Paakat
Revista de tecnología y sociedad

Catalog of digital techniques for research with content generated by social media users

Repertorio de técnicas digitales para la investigación con contenidos generados en redes sociodigitales

http://dx.doi.org/10.32870/Pk.a10n19.498

Gabriela Elisa Sued*
https://orcid.org/0000-0002-4516-678X
Technological Institute of Monterrey, México

Received: January 30, 2020
Accepted: March 3, 2020

ABSTRACT

This work makes a systematic classification and description of digital techniques applied to the study of data produced by users in social media. The institutional actors that produce and promote them are identified, the available tools and their scope are evaluated, and examples of studies that use them are presented. The techniques are classified according to the place they occupy in a research sequence formed by four stages: data collection, cleaning, processing and visualization. It focuses on both textual and image processing techniques. In the first case it addresses textual analytics, network analysis and sentiment analysis, while in the second case it focuses on the visual analysis of photographs and online video. The quantitative analysis of reactions, such as likes and shares, is also addressed. In the conclusions a critical evaluation of its scopes is carried out, among which are its effectiveness for the appreciation of extensive data sets produced in the dynamic and fluid context of social media, and the possibility of identifying patterns and recurrences in them.

Keywords
Electronic media; data collection; data processing; automatic text analysis; data visualization

RESUMEN

Este trabajo efectúa una clasificación y una descripción sistemática de las técnicas digitales aplicadas al estudio de datos producidos por usuarios en redes sociodigitales. Se identifican los actores institucionales que las producen y promueven, se evalúan las herramientas disponibles y su ámbito de aplicación, y se presentan ejemplos de estudios que las utilizan. Se clasifican las técnicas en función del lugar que ocupan en una secuencia de investigación formada por cuatro etapas: recolección, limpieza, procesamiento y visualización de datos. Asimismo, se enfoca tanto en técnicas de procesamiento de textos como de imágenes; en el primer caso se aborda el análisis cuantitativo de contenidos, el análisis de redes y el análisis de sentimientos, mientras que en el segundo caso se centra en la analítica visual de fotografías y video en línea. También se incluye el análisis cuantitativo de reacciones, como likes y compartidos. En las conclusiones se evalúan de manera crítica los alcances de este estudio, dentro de los que se encuentran: su eficacia para la apreciación de conjuntos extensos de datos producidos en el contexto dinámico y fluido de las redes sociodigitales, y la posibilidad de identificar patrones y recurrencias dentro de estos.

Palabras clave
Medios sociales; recopilación de datos; procesamiento de datos; análisis automático de textos; visualización de datos

* PhD in Humanities, Technological Institute of Monterrey. She is specialized in digital culture. e-mail: gabriela.sued@gmail.com

Introduction

Over the last decade, there has been a strong concern for the use of contents generated by users of socio-digital networks for social research. This has underscored its dual role: on the one hand, they comprise a primary source that enables the study of trends made by public opinion, generally, through different social platforms, in addition to the fact that they enable the understanding of the complex logic of social platforms as a new means of communication (Rogers, 2009).

In this context, several research programs arise, within which the initiative of digital methods (Rogers, 2013, 2019) and the initiative to study software (Manovich, 2016) stand out. Both initiatives focus on the study of digital objects using digital methods; therefore, they are located at the intersection between media studies and IT sciences.

In relation with previous analysis about digital practice and environments, such as digital ethnography (Hine, 2004; 2015), these methodologies have changed the focus of study of users' practices to the objects they produce (Marres, 2017). In addition to this movement, digital methods assume the incorporation of new research concepts, skills, and techniques. To Rogers (2009), studying digital objects generated in the web ought to be done by using digital methods and techniques related to the logic of the digital environment, therefore, a change in method needs a change in techniques.

Implementing digital techniques for social research and movement from the practice to objects does not only represent a change in procedures, but it also means, on the one hand, the likelihood to understand the logic of social platforms and the manner they model public expression and current culture (Nieborg & Poell, 2018) and, on the other hand, the opportunity to understand transformations produced by technical innovations inside social sciences as they assume the challenges of study and understanding of new digital objects. To social, non-academic disciplines, such as marketing, application of digital techniques to analyze data is not new. Analytical techniques model likes, decisions and online consumption, therefore, it is necessary for academic research to adopt new inputs that enable understanding these new sociocultural trends.

This work makes a systemic classification and description of digital techniques applied to the study of data produced by users of socio-digital networks, and evaluates available tools and their field of application. We will consider digital techniques as ways to do, such as digital tools or software for studying digital objects, whether they are generated by the web or that are initially analogic and then digitalized (Rogers, 2015). We classified the techniques per function in the research process to gather, process and analyze data.

Opening the black box of digital techniques

With the purpose of having a heterogeneous and dynamic program systematized, it is important to make a distinction between techniques and tools. Scientific research tools are procedures validated by practice, generally oriented to obtaining and transforming useful information to solve problems related to the knowledge of scientific discipline (Rojas, 2011); these operate at a conceptual level, while tools work at a concrete level.

In respect to digital techniques, software is the instrument or tool of technique. Tools are software programs, algorithms developed by one person or a group of people; they are available to the public or in private, they depend on the context of use or application. Furthermore, techniques may be more stable than tools. A tool may disappear, but another one may substitute it. Generally, there is more than one tool for a technique. Selecting the most proper one should answer to a set of parameters among which is the field of development, the type of licensing, complexity of the interface, among others.

As they are based on writing complex algorithms which escape the knowledge of social scientists, digital tools are black boxes: devices that process input information and produce output information, whose operation is unknown (Latour, 1992). If, as social researchers, we elect to use them, it is necessary to delegate scientific practices to third parties, whether they be algorithms, web interfaces, software or human programmers; for this reason, Marres & Gerlitz (2016) dub them as interface methods.

Currently, a number of different social actors take part in the production of digital tools to analyze data; not all of them, but a significant number, are members of academic communities. There are tools from marketing, data sciences or, even, developed by social platforms that may be used for social research. On the contrary, there are undertakings of academic origin that have turned into entrepreneurial organizations, such as the data visualization package Tableau, a commercial detachment of a research performed by Stanford University (Solomon, 2016).

These production conditions distinguish digital research from usually used methods for social sciences, where the researcher is the person who designs his/her data collection and analysis instruments. A result of this opening is that part of the research has to be delegated to an artifact built by third parties. Rogers (2009) dubs the use of tools for academic purposes, which have not been especially designed for research, as reorientation or repurposing. We ought to consider that the less the technological knowledge is demanded by the tool, the greater the knowledge we will delegate to it.

