden, and big data: Capacities, consequences, critique. *Big Data & Society*, 1(2), 2053951714541861
- Mager, A. (2012). Algorithmic ideology: How capitalist society shapes search engines. *Information, Communication & Society*, 15(5), 769–787.
- Mager, A. (2014). Defining algorithmic ideology: Using ideology critique to scrutinize corporate search engines. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 12(1), 28–39.

- Mahrt, M., & Scharrow, M. (2013). The value of Big Data in digital media research. *Journal of Broadcasting & Electronic Media*, 57(1), 20–33. doi:10.1080/08838151.2012.761700
- Manovich, L. (2012). Trending: The promises and the challenges of big social data. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp. 460–475). Minneapolis, MN: University of Minnesota Press.
- McChesney, R. W. (2013). *Digital disconnect: How capitalism is turning the Internet against democracy*. New York and London: The New Press.
- McGregor, S. C., & Mourão, R. R. (2016). Talking Politics on Twitter: Gender, Elections, and Social Networks. *Social Media + Society*, July-September, 1-14. doi: 10.1177/2056305116664218
- Murthy, D., & Bowman, S. A. (2014). Big Data solutions on a small scale: Evaluating accessible high-performance computing for social research. *Big Data & Society*, 1(2), 1-12. doi: 10.1177/2053951714559105.
- Murthy, D. (2013). Ethnographic Research 2.0: The potentialities of emergent digital technologies for qualitative organizational research. *Journal of Organizational Ethnography*, 2(1), 23-36.
- Neuman, W. R., Guggenheim, L., Mo Jang, S., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication*, 64(2), 193-214.
- Panger, G. (2016). Reassessing the Facebook experiment: critical thinking about the validity of Big Data research. *Information, Communication & Society*, 19(8), 1108-1126.
- Park, J., Baek, Y. M., & Cha, M. (2014). Cross-Cultural Comparison of Nonverbal Cues in Emoticons on Twitter: Evidence from Big Data Analysis. *Journal of Communication*, 64(2), 333-354.

- Parks, M. R. (2014). Big data in communication research: Its contents and discontents. *Journal of Communication*, 64(2), 355-360.
- Puschmann, C., & Burgess, J. (2014). Metaphors of big data. *International Journal of Communication*, 8, 1690-1709.
- Quail, C., & Larabie, C. (2010). Net neutrality: Media discourses and public perception. *Global Media Journal*, 3(1), 31-50.
- Rubinstein, I. & Good, N. (2013). Privacy by design: A counterfactual analysis of Google and Facebook privacy incidents. *NYU School of Law, Public Law Research Paper No. 12-43*. Available at SSRN: <http://ssrn.com/abstract=2128146> or <http://dx.doi.org/10.2139/ssrn.2128146>
- Schutt, R. K. (2009). *Investigating the social world: The process and practice of research (6th ed)*. Thousand Oaks, CA: Sage.
- Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Big data, digital media, and computational social science: Possibilities and perils. *The ANNALS of the American Academy of Political and Social Science*, 659, 6-13. doi:10.1177/0002716215572084
- Shahin, S. (2016a) When scale meets depth: Integrating natural language processing and textual analysis for studying digital corpora. *Communication Methods and Measures*, 10(1), 28-50, doi: 10.1080/19312458.2015.1118447
- Shahin, S. (2016b). Right to Be forgotten: How national identity, political orientation, and capitalist ideology structured a trans-Atlantic debate on information access and control. *Journalism & Mass Communication Quarterly*, 93(2), 360-382. doi: 10.1177/1077699016638835

- Simon, A. F. (2001). A unified method for analyzing media framing. In R. P. Hart & D. R. Shaw (Eds.), *Communication in U.S. elections: New agendas* (pp. 75–89). Lanham, MD: Rowman and Littlefield.
- Sjøvaag, H., Moe, H., & Stavelin, E. (2012). Public service news on the Web: A large-scale content analysis of the Norwegian Broadcasting Corporation's online news. *Journalism Studies*, 13(1), 90–106. doi:10.1080/1461670X.2011.578940
- Stoycheff, E. (2016). Under surveillance examining Facebook's spiral of silence effects in the wake of NSA internet monitoring. *Journalism & Mass Communication Quarterly*, 93(2), 296-311.
- Su, L. Y. F., Cacciatore, M. A., Liang, X., Brossard, D., Scheufele, D. A., & Xenos, M. A. (2016). Analyzing public sentiments online: Combining human-and computer-based content analysis. *Information, Communication & Society*, 1-22. doi: 10.1080/1369118X.2016.1182197
- Suthaharan, S. (2014). Big Data classification: Problems and challenges in network intrusion prediction with machine learning. *SIGMETRICS Performance Evaluation Review*, 41(4): 70–73. doi:10.1145/2627534.2627557
- Tene, O., & Polonetsky, J. (2012). Privacy in the age of big data: A time for big decisions. *Stanford Law Review Online*, 64, 63-69.
- Van Atteveldt, W. (2008). *Semantic network analysis: Techniques for extracting, representing, and querying media content*. Charleston, SC: Book-Surge Publishers.
- Van Eeten, M. J. G., & Mueller, M. (2012). Where is the governance in Internet governance? *New Media & Society*, 15(5), 720-736. DOI: 10.1177/1461444812462850

Vargo, C. J., Guo, L., McCombs, M., & Shaw, D. L. (2014). Network issue agendas on Twitter during the 2012 US presidential election. *Journal of Communication*, 64(2), 296-316.

Westwood, S. J., Weiss, R. J., & Iyengar, S. (2013). All the news that is fit to print? Gatekeeping effects in newspaper coverage of international affairs. Paper presented at the 63rd annual conference of the International Communication Association in London.