
From Data Analytics to Data Hermeneutics

Online Political Discussions, Digital Methods and
the Continuing Relevance of Interpretive Approaches

Paolo Gerbaudo

Abstract

To advance the study of digital politics it is urgent to complement data analytics with data hermeneutics to be understood as a methodological approach that focuses on the interpretation of the deep structures of meaning in social media conversations as they develop around various political phenomena, from digital protest movements to online election campaigns. The diffusion of Big Data techniques in recent scholarship on political behavior has led to a quantitative bias in the understanding of online political phenomena and a disregard for issues of content and meaning. To solve this problem it is necessary to adapt the hermeneutic approach to the conditions of social media communication, and shift its object of analysis from texts to datasets. On the one hand, this involves identifying procedures to select samples of social media posts out of datasets, so that they can be analysed in more depth. I describe three sampling strategies – top sampling, random sampling and zoom-in sampling – to attain this goal. On the other hand, “close reading” procedures used in hermeneutic analysis need to be adapted to the different quality of digital objects vis-à-vis traditional texts. This can be achieved by analysing posts not only as data-points in a dataset, but also as interventions in a collective conversation, and as utterances of broader “discourses”. The task of interpretation of social media data also requires an understanding of the political and social contexts in which digital political phenomena unfold, as well as taking into account the subjective viewpoints and motivations of those involved, which can be gained through in-depth interviews, and other qualitative social science methods. Data hermeneutics thus holds promise for a closing of the gap between quantitative and qualitative approaches in the study of digital politics, allowing for a deeper and more holistic understanding of online political phenomena.

Keywords: Big data; digital politics; hermeneutics; data analytics
qualitative methods; random sampling; close reading.

Introduction

The digital transformation of our societies has not just transformed the “ontology” of politics, i.e. the nature or essence of contemporary political phenomena in a variety of domains, from social movements using data leaks as protest tactics to political campaigns mastering the art of data-driven targeted advertising. It has also transformed the “epistemology” of political research, i.e. the methods used to analyse political phenomena as they unfold online. If opinion surveys, statistical studies of electoral behaviour, content analysis of political speeches and news broadcasts have long been the tools of the trade for political analysts, we are now witnessing the development of an array of digital methods in political research – sophisticated computational techniques that are employed to sift the ocean of information contemporary politics is immersed in.

The most popular method in the emerging field of digital politics – the fledgling field of scholarship that explores the variety of political phenomena enabled by digital technology – is undoubtedly data analytics. Data analytics can be described as a series of techniques for statistical and computational analysis of various types of digital datasets (Kambatka 2014). Social scientists have used various kinds of econometric tools to analyse the different aspects of social media conversations, typically including: their network structure (see for example, González-Bailón et al. 2013); the temporal evolution of conversations (Conover et al. 2013); the process of information diffusion (Theocharis 2013); and the correlation of various popularity metrics, such as the number of likes or retweets (Dubois et al 2014).

Data analytics has enriched research on digital politics, by providing sophisticated ways to study and visualise online political dynamics. However, it has fast become a sort of methodological orthodoxy in the field, leading some researchers to overlook its manifold limits and shortcomings (Tufekci 2014; boyd/Crawford 2014). More importantly, its quantitative bias has contributed to marginalizing questions of cultural meaning and social motivation, which are fundamental to understand the content of social media conversations (Couldry/Stephansen 2014).

This article aims at fleshing out an alternative, qualitative approach to the study of online political phenomena that addresses some of the problems with data analytics. This is what I term “data hermeneutics”, an updating of the interpretive methods originating from a number of disciplines including phenomenological philosophy, literary criticism, qualitative sociology and cultural anthropology – to deal with the specific properties of digital communication and social media datasets. Political research cannot content itself with ever more sophisticated forms of computational analysis of political behavior online. It also needs to answer qualitative questions about the “who”, “what”, “how” and “why” of digital political phenomena. If it has to have any real explanatory power, it has to pay attention to the meaning and subjective viewpoints inherent in social media conversations and the contexts in which these conversations occur.

The tradition of hermeneutics with its focus on structures of deep meaning and subjective worldviews reflected by them, provides with an inspiring but outdated blueprint on how to pursue such endeavour. To make this approach current again it is necessary to revise it to match the peculiar conditions of digital communication, which has a non-linear, dialogic, extemporaneous, and interactive character. Whereas traditional hermeneutics has dealt with the analysis of various traditional “texts” (novels, films, political speeches, news articles), data hermeneutics needs to find ways to analyse online content, approaching data as “inscriptions” (Ricoeur 1971) or recorded traces of a peculiar form of social text: social media conversations.

I discuss two practical aspects of this digital adaptation of hermeneutics: 1) the development of qualitative sampling procedures geared at reducing the size of social media datasets; 2) the development of “close data reading” procedures that may help interpret in the context of larger conversations and in relation to a number of connected discourses, narratives, motivations and worldviews.