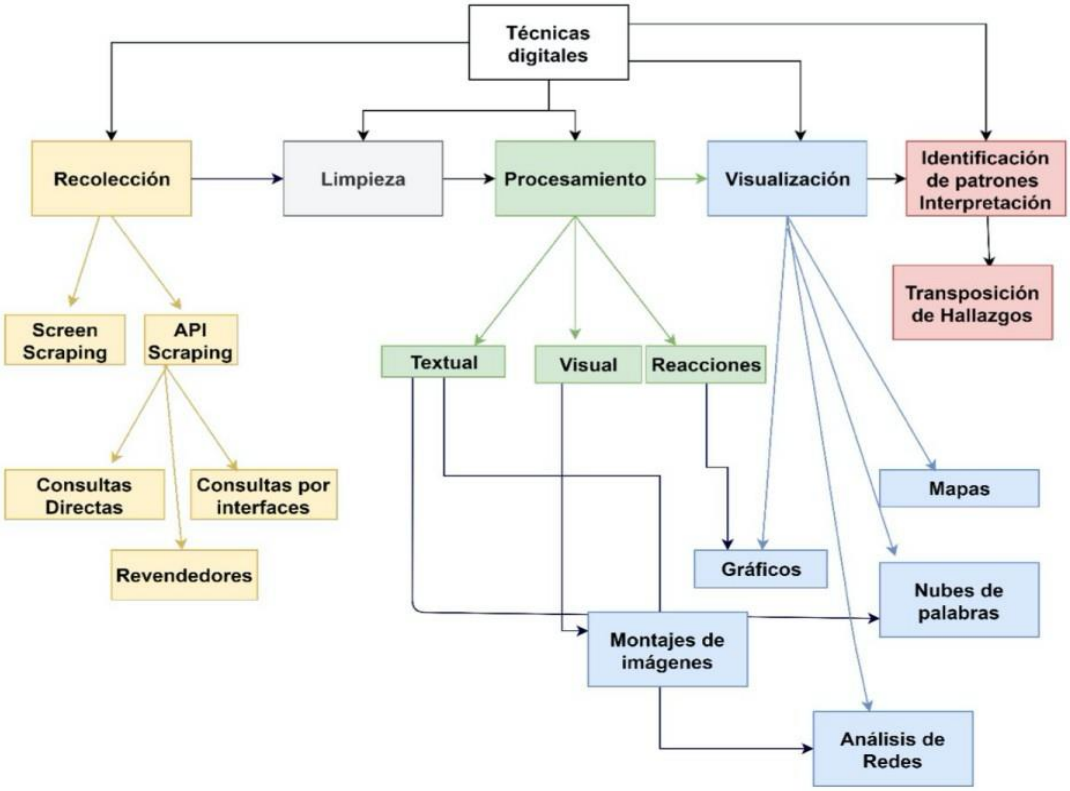
In view of the change of social and humanistic studies towards digital and digitalized objects, different university research centers (some of which have renewed their names as Media Lab or media laboratories) have developed collection, analysis and data visualization instruments. These were made, to a greater or lesser extent, available to academic communities. Some of these tools gather, process and visualize data, and deliver a finished product that combines with three technique types; there are others which deliver data and metadata in plane formats. Afterwards, researchers shall have to use different techniques and instruments to organize the data for meaning and to answer their research questions.

Social research digital techniques have been thoroughly used in English literature. Textual mining systematizations of Moreno & Redondo (2016) stand out, data analytics systematizations of Gandomi & Haider (2015) and the handbook of online research methods edited by Fielding, Lee & Blank (2016); however, literature in Spanish is scarce. This work is a contribution in that sense, as it analyzes techniques related to their role in the research process.

Typology of digital techniques

At the heterogeneous and changing scenario where techniques and tools are presented, it is not about renouncing to categorization, but to work on a number of manners to instrumentally classify digital techniques. In this case, we will focus on the role of research process techniques, but there are other possibilities: in accordance with the materiality of the digital object, the features of the software to be used and the object as part of a platform. Figure 1 organizes this information, in accordance to the place of a research sequence made by four sequential stages which include data collection, cleaning, processing and visualization.

Figure 1. Classification of digital techniques



Source: developed by the author.

Scraping as a data collection technique

Scraping is a specific technique of the digital media which enables automatic online data collection. This is one of the most representative techniques of current digital methods, as it enables research based on date in the digital media (Marres & Weltevrede, 2013). With this technique hard data are recovered, although they are structured in a website they were part of originally, they would need to be organized again so that they may be read and interpreted and become relevant information for the research (Carvajal, 2013). By means of executing a software developed for this purpose, and from the purpose of this study, we are able to recover different data types: from those in a PDF report to pour them into a spreadsheet, a set of images posted in a social network, tweets written on an occurrence or metadata of a set of videos stored in YouTube.

There are three types of scraping: screen and interface scraping, crawler scraping and API scraping (Elmer, 2015). The first one is the oldest; this is an automatic method used to extract data that have been designed to be primarily seen by humans. The program reads the screen and simulates a human being, and collects interesting data in a list that may be processed automatically (Fulton, 2014). It is based on data extraction from the HTML code visualized on the user interfaces, which means that the data are formatted and personalized for specific users.

A screen scraping example is the Bulk Media Downloader extension, a supplement mounted on the Firefox browser. After it has been installed, the user may download all the contents he/she sees on screen and store them in his/her computer. But, since searches in social platforms and websites are personalized, which the user collects, they are biased by his/her experience.

Crawler scraping extracts the structure of a website, in addition to data that may be contained by databases associated to the site. The third type of scraping, usually the most used, is based on application programming interface inquiries, known as API (Application Programming Interfaces). In this case, data are structured in a database as a function of the interests of the platforms they are stored in. For this reason, we have to use API; they consist of a set of algorithms, functions and procedures offered by a platform to be used by another software that will request information. They are a central tool in the process, as they represent communication capability between the database of the platform providing the information and the collection program.

If the platform does not enable its API to collect information, we will not be able to obtain its data, unless it is done through it. Generally, platforms enable their API for independent developers to create new applications for data, but not to extract information with the purpose of analyzing it, for example, Facebook enables its API for the creation of videogames, surveys, and advertising and marketing applications, but not for exporting data.

Cleaning the database

Cleaning the database is an intermediate stage between data collection and processing. This implies time and effort, but it is the basis to obtain a *corpus* that may be managed and visualized correctly. A clean database has a clear, reliable and well organized data structure, in such a way that a brief exploration will make sense (Wickham, 2014).

Data structure corresponds to a semantic organization. In a clean database, each column is a variable, each row is a remark containing one or more values, usually made up by numbers or words, and each set of remarks is a table. The order of a database is not part of its structure, but exploring data before processing them is made easier. The standard order, first off, places fixed variables and, secondly, numerical variables; related variables shall be placed next to each other and the rows may be ordered for the first variable.

