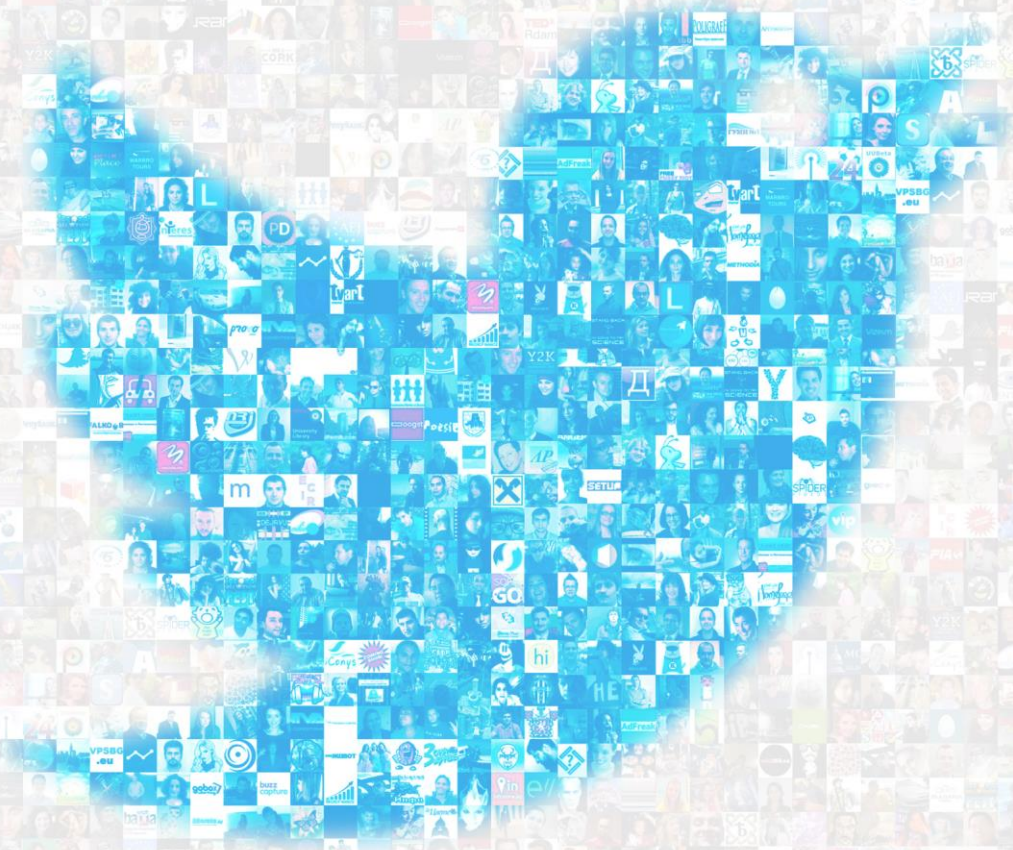
Holistic Twitter Research

Explorative study on the methods and practices, involved in the production of knowledge with social media data



Kalina Dancheva
3721329

Master Thesis New Media & Digital Culture
Faculty of Humanities, Utrecht University
Supervisor: Dr. Sybille Lammes
Second reader: Dr. Mirko Tobias Schäfer

'Would you tell me, please, which way I ought to go from here?'

'That depends a good deal on where you want to get to,' said the Cat.

'I don't much care where – 'said Alice.

'Then it doesn't matter which way you go', said the Cat

'- so as long as I get SOMEWHERE', Alice added as an explanation.

'Oh, you're sure to do that,' said the Cat, 'if you only walk long enough."

— Lewis Carroll, Alice in Wonderland

… a construction is not a representation from the mind or from the society about a thing, an object, a matter of fact, but the engagement of a certain type of world in a certain kind of collective.

— Bruno Latour 1997, xiii–xiv

Acknowledgements

The completion of this master thesis would not have been possible without the help and support of many people, who I would like to acknowledge in this section.

First, I would like to express my deepest gratitude to my tutor and supervisor Sybille Lammes. Her guidance, advice and critical mind have been crucial throughout the whole process and I cannot thank her enough for staying by my side, despite her relocation to the other shore of the North Sea. Furthermore, I would like to thank Alex Gekker for the many long hours of theoretical discussions, Johannes Paßmann, whose insights have greatly enhanced my view and Maria Koomen, for polishing the writing style of this work. I would like to acknowledge with much appreciation my teachers and classmates in the master programme New Media and Digital Culture at Utrecht University. Their work has been a great inspiration and I feel grateful and honoured for being part of this progressive academic community.

I sincerely thank my parents and family for their endless love, support and encouragement. Many thanks to my dear friends who kept my spirit high during the periods of social isolation.

And to Maarten - thank you for your patience, trust and unconditional support.

Abstract

With the establishment of social media sites as part of the daily life of online users, scholars in humanities and social sciences are raising the debate about the methods for researching these environments. The vast amount of data that derives from social networks challenges the established research techniques and invites academics to use software tools and algorithms as part of their practices of knowledge production. This thesis aims to contribute for the discussions on online research methods, by suggesting the concept of a holistic approach to the study of social media. This idea argues that data, online platforms and tools cannot be perceived as isolated objects, but as assemblages of heterogeneous agents. Employing this view, I build a case on research with data from the microblogging site *Twitter*, analysing the processes of knowledge-making afforded by the platform, the APIs, the database and the software tool for data visualisation *Gephi*. As a result of this empirical exploration, I argue that social interactions are inherently implicit and appropriated by *Twitter* users and thus, scholars need to develop tactics that would allow them to look beyond the logic of the database. Moreover, I show that social media research takes place on several levels and on each stage, the method of research is shaped by the actors. Thus, I argue that the concept of online research method cannot be confined within the traditional frames, but instead, it should be perceived as an evolving process, being in a state of constant flux, where heterogeneous actors influence the research decisions, mobilise traditional and new methods and negotiate the research choices of scientific exploration. The main argument of this thesis is that social media studies are not only a technological accomplishment, but should be understood as a complex holistic process.

Key words: online research methods, holistic approach, *Twitter*, big data, actor-network theory, software tools.

Table of contents

Introduction

Prologue

Research methods have been a fundamental part of science making in both practical and philosophical terms. *Methods* are essential as they provide the "logical coherence" (Little 1995, 3), techniques or processes (Kinash 2012, 3) for observing, testing, measuring and building scientific conclusions about phenomena. In positivistic perspective, they are the "magnifying glass" (Saukko 2003, 9) between scientists and reality and they serve as the practical tools which ensure the establishment of truth and credibility of scientific reasoning (Little, 1995, 3). Technical instruments have been an embedded part of the methods of modern science-making as early as its consolidation in the seventeenth century - from Robert Boyle's air pump, invented in early 1660s to validate scientific claims (Shapin & Shaffer 1989), or the Camera Obscura used for depicting objects in nineteenth century scientific atlases (Daston and Galison 2010), to contemporary powerful laboratory equipment.

With innovations in the field of information and computer technologies, however, advanced software tools and algorithms are progressively being employed not only by natural scientists, but also by scholars in humanities and social sciences as means for research. More recently, since the rise of social networking sites in the mid-2000s, academics are turning to computer tools for monitoring, data mining, analysis, and the visualisation of data deriving from social media, using approaches which challenge the established methods of academic exploration. As management scholar Dan Farrell and sociologist James Petersen (Farrel and Petersen 2010) argue, the traditional sociological methods are being replaced by the widely adopted web-based techniques, including the gathering of data from social media:

> The emergence of Facebook, MySpace, blogs, Twitter, and other forms of electronic networking have created new sources of potential data for sociological research. (Farrel and Petersen 2010, 2)

With its possibilities for retention of large datasets, social media studies are related to the practices of the so called "big data" research (boyd and Crawford 2011; Manovich 2011), and are rhetorically hyped in popular discourse as yielding "revolution" and radical transformation (Anderson 2008; McMillan 2013). In The New York Times for example, Markoff wrote that social scientists mined vast amounts of data from *Twitter*, *Facebook* and

blogs, which allowed them to forecast "political crises, revolutions and other forms of social and economic instability" (Markoff 2011). As a recent report shows, although the research defined as "big data" has existed since the 1970s, the last decade has been marked by an exponential growth in scholarly publications that explore large datasets (Halevi and Moed 2012). Now prevailing in academic institutions, corporations, governmental institutions and the media alike, data is argued to be "the oxygen we breathe and the carbon dioxide that we exhale" (boyd and Crawford 2011, 2).

The social media platform *Twitter*, in particular, has gained remarkable attention by scholars, media and companies, looking to study social phenomena behind massive amounts of online data. Analysing social events, for instance, is one of the fields, in which the microblogging site has been argued to be of major importance as user activities can be tracked in real-time. For example, media scholars Axel Bruns and Burgess state that hashtags played a crucial role in managing the 2011 flood crisis in Australia by serving as a channel for the dissemination of information by authorities and citizens (Bruns and Burgess 2011). In 2011, *The Guardian* published an extensive study, conducted by researchers, who extracted and illustrated the data around the spread of rumours on *Twitter* during the 2011 UK riots, thereby revealing the peaks and troughs in frequency of the stories as they unfolded in real-time (Procter et al. 2011). The increasing hype about research with *Twitter* data has even stretched to arguments suggesting that it offers predictive potential for research, such as forecasting outbreaks of influenza (Collier et al. 2011) or detecting fluctuations in the values of the stock market (Bollen et al. 2011).

While *Twitter* as a source of data increases in popularity among academics, are scholars able look beyond the hype and examine critically the deeper implications of social media research for the practices of knowledge-making?

The epistemological question, "how do we know what we know", has been explored for decades by philosophy of science scholars from the rank of Emile Durkheim (1982), Karl Popper (2002) and Michel Foucault (2002). When it comes to conducting research with *Twitter* data, however, questions pertaining to the governing principles behind this type of study are rare. Scholars may come across vast amounts of academic work and software tools for the monitoring, analysing and visualising of social data, yet the emphasis is on the *result* of the scientific exploration, instead on the *process* of science making. Algorithms and tools are being employed to produce complex and eye-catching visualisations, but instead of achieving more transparency, research seems to be getting more *opaque* and *black-boxed*.