First, data hermeneutics requires sampling procedures aimed at reducing the size of the datasets to a scale amenable to qualitative analysis. I discuss three such procedures: top sampling, random sampling and zoom-in sampling. In the first case the database is filtered for the top messages based on a number of popularity metrics, such as number of retweets or of likes. In the second case the sample is obtained by selecting a random set of posts, (tweets or Facebook posts), which can be considered representative of a conversation. In the third case, the researcher zooms in on a particular point in the conversation deemed to be particularly significant – for example a spike in user engagement. These three procedures, which can be used in combination, provide researchers with a “select social media dataset” that can then be analysed in more depth.

Second, “close reading” and “in-depth analysis”, expressions that condense the *modus operandi* of hermeneutics, need to be revised to match the properties of social media data vis-à-vis traditional texts. I propose a three-step “close data reading” procedure: reading posts as rows in a dataset; as part of the conversation; as part of a certain social discourse. First, the researcher reads posts as data-points in a dataset, paying attention to specific content and stylistics of a given post. Second, she approaches posts as exchanges in the conversation they were “scraped” from, paying attention to their embeddedness in a dialogic communication. Third, a full understanding of the meaning of social media posts and conversations requires interpreting their themes in light of the cultural, social and political contexts they navigate. To this end researchers need to take into account the subjective viewpoints and motivations of participants, by using more traditional qualitative methodologies, such as in-depth interviews.

The article begins by reviewing the current debate on digital methods in digital politics research, focusing on a number of critical issues, and in particular the neglect for issues of meaning. The second part of the article moves instead to flesh out in the positive the data hermeneutics approach, by presenting connected

sampling procedures and analytical techniques. The conclusion sums up the content of my methodological proposal and considers some of the questions and challenges ahead for digital politics research.

Beyond the quantitative bias of data analytics

Methodology in the emerging field of digital politics is dominated by “data analytics” (Kambatka et al. 2014), a set of statistical and computational techniques for the analysis of large social media datasets. The development of data analytics has been led by major internet firms as Facebook, Amazon and Twitter, which in recent years have created data science teams to develop sophisticated market intelligence (Chen et al. 2012). Social scientists, and more specifically political researchers, have been fast to follow this trend, using data analytics to study what happens in social media conversations, relevant to various political phenomena, from social movements to electoral campaigns and political debates (Theocharis 2013; González-Bailón et al. 2013; Conover et al., 2013). Data obtained by “mining” from the APIs of social network sites is subject to various forms of statistical and computational analysis. Among the most typical issues analysed via data analytics feature the structure of social media conversations often pictured through network diagrams; their clustering into distinct groupings sharing common features; the distribution and correlation between different social media metrics (followers, likes, shares, retweets etc.); the temporal evolution of conversations (peaks, lows, etc.); the frequency of key terms used in conversations.

Data analytics has provided innovative and sophisticated techniques to renew and advance sociological methods, beyond the current crisis of empirical sociology (Burrows/Savage 2007). It can be understood as a necessary update to traditional quantitative social science data such as surveys (Kitchin, 2014; Tinati, et al. 2014), facing up to the new methodological challenges posed by the abundance of social media data, and allowing to explore new social and political phenomena that have evident interest for research. Furthermore, data analytics offers great advancements the detail and sophistication of quantitative analysis, allowing researchers to conduct studies “at scale” (Kambatla 2014), examining the entire “population” of a given phenomenon or a close approximation.

These positive elements notwithstanding, data analytics has demonstrated significant shortcomings, which call for the development of alternative approaches. After an initial phase of naïve enthusiasm, in fact in recent years this methodological approach has begun to be criticized by a number of scholars, who have highlighted how data analytics frequently falls short of the standards expected in quantitative research (Tufekci 2014; Tinati et al. 2014; Lazer et al. 2013). Zeynep Tufekci has argued that in many circumstances there has been a “lack of clarity with regard to sampling, universe and representativeness” (Tufekci 2014: 324). Similarly Kate Crawford and danah boyd have highlighted that “claims to objec-

tivity and accuracy are often misleading” and that it is erroneous to think that “bigger data is always better data” (2014: 217).

There is however a more fundamental critique that can be made against data analytics: its quantitative bias and its neglect of issues of meaning. By focusing on the mathematical *form* of social media conversations – its structure and dynamics – data analytics tends to overlook their *content*, and the deep meaning structures expressed in them and the motivations they betoken. Data analytics can reveal with great sophistication the mathematical properties of datasets, but it is not well equipped to answer qualitative questions about the “what”, the “how” and the “why” of social media conversations. Also when content is analysed, as in computer-assisted textual analysis (Grimmer/Stewart 2013), the logic remains one of “counting occurrences”, for example measuring frequency of certain terms, or the emotional tone of a conversation. This quantitative bias is exacerbated by the fact that data analytics-driven politics research often lacks in context (Tufekci 2014; Crawford/boyd 2014). To quote Clifford Geertz sometimes it is as if data analytics researchers seem to think that it is possible to understand phenomena without *knowing* them (1957: 64).