There are common problems to clean the databases that may be solved with the usual data storage programs or with a specialized software. Open Refine (free on Google), is mostly used in this category.

Processing techniques: data analytics

Once we have captured the data, we have to process them to produce meaning. Data science names techniques used to produce meaning to captured information as *analytic* (Gandomi & Haider, 2015). Other authors in the same field prefer the term data mining (Han, Kamber & Pei, 2011). Both concepts include the analysis of any type of databases, for example, governmental, commercial, entrepreneurial, and academic. Although they did not originate in the field, these denominations have been adopted by social research. In accordance with the data type in question, they are classified as textual, audio, image, video, social networks and predictive analytics.

a) Textual analytic technique

The denomination of textual analytic or textual mining arises from computer science and it is a derivative of the data mining concept. From data science, Gandomi & Haider (2015) and Moreno & Redondo (2016) defend it as the discovery of new information automatically extracted from different written sources. In order to carry out this process, programs and algorithms are employed to make several operations on the texts, which the authors call textual analytic techniques. This procedure is described as a transformation of unstructured and qualitative information to structured and quantitative information. The development of natural language processing has had an advance in the past few years, and the content produced by users of networks is an area of special interest for its study.

Specialized literature mentions among the main techniques: information extraction, text summary, answers to questions made in natural language (Gandomi & Haider, 2015), contents analysis (Popping, 2016), sentiment analysis (Paltoglou & Thelwall, 2012) and visual data mining (Moreno & Redondo, 2016). From these, the latter three are employed in the analysis of objects produced in social platforms.

Popping (2016) defines content analysis as the reduction of an unstructured and qualitative text flow to a set of manageable and quantitative symbols for a valid text reference towards its context. Generally, these symbols represent values of presence, intensity and frequency. One of the basic forms of content analysis is the one that analyzes how often a word appears in a text (Robichaud & Blevins, 2011), and it is mostly used in studies based on data. Word count is a simple technique, but it seems to provide relevant results in the case of analyzing textual objects of the social web. This may be due to the fact that posts in social networks usually are short, plain and lack discursive complexities (Paltoglou & Thelwall, 2012). Underwood (2013) considers that this technique is plain and effective; although it is centered in the most simple unit of language, the word, this is, *per se*, a sufficiently complex linguistic element.

Moreno & Redondo (2016) emphasize on the capability of visual mining to place large volumes of data in a visual hierarchy, a map or a chart, which may be inspected in an interactive mode. This may be useful to explore a large number of documents and relations between subjects and topics. Examples of this category include word clouds, subject tree diagrams, or word networks. In this sense, this category is not completely autonomous, as this is the visual expression of another type of textual mining: word clouds visually present the analysis of word frequency, just as topic trees present the analysis of related topics, and network charts imply an expression of at least two entities which establish a relationship between each other.

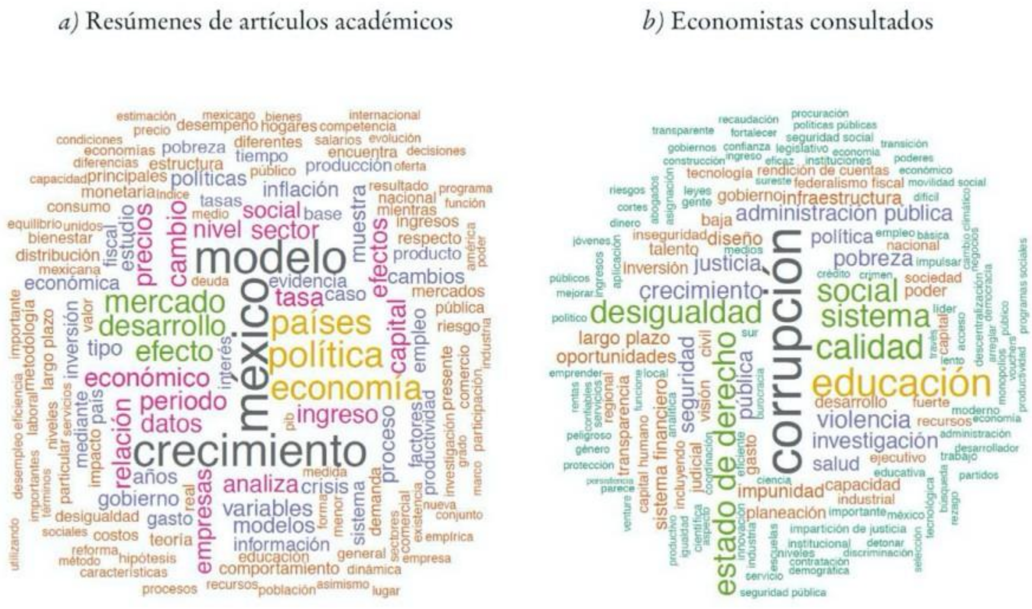
Voyant-Tools[1] is one of the most famous visual mining tools. It has been developed since 2003 by Stefan Sinclair and Geoffrey Rockwell (of the University of Alberta and McGill, respectively); it is free and of a free code. It is oriented for academic analysis in the field of digital humanities (Voyant Tools, 2018). Over the years, it has incorporated new modules whose functionalities make the initial word frequency analysis complex. The current full environment includes the analysis of topics, network charts and bubbles, phrase analysis, among other applications.

The tool has begun to be used in different social research areas, not only in the area corresponding to the analysis of social media. In the economy area, for example, Campos-Vazquez & Lopez Araiza (2018) use it to compare a set of words that are frequently used, which are extracted from articles published in the main issues of economy in Mexico, which contain concepts used by a group of economists interviewed on the priorities of the research of Mexican economy (see figure 2).

In the area of knowledge management, Cortes (2018) uses this to make a comparative examination of the phrases of the mission of a number of universities worldwide. From digital humanities, Lacalle & Vilar (2019) apply it to a *corpus* of academic issues on literature with the purpose of identifying mostly studied topics. Regarding social media analysis, Sued (2018) uses it to identify semantic networks among the most frequent words used for the description of a set of photographs posted in Instagram.

Paltoglou & Thelwall (2012) call sentiment analysis to the automatic process which determines whether a textual segment contains emotional or subjective expressions, whether there is positive or negative polarity. It was created in the 1990s to process the evaluations made by customers and critics regarding online product sales, it became more important over the past few years, from the interest in analyzing contents published in blogs and social networks.

Figure 2. Comparison between most frequently used words in academic articles and by inquired economists.



Source: Campos-Vázquez and López-Araiza (2018). Image reproduced with permission of the author and publisher.