As media scholars Barocas et al. (2013) argue, the access to an algorithm is not a guarantee for a scientific breakthrough: "a better sense of *how* discoveries can be made would be of little help in assessing *what* discoveries can be made" (ibid, 4). Instead of focusing on data and tools, one should look at the practices of science making, or as anthropologists Bruno Latour calls it, "science-in-action" (Latour 1987). In a Latourian sense, figures that visualise big data can be defined as *inscriptions* which have been "extracted from the instruments in [a] room, cleaned, redrawn, and displayed" (ibid, 65). In other words, when discussing *Twitter* research, this thesis focuses not only on the particular tools and algorithms, but on the *type of questions* that scholars should ask themselves, before using *Twitter* data for their own research purposes.

Research goal and methodology

With regards to this issue, the main research goal of my thesis is to investigate the kind of research approach, which will help scholars from humanities and social sciences address the process of knowledge production with data, deriving from social media sites. I particularly focus on examining empirically the microblogging site *Twitter* due to the growing scholarly interest in this platform, as part of the academic trend of "big data" research.

I will base my methodology on principles, deriving from the approach of Actor-Network Theory (ANT), developed by researchers of science and technology studies (Akrich 1992; Callon 1986; Latour 2005; Law 1992). ANT diverts from the traditional approach by social scientists, who explore relations between people or groups, and instead suggests that the social is composed by *associations* of human and non-human actors. To explore these associations, scholars should follow the relations they form, instead of imposing a preliminary defined frame on them. As Latour describes it: "[f]ollow the actors in their weaving through things they have added to social skills so as to render more durable the constantly shifting interactions" (Latour 2005, 68). In this heterogeneous social network, an actor (or rather *actant* (ibid, 54)) possess agency when it is able to *transform*, *translate* and "*modify* a state of affairs by making a difference" (ibid, 71). Moreover, an important notion in ANT view is the idea that artefacts have a "script" and an "affordance" (Latour 1994, 31). The term "affordance" (Gibson 1986; Norman 2002[1988]) is defined by design scholar Donald Norman as "the possible actions a person can perform upon an object" (Norman 2002[1988], 228). For Latour, affordance is the potential of artefacts to allow or forbid certain actions and the ability to force humans to "play roles" (ibid, 31).

Following ANT principles, in the empirical part of this thesis, I am going to build up a case of a *Twitter* research, adopting the position that this type of academic study is a result of a constructed process, where human and non-human actors contribute for the creation of knowledge. In the presented study, I am going to follow the actants involved in the process of science-making with data from *Twitter* and will critically discuss the *afforded practices, skills and methods*. In the analysis of the actors, I am going to give examples from research, conducted with *Twitter* data and will reflect on my personal experience with the platform and with the visualisation tool *Gephi*.

Theoretical framework and empirical case

My thesis is composed out of two main parts – a theoretical framework and an empirical case. In the first part I would like to present a theoretical exploration on the concepts of *online research method*, as discussed by scholars in ethnography, sociology, computational social science and new media. I begin by presenting the notions of *virtual ethnography*, theorised by ethnographer Christine Hine (1994; 2000; 2005) as well as the notion of *online sociological methods*, used by traditional sociologists (Mann and Steward 2000) and scholars of social network analysis (Garton et al 2006). By investigating their practices, I show that these scholars rely on existing methods which have been adapted to specific of online environments. My theoretical review continues in presenting the state of online research methods according to scholars of digital humanities (Manovich 2012a; 2012b), computational social sciences (Lazer et al. 2009), digital sociology (Marres 2012) and new media (Rogers 2009). By elaborating on these academic views, I argue that the discussion about online methods is predominantly focused on specific techniques and software tools, instead on the processes and principle of knowledge production. This discourse, in combination with the accessibility of software applications and algorithms, invites researchers to think first about the availability of tools instead of their hypotheses and goals.

As a conclusion of this theoretical exploration, I then suggest the concept of a *holistic approach*. The idea behind this notion is the understanding that data, social platforms and tools cannot be perceived as isolated objects, but as *assemblages of heterogeneous agents*, constructing the research process. The concept of a holistic approach in this thesis can be situated within the larger movement of holism in science. The roots of holism can be traced back to Aristotle's belief that "The whole is greater than the sum of its parts". In

science, one can differentiate holism from reductionism - a concept held by philosophers Oppenheim and Putnam (1958), who argue that all scientific fields can be reduced to a single vocabulary and one "over-arching meta-scientific hypothesis" (ibid, 4) and thereby construct a unified body of knowledge. Holism can be easily described by the philosophical view of *Gestalt psychology* (Koffka 1922; Köhler 1970[1929]; Wertheimer 1912), which argues that human perception is a complex whole, composed by separate atomic elements, experienced in their entirety. In other words, instead of analysing disintegrated pieces, I am adopting a research position which presumes that there are processes between the parts which are determined by the whole (Wertheimer in Ellis (ed) 1938 [1925], 5). In this perspective, in order to understand the analysed object as a whole we need to investigate the role and function of each piece within it.

In the second part of my thesis, I am going to present an empirical exploration on the process of knowledge-making afforded by the platform *Twitter*, by the *Twitter* APIs and data and finally, by the software tool for data visualisation *Gephi*. By critically discussing the role of each actor within the research, I will then illustrate the results of the process and the steps that I took when employing the holistic research approach to *Twitter*. With this empirical investigation I would like to contribute for the better understanding of *Twitter* based research and to provide guidance about the type of questions that scholars in social sciences and humanities need to ask before starting an explorative study in this environment.

Part 1.
State of online research methods

There is nothing more necessary to the man of science than its history, and the logic of discovery . . . : the way error is detected, the use of hypothesis, of imagination, the mode of testing.

— Lord Acton (in Popper 2002, XVII)

Twitter was launched in 2006 and as a platform of research it is still one of the newest fields of academic investigation. In order to situate my approach to *Twitter* based research, I would therefor like to begin this work by presenting an explorative investigation about some of the leading practices of academics considering the use of online spaces as a source and site of research. A desktop search[1] on the topic of "online research method" shows that the subject has received great attention by both academics and practitioners alike. In this study, I focused my research on analysing the predominant views on online research from the stand points of ethnography, sociology and digital humanities, computational social sciences, digital sociology and new media.

1.1. Early views on online research methods

One of the first scholars to suggest methods that can be used for the purposes of online exploration is ethnographer Christine Hine (1994; 2000; 2005). In her view the Internet provides new sites for field research but in order to tap into this potential, scholars need to *adapt* their methods to the specifics of the online setting. She introduces the approach of "*virtual ethnography*", which aims to discover the effect of online environments on people's everyday practices. As she puts it:

> An ethnography of the Internet can look in detail at the ways in which the technology is experienced in use... The aim is to make explicit the taken-for-granted and often tacit ways in which people make sense of their lives. (Hine 2000, 4-5)

She argues that this form of research has developed in response to the growing number of people who use computer mediated communication as part of their daily routine and the need of ethnographers to develop tactics for studying online communities (Hine 1994).

---

[1] Query on the key term "online research methods" reveals the amount of some 1,121 results on Amazon.com and 151 M. on Google.
Search query of Amazon.com, made on 12 April 2013
<http://www.amazon.co.uk/s/ref=nb_sb_noss_1?url=search-alias%3Daps&field-keywords=online%20research%20methods&sprefix=online+research+me%2Caps&rh=i%3Aaps%2Ck%3Aonline%20research%20methods>.

Search query of Google.com, made on 12 April 2013
<https://www.google.com/#hl=nl&output=search&sclient=psy-ab&q=online+research+methods&oq=online+research+methods&gs_l=hp.3..0j0i30l3.834.9906.0.10608.2.2.0.0.0.0.118.208.1j1.2.0...0.0...1c.1.9.psy-ab.frjVbkFT2FQ&pbx=1&bav=on.2,or.r_cp.r_qf.&bvm=bv.45175338,d.d2k&fp=7a90ef4a29d3e8a8&biw=1241&bih=584>.

Hine proposes that the online space can be studied in two perspectives – as a *cultural artefact* and as *culture on its own* (ibid, 14). What is important to be noticed here is that each of them suggests different approaches and techniques. The first one implies that a researcher views online services as a product of culture, shaped by social processes. In other words, sites such as newsgroups are recognised as "a cultural achievement" (ibid, 39) and they influence society in a medium-specific way, bringing "different meanings for different groups of people" (Hine 2005, 9). In this approach, the academic study might evaluate what a particular online site means to a social group in comparison to other communication media such as newspapers, radio and TV.

In comparison to this method, to analyse the Internet as a culture of its own right suggests that online environments *afford specific user practices* to take place – such as email correspondence, the creation of websites, online chat or the formation of virtual communities in news groups and bulletin boards. The role of the researcher then is to observe these practices and their effect on the participants. As a technique of research, Hine suggests activities such as measuring the performance of a task that is executed via an online application and also in a face-to-face communication. The method of observation was also discussed by Swedish ethnographer Jörgen Skågeby (Skågeby 2010) who defines it as a *real-time observation* of user practices in particular online environment (for instance chat rooms), where the researcher is actively engaged as a participant.

Parallel to the qualitative ethnographic methods, the *sociological* approach to online research sees online users, not only as objects for observation, but as active participants in the research process, who can be researched by adjusting traditional sociological techniques to digital environments. Examples of such research devices are *web-questionnaires*, in which researchers create an online tool which is embedded with the principles of the method and the participants take part in the research through a software interface. For example, human geographer Clare Madge and sociologist and Henrietta O'Connor (Mange and O'Connor 2004) conducted a reflexive research project in which they investigated the advantages and disadvantages of the use of the formats of web-based questionnaires and synchronous online group interviews. In a similar vein, management scholar Dan Farrell and sociologist James Petersen (Farrel and Petersen 2010) argue that web-based surveys can be successfully used as tools for sociological research, replacing the dominant format of the telephone interview. Communication scientists Chris Mann and Fiona Steward (Mann and Steward 2000) also explored the impact of online technology on qualitative research methods. They

even went one step further in the discussion of online tools by introducing the use of a specially developed software for interviewing focus groups (ibid, 101). This technological solution requires both the researcher and the participants to obtain technical skills, which was noticed as crucial for the success of the research.