Looking at social media data as texts

The shortcomings of data analytics should be seen by qualitative researchers as evidence for the continuing relevance of interpretive methods. Social media data is not inherently hostile to a qualitative research agenda, and potentially offers a treasure trove for qualitative research. The abundant and fine-grained information that can be retrieved from online conversations, including textual information, pictures, videos, and other digital objects is a resource that waits to be tapped from a qualitative angle. Indeed some qualitative researchers have begun working in this direction, as seen in the field of online discourse analysis (Androutsopoulos/Beißwenger 2008; Kelsey/Bennett 2014). Yet, much still remains to be done in devising effective strategies and procedures for the analysis of social media data that can make qualitative methods relevant once again.

Data hermeneutics is a notion that proposes a radically different orientation from the one pursued by data analytics, by giving a new lease of life to one of the most old-fashioned of interpretive methods: the hermeneutic approach. Where data analytics operates according to a purely objectivist view of online political conversations, conceived of as forms of collective behaviour which can be objectively measured, data hermeneutics – alike all interpretive methodologies – operates with the idea that these conversations are first and foremost symbolic interactions which cannot be understood without taking into account the subjective viewpoints of those involved. Where data analytics – as in fact all forms of analytics – involves various forms of numerical analysis, drawn from the field of statistics and computation, data hermeneutics centers on the symbolic analysis

of the meaning structures of online conversations, in light of connected social discourses and motivations. Finally, where – as the very etymology of the term analytics suggests – data analytics is mostly interested in *analyzing* that is in “breaking down” a given phenomenon into basic units and variables for statistical analysis, data hermeneutics’ chief concern is the synthetic aim of interpreting, reconstructing and explaining the overarching narratives that underpin social media conversations.

Data hermeneutics is a digital adaptation of hermeneutics, a term that derives from the ancient Greek word “hermeneuo” which means to “understand” or to “interpret” and can be described as a broadly developed set of methodologies for interpretation (Szondi 1975). The origins of this approach hark back to Greek antiquity and medieval philosophy, with its interest in the different literal and allegorical layer of meaning of sacred texts. In modern times hermeneutics has been associated with phenomenological philosophy, and in particular the work of Edmund Husserl (1970), Martin Heidegger and his pupil Hans Georg Gadamer (2004), and has influenced a wide array of social and political theorists, including Max Weber (1978; 1981), Walter Benjamin (1977), Fredric Jameson (1981) and Anthony Giddens (1984). The hermeneutic interpretive approach has also profoundly shaped research methods in the humanities and social sciences, as most evidently seen in the context of new criticism, qualitative sociology and social anthropology (Bleicher/Bleicher 1980).

In the context of literary criticism (Bressler 1999), hermeneutic approaches have informed the development of “close reading” procedures (Wolfreys 2000). Close reading can be described as a process of deep analytical engagement with a text – a novel, a poem, but also a film or any other similar artefact – with the aim of exploring the complex network of meaning that underpins it. Thus, for example, in analyzing a novel, a film, or a political speech, literary scholars look at their content and formal characteristics, such as the language, tone, imagery, and rhetorical figures. While sometimes – as in the case of New Criticism and semiotics – this analysis can take an purely formalistic character, close reading has also been used in the context of more sociologically minded discourse analysis approaches, as performed for example in the context of cultural studies and sociology (Wodak/Krzyzanowski 2008), looking for the connection between specific texts and broader discourses.

In social science, the hermeneutic approach has been popular among both sociologists and anthropologists. Key in this respect has been Max Weber’s view of sociology as an interpretive science, different from the natural sciences, because the subjects and objects of analysis are both human beings, bestowed with consciousness and reflexivity (Tucker 1965). The priority for the social sciences, vis-à-vis the natural sciences, is *Verstehen*, the in-depth understanding of the subjective viewpoints, motives and worldviews that inform social action, rather than action as just an externally observable behaviour (Weber 1978; 1981). Anthony Giddens expanded on this idea by coining the notion of “double hermeneutics”,

which highlight how social science studies does not study only people's behaviour but also people's interpretation of the social world and of their social action, and thus revolves around an interpretation of already existing interpretations, or the scholarly interpretation of lay conceptions (1984: 20).

The hermeneutic approach has also been taken up by social anthropologists such as Clifford Geertz, who argued that in producing "thick descriptions" of communities and social practices, anthropologists should take into account "the interpretations to which persons of a particular denomination subject their experience" (1973: 15). He argued that interpretive methods required an effort to "find one's feet" (Geertz 1973: 18) in the phenomena, or to use another metaphor "put oneself in the shoes" of the actors and communities analysed, looking at people as conscious and creative subjects rather than objects prey to social forces they cannot control. This interpretive orientation has been at the heart of qualitative methods such as in-depth interviews and focus groups and has informed the development of the now popular "grounded theory" approach (Strauss/Corbyn 1998) in which researchers are expected to develop their interpretation of social phenomena in a bottom-up manner, rather than testing a-priori hypotheses.