In the context of social platforms, sentiment analysis is oriented to identify whether the interactions contain expressions regarding the state of mind and whether they express positive (such as enthusiasm or joy) or negative sentiments (such as disagreement or irony). It is based on lexical classifiers which estimate the level of emotional valence in order to make a prediction. Classifiers are stored in dictionaries that may be produced automatically or created by users. The software may read texts and uses an algorithm to produce an estimated dimension of the content of sentiments.

The main problem of its application on social network contents is that interactions, mainly on political and controversial topics, usually contain irony and sarcasm. These forms are not so easily identified by algorithms (Thelwall, 2017); therefore, uncontrolled use may give incorrect results. Another problem lies in recognizing language, as most of them are developed for the English language, which poses a problem to automatic recognition of Spanish structures and of other languages.

b) Visual network analysis

Visualizing and interpreting phenomena as networks is part of current digital cultures, structured as interaction spaces where the information is flowing and compared with no order or apparent hierarchies. The general research strategy designed on the basis of network structures is called visual network analysis (VNA). The network charts disclose relationships among the elements which apparently are dispersed (Venturini, Jacomy & Pereira, 2015).

VNA is centered in two essential items: nodes and edges. The former are interrelated entities; the latter are elements that build relations. In addition, they may be characterized per weight or relevance in the network. Nodes establish more relations with the others, for example, the most tweeted by others may be seen bigger, whereas the edges, such as retweeted tweets may be seen thicker (Goldbeck, 2013).

Venturini, Jacomy & Pereira have identified three basic elements for interpreting network charts, the position of nodes, size and color. This location is assigned by one of the specialization algorithms included in the user interface. The more commonly used ones are those of the force-vector, which give rise to nodes with more connections organizing the nodes in the space and minimize edge crossing, and it places them close to the nodes with more connections among them, and those disconnected are placed in the peripheral section. Thus, most of the charts will disclose an area where many nodes are assembled and there are others almost empty. Proximity and remoteness that build unequal density areas are disclosed automatically by specialization algorithms.

Network clusters are the biggest agglomeration areas. Spatializations opposite to clusters are called structural holes; their presence is an indication of disconnection among clusters, while their absence may be interpreted as a sign of an opposition. The cluster size is defined in accordance with the number of nodes it contains, and its density is from the number of edges linking the nodes. The denser a cluster is, the more consistency among its members. Clusters are tight when there are many edges and are looser when there are few of them; in the latter, what is interesting is not what joins the nodes, but what separates tight clusters and adds distance among them. In addition to clusters, nodes and edges may also be of a different size; this depends on the number of edges they generate or receive; the largest are dubbed as “authorities” of the network.

Colors are node classification operations, communities, clusters or groups among nodes that may be found in a network, which is determined by the modularity algorithm. The closer the modularity coefficient is to 1, the more community structures are contained in the network (Blonder *et al.*, 2008). There are various algorithms embedded in Gephi’s interface measuring the relationship between the nodes and the links; here, we have suggested the essentials to produce and interpret network charts.

In the survey of Twitter, VNA has been specifically used to model conversations between users through retweets and mentions, the former are considered as nodes and the latter as edges. As the centrality algorithm is applied central users of the network are identified, who may be both those who tweet more and those who are mostly retweeted.

VNA makes visible who has more conversations on the networks, how much they do and with whom; it does not identify the topic of the conversation. If the discussion derives into a dispute of senses and have an influence on media agendas, it is necessary to know, not only who exerts such influence and in what direction it is spread, but also what feelings are created.

In the survey on social platforms, several case studies have used VNA. In the traditional social survey, it has been applied to the study of non-virtual social movements, and then, it was incorporated to online movements. Tremayne (2014) analyzes the Occupy Wall Street movement which took place in New York in 2011. In Mexico, Monterde *et al.* (2015) have applied it to the *#yosoy132* (I am 132) Mexican case; Reguillo (2017) uses it to make a reflection on 2.0 social movements; Pedraza & Cano (2019) use it to map female organizations from data collected on Facebook; with the novelty format of the media lab, ITESO's Signa_Lab[2] uses it to study public conversation on topics related to the political situation.

Another way to implement VNA is to apply it to concurrent hashtags (this is when two or more hashtags are in the same publication). It is understood that these are macro terms that crystalize or synthesize discourse; the analysis of networks is combined, in this case, with the textual analysis of content, and a component is included to enable the collection of a variety of topics that are dealt with in a set of publications (Borra & Rieder, 2014).

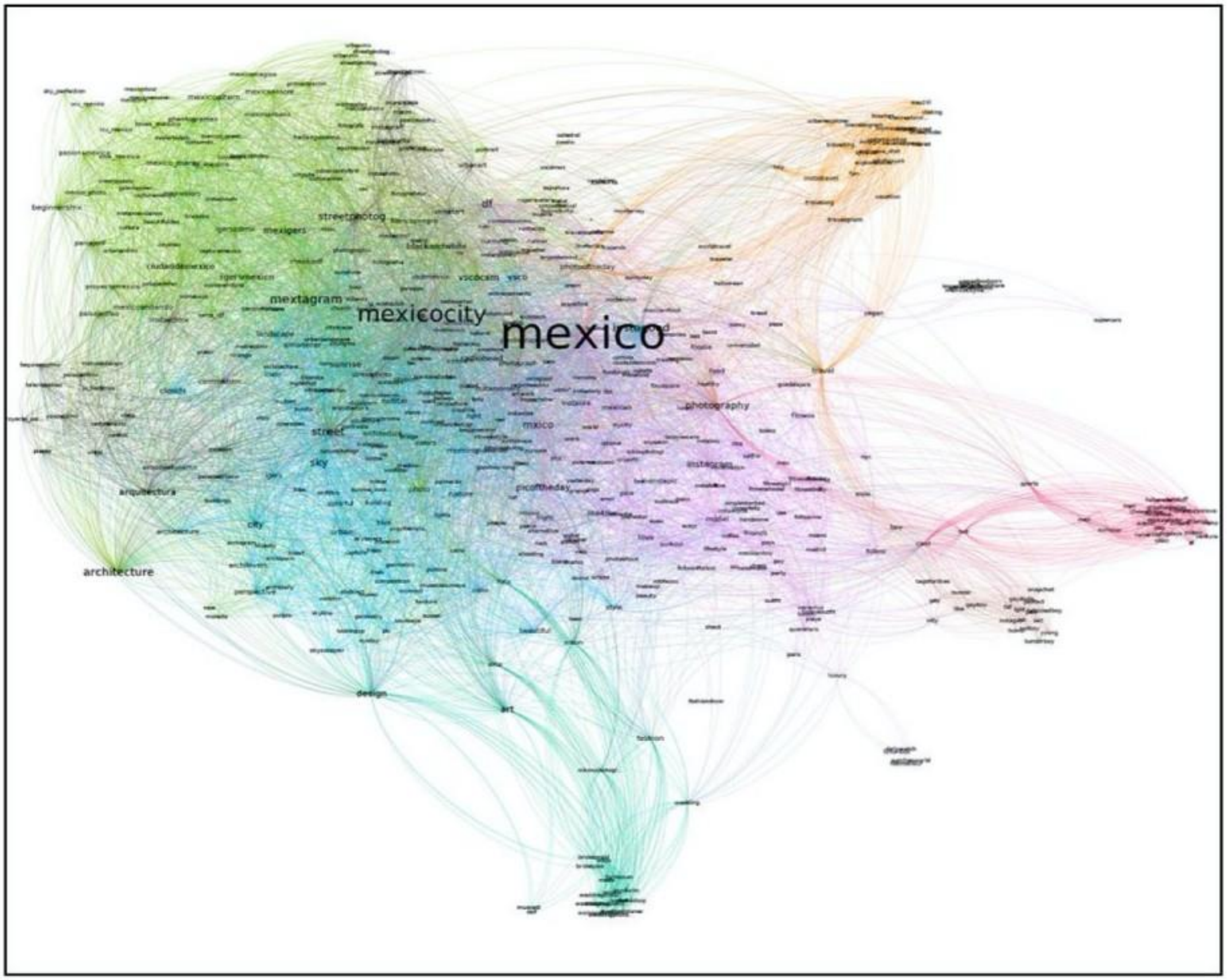
Using concurrent hashtags is more customary in Instagram than on Twitter. Sued (2018) uses it to map a network of hashtags associated to a set of 400 photographs of Mexico City published in Instagram, which enabled the determination of different thematic clusters for the images (figure 3) by means of concurrence.