A third type of the early approaches to online research has been introduced by scholars of S*ocial Network Analysis* (SNA). This division of sociology adopted methods that derived from classical sociology as well as from the study of *graph theory*, developed by academics in computer sciences, physics and applied mathematics (Barabàsi 2003; Mitchell 2006; Newman et al. 2006). In the field of social sciences, the development of graphical network analysis for the exploration of social networks was introduced by Jacob Moreno (1953). Scholars of social network analysis perceive networks as the primary building blocks of society (Marin and Wellman 2011) and the role of the researcher is to trace the "relations and patterns" (Wellman and Marin 2010, 1) between people or groups with emphasis on their interactions within "whole" or "ego" networks (Knox 2010, 118).

For SNA scholars the Internet is a medium where social relations form and researchers can investigate the hidden patterns that occur on a micro-and-macro level. Data for the network graph is gathered by a combination of methods that might include traditional sociological techniques, such as a survey or an interview, as well software tools that are capable of retrieving data from communication networks and thereby reveal one's connections to the group (Garton et al 2006). An example of the latter is a study of sociologists Garton and colleagues who explore the use of email and video communication between members of an organization. The scholars visualised the communication networks between the employees in a sociogram (Fig. 1), where data was collected from an electronic log file (Garton et al. 2006).

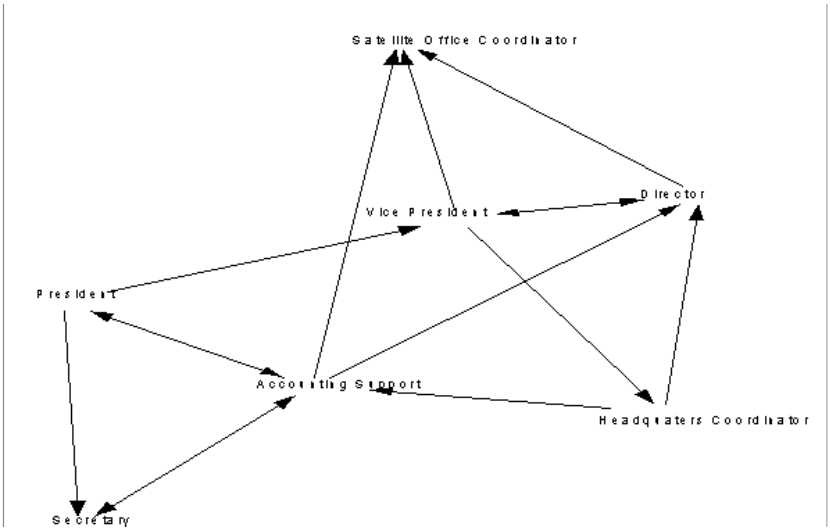


Figure 1: Work Interaction by CMS (Garton et al. 2006).

To sum up, the above presented methods of online research perceive the Internet as a space where social relations between people and groups are formed and exist in a complex interplay between computer networks, computer-mediated communication and social processes. In their approach, ethnographers and social scientists explore users by employing *existing techniques, which are re-fashioned* to be compatible to the specifics of online environments. Data for these types of studies is gathered either by observation or by self-reported forms. Although some automatic devices for data gathering are used, the data they gather is usually on a small scale (for example within an organization) and these devices are not used as a main research technique.

1.2. The rise of computational methods in social research

The second view on online research methods among social sciences, are approaches which rely on data retention and software tools for data visualisation as a main part of the research process. Media scholar David Berry (Berry 2011; 2012) defines the increasing role of technologies in humanities and social sciences as the "the computational turn" (Berry 2011; 2012), which can be seen in the development of *digital humanities* (Berry 2011; 2012; Hayles 2011, Schnapp and Presner 2009), *computational social sciences* (Lazer et al. 2009), *digital sociology* (Marres 2012) and *digital methods* (Rogers 2009).

David Berry describes *digital humanities* with a terminology that comes from software and media studies and argues that this type of research assigns agency to the medium as part of the production of knowledge. His definition of this trend among researchers is formulated as: "...to look at the digital component of the digital humanities in the light of its medium

specificity, as a way of thinking about how medial changes produce epistemic changes" (Berry 2012, 4). Literary scholar and science and technology theorist N. Katherine Hayles also suggests a definition of *digital humanities*, but she emphasises on the collaboration of methods, or as she puts it: "the diverse field of practices associated with computational techniques" (Hayles 2012, 45). She suggests that research techniques such as text encoding and analysis(ibid). In her interpretation of the this field, despite the transformations that digital technologies bring in the process of research, the *tradition of humanities* is *not generally* transformed, but it is in an *"active interplay"* with the technological means of research (ibid, 60).

Media theorist and founder of the Software Studies Initiative in the University of California, Lev Manovich (Manovich 2012a; 2012b) is also an important voice in the domain of digital humanities. His point of view stems from the presentation of the field in the book *Software Studies* edited by Mathew Fuller (Fuller 2008). In this collection of critical texts, the authors present the notion that the technical characteristics of software programmes should be studied from the critical perspective of humanities, which can reveal their embedded conceptual aspects and implications on cultural practices. In the same vein, Manovich elaborates on methods that enable researchers in humanities and social sciences to perform *cultural analysis* with big data, deriving from software-driven media, which includes, but not only, web 2.0 sites. In his analysis *How to follow software uses,* Manovich emphasises that in software-driven media, interactions are constantly produced, monitored and tracked by applications. For the purposes of research and analysis of these interactions, Manovich argues:

> We need to be able to record and analyse interactive experiences, i.e. concrete temporal interactions of particular users with the software – as opposed to only analyzing media "documents". (Manovich 2012a)

In terms of tools that can be used, Manovich suggests an *integrated approach* between classical cultural scholars and software tools – as he puts it: "using the same methods, but for different purposes" (ibid, 14). For the scholar, the potential of this type of research is that it takes into account the visible interactions as well as the ones that are hidden in the layered structure of software objects. The recorded data can be data about user activities as well as meta-data, which is embedded in the code of online artefacts. As an example of this

type of analysis, he suggests the software *Google Analytics* which tracks the visits of a website and presents the aggregated results in a dashboard with dynamic graphs and charts.

Another method within the scope of *digital humanities* is the visualisation of images, which Manovich defines as *media visualisation* (Manovich 2012b). This technique suggests that large image collections can be plotted on a digital canvas and form a new representation which a researcher can observe and discover patterns. For example, the *Software Studies Initiative* conducted a study in which the researchers displayed 4535 *Time* magazine covers, published between 1923 and 2009 years and plotted them within a single high-resolution image (Fig 2). This method also aims to serve the purposes of cultural analysis with big data, which is particularly relevant to online environments, where users share massive amounts of images. For instance, Manovich suggests that the same software, which was used to visualise the covers of *Time* magazine, can also be employed to display collections of images which are extracted from social media sites, such as *Flickr*.

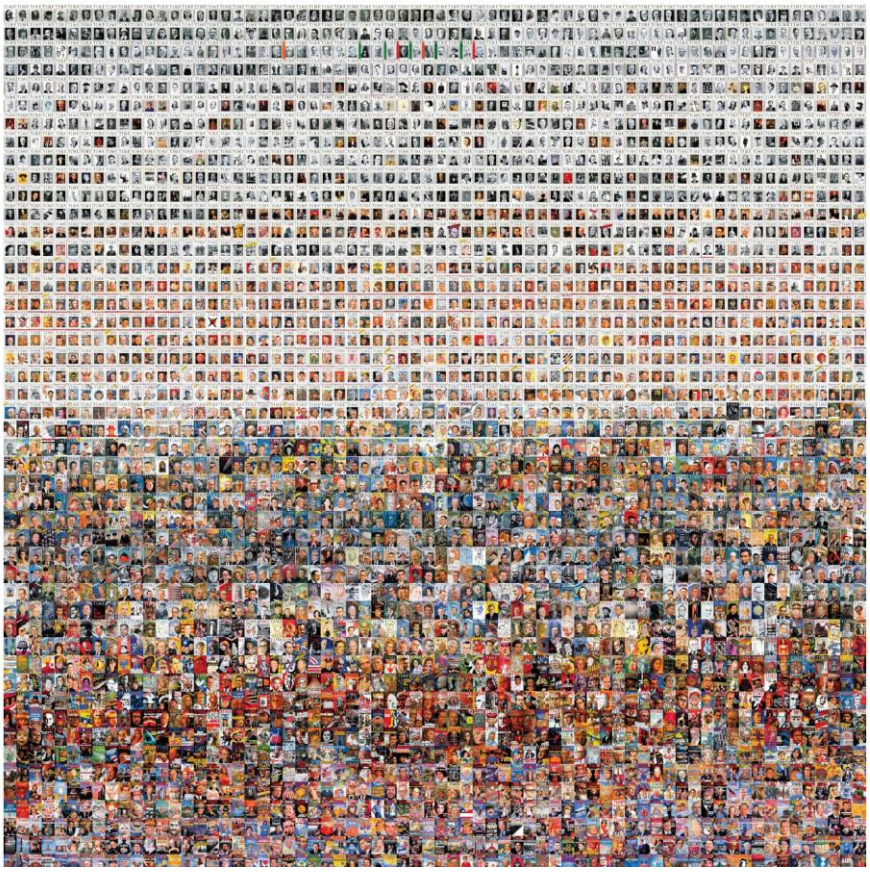


Figure 2: "Mapping Time": A visualisation of 4535 covers of Time Magazine, issued in the period 1923-2009 by Manovich and Douglass (Manovich 2008b).