In recent years there have been some inklings of a digital adaptation of hermeneutics, as signaled by terms as "computational hermeneutics" (Harnad 1990; Mohr et al. 2013) and "digital hermeneutics" (Capurro 2000). Rafael Capurro for example has argued that hermeneutics needs to face up to the challenge of digital technology, and develop an "understanding [of] the foundations of digital technology and its interplay with human existence" (ibid: 37). This article contributes to this emerging line of methodological reflection and practice, by exploring specific strategies and procedures for the specific purpose of social media analysis and digital politics research.

A digital adaptation of hermeneutics, does not simply entail saying that hermeneutics needs to "find its own feet" in the digital world, but also that to understand the digital world it is necessary to recuperate the concern with interpretation which is ultimately hermeneutics' *raison d'être*. This assertion is a highly contentious one, due to the anti-interpretive character of the ideology of Big Data, or "dataism" (Van Dijck 2014) – and the idea that data is already a ready-made form of knowledge which does not require active interpretation. This persuasion has been put forward most explicitly in a famous article by *Wired* magazine editor Chris Anderson, where he argued that in the present data deluge "[c]orrelation is enough. We can stop looking for models. We can analyze the data without hypotheses about what it might show" (2008). This intervention has been criticized as going too far by other Big data experts (Bollier/Firestone 2010; Cukier/Mayer-Schönberger 2013: 72). Yet, it is interesting precisely because it reveals in condensed form the overly positivist and anti-hermeneutic stance of data science.

Anderson's prophecy about the end of theory and interpretation neglects a number of facts. First, there is no such a thing as "raw data", which is in fact an "oxymoron" (Gitelman 2013), since data is always structured in higher order

categories that reflect various biases and assumptions as expressed in the DIKW (Data, Information, Knowledge, Wisdom) hierarchy used in Informatics (Rowley 2007). Secondly, datasets display a number of social and political biases that reflect and sometimes amplify social inequality (O’Neil 2016), and can be identified only through processes of in-depth interpretation. Thirdly, the overabundance of data makes the task of interpretation particularly important. As argued by Alyssa Wise and David Shaffer “with larger amounts of data, theory plays an ever-more critical role in analysis” (2015: 5). Therefore, rather than hastily throwing interpretation out of the window, it is urgent to revive and revise interpretive methodologies to match the conditions of a digital era.

The main challenge data hermeneutics is to shift from texts to data as the main object of analysis; or better to find ways to read data as text, that is as a partly coherent and discrete web of meaning. Interpretive approaches have traditionally been concerned with analyzing texts – novels, poems, films, speeches, interviews, field-notes, etc. – by examining them in great depth sentence by sentence, one might say, or even word by word, as signified by notions as “close reading” and “in-depth analysis” frequently referred to as methodological short-hands. Social media conversations may indeed analysed in ways similar to the analysis of traditional texts, such as by exploring their language, imagery, tone, and other stylistics. However, significant modifications are necessary due the specific nature of social media as objects of analysis.

With their non-linear, extemporaneous, and interactive nature, social media conversations are radically different from a novel, a film, or an ethnographic field-note. Consider for example the way in which social media resemble more oral conversations, rather than written texts; the way in which each tweet or Facebook message can hardly be understood in isolation from other messages; the sheer quantity of social media messages and the connected risk of information overload; the speed and instantaneity of conversations; their fluid and networked character; or the way in which the various interactions available on social media (such as liking, retweeting, or favouriting) add another layer of meaning that was unknown in pre-digital texts (Van Dijck/Poell 2013). These idiosyncrasies of online communication pose serious challenge to interpretive approaches and require significant adaptations. In the continuation of this article I focus on two practical issues relevant in the development of data hermeneutics: a) “small data” sampling methods; b) and “close data reading” procedures.

Sampling social media datasets for qualitative analysis

The main obstacle for data hermeneutics lies precisely in the “Big-ness” of Big Data, in the vastness of datasets available to researchers. While this is the aspect that makes social media datasets so interesting for quantitative researchers due to their great level of detail and the possibility to study conversations “at scale”,

it is also the element which is most problematic for qualitative researchers who are used instead to work with “small data” (Couldry/Stephansen 2014). Qualitative researchers are expected to engage at length with research material, exploring the fine-grained meaning structure of texts and connected discourses. This approach limits the amount of evidence which can be analysed. The employment of textual analysis software such as NVivo or ATLAS.ti (Friese 2014), provides only partial solace. Ultimately effective close reading continues to imply a great deal of “manual” coding by “human operators”. From this situation it follows that the main challenge for data hermeneutics is one of focus: reducing the amount of data to analyse, to a selected sample which can still be considered significant and representative of a given aspect of the conversation.

Three sampling procedures can be used to perform this task: top sampling, random sampling, and zoom-in sampling. First, one may decide to sample for top, by focusing on the messages which – based on a number of popularity metrics (likes, retweets, shares) – can be considered as the most visible or important in a given conversation. Second, a different strategy involves random sampling, selecting by chance a subset of messages from a given conversation using appropriate software. Third, “zoom-in sampling” involves concentrating on a particular period of time in the conversation, which for whatever reason is considered particularly significant (start dates of protest waves, election days, etc.). Each of these sampling procedures provides a different approximation to a data sample for qualitative analysis, and will therefore benefit different research questions and designs.