c) Visual analytic techniques

In spite of the fact that social platforms are full of images, visual culture has not been the central object of digital study, especially on what is done when experimenting and disseminating technologies related with the study of themes and style linked to platforms (Highfield & Leaver, 2016).

The visual analytic concept, renamed by Manovich (2009) as cultural analytic, refers to the study of images created and uploaded by individuals on shared photography platforms such as Instagram, Flickr, Pinterest, Tumblr and even Facebook and Twitter (Niederer & Taudin, 2015). Massive publication of images occurring on these platforms represents a change of scale in contemporary visual culture and resignification both of the concept of photography and its production, circulation and consumption modes (Fontcuberta, 2011). Visual analytic studies static and dynamic, always at a great scale. The number of objects studied is important because this enable identification of patterns and regularities more easily.

Figure 3. Network of co-hashtags in #cdmx



Source: Sued (2018). Note: clusters by color. Light green: global cities. Pink: fashion, lifestyles, art and decoration. Dark green: urban art, graffiti and local tourism. Sky-blue: food and drinks. Orange: global Instagram tourism, travel and collectives. Turquoise: local tourism and local Instagram groups.

The image processing program Image-J is the instrument used by the technical phase of cultural analytic. Image-J is an open code application and was developed by the U.S. National Institute of Health, readapted for use in the field of digital humanities and the study of media by means of the Image Plot extension. The program works in two phases: in the former, it measures brightness, saturation and hue of images; in the latter, it reconstructs images as an assembly or collage, which changes them into miniatures and organizes them in accordance to measure or variable chosen by the user.[3]

Aesthetic patters of color, brightness and saturation may be identified in these assemblies or collages, although there are manners to make assemblies in accordance with previously determined content categories, the scopes of the proposal are limited to identifying aesthetic patters and, occasionally, temporary. There are other easier to use tools that may provide similar benefits, for example Image Sorter.[4]

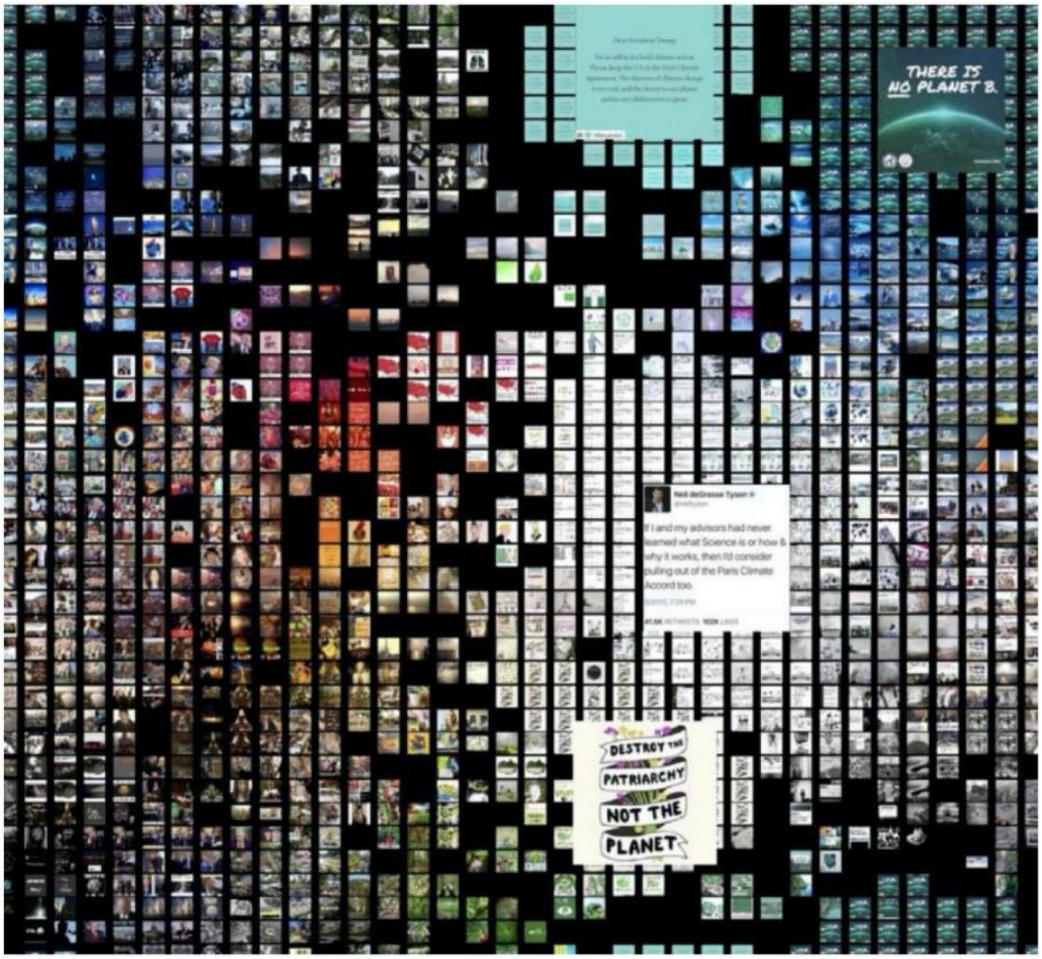
There is a paradigmatic article on the use of cultural analytic in the study of social media written by Hochman & Manovich (2013), where the visual values are compared of millions of pictures of thirteen cities, published on the Instagram platform. A more recent study is that of Pearce *et al.* (2018), who employ multi-platforms in their analysis of the content created by users on climate change.

Figure 4 shows a photograph assembly created by using the Image Sorter tool (Berthel, 2006), consisting of 1,203 images collected from the hashtag #parisagreement of Instagram. In the assembly, the images are organized per color, which enables the researcher to identify homogeneous clusters; for example, the group of white images represents scientific images, whereas the group where the red color is predominant represents empirical research images which alert about the dangers of climate change. Larger size images are mostly shared in several platforms.

In addition to visual analytic, Rose (2016) mentioned another relevant technique to analyze a large number of images: contained visual analysis, with a quantitative basis originally developed for studying verbal text. This consists of the quantification of a universe of photographs that may be classified in categories from a manual labeling procedure, where the analyst is the person who determines, in accordance with his/her research questions, the relevant categories, and then he/she is in charge of labeling and recording the procedure in a file which, in a second stage, may be presented as a clear analysis of category frequency.

Progress of the last three years in the field of computer vision has enabled the development and commercialization of automatic classification programs of contents (Sightengine, 2018). The purpose hereof is that computers may understand the world as they extract useful information from digital signals to perform complex reasoning (Wang, Komodakis & Paragios, 2013). Computer vision systems applied to image recognition are based on algorithms that employ the deep-learning technique for image recognition. Three main technological companies such as Google, Amazon and Microsoft have developed the Vision APIS application of free access for a limited time and use; however, managing the interface is still complex for large data collection (Vitale, 2015).

Figure 4. Photograph assembly: 1,203 images published in Instagram, #parisagreement



Source: Pearce *et al.* (2018). (Image design: Federica Bardelli, Carlo de Gaetano and Michele Mauri, reproduced with permission of the authors and publishers).

Online video analysis is an interesting possibility to understand participatory culture and the links with social and cultural practice, mainly because the YouTube API is available for the collection of information, which makes it possible to perform a data and metadata analytic. This is a work where the participatory culture that is built on this platform is analyzed quantitatively and qualitatively. Burgess & Green (2018) have designed a mixed methodology for the online video. The authors make a distant reading based on a *corpus* of 4,320 videos collected during three months in 2007.

The methodology does not consist in watching the videos one after the other as a close reading, on the contrary, the proposal lies in the construction of a database with video metadata. From this analysis, the authors make an interpretation of the practice of participatory cultures in YouTube, and set their eyes on the categorization of video producers, as well as on the different manners on how users express their reactions to them, and make a distinction between the actions of watching, commenting on and sharing a video.

d) Reaction metrics

Reaction metrics are defined as the way to measure how a publication on social platforms is received and put on circulation. They are based on reception indicators of users, such as *likes*, comments, shares, clicks to a web address, or the number of visualizations of content. These objects are generally presented with a numerical value that may be linked to an affective reaction (as is the case of Facebook) and, therefore, receives a quantitative treatment that answers to the logic of selecting publications from the platforms; here, it would not be relevant to ask about who saw or made comments on a specific content, but how many persons saw it, how many liked it, or how many times it was shared.

Marketing literature and digital businesses refers to reactions with the English term *engagement*, which, in these areas, is not so much a sign of user commitment as a sign of success of an online communication strategy. In this sense, the top commitment implies that the user gets involved in the publication in such a manner that it affects his/her life offline; for example, that he/she decides to visit a tourist site from the images he/she saw in a social network; and the minimum level is that he/she has only seen the picture.

Research of digital methods understands that reaction metrics is valuable to understand the interactions of social media, but it proposes a reorientation which consists of identifying what the level of interest is regarding the use on a specific topic and how this interest is expressed and constructed in terms of conversations and connections among users.

Rogers (2018) emphasizes on the social productivity aspect of networks in relation with the construction of networks of influence that justify the need to study them in terms of *engagement*. In line with a general proposal of digital methods, publication of social contents as a space where different social issues are expressed, is understood. This aspect is stressed in the use of networks for social activism as they have an influence on agendas or, in political terms, with an effect on citizens' thinking.

Gerlitz & Helmond (2013) have made an analysis on reactions from symbolic exchanges occurring among users and platforms which lead them to formulate an economy of the word *like*; while for the former the word *like* has an emotional value, to the latter, this shows a commercial value in the subject of marketing and personalized consumption. Therefore, platforms encourage the use of *like*. In accordance with the logic of economy, publication strategies of users are oriented to attain visibility and reactions, as an operation which is reproduced in social platforms.

The number of reactions from a *post* may be gathered by means of the API scraping technique or, if it is about a digital object such as a newspaper article in circulation in social networks, social monitoring tools may be used; for example, Crowdtangle[5] allows reactions to be monitored of links in circulation in Facebook and Twitter.

Data visualization techniques

The production of visualization consists in coding information in a set of basic elements such as size, form, color and position of every item included (Cairo, 2018). The author identifies three basic visualization elements: a framework, one or more visual codifications and annotations. Visual codifications are the most important elements and also the most difficult to use, they may be: height and width of the elements, their position, colors and hue variations, area, line thickness, among others.

For purposes of the techniques under review in this article we may consider that the processing and visualization stages are combined in several of them. This is what happens with word clouds produced by text mining, with photograph assemblies resulting from the visual analytic and with graphics made by using VNA. In other cases, there is a software to visualize data, generally of a quantitative nature (but not exclusively) in the form of charts. There are different types of charts: bar, lines, pie, area, Gantt diagrams, among others. Maps which include extra-geographical information and produced for specific purposes are also considered as charts.

There are several tools at hand to prepare charts. Generally, they may be developed by means of common calculation template chart tools. In addition, the Tableau Software[6] commercial package is frequently used, as well as the RawGraphs,[7] application designed by Density Design Lab of the Polytechnic of Milan; it works online, it is an open code, free and employed in the educational field. It combines standard chart types with the style based on their graphic design.

Another academic venture is Wrangler,[8] designed by the visualization group of Stanford University, used in the newspaper data area; it is also available online, it is not an open code and gives rise to a commercial visualization venture called Trifacta. Regarding the production of charts based in models, the Draw.io[9] extension for browsers is free and simple to use.

Conclusions. Scope and limitations of digital techniques

A repertoire of digital techniques was reviewed in this article for the study of contents generated by users of socio-digital networks. Taken for their production, different fields, academic and non-academic actors have developed techniques and tools to process data produced in these networks. If the techniques and tools are the result of digital marketing or the entrepreneurial environment, they may be reoriented towards academic research.

In the opposite direction, the digital tools generated at the university may be installed afterwards in extra-academic contexts; however, being software its comprising materiality, it is necessary to know that these tools work as black boxes: on the one hand, the researcher, who is not the creator but the user, delegates part of the knowledge about the operation and, on the other hand, he/she must be limited to process the information collected by the tool.

Digital techniques receive different classification modes: in accordance with the place they have in the research process, in collection techniques, cleanliness, processing and visualization; in addition, they may be distinguished in accordance with the type of object to be analyzed: text, image, video or reactions. In the information collection stage, the scraping technique allows the collection and structure of a set of unstructured data and to make a *corpus* of objects that may then be processed by using one or more techniques.

As information is processed, in accordance with the literature under analysis and the cases considered as examples for the use in this work, digital techniques presented herein are efficient to explore extensive sets of data in a quantitative manner. These techniques are necessary to make variations and recurrences of visible data sets, they enable identification of patterns and trends and, an interpretation may be given to them.

If the basic function of a technique is to operate on a set of data to extract information from them, the techniques presented herein are efficient, although not all of them have the same capacity of access and use. From those presented herein, textual mining and textual analytic techniques may be used by means of simple use tools which do not need further training for researchers; others, such as visual analysis of networks or computer vision, require of specific training in the first case, and, of interdisciplinary work with the science of data, in the second. Future works may contribute, through diverse methods, to the knowledge regarding the type of digital techniques and their research contexts in Latin America.

REFERENCES

Berthel, K. U. (2006) Image Sorter v. 2.01 [software]. Berlín: Mensch-Maschine-Kommunication.

Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R. & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), 10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Borra, E. & Rieder, B. (2014). Programmed method: Developing a toolset for capturing and analyzing tweets. *Aslib Journal of Information Management*, *66*(3), 262-278. https://doi.org/10.1108/AJIM-09-2013-0094

Burgess, J. & Green, J. (2018). *YouTube: Online video and participatory culture. Second Edition*. Cambridge: Polity Press

Cairo, A. (2018). *The elements of visualization*. Trabajo presentado en Data Visualization for Storytelling and Discovery. Curso en línea de la Fundación Knight. https://journalismcourses.org/courses/DE0618/Module1_Video2_Presentation.pdf

Campos-Vázquez, R. & López-Araiza, S. E. (2018). El estatus de la ciencia económica en México. *El Trimestre Económico*, *85*(340), 683-700. https://doi.org/10.20430/ete.v85i340.771

Carvajal, R. (2013). *¿Qué es el scraping y cómo hacerlo bien? Manual de Periodismo de Datos Iberoamericano*. http://manual.periodismodedatos.org/rigoberto-carvajal.php

Cortés Sánchez, J. D. (2018). Mission statements of universities worldwide: Text mining and visualization. *Intangible Capital*, *14*(4), 584-603. https://doi.org/10.3926/ic.1258

Elmer, G. (2015). Scraping the first person, en *Compromised Data: From Social Media to Big Data* (pp. 112–125). New York: Bloomsbury Academic.

Fielding, N. G.; Lee, R. M. & Blank, G. (2016). *The SAGE Handbook of Online Research Methods*. SAGE.

Fontcuberta, J. (2011). Por un manifiesto postfotográfico. *La Vanguardia*. http://www.lavanguardia.com/cultura/20110511/54152218372/por-un-manifiesto-posfotografico.html

Fulton, K. (2014). Screen scraping: How to stop the internet's invisible data leeches. *TechRadar*. http://www.techradar.com/news/internet/web/screen-scraping-how-to-stop-the-internet-s-invisible-data-leaches-1214404

Gandomi, A. & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gerlitz, C. & Helmond, A. (2013). *The like economy: Social buttons and the data-intensive web. New Media & Society*. https://doi.org/10.1177/1461444812472322

Goldbeck, J. (2013). *Analyzing the Social Web* (1st Edition). Morgan Kaufmann. https://www.elsevier.com/books/analyzing-the-social-web/golbeck/978-0-12-405531-5

Han, J.; Kamber, M. & Pei, J. (2011). *Data Mining: Concepts and Techniques*. Waltham, United States of America: Morgan Kauffman/Elsevier.

Highfield, T. & Leaver, T. (2016). Instagrammatics and digital methods: Studying visual social media, from selfies and GIFs to memes and emoji. *Communication Research and Practice*, *2*(1), 47-62. https://doi.org/10.1080/22041451.2016.1155332

Hine, C. (2004). *Etnografía virtual*. Barcelona: Editorial UOC.

Hine, C. (2015). *Ethnography for the Internet: Embedded, embodied and everyday*. London: Bloomsbury Academic, an imprint of Bloomsbury.

Hochman, N. & Manovich, L. (2013). Zooming into an Instagram City: Reading the local through social media. *First Monday*, *18*(7). https://doi.org/10.5210/fm.v18i7.4711

Lacalle, J. M. y Vilar, M. A. (2019). Estudios literarios y lectura distante: Un primer acercamiento a la actualidad de la investigación en las revistas académicas argentinas. *Anclajes*, *23*(1), 19-40.

Latour, B. (1992). *Ciencia en acción: Cómo seguir a los científicos e ingenieros a través de la sociedad*. Barcelona: Labor.

Manovich, L. (2009). Cultural Analytics: Visualizing Patterns in the era of more media. http://manovich.net/content/04-projects/063-cultural-analytics-visualizing-cultural-patterns/60_article_2009.pdf

Manovich, L. (2016). The Science of Culture? Social Computing, Digital Humanities and Cultural Analytics. *Journal of Cultural Analytics*. https://doi.org/10.22148/16.004

Marres, N. & Gerlitz, C. (2016). Interface Methods: Renegotiating Relations between Digital Social Research, STS and Sociology. *The Sociological Review*, *64*(1), 21-46. https://doi.org/10.1111/1467-954X.12314

Marres, N. & Weltevrede, E. (2013). Scraping the Social? Issues in live social research. *Journal of Cultural Economy*, *6*(3), 313-335.

Marres, N. (2017). *Digital Sociology: The Reinvention of Social Research*. Cambridge, United Kindom: Polity Press.

Monterde Mateo, A.; Carrillo Martin, R.; Esteve del Valle, M. & Aragón Aragón, P. (2015). #YoSoy132: ¿Un nuevo paradigma en la política mexicana? IN3 Working Paper Series. http://in3-working-paper-series.uoc.edu/in3/es/index.php/in3-working-paper-series/article/view/2066.html

Moreno, A. & Redondo, T. (2016). Text Analytics: the convergence of Big Data and Artificial Intelligence. *International Journal of Interactive Multimedia and Artificial Intelligence*, *3*(Special Issue on Big Data and AI), 57-64. https://doi.org/10.9781/ijimai.2016.369

Nieborg, D. B. & Poell, T. (2018). The platformization of cultural production: Theorizing the contingent cultural commodity. *New Media & Society*, *20*(11), 4275-4292. https://doi.org/10.1177/1461444818769694

Niederer, S. & Taudin Chabot, R. (2015). Deconstructing the cloud: Responses to Big Data phenomena from social sciences, humanities and the arts. *Big Data and Society*, *2*(2). http://journals.sagepub.com/doi/full/10.1177/2053951715594635

Paltoglou, G. & Thelwall, M. (2012). Twitter, MySpace, Digg: Unsupervised Sentiment Analysis in Social Media. *ACM Trans. Intell. Syst. Technol.*, *3*(4), 1-19. https://doi.org/10.1145/2337542.2337551

Pearce, W.; Özkula, S. M.; Greene, A. K.; Teeling, L.; Bansard, J. S.; Omena, J. J. & Rabello, E. T. (2018). Visual cross-platform analysis: Digital methods to research social media images. *Information, Communication & Society*. https://doi.org/10.1080/1369118X.2018.1486871

Pedraza Bucio, C. I. & Cano Rodríguez, C. A. (2019). Resistencias sumergidas. Cartografía de la tecnopolítica feminista en México. Teknokultura. *Revista de Cultura Digital y Movimientos Sociales*, *16*(2), 197-212. https://doi.org/10.5209/tekn.64163

Popping, R. (2016). Online Tools for Content Analysis, en *The Sage Handbook of Online Research Methods* (pp. 329–343). London, New York: SAGE.

Reguillo Cruz, R. (2017). *Paisajes insurrectos: Jóvenes, redes y revueltas en el otoño civilizatorio*. Barcelona: NED Ediciones.

Robichaud, A. & Blevins, C. (2011). 4: Content-Based Analysis » Tooling Up for Digital Humanities. Tooling Up Digital Humanities. http://toolingup.stanford.edu/?page_id=205

Rogers, R. (2009). The End of Virtual. Digital Methods. Amsterdam: Vossiuspers UvA. http://www.govcom.org/rogers_oratie.pdf

Rogers, R. (2013). *Digital methods*. Boston: MIT Press.

Rogers, R. (2015). Digital Methods for Web Research, en *Emerging Trends in the Social and Behavioral Sciences*. https://doi.org/10.1002/9781118900772.etrds0076

Rogers, R. (2018). Otherwise Engaged: Social Media from Vanity Metrics to Critical Analytics. *International Journal of Communication*, *12*(0), 23. https://dare.uva.nl/search?identifier=e7a7c11b-b199-4d7c-a9cb-fdf1dd74d493

Rogers, R. (2019). *Doing Digital Methods*. London: SAGE.

Rojas Crotte, I. R. R. (2011). Elementos para el diseño de técnicas de investigación: una propuesta de definiciones y procedimientos en la investigación científica. *Tiempo de educar*, *12*(4), 277-297.

Rose, G. (2016). *Visual methodologies: An introduction to researching with visual materials* (4ta. edición). Londres: SAGE.

Sightengine. (2018). Benchmarking Google Vision, Amazon Rekognition, Microsoft Azure on Image Moderation. *Medium*. https://medium.com/sightengine/benchmarking-google-vision-amazon-rekognition-microsoft-azure-on-image-moderation-73909739b8b4

Solomon, B. (2016). How Tableau Built A $3 Billion Data Empire on Top of Beautiful Charts. *Forbes*. https://www.forbes.com/sites/briansolomon/2016/05/04/how-tableau-built-a-3-billion-data-empire-on-top-of-beautiful-charts/

Sued, G. (2018). Métodos digitales para el estudio de la fotografía compartida. Una aproximación distante a tres ciudades iberoamericanas en Instagram. Empiria. *Revista de metodología de ciencias sociales*, *0*(40), 15-39. https://doi.org/10.5944/empiria.40.2018.22009

Thelwall, M. (2017). Heart and soul: Sentiment strength detection in the social web with SentiStrength. In J. Hlys (ed) Cyberemotions: *Collective emotions in cyberspace* (pp. 437–445). http://sentistrength.wlv.ac.uk/documentation/SentiStrengthChapter.pdf

Tremayne, M. (2014). Anatomy of Protest in the Digital Era: A Network Analysis of Twitter and Occupy Wall Street. *Social Movement Studies*, *13*(1), 110-126. https://doi.org/10.1080/14742837.2013.830969

Underwood, T. (2013). *Wordcounts are amazing. The Stone and the Shell*. https://tedunderwood.com/2013/02/20/wordcounts-are-amazing/

Venturini, T.; Jacomy, M. & Pereira, D. (2015). *Visual Network Analysis*. http://www.tommasoventurini.it/wp/wp-content/uploads/2014/08/Venturini-Jacomy_Visual-Network-Analysis_WorkingPaper.pdf

Vitale, A. (3 de diciembre 2015). What Google Cloud Vision API means for Deep Learning Startups. *Medium*. https://medium.com/google-cloud/what-google-cloud-vision-api-means-for-deep-learning-startups-cd39226922e5

Voyant Tools. (2018). En Wikipedia. Wikipedia. https://en.wikipedia.org/w/index.php?title=Voyant_Tools&oldid=845947425

Wang, C.; Komodakis, N. & Paragios, N. (2013). Markov Random Field modeling, inference & learning in computer vision & image understanding: A survey. *Computer Vision and Image Understanding*, *117*(11), 1610-1627. https://doi.org/10.1016/j.cviu.2013.07.004

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, *59*(10), 1-23.

[1] See http://voyant-tools.org/

[2] See https://signalab.iteso.mx/

[3] Tutorial about Image Plot at http://lab.softwarestudies.com/p/imageplot.html

[4] See http://www.tucows.com/preview/510399/ImageSorter

[5] See https://apps.crowdtangle.com/chrome-extension

[6] See https://tableausoftware

[7] See https://rawgraphs.io/

[8] See https://www.trifacta.com/start-wrangling/

[9] See https://www.draw.io/