Another filed in which the notion of online methods has been discussed is *computational social science*s. Network scientist David Lazer and colleagues (Lazer et al. 2009) define it as a field that "leverages the capacity to collect and analyze data with an unprecedented breadth and depth and scale" (ibid, 722). This type of research is also employed by companies such as *Google* and *Yahoo* and in this regard, scholar Duncan Watts argues that academics should cooperate with these companies, which currently dominate the collection of data (Watts 2007). For these scholars, the research of online social networking and the large data that derives from them is perceived as an advanced method for the study of society. An example that Lazer provides for this kind of research is the exploration of political blogs in which a visualisation is used to reveal the bias within the network (Fig. 3).

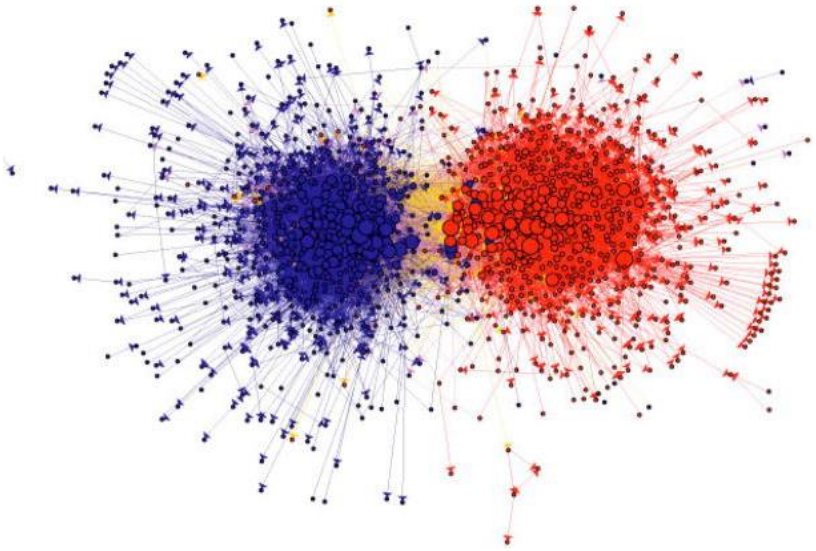


Figure. 3: A link structure within a community of political blogs (from 2004), where red nodes indicate conservative blogs, and blue liberal. Orange links go from liberal to conservative and purple ones from conservative to liberal. (Lazer et al. 2009, 721)

Finally, I would like to present the notion of *online methods,* employed by new media scholar Richard Rogers and digital sociologists Noortje Marres.

Scholar Richard Rogers is one of the first to discuss the topic of methods that use data from Web 2.0 sites. He coins this approach as digital methods (Rogers 2009) and defines it as "research practice which grounds claims about cultural change and societal conditions in online dynamics" (ibid, 5). In contrast to methods employed previously by scholars in humanities and social sciences, these methods do not have an ancestor in any traditional academic domain, but are *born* in the digital environment. For researchers using digital

methods, traditional techniques are limiting the potential of researchers to understand and explore online environments. Rogers discusses this shift of research approach in the following way:

> I would like to suggest inaugurating a new era in Internet research, which no longer concerns itself with the divide between the real and the virtual. It concerns a shift in the kinds of questions put to the study of the Internet. The Internet is employed as a site of research for far more than just online culture. The issue no longer is how much of society and culture is online, but rather how to diagnose cultural change and societal conditions using the Internet. The conceptual point of departure for the research program is the recognition that the Internet is not only an object of study, but also a source. (Rogers 2009, 8)

The methods for studying digital objects can be understood as ones that "follow the medium" which the scholar argues is the methodological distinction from previous ethnographic and sociological methods. The media-specific perspective *following the medium* means that a researcher explores how the online environment treats, orders and transforms digital objects. This approach to research can be traced back to the 1960s, when media theorist Marshal McLuhan presented his understanding that the medium is an embedded part of the message: "it is the medium that shapes and controls the scale and form of human association and action" (McLuhan 1994, 5).

In a much similar vein, in *The End of the Virtual: Digital Methods*, Rogers (2009) argues that researchers gain knowledge from online environments by conducting research that focuses on exploring the medium-specific Internet artefacts and consider the Internet in a threefold manner: "as a source of data, method and technique" (ibid, 13). In his work, he presents a series of case studies, conducted under the Digital Methods programme. For instance, one of the research projects explores the page rank of a website[2], which investigates the attacks on the World Trade Centre in New York in 2011. By systematically tracking the position of the site with a dedicated tool, the research team found a sudden drop around the anniversary of the tragedy. The researchers then argued that they have detected a case of apparent removal of a site in the search engine *Google* and raised questions about the

[2] <911truth.org>.

volatility of search results and the intrusions that might manipulate the opinion of online users.

In another case, Rogers shows an example of a "post-demographic analytical practices" (ibid, 26) in which the research team uses data from the networking platform *MySpace*. The scholars used the architecture of this platform, which allowed a user to view the friends of another user and their interests, without being directly connected to any of them. Using these afforded opportunities the research focused on the profiles of US presidential opponents in 2008 Barack Obama and John McCain and employed software that could extract the interests, which the friends of the two politicians had mentioned in their profiles. The results of this exploration showed that the two groups had distinctive differences in their preferences in terms of movies, music, books and TV shows (Fig. 4). This type of study is traditionally conducted by sociologists who explore the demographic characteristics of TV viewers, by asking preliminary selected group of people to report their behaviour and preferences. In the case of the method used by Rogers, this kind of analysis was possible through the affordances of the social media platform and suitable software for data aggregation and comparison.

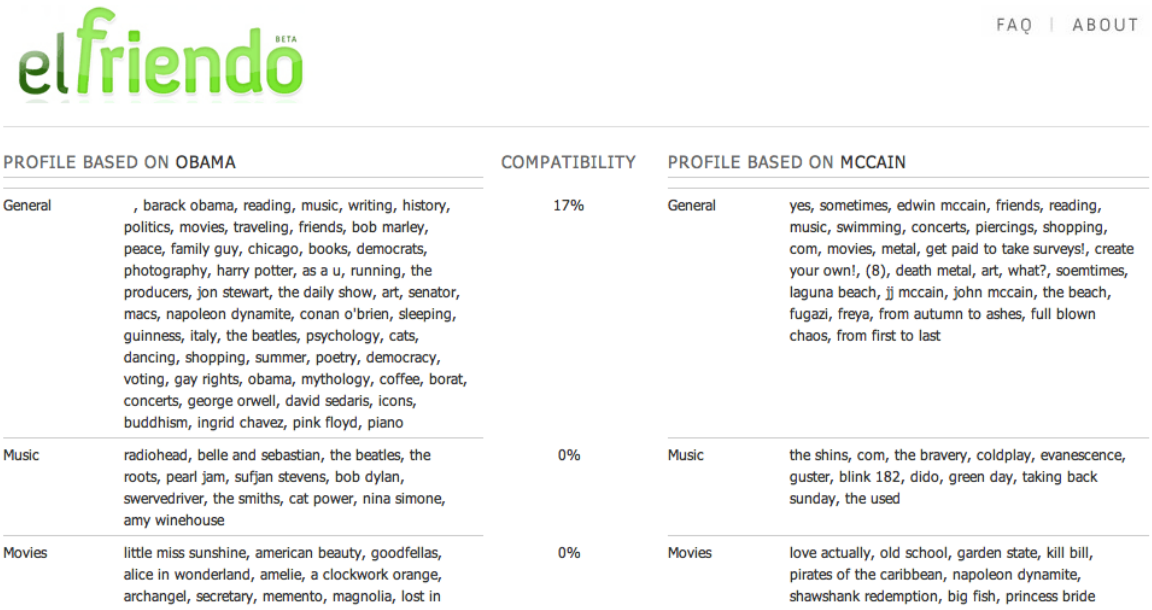


Figure 4: Post-demographic research with the software Elfirendo, conducted by Digital Methods Initiative. (Andrews et al. 2008).

Closely related to the research of Rogers is the work of digital sociologists Noortje Marres (2012). In her view, online methods should not be considered only as brand new ones, as

Rogers suggests, but also as traditional methods which have been "remediated" (ibid, 152) to digital environments. Examples of this notion are the tools *Issue Crawler* and *Co-Word Machine*. The first tool, the *Issue Crawler*, was developed to perform an analysis and visualisation of hyperlinks on the web, but it uses an algorithm that was employed in the study of co-citation of academic papers. In this case, the connection between the traditional object of research and the new one comes from the fact that the *Pagerank* algorithm of *Google* was also built on principles of academic citation analysis (Page et al., 1999). In a similar vein, the *Co-Word Machine* is embedded with classical models for co-word text analysis and was used as a method for identifying issue dynamics on *Twitter*. In both cases, Marres shows that online sociological research can be conducted with tools which carry traditional methods, *adapted* to the specifics of the online environment.

Furthermore, Marres introduces a broader notion to digital social research methods, understood as a "distributed accomplishment" (ibid, 140) of online platforms, users, devices and informational practice (ibid). As she argues:

> [W]e should examine how, in the context of digitization, the roles of social research are being distributed between a range of different actors: between researchers, research subjects, digital technologies, and so on. Moreover, the concept of redistribution directs attention to a possible implication of digitization for social research: digitization may be unsettling established divisions of labour in social research. If we use blogs in social research, does this mean that we are partly delegating the task of data collection to bloggers? (Marres 2012, 140-141).

To sum up, in this part I presented an explorative study on the methods used by academics in the disciplines of digital humanities, computational social sciences, digital sociology and digital methods. For both groups of scholars the practices of data retention, calculation and visualisation are established as standard in academic research. This theoretical investigation showed that academics are currently holding opposing views about the methods, which can be used in the exploration of online phenomenon – from traditional discipline-specific methods which are adjusted to digital environments, to the notion of brand new natively digital methods. However, despite the differences in their views, for both groups online research methods are perceived with positive notions about the great research potential,

enabling them to explore hidden relations and to discover patterns that would otherwise be beyond the reach of the practices of scholars in humanities and social sciences.