Top sampling is a strategy that has already been utilised by researchers, especially those interested in the behaviour of “power users”, user that have disproportionate level of influence on online conversations (Cha et al. 2010), on the basis of a number of popularity metrics (likes, retweets, favourites etc.) Practically, sampling at the top is fairly easy and can be performed by using standard spreadsheet software such as Microsoft Excel, selecting the column of the variable one takes as the most indicative of the popularity of messages, ordering in descending order, and then filtering the top 50, 100, 200, 1000 messages. The viewpoint over a conversation offered by this procedure is evidently biased. It only affords an understanding of what happens “at the top” of a conversation and its most influential users and messages. This procedure is particularly suitable when it comes to highly public and visible conversations which tend to have a strong power law distribution and few communication centres (Gonzalez-Bailon et al. 2011; 2013). However, it is less desirable when analysing less topical conversations, and cannot be considered representative of the average of a given conversation. It is good to look at the peak of conversations, not to explore explore the “base”.

The second sampling procedures is random sampling. Random sampling is an already well-rehashed sampling strategy in the social sciences (Patton 2005) which involves selecting by chance a sub-set of a given population. This approach can be updated in a digital context by using a number of digital tools such as T-CAT tool developed by the Digital Methods Initiative (DMI) at the University of

Amsterdam. This type of sampling strategy has many advantages. It can return a sample that can be considered representative of the totality of messages contained in a given dataset. As it is the case with random sampling more generally, the representativeness of the sample depends on the ratio between the size of the population analysed and the size of the dataset: the greater the ratio, the greater the risk the sample may not be truly representative (Marshall 1996). This approach can be used if one is interested in getting a general sense of the type of messages to be found at the “base” of a conversation, including messages with relatively low popularity.

The third sampling strategy is zoom-in or peak sampling, a sampling procedure that focuses on a given time in the conversation that is of particular interest to the researcher. These may include online reflections of “real world” events (a protest event, or an election) or moments of high user engagement on social media, or any other event considered of particular significance to understand the dynamics of the conversation. In my own research I adopted this strategy to look at the online preparation of major protest events in the Arab Spring and the Indignados, where one could see the build-up of “digital enthusiasm” (Gerbaudo 2016). The advantage of zoom-in sampling is that it concentrates on moments that can be particularly revealing of a number of digital political dynamics, such as the nexus between social media and mobilization, or the social media reflection of offline events. Its main disadvantage is obviously its temporally limited coverage, and the fact that it thus returns a selective image of the conversation.

These three sampling procedures may be used in combination in the design of concrete research projects. For example, if one is interested in the way in which key top social media accounts reacted to a certain incident, zoom-in sampling and top sampling may be utilized in concert. The combination of two or more sampling procedures can also help more easily achieve the aim of “dataset reduction” which – as we have previously seen – is a condition of possibility for data hermeneutics. As a rule of a thumb, based on my own experience conducting various digital politics research projects, when sampling from large datasets researchers should aim for a dataset numbering between 1000 and 100,000 words, which roughly equates between 40 and 4,000 tweets. This corpus is comparable in size to the ones traditionally studied by qualitative researchers – novels, films, political speeches, and the like – and small enough to be analysed in-depth by a “human operator” without falling prey to information overload.

Data close reading

Besides reducing the size of datasets, data hermeneutics entails a rethinking of the procedures traditionally used to analyse texts, for the purpose of adapting them to social media data analysis. This is what can be described as “close reading of data” or “data close reading”, an adaptation of close reading procedures to the

specific conditions of social media communication. Hermeneutic researchers in the humanities and social sciences have typically studied texts, where the notion of text is not limited to written texts, but to all compositions, artworks, and social performances that can be understood as relatively discrete and coherent symbolic objects. In the case of literary criticism, typical texts have included novels, poems and films; in the social sciences, everyday behaviour, public performances and similar phenomena can also be read as texts, often by producing textual accounts of them, such as ethnographic field-notes. In approaching these texts qualitative researchers have typically aimed for an in-depth engagement with the object of analysis. This stance is most clearly revealed by the term “close reading” used in literary criticism, which highlights how researchers are expected to explore texts in great detail, deciphering their complex and largely invisible deep meaning structures, and their connection to broader narratives and discourses.

The analytical process in qualitative research typically involves the use of various “coding” procedures whereby the researcher marks out certain portion of text as belonging to certain overarching themes, which can then be organized in broader categories and narratives. As argued by Johnny Saldana, a code in this context can be described as “a word or short phrase that symbolically assigns a summative, salient, essence-capturing, or evocative attribute for a portion of language-based or visual data” (2015: 4). This methodology has been applied to such diverse types of data as interviews, ethnographic observations, films, and newspaper articles. It is easy to understand how such approach can be used to analyse traditional texts, or social events and rituals that are relatively circumscribed, coherent and mostly linear in their form. But how does close reading functions when studying social media data?

Close data reading needs to approach social media data as “texts”, considering data-points, e.g. Facebook posts or tweets, as meaningful messages. Reflecting back on my experience conducting digital politics research I propose that close data reading should proceed in different steps, allowing to progressively “close in” on the deep meaning structures of posts and conversations. Selected social media datasets, obtained through the sampling procedures previously described, can be analysed in three steps which all imply a different “gaze” on research data: reading posts as rows in a dataset; as part of the conversation; as part of a certain social discourse.

First, Facebook posts and Twitter posts can be read as rows in a dataset, the practical form in which they manifest themselves to qualitative researchers at the start of a project. Researchers typically view such datasets in the form of a long table of rows, to be browsed through via spreadsheet software or via qualitative analysis software such as NVivo or ATLAS.ti. Thereby posts appear in their rawest form as mere rows listed in a table, chunks of interactions now stripped from their surrounding context of a live conversation, and which now appears in the dead form of data-points. A number of elements can however already be identified at this stage. Researchers can explore the topics discussed in each post, as well

as the form they are expressed in, such as the use of a certain type of language, imagery, tone, or specific rhetorical figures. In the case of my own research on the 2011 protest movements, this step of analysis already allowed to identify a number of significant features of online protest discourse including: the use of a conversational and exhortative language (2016); the popularity of different memes (2015); the adoption of the entire gamut of typical social media tropes from emoticons, to sloganeering shortened sentences; and in terms of content the abundance of reference to unifying subjects, as “the people”, the citizenry, or the 99% (2014).

The second level of analysis involves repositioning posts in their original environment, reading them in the “live” context of the conversations. This step approaches social media data as the traces or inscriptions of a specific type of social text: a social media conversation. Yet, to better understand the meaning of posts it is necessary to approach these data-points not as texts in and of themselves, but rather as traces or “inscription” (Ricoeur 1972), that is the partial and largely arbitrary recordings of live social media conversations. To get a sense of how a message was perceived by internet users it is necessary to read it in the context of the conversations in which it was uttered. Practically, this can be done by browsing the associated web address and exploring the conversations it contributed to. It is useful for this purpose to create a folder of screenshots of conversations, one can then refer back to during the analysis. In my research, this step of analysis helped me to understand the intense emotional dialogue developed in such movements as the Egyptian 2011 uprising and the Spanish Indignados, and the way users reinforced positive messages channeled by activist social media, thus fuelling a wave of “digital enthusiasm” (2016). By reading posts in their lived contexts, we can understand the degree to which they are in-tune with the mood of internet communities, and the content of the dialogic discourse that emerges out social media interactions.

The third step for the close reading of data – the one which is more similar to traditional coding procedures – explores the discourses and deeper structures of meaning of a given post. Like with step two, the gist is to avoid reading posts in isolation. However, in this case the connections one needs to pay attention to are not just the ones with other messages within a specific conversation, but the links between the message and broader discourses that act as background and source of meaning for a given message. Important to this end is that researchers acquired an understanding of the context in which digital political phenomena operate and the subjective motives of participants as it can be secured by tapping into more traditional qualitative methods. In the case of my own research the most important traditional methods used in tandem with data hermeneutics were in-depth interviews with protest movement participants active on social media which allowed to gain an in-depth understanding of their subjective viewpoints and backgrounds. Further contextual sources of information that can aid interpretation include ethnographic observation, archival documents, and similar sources of background information about a given phenomenon.

A practical example of how this cultural understanding can aid the work of social media interpretation can be offered considering the famous “Bring tent” tweet by the Canadian countercultural magazine *Adbusters* that first launched the Occupy Wall Street movement. In and of itself this tweet seems to have quite limited meaning, too short and cryptic to allow for analysis. Yet, seen in the cultural context of contemporary social movements, of post-modern neo-anarchism and of a nascent Occupy Wall Street movement, some interesting elements can be inferred from it. For someone who knows the context of this tweet, the nature, motivations and aims of the Occupy movement, the message resonates for a number of motives, including the appeal to participate in the Occupy movement; the action format of the occupation; the pragmatism of the movement; its distrust in traditional ideologies, and its connected emphasis on the political significance of concrete and practical activities, such as the act of bringing a tent and setting it up in a protest camp in a central public space.

This three-step of analysis allow researchers to move deeper and deeper in the web of meanings of social media conversations. At each step the research progressively “closes in” on the interpretation of posts and social media conversations, while at the same time broadening the perspective of analysis, and paying attention to the general context. This procedure is evidently not an exact science. To quote Clifford Geertz is more of an artisanal process of “guessing at meanings, assessing the guesses, and drawing explanatory conclusions from the better guesses” (1973: 20). Researchers should verify the validity of their emerging interpretation of meanings, by “triangulating”, that is comparing and contrasting the findings emerging from different posts, and progressively refining their interpretation.

Conclusion

As I have sought to demonstrate in this article, a digital update of the hermeneutic method – data hermeneutics – is urgently needed to overcome the limits of data analytics. Crunching numbers in ever more powerful and sophisticated ways is not enough, if one is not able to fully explain the categories of analysis, and the significance, ramifications, and implications of findings. While “counting” – the core logic of data analytics – can no doubt be useful in gaining an overview of the structure and dynamics of conversations, a real understanding of their motivations and meanings can be achieved through the sampling and close reading procedures proposed in this article. To make sense of online political phenomena, we cannot approach them as merely structures of behaviour to be studied mathematically. We also need to approach them as texts, webs of meanings which researchers need to slowly acquaint themselves with before they can claim to know them and comprehend them. Research based on data metrics has been weak in

depth of understanding and contextual knowledge, and it is precisely in these areas that data hermeneutics can give a timely contribution.

The idea of data hermeneutics put forward in this article is in part a polemical response to the current dominance of data analytics in social science, and a reassertion of the importance of qualitative methods. However, when it comes to designing concrete research projects, data hermeneutics should be understood as non-exclusive. Data analytics and data hermeneutics should often be used in tandem, as part of a “quanti-qualitative” approach (Venturini/Latour 2008), with various iterations between the two. Data analytics is particularly precious in the initial scoping of a research project, since it allows to gain an overview of conversation structures, from which researchers can then turn towards data hermeneutics, looking in more depth at the content of specific messages, and specific excerpts of a given conversation. Data hermeneutics can improve on quantitative approaches by providing a clearer understanding of various categories utilized in such analyses. For example in analyzing the language utilized in a given conversation one can combine qualitative and quantitative methods in effective ways, by identifying the most recurrent terms and then looking at how these terms are concretely used in a number of expressions, and finally investigating the motivations underlying these expressions. Furthermore, statistical testing procedures from data analytics can provide indications on where to go deeper, for example by highlighting the peaks of activity in a certain dataset and thus suggesting where a more in-depth quantitative analysis should be conducted.

Thus, what is required is not the wholesale substitution of data analytics with data hermeneutics, but a methodological rebalancing the ultimately can benefit both qualitative and quantitative research, and more generally allow for a more holistic and better contextualized understanding of contemporary politics.

References

- Anderson, Chris (2008): “The end of theory: The data deluge makes the scientific method obsolete.” In: *Wired magazine* 16/7 (<https://www.wired.com/2008/06/pb-theory/>).
- Androutsopoulos, Jannis/Beißwenger, Michael (2008): “Introduction: Data and methods in computer-mediated discourse analysis.” In: *Language@Internet* 5/2, pp. 1–7.
- Batrinca, Bogdan/Treleaven, Philip C. (2015): “Social media analytics: a survey of techniques, tools and platforms.” In: *AI & Society*, 30/1, pp. 89–116.
- Benjamin, Walter (1977): *The origin of German tragic drama*, London: NLB.
- Bleicher, Josef (1980): *Contemporary hermeneutics: Hermeneutics as method, philosophy and critique*, London: Routledge & Kegan Paul.
- Bollier, David/Firestone, Charles M. (2010): *The promise and peril of big data*, Washington, DC: Aspen Institute.

- boyd, dana/Crawford, Kate (2012): "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon." In: *Information, Communication & Society* 15/5, pp. 662–679.
- Bressler, Charles E. (1999): *Literary criticism: An introduction to theory and practice*, Upper Saddle River, NJ: Prentice Hall.
- Burrows, Roger/Savage, Michael (2007): "The coming crisis of empirical sociology." In: *Sociology* 41/5, pp. 885–899.
- Capurro, Robert (2010): "Digital hermeneutics: An outline." In: *AI & Society* 25/1, pp. 35–42.
- Cha, Meyung/Haddadi, Hamed/Benevenuto, Fabricio/Gummadi, P. Krishna (2010): "Measuring User Influence in Twitter: The Million Follower Fallacy." In: *ICWSM* 10/30, pp. 10–17.
- Chen, Hsinchun/Chiang, Roger H./Storey, Veda C. (2012): "Business Intelligence and Analytics: From Big Data to Big Impact." In: *MIS Quarterly* 36/4, pp. 1165–1188.
- Conover, Michael D./Ferrara, Emilio/Menczer, Filippo/Flammini, Alessandro (2013): "The digital evolution of occupy wall street." In: *PloS one* 8/5.
- Dubois, Elizabeth/Gaffney, Devin (2014): "The multiple facets of influence: Identifying political influentials and opinion leaders on Twitter." In: *American Behavioral Scientist* 58/10, pp. 1260–1277.
- Friese, Susanne (2014): *Qualitative data analysis with ATLAS.ti*, London: Sage.
- Gadamer, Hans-Georg (2004): *Truth and Method*, New York: Bloomsbury Publishing.
- Giddens, Anthony (1984): *The constitution of society: Outline of the theory of structuration*, Berkeley: University of California Press.
- Gerbaudo, Paolo (2012): *Tweets and the streets: Social media and contemporary activism*, London: Pluto Press.
- Gerbaudo, P. (2014): "The 'Movements of the Squares' and the Contested Resurgence of the 'Sovereign People' in Contemporary Protest Culture". In: SSRN 2439359.
- Gerbaudo, Paolo (2015): "Protest avatars as memetic signifiers: political profile pictures and the construction of collective identity on social media in the 2011 protest wave." In: *Information, Communication & Society* 18/8, pp. 916–929.
- Gerbaudo, Paolo (2016): "Rousing the Facebook Crowd: Digital Enthusiasm and Emotional Contagion in the 2011 Protests in Egypt and Spain." In: *International Journal of Communication* 10/20, pp. 254–273.
- Gibbs, Graham R. (2002): *Qualitative data analysis: Explorations with NVivo*, Buckingham: Open University Press.
- Gitelman, Lisa (2013): *"Raw data" is an Oxymoron*, Cambridge, MA: MIT Press.
- González-Bailón, Sandra/Borge-Holthoefer, Javier/Rivero, Alejandro/Moreno, Yamir (2011): "The dynamics of protest recruitment through an online network." In: *Scientific Reports* 1/197.

- González-Bailón, Sandra/Borge-Holthoefer, Javier/Moreno, Yamir (2013): "Broadcasters and hidden influentials in online protest diffusion." In: *American Behavioral Scientist* 57/7, pp. 920–942.
- Grimmer, Justin/Stewart, Brandon M. (2013): "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." In: *Political Analysis* 21/3, pp. 267–297.
- Harnad, Stevan (1990): "Against computational hermeneutics." In: *Social Epistemology* 4, pp. 167–172.
- Husserl, Edmund (1970): *The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy*, Evanston: Northwestern University Press.
- Jameson, Fredric (1981): *The political unconscious: literature as a socially symbolic act*, London: Methuen.
- Kelsey, Darren/Bennett, Lucy (2014): "Discipline and resistance on social media: Discourse, power and context in the Paul Chambers 'Twitter Joke Trial'." In: *Discourse, Context & Media* 3, pp. 37–45.
- Kitchin, Robert (2014): *The data revolution: Big data, open data, data infrastructures and their consequences*, London: Sage.
- Lazer, David/Kennedy, Ryan/King, Gary/Vespignani, Alessandro (2014): "The parable of Google Flu: traps in big data analysis." In: *Science* 343/6176, pp. 1203–1205.
- Marshall, Martin N. (1996): "Sampling for qualitative research." In: *Family practice* 13/6, pp. 522–526.
- O'Neil, Cathy (2016): *Weapons of math destruction: How big data increases inequality and threatens democracy*, New York: Crown.
- Patton, Michael Quinn (2005): *Qualitative research*, Chichester: John Wiley & Sons.
- Ricoeur, Paul (1971): "The model of the text: Meaningful action considered as a text." In: *Social research* 38, pp. 529–562.
- Rogers, Richard (2013): *Digital methods*, Cambridge, MA: MIT Press.
- Rowley, Jennifer E. (2007): "The wisdom hierarchy: representations of the DIKW hierarchy." In: *Journal of Information Science*, 33/2, pp. 163–180.
- Saldaña, John (2015): *The coding manual for qualitative researchers*, London: Sage.
- Stephansen, Hilde C./Couldry, Nick (2014): "Understanding micro-processes of community building and mutual learning on Twitter: A 'small data' approach." In: *Information, Communication & Society* 17/10, pp. 1212–1227.
- Strauss, Anselm/Corbin, Juliet (1990): *Basics of qualitative research*, Newbury Park, CA: Sage.
- Theocharis, Yannis (2013): "The wealth of (occupation) networks? Communication patterns and information distribution in a Twitter protest network." In: *Journal of Information Technology & Politics* 10/1, pp. 35–56.

- Tinati, Ramine/Halford, Susan/Carr, Leslie/Pope, Catherine (2014): "Big data: methodological challenges and approaches for sociological analysis." In: *Sociology*, 0038038513511561.
- Tucker, William T. (1965): "Max Weber's 'Verstehen'." In: *The Sociological Quarterly* 6/2, pp. 157–165.
- Tufekci, Zeynep (2014): "Big questions for social media big data: Representativeness, validity and other methodological pitfalls." In: *arXiv*, 1403.7400.
- van Dijck, José (2014): "Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology." In: *Surveillance & Society* 12/2.
- Venturini, Tommaso/Latour, Bruno (2010): "The social fabric: Digital traces and quali-quantitative methods." In: *Proceedings of Future En Seine 2009*, pp. 87–101.
- Weber, Max (1978): *Economy and society: An outline of interpretive sociology*, Berkeley: University of California Press.
- Weber, Max (1981): "Some Categories of Interpretive Sociology." In: *The Sociological Quarterly* 22/2, pp. 151–180.
- Wise, Alyssa F./Shaffer, David W. (2015): "Why theory matters more than ever in the age of big data." In: *Journal of Learning Analytics* 2/2, pp. 5–13.
- Wodak, Ruth/Krzyzanowski, Michal (2008): *Qualitative discourse analysis in the Social Sciences*, London: Palgrave Macmillan.
- Wolfreys, Julian (2000): *Readings: Acts of close reading in literary theory*, Edinburgh: Edinburgh University Press.