1.3. Analysis of the theoretical views on online methods

To conclude this chapter, I would like to present a critical analysis on the theoretical overview of the concepts of online methods of research by scholars in social sciences and humanities.

First of all, the above addressed theoretical notions reveal that the concepts of *Internet methods* have been evolving in a rather short period and are currently in a state of *formation* and *transformation*. When we compare the early views with the notions developed over the last five years, we can see that these explorations have undergone a change from a perspective which understands the online as a *facilitating medium* for research, to a view which uses digital environments as a *source* and *space of research*. Parallel to this process, the practices of automatic data gathering have evolved not only in terms of scale but also in terms of its role in the research process. While earlier ethnographic and sociological studies used automatic extraction as a complementary source of data, nowadays, the prime focus of academics is on big data retention and algorithms that help them gain insights. However, we should notice that traditional methods do not cease to exist but continue to be used parallel to the newly developed ones. For instance, if we focus on the research technique of web-survey, we will find that this type of a tool is currently popular and it has adopted new forms, suited for the Web 2.0 environment. Examples of such devices are services such as *Survey Monkey* and *Google Forms* which have become popular tools for academic and corporate research. These services provide web-based questionnaires, which have easy to use interfaces, afford quick and easy distribution and are embedded with built-in statistics that help researchers analyse the results.

Second, the debate about how scholars in humanities and social sciences conduct online research is evolving from a discussion about general practices, such as observation, interview and questionnaire, to a discussion focused on *specific techniques* and *software tools*. Although the first are perceived by some as outdated or limiting the researchers, they do provide a clearer guidance about the principles of knowledge production, than the software driven approaches. The emphasis on the tools, discussed by scholars of digital humanities, computational social sciences and digital methods shows a practice of reducing the results of a research to the possibilities of a tool. Moreover, digital research tools are often *specially*

built to explore a very particular research question or a phenomenon. In ANT sense, software tools for data analysis are embedded with the methods, goals and functions of their creators (Latour 1994, 31). In this perspective, we can argue that the tools can exert agency and prescribe actions, roles and skills back to human actors. My critique to the above notions of online methods is that they invite researchers to analyse online phenomena through the perspective of software tools, and thereby adopting the method and logic embedded by their creators. An example of such applications is the software *Google analytics,* which provides insights about the interactions that have taken place on a given website. However, when using the tool, researchers are confined to rely on the predefined categories in which data is being presented. Unless scholars are provided with the API of the system and extract the data in their own database, they are not allowed to do a more precise analysis, customised to their research needs. Thereby, the fact that some tools and algorithms are freely accessible has two sides – on the one hand it invites academics to conduct more research with online data, but on the other, it suggests scholars to think first about the availability of tools and then about their hypothesis and goals. The problem here is similar to the situation in Lewis Carroll's *Alice in Wonderland*—researchers might first ask about a particular road, before deciding where they want to go.

1.4. The holistic approach to online research

In the midst of the debates about existing methods and new digital tools, I would like to suggest a *holistic approach* to the processes involved in researching online environments. This holistic view implies that data, social platforms and tools cannot be perceived as isolated objects, but as assemblages of heterogeneous agents, both human and non-human.

The notion of a *holistic approach* to online methods is based on terminology deriving from ANT scholars. In epistemological sense, ANT is a constructivist approach, which implies that scientific knowledge is conceived as a result of multiple events and parties involved. Sociologist Karin Knorr-Cetina theorises this epistemological stance as: "the products of research are fabricated and negotiated by particular agents at a particular time and place" (Knorr-Cetina 1981, 33). According to this understanding, truth as an absolute does not exist, but it is rather a constructed position, considered as valid within a certain context and historical period. This view can be also be related to Michel Foucault's notion of *episteme*, which he defines as: "the total set of relations that unite, at a given period, the discursive practices that give rise to epistemological figures, sciences, and possibly formalized systems" (Foucault 2002, 211). The constructivist position is thus providing an alternative

perspective to the classical positivist view, employed by sociologists, who understand scientific phenomenon as something that is "out there" and scientists can objectively study it without being affected by the circumstances occurring in the process of research (Willig 2008, 3). As sociologist John Law notes, *knowledge* in ANT perspective is seen as embodied in *multiple agents* where social and technical are converted into a holistic whole (Law 1992, 2). Thereby science-making in the view of ANT is understood as a *distributed practice* between heterogeneous actors that takes place in a particular time and context.

Furthermore, according to sociologist John Law, contemporary societies and events are too unpredictable, chaotic and messy and scientists cannot succeed in exploring them with the traditional notions of a *method* as a structure with fixed *norms* and *rules* (Law 2004). Instead, he argues, scholars should think about methods in a way which is "broader, looser, more generous" (ibid, 4). When it comes to researching online phenomena, the situation is even more complex as in digital environments humans and technical objects are intertwined and contribute for the pattering of the social. Thereby, we need to develop a wider and more fluid understanding about the concepts of online research methods. Actor-Network-Theory allows us to employ such a broad perspective by suggesting that we should *follow the actors* and reflect on the practices and processes, which are afforded by the actors involved. Here, it is important to mention that both Richard Rogers (2009) and Noortje Marres (2012) base their concepts on ANT philosophy. Marres, in particular, does bring a much broader and more holistic approach to the notion of online methods, than academics from digital humanities and computational social scientists. However, both scholars seem to use data as a single fact that can be extracted and computationally processed, without focusing on the view of ANT which argues that actors come into being as a result of associations. With the idea that "data is not a fact", I imply the notion that data is formed in a process, in which user behaviour can be prescribed by the stricture of a platform or appropriated by the users themselves. For example, a research project by new media scholars Paßmann et al. (Forthcoming) shows that the use of the Favstar button in *Twitter*, which is originally created as a way of bookmarking, has been appropriated by German *Twitter* users as a form of reward. This reward then becomes a sign of prestige and gives status to the members who have received the most favstars. Thereby, in my use of a holistic approach, I argue that we need to explore the context in which data is used before using it directly in academic research.

The idea of a *holistic approach* in the analysis of online phenomena was also inspired by recent debates among anthropologists and social scientists which argue that social, cultural and economic processes should be understood in a much more elastic and interrelated fashion. This view can be found in the notion of scapes, introduced by anthropologist Arjun Appadurai (2005). Interested in the area of globalization, Appadurai argues that social processes can be better understood as scapes of "people, machinery, money and images and ideas" (ibid, 33). These five scapes are not isolated from each other, but are understood global as cultural flows, which constantly leak into each other and "spill over traditional paradigmic as well as geographic boundaries" (Saukko 2003, 6). In a similar vein, online platforms reflect the multi-scape dynamics by bringing together social, cultural and technological factors. Thus, when analysing social networks the holistic approach implies that we should take into account both the user activities as well as the technical factors which assemble digital platforms.

In the following part, I am going to build a case of a *Twitter* research, employing the holistic point of view towards the actors and reflecting on the practices involved in this exploration.

Part 2.
Holistic approach to Twitter research

2.1. Explicit and implicit characteristics of Twitter users and interactions

Twitter is a social networking platform which allows users to post status updates of utmost 140 characters (called "tweets") to a network of associates ("followers") via the web-based interface, mobile application, mobile text message or a desktop client[3]. The messages are displayed on one's profile in a stream, which by default, is set as open (scholars boyd and Ellison (2007) call it "public") and thereby enabled to be viewed, searched and aggregated by online users, both members and non-members of *Twitter*. Researchers Zhao and Ronsson (2009) define the main technical characteristics of *Twitter* as brevity, mobility (pervasiveness) and broadcasting communication nature. The platform also allows private one-to-one computer mediated communication via direct messages between the users that remain private. These characteristics of *Twitter* define its architecture as a hybrid between a *social network* and of a *microblogging service,* borrowing some features of both systems. The first characteristic allows users to build a network of fellow users (called "friends"), while the microblogging features refer to the broadcasting nature of the tweets.

In the analysis of *Twitter*, I am going to investigate the ways in which the design and functions of the platform invite users to adopt particular practices and behaviour. As media scholar Alexander Galloway (2004) or law professor Lawrence Lessig (2006) might argue, it is important to look at the technical architecture of online objects, because it defines the rules in which information flows and it regulates user activities online. Thus, in order to understand the properties of *Twitter* and how they would affect the research process, we need to explore its structural characteristics and the embedded rules of interaction. In the following part, I am going to argue that the platform *Twitter* should be understood in terms of two types of behaviour – on the one hand is the *prescribed behaviour* by the platform's affordances, and on the other hand are practices that users develop in terms of self-presentation and interaction, which are much more *implicit, obscure and multi-dimensional*. This argument might have been supported by Galloway's understanding about *protocol* as an agent that does not produce actions, but provides a "set of object dispositions" (Galloway 2004, 75). However, although the notion might seem similar to the idea of *affordances*, Galloway's focus is on the concepts of power and control, exerted from the structure,

[3] Desktop clients are software applications which allow users to manage their Twitter account, instead of the web-based twitter interface. These types of tools usually provide some extra features such as: management of multiple accounts, desktop notifications, and statistics.

while this thesis aims to explore the process of construction as an association between human and non-human actors, and thus, the Latourian perspective will be employed.

2.1.1. Prescribed properties of Twitter users and interactions

The first aspect on which I will focus this exploration, are the particular characteristics of *Twitter* users and their profiles. New media researcher danah boyd argues that the technologies and architecture behind social media sites construct specific properties of online users, which she defines as "network publics" (boyd 2010). As such, she states that they are both the space that is created by network technologies and the imagined collective that emerges as a result of the intersection of people, technology, and practice (ibid, 39). User *profiles* are the spaces where users are asked to provide information about them and which allow them to create a particular self-presentation (Papacharissi 2009). They are a fundamental part of social networks, but each platform defines the type of information that users can submit about them. On *Twitter,* in order to create a profile, users are required to provide a user name, a name (real or not) and an email address. In addition to that, users can add a profile picture, background image, short description, link to a website and select their language, time-zone, and location. In terms of visibility, the profiles are by default open to any online users but *Twitter* accounts can manage this property and allow only the users from their personal network to view their tweets her and profile[4]. Online users can generate more than one *Twitter* account, which might be a personal profile or a profile, moderated by an organisation. Moreover, part of the *Twitter* ecosphere are also fake computer generated profiles, known as "bots". The latter are used for expanding one's network of followers, for distributing paid content and promoting certain users.

When it comes to researching the profiles of *Twitter* users, scholars might come across two main issues: *authenticity* and *segmentation*. Not only are there accounts which belong to organisations, but computer generated accounts can be programmed to act in similar ways to human users so that distinguishing between both requires specially built algorithms, a close look at the data or an ethnographical exploration of the users. In large datasets, this process might be impossible for manual investigation and thus the analysis can potentially include human and non-human fake users. For example, during my internship I identified an account which initially seemed to be a profile of a travel agency. However in one particular period, the user was reported to had sent more than a thousand tweets addressed to other *Twitter* users, which revealed that the profile was driven by a non-human agent. In

[4] A recent study calculated that on average, 12% of the Twitter users close their profile (Beevolve 2012).

this research, I based my analysis on the data, in the way it was provided in the database, and I did not conduct any ethnographical study of the actors. Thus, creating tactics for detecting user behaviour, both manual and facilitated by tools, was of prime importance for the analysis.

A second important feature of *Twitter* is the way in which users build their *network of relations*. After a user is added to a one' network, the latter begins to receive updates from the users she follows on her personal stream. Both, the users who one follows and the ones who follow the user, are gathered and counted in two separate lists which for public accounts are accessible to other online users as well. In contrast to the social networks *Facebook* and *LinkedIn*, the process in which *Twitter* users add a relation to the personal network is one-way, or non-reciprocal. Thus, one can be potentially followed by many or by none and follow many or nobody. This specific property has been observed by computer scientists Java et al. (2007) who argue that the network of *Twitter* exhibits properties of a "scale-free network" (Barabási 2003) and it is characterised by a power law distribution of connections. This structural peculiarity is important for scholars because it shows that while the platform allows users to distribute their messages on a global scale, in practice, we can expect that the ones that will receive the most attention will be a small percentage of the total number of users, and the majority will be followed by much less than the average distribution of connections. The power law principle on *Twitter* is also depicted in the research of Paßmann et al. (Forthcoming) on the German-speaking *Twitter* sphere – on Figure 5 below, the size of the nodes represents the number of retweets that each node has received, showing clearly the few most prominent users.

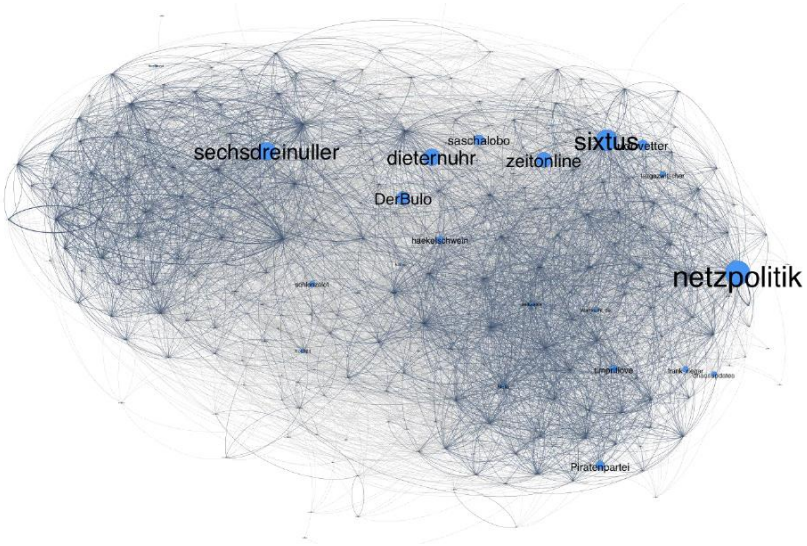


Figure 5: Retweets in the Favstar scene German-speaking scene (Paßmann et al. Forthcoming)

Last, but not least, *Twitter* affords specific *conventions for communication and interactions* between users. The main communication means on the platform are *tweets*, which, as mentioned above, are status updates, restricted to a length of 140 characters. Tweets are by default open to be viewed, searched and retrieved by online users and data mining machines, but they might be also protected which would require users to be given a permission to see the tweets of these accounts. A tweet can be directed to a specific user, when it begins with the addressed username (@username) and in this case, it is called a *reply*. Furthermore, a tweet can be forwarded by a user to her network of followers. This type of interaction is called *retweet* and it is distinguished by the use of the original message, the symbol "RT" and the name of the author (RT @username). Users can also express their preference to a particular tweet by marking it as a *favourite* (favstar) and they can communicate privately by exchanging tweets, called *direct messages*. Last, but not least, a tweet can contain a word prefixed with the symbol *hashtag* ("#") in front of it, and in this case all tweets having a particular key word with a # will be gathered in a common stream of messages. Hashtags are important part of *Twitter* because they provide an additional context of a tweet. *Twitter* recommends users to employ no more than two hashtags per tweet (Twitter 2013d) and to use them "only on Tweets relevant to the topic" (ibid), suggesting that they are a tool for users to gather around common matters and interact with like-minded users. Moreover, additional motivation for users to use hashtags is that the most popular hashtags become part of *Twitter's Trending Topics* – a category of keywords, which are displayed in the interface of *Twitter* and show the most popular keywords in a specific location on a given day. For academics, these particular conventions of the interactions on the platform are particularly important when researchers aim to investigate *communication* activities between users from the same network as well as the spread of tweets (also known as "word of mouth phenomena[5]").

2.1.2. Implicit behaviour and multi-dimensional interactions

Parallel to the clearly afforded properties of user profiles and activities, *Twitter* users communicate in ways, which are much more ambiguous, unstructured and implicit. In this part I am going to discuss few of these processes and the implications for scholarly research.

[5] "Word of mouth"" is defined by the Word of Mouth Marketing Association as "[t]he act of consumers providing information to other consumers (WOMMA 2010).

First of all, the functions of the profiles allow users to provide information about them which is not structurally defined by the platform. An example of this is the *demographic* information about users, such as age, gender and ethnics. Although these attributes are *not afforded* within the profiles, users have found *strategies to communicate* them and studies have shown that demographic characteristics can be detected by a close look at the data or the use of additional machine algorithms. For example, research conducted by computer scientists John Burger and colleagues discovered that the attribute "gender" can be identified by exploring the self-reported description, where some users might specify their gender with key words, for instance, as "a mother"(Burger et al. 2011, 1305).

Another example of an *obscure behaviour* is the creation of private *Lists*. This functionality enables a user to gather a set of other users in a common stream, but without being required to add them explicitly to her personal network of followers. *Lists* can be private or public, and in the latter case other users are able to view the group and subscribe to it. Research has found that lists infer common characteristics and interests of the subscribed users (Kim et al. 2010). This feature of *Twitter* is often used during workshops or conferences, where the hosts creates a list for the event and adds the participants with a *Twitter* profile to it and thereby enabling the members to connect and share information. For researchers, it is important to notice that this function allows users to maintain more than one network of relations. This means that while some users appear to follow no one (like the rapper Eminem or Dutch politician Geert Wilders), they might privately follow other *Twitter* users in a *List* and thereby creating a *secret network* of users, which they follow.

Furthermore, users are allowed to *appropriate* some of the provided data fields and use them for self-representation. For instance, in the case of location data, users can submit the place where they tweet from, but the interface does not provide any verification of this text field. This property allows users to either ignore it or appropriate it for other purposes. A study, conducted by Hecht et al. (2011) shows that over one third (34%) of *Twitter* users does not provide valid geographical information about themselves. About 18% of these users do not give any information, while the rest of the users (16%) provide sarcastic comments or non-existent location data, such as "in Justin Bieber's heart" (ibid, 4). Based on their findings, the scholars argue that in order to detect faulty location information, academics have to explore the statuses of the users, where this type of information might be implicitly embedded. Another practice that shows how the appropriation of the interface design can

create meaning is presented in the previously mentioned research by Paßmann et al (Forthcoming). In their analysis, the scholars use a combination of demographic methods (interview, observation) and quantitative visualisation and thereby define two distinct clusters of *Twitter* users–one, which is involved in politics, and another one, which is focused on the exchange of favstars (the *Favstar* scene). With this project, the scholars prove that the *Twittershpere* is not homogenous but composed by sub-spheres, each of which expresses different behaviour, characteristics and perception about what they consider to be a meaningful communication.

Finally, I would like to state that the means for communication between users afford more than one context to be embedded within a tweet and thus, academics need to employ *a multimodal interpretation* of the communication flow. Here, I would like to give as an example the use of *hashtags* which is a popular research object in *Twitter* studies and has been used by both academics and journalists in the investigation of users' opinion on social issues. Although *Twitter* recommends the use of hashtags only for "relevant topics" (Twitter 2013d), scholarly work has shown that the reality is much more complex. Research conducted by new media scholars Bruns and Burgess (2011) shows that hashtags serve as *ad hoc issue publics*, where users gather, discuss and coordinate activities around local or global topics. Thus, as a method of community research, a scholar might decide to explore the conversation within a particular hashtag. However, Bruns and Burgess also notice that hashtags can be used in tweets which are not related to the discussion, but the authors aim to get visibility of their message by using a popular hashtag word. Another issue with the use of this specific communication symbol is that some key words might be used by another *Twitter* group which would result in merging the two different discussions in a common stream of tweets. Finally, users embed hashtags to express emotion and in this case, although messages are gathered in a stream, no discussion takes place. These specifics of the tweets containing hashtags show that academics must *employ tactics* that would allow them to distinguish the particular use of a hashtag so that the analysis of the data would include the relevant type of messages and users.

To conclude this part of my thesis, I would like to argue that research which is based on *Twitter*, suggests that data, provided by the users should be explored not only from the perspective of the platform's features, but also within a broader context of the user activities. *Twitter* data should be understood as a result of a *process* in which both the

prescribed *affordances of* the platform and the *appropriation* of the users are shaping the data. Moreover, the specifics of *Twitter* profiles challenge some of the main requirements of a traditional sociological or ethnographic research, such as authenticity and demographic attributes. Thus, in order to develop a thorough understanding about their data, researchers need to *develop tactics* that would help them see beyond the visible properties. Often, this means that scholars should create research tools, requiring them to either possess coding skills or work in collaboration with coding experts. Furthermore, academics might need to employ traditional ethnographic or sociological techniques which would look *out of the database* and would help them reveal the implicit behaviour and dynamics of *Twitter* users.

2.2. Afforded practices for data retention – the role of Twitter's APIs

After a researcher is familiar with the specifics of the platform and its users, the second step of the process is the gathering of data. On *Twitter*, exploring data on a small scale can happen through the web-based interface of the platform (via *Twitter* search) which allows a researcher to get an overview of what is tweeted with a limitation of 1500 messages, created in the last fifteen days of the search[6]. This approach has limitations, not only in terms of scale but also with regards to usability. The messages in *Twitter* search can be only viewed and copied and the interface does not allow any further actions, such as filtering, to be performed. Bearing in mind the vast amount of data that is generated in popular discussions, research projects rely on an automatic data retention which allows for larger data sets and flexible database interface. Obtaining data from a social media site can take place either through the provided application programming interface (API) or through the practice of scraping. Data retention through the official APIs of a platform is considered the "polite" (Helmond 2010) way of data gathering because it is enabled and maintained by the social platform itself. In contrast, scrapping is an approach for data collection which might not obey the terms and conditions of the APIs and thus overcoming some of rules and its limitations (Marres and Welterverde 2013). Since this present work is investigating the role of the affordances of the platform, I am going to focus on exploring the APIs of *Twitter* and will not discuss the practice of scrapping. In the following part, I am going to present briefly the basic specifics of data mining with the use of and API and will then

[6] This restriction means that if a topic is widely discussed, users can track the activity till the 1500 message. When the same discussion ceases, the messages will appear only if there is a message in the last fifteen days prior to the search.

continue with a study on the particular APIs by *Twitter* and their role in the methods of research.

An API has often been described as "a source code-based specification intended to be used as an interface by software components to communicate with each other. It allows a developer to create a program that will interact with another program in a predictable manner" (Polycom 2013). It is an important non-human actor because it describes and prescribes the expected behaviour which will allow developers to access structural data (boyd and Crawford 2011, 7). Data is retained by the use of a graphic user interface which makes *requests* (or *calls*) to *Twitter* and as a result, the specified type of data is being returned. APIs can be characterized as a general part of the ecology of social media sites. As computer scientists Gavin Bell argues "(t)oday, a website is more than a brochure; it is a data repository with multiple interfaces to the content" (Bell 2009, 331). The social networking site *Facebook*, the photo-sharing platform *Flickr*, *Google* or even the *Guardian* all provide their own API.

As a social platform, *Twitter* also offers an API for the purposes of data retention. In fact, there are three types of API – two discrete APIs (Search and REST) and a Streaming API (*Twitter* 2012). The Search API enables the extraction of a maximum of 1,500 tweets for a single query and is designed to make query for *Twitter* content (specific keywords, or tweets that refer to a particular user). The REST API provides data about a particular user (such as name, profile picture, network of relations) and allows a maximum of 3200 statuses to be retrieved. *Twitter* explicitly states that these limitations are established in order to "to protect *Twitter* from abuse" (ibid). The Streaming API establishes a continuous connection with *Twitter* and provides constant updates of public tweets into a user's database. It is suitable for "following specific users or topics, and data mining" (*Twitter* 2009). The Streaming API, offers three types of data-streams, which provide different percentage of public data. These are called the "Spritzer", the "Gardenhose" and the "Firehose".

The "Spritzer" extracts about 1% of all streamed tweets, the "Gardenhose" – around 10% and the "Firehose" is been reported to provide the full access to public data (*Twitter* 2013). While the Search and REST API are easily accessible by developers, the Streaming APIs requires permission by *Twitter*. Apart from the different limitations of data calls, it should be emphasised that the API of *Twitter* is a proprietary source software, and as such, it

imposes a list of principles and rules, which serve as a contract between *Twitter* and the developers. If the rules are not obeyed, than the developer's API access might be blocked. Here is a part of the *Twitter*'s API rules:

> Your use of the *Twitter* API and *Twitter* Content are subject to certain limitations on access, calls, and use as set forth in the Rules, on dev.*Twitter*.com, or as otherwise provided to you by *Twitter*. If *Twitter* believes that you have attempted to exceed or circumvent these limitations, your ability to use the *Twitter* API and *Twitter* Content may be temporarily or permanently blocked. (*Twitter* 2012)

The policies of *Twitter* towards data retention have been strongly criticized as *unfavourable* for academic research in several aspects. One of the main reasons is that *Twitter* does not clarify the criteria, under which data provided by the "Spritzer" and "Gardenhose" APIs has been *sampled*. The only type of data that is known to be excluded is the private tweets, but the percentage that is provided from the public ones remains a black-box. What is more, boyd and Crawford (2011) state that even in the case of "Firehose" API access, some public data is also reported to be missing (ibid, 7).

The limitations of the Rest APIs and the two streaming types, forces researchers to turn to the "Firehose" as the only option for a valid research. However, as argued by scholars boyd and Crawford (2011), the access policy of *Twitter* to this type of API is *granted to few companies* or to researchers with sufficient funding. The access to the full data is currently provided by partners of *Twitter* (Datasift[7], Gnip[8] and Topsy[9]) and these companies resell the data, the cost for which requires a solid budget. As scholar Anne Helmond states, the API policies "bring back the notion of scarcity in the digital age which is often considered to be the domain of abundance" (Helmond 2010). To this, new media scholar Lev Manovich adds the issue of accessibility:

> Only social media companies have access to really large social data – especially transactional data. An anthropologist working for Facebook or a sociologist working for Google will have access to data that the rest of the scholarly community will not. (Manovich 2011, 5)

[7] <http://datasift.com>
[8] <http://gnip.com>
[9] <http://about.topsy.com>

Furthermore, researcher Anne Helmond expresses a critique towards the *changing policies* for the use of the *Twitter*'s APIs. In her analysis she argues that since developers use third party applications on top of the platform using the API, any change can cause a chain reaction of tools which fail to perform and need to be adjusted accordingly. The scholar also proves that on several occasions these changes took place without any prior notification which has caused a negative reaction from the developers. The implication for the research methods, Helmond argues, is that this ephemerality makes methods highly *volatile* and researchers need to develop methods, which monitor constantly the occurring changes.

Finally, I would like to argue that the processes of data mining with the API of *Twitter* invites scholars in humanities and social sciences to adopt new skills and work in cooperation with academics, who have at least basic experience with coding languages. Here I would like to emphasise that there is a crucial difference between a company that provides its data and a company that provides the API for data gathering. The first implies that a researcher would receive the data in a database or, directly in a form of analysis and reports. The later means that access to the data is granted but the extraction of this data (partial or full) requires an additional software programme which makes calls to *Twitter* and provides an interfaces for it. Only after that, data can be available for analysis. On *Twitter*, the documentation for the set-up of the API is provided by the platform and several communities of developers have created manuals for implementing the API in different programmable languages. Examples include the *Python-Twitter*[10], *Twitter4Java*[11] and *Twitter R Package*[12]. These documentations (examples on figures 6 and 7) are provided free of charge, but they do require coding *skills* and *experience* with at least one programming language. Moreover, since each API provides different types of data (tweets, profiles' connections, etc) and different percentage of the total public data, one needs to be able to evaluate the kind of API that is required for the purposes of research. For scholars in the humanities or social sciences, the expertise that involves the actual work with the API implies that in order to get the data that they need to work in cooperation with a computer scientists or coding specialist who is able to implement the API and is acquainted with the kind of data request that should be made. However, a researcher should have at least some basic knowledge regarding the purposes of the different APIs and their limitations, as this will directly influence the data gathered and the further analysis.

[10]<https://code.google.com/p/python-twitter>.
[11] <http://twitter4j.org/en>.
[12] <http://cran.r-project.org/web/packages/twitteR/vignettes/twitteR.pdf>.

```
$ pydoc twitter.Status
$ pydoc twitter.User
$ pydoc twitter.DirectMessage
```

Figure 6: Part of the model for extracting Tweets in Python <https://code.google.com/p/python-Twitter>.


```
{
  "completed_in":0.031,
  "max_id":12207846184098 2016,
  "max_id_str":"122078461840982016",
  "next_page":"?page=2&max_id=122078461840982016&q=blue%20angels&rpp=5",
  "page":1,
  "query":"blue+angels",
  "refresh_url":"?since_id=122078461840982016&q=blue%20angels",
  "results":[
    {
      "created_at":"Thu, 06 Oct 2011 19:36:17 +0000",
      "entities":{
        "urls":[
          {
            "url":"http://t.co/L9JXJ2ee",
            "expanded_url":"http://bit.ly/q9fyz9",
            "display_url":"bit.ly/q9fyz9",
            "indices":[
              37,
              57
            ]
          }
        ]
      },
      "from_user":"SFist",
      "from_user_id":14093707,
      "from_user_id_str":"14093707",
      "geo":null,
      "id":122032448266698752,
```


Figure 7: Part of the Search API documentation, provided by Twitter <https://dev.Twitter.com/docs/api/1/get/search>.

To sum up, in this part I explored the conditions under which researchers can obtain data thought the provided APIs by *Twitter*. By examining the policies for data retention, I argued that the lack of transparency about the criteria for inclusion or exclusion of data is a methodological concern for scholars because without a clear guidance about the kind of data that has been extracted, researches are not able to make a proper analysis. The only option for scholars becomes the "Firehose" API access but the high barrier to entry means that researchers need to have ensured funding before they start their scholarly work. The technical skills required for the implementation of the API is yet another challenge, especially for scholars in the humanities and social sciences who are not experienced with coding and thus, need to work in cooperation with a computer scientists in order to gather the data for their research. In this aspect, *Twitter* research significantly *limits the independence* of researchers who are used to traditional research methods, such as interviews or questionnaires. Based on these findings I would like to state that although the platform allows the extraction of large amounts of data, the basic principle of the API access is not abundance, but *control* and *sampling* of the provided data.

2.3. Twitter data as an actor in the research process

After the type of *Twitter* API is selected and the access to it is ensured, the research process continues with the exploration of the database.

The first point that we need to mention in this process is that research that is conducted by academics with *Twitter* often implies that scholars have to deal with massive amounts of data. We can see this in studies by Kwak at al., who obtained more than 41 million user profiles (all *Twitter* users at the time of their research), Cha et al. , who analysed the interactions of 52 million users or by Anger and Kittle (2011), who explored more than 175 million *Twitter* accounts. Online data of this scale hides *risks* such as possible errors in the database, lack of the context of the interactions between users as well as ethical considerations about the use of data (boyd and Crawford 2011).

Furthermore, data is provided in several formats (XML, JSON, RSS and Atom), which vary per programme language and a software expert should explore these afforded possibilities. What is important for the process of research is that data extracted via the API comes in a *structured* fashion. Practically, this means that data is segmented, arranged in categorical fields and each field is labelled in ways, defined by *Twitter*. In order to explore the way data is being organised, I examined the specifications, provided by *Twitter* in their site for developers (https://dev.*Twitter*.com). Here, I am going to present my observations, which I made from the perspective of a master student in new media studies, without background in computer programming or related expertise in the processes of software development. Due to the scope of this research project, I analysed a selected quantity of the afforded types of data and drew the following conclusions, with regards to the research process.

The first and most obvious property is the remarkable number of categories of data types (called *fields*) which *Twitter* provides for retention. However, a closer look at the specifications shows that not all fields are available in the three API types. For instance, software engineer Curtis Chen reports that in 2010, *Twitter* added few additional data fields which were made available only in the *Streaming API*, which requires permission for access (Chen 2010). The platform groups the fields into four data categories: tweets, users, entities and places (*Twitter* 2013c) which together contain more than 100 fields. Here below I present some of the fields and their descriptions, as defined on the website of *Twitter* (Twitter 2013c):

- Screen_name - The screen name of the user who contributed to this Tweet.
- User - The user who posted this Tweet.
- Text - The actual UTF-8 text of the status update.
- Created_at - time when this Tweet was created.
- Retweeted - Indicates whether this Tweet has been retweeted by the authenticating user. The data corresponds to a retweets via the afforded Retweet Button.
- In_reply_to_screen_name - If the represented Tweet is a reply, this field will contain the screen name of the original Tweet's author.
- Entities - Entities which have been parsed out of the text of the Tweet. For instance, hashtags and URLs.
- Favourited – Indicates whether this Tweet has been favourited by the authenticating user (true or false).
- Retweet_count - Number of times this Tweet has been retweeted.
- Favorite_count - Indicates approximately how many times this Tweet has been "favorited" by *Twitter* users.
- Followers_count - The number of followers this account currently has.
- Listed_count - The number of public lists that this user is a member of.

What are the implications of the afforded indicators for the process of research?

The fact that data comes in pre-categorised fashion directly shows that *Twitter* does not provide data in a raw format. By raw, I follow the definition of scholars Tergan and Keller (2005) who argue that raw data does not answer question regarding people, places, time or purposes. If it does, then it is cannot be defined as raw data, but as *information*. Information can be then categorised as *primary*, which is information in its original form, and *secondary,* which has been already analysed or translated in some way (ibid). In the case of *Twitter*, when exploring its data fields, we can observe that the API provides fields which correspond to primary information and fields which can be defined as secondary information.

2.3.1. Primary type of information

The first group of information fields are parameters such as screen name, text of the tweet, location, hashtags. These categories are deriving from the affordances of the platform, mentioned in the previous part of this writing which suggests that a scholar should be familiar with them, before starting the research. However, here I would like to present a case, which shows that not all types of interactions that exist on *Twitter* are labelled and separated in categories, as one might expect.

The example derives from my research internship, were part of my tasks was to explore the way Dutch *Twitter* uses discuss the key word "flight ticket" (*vliegticket* in Dutch). For this purpose, I had to analyse the different types of communication between *Twitter* users (tweet, mention, reply, retweets). The issue that I faced was that the database contained the field *text* which gathered all tweets, regardless of their contextual type. After reading discussions on this issue between developers, I found that this segmentation of data was not afforded by the API. Thus, in order to continue my research, I had to filter the results, which required an understanding about how these types of tweets differ from each other in a way of structuring. For instance, in the case of retweets, users can use the provided button in the interface or retweet manually. In the database, the difference will be marked in terms of the position of the retweet's identifier "RT @username" in the status update. If users apply the automated retweet, then this message will contain the identifier in the beginning of the message. This type of interaction will be able to be extracted through the API as well (through the data field - retweet_count). However, if the retweet has been done manually, then the position of the identifier can occur in the middle or in the end of the tweet. Since *Twitter* does not distinguish these tweets as retweets, a researcher should additionally filter and segment the database. In my internship research, I separated the different kinds of tweets by conducting several filtering procedures using the software *Microsoft Excel*. Figure 8, shows a sample of the database before and after segmentation.

Before filtering
Text
En iedereen zegt wel: je mag met mij mee op vakantie! Maar ik meen het serieus. Me moeder betaald vliegticket en alles. Haha
Als een trein een [vliegticket] boeken #beenthere RT @Bezoekthailand: Trein verkeer ligt stil na ongeluk met vrachtwagen ri noord Thailand
Ik heb net mijn [vliegticket naar Eindhoven geboekt via @CheapTicketsNL! | www.cheaptickets.nl
@jaspergaarthuis @sovisch Kunnen wij misschien helpen met het regelen van het [vliegticket?
#Suriname Voorschool Nieuw Aurora boekt goede resultaten: In het dorp Nieuw Aurora (Boven S... http://bit.ly/gGxDZq #[Vliegticket #Woning
Handig! RT @fflekkerweg: Stel eenvoudig en snel een flight alert in voor een [vliegticket] http://t.co/eliPGhG
@sovisch Uhm nee. Oja goed nieuws, moeder wil graag mijn [vliegticket betalen:) dus ik zeg morgen boeken\;)
RT @olchert: heeft nog startbewijs over voor halve marathon van Berlijn komende zondag. Ook [vliegticket (vr-ma) beschikbaar. Iemand inte ...
Vanavond [vliegticket boeken voor vakantie met @nikkipinda #yeahhhhhh:D
RT @AnassMeknesi: Reserveer je kaarten voor 'De IsLama's' en maak kans op een gratis [vliegticket naar Marokko! Mede mogelijk gemaakt doo ...
After filtering
Tweet
En iedereen zegt wel: je mag met mij mee op vakantie! Maar ik meen het serieus. Me moeder betaald [vliegticket en alles. Haha
#Suriname Voorschool Nieuw Aurora boekt goede resultaten: In het dorp Nieuw Aurora (Boven S... http://bit.ly/gGxDZq #[Vliegticket #Woning
Retweet
Als een trein een [vliegticket] boeken #beenthere RT @Bezoekthailand: Trein verkeer ligt stil na ongeluk met vrachtwagen ri noord Thailand
Handig! RT @fflekkerweg: Stel eenvoudig en snel een flight alert in voor een [vliegticket] http://t.co/eliPGhG
RT @olchert: heeft nog startbewijs over voor halve marathon van Berlijn komende zondag. Ook [vliegticket (vr-ma) beschikbaar. Iemand inte ...
RT @AnassMeknesi: Reserveer je kaarten voor 'De IsLama's' en maak kans op een gratis [vliegticket naar Marokko! Mede mogelijk gemaakt doo ...
Reply
@jaspergaarthuis @sovisch Kunnen wij misschien helpen met het regelen van het [vliegticket?
@sovisch Uhm nee. Oja goed nieuws, moeder wil graag mijn [vliegticket betalen:) dus ik zeg morgen boeken\;)
Mention
Ik heb net mijn [vliegticket naar Eindhoven geboekt via @CheapTicketsNL! | www.cheaptickets.nl
Vanavond [vliegticket boeken voor vakantie met @nikkipinda #yeahhhhhh:D

Figure 8: Twitter database before and after segmentation.

This example illustrates that despite the great number of data fields, the database possesses *limitations* that on the one hand might *mislead the researcher* by excluding certain types of content (manual tweets) and on the other hand, might require *additional software for segmentation and filtering* of the data. Moreover, this fact challenges the reliability and trust in paid tools which provide analysis, based only on the afforded data fields and thus limiting the results to the activities, which are being enabled by the API.

2.3.2. Secondary type of information

The secondary type of information that *Twitter* provides, is information about the presence (or absence) of particular activities as well as insights about the number of times in which some actions occur. For example, the field *Listed* provides insight whether an account has been included in a list, while *Listed_count* gives the qualitative indicator in case a user is included in one or more lists. The fact that *Twitter* includes these types of calculations shows that the platform provides data which is has already been *analysed*. Although at the current moment of this writing, *Twitter* does not provide a special interface with analytical data like, for instance *Google* or *Facebook,* we can see that these calculations are afforded through the data calls. By this way, the platform implicitly triggers particular movements of associations, suggesting certain *methods* in which users and interactions can be evaluated. This analysis is particularly relevant for online marketing experts and social media

specialists, who often evaluate their campaign based on the amounts of followers and interactions that their campaigns have received.

In order to explore the extent to which these quantitative indicators are embedded into the research methods, I analysed a corpus of research papers that conduct research on *Twitter* using an API. With regards to the scope of this research, I focused this study on the topic of *Twitter influence*. Within this research corpus, we can see that scholars suggest different approaches for defining the leaders on the platform. Bashky et al. (2011) created a model in which they defined *Twitter* influence from the perspective of "user's ability to post URLs which diffuse through the *Twitter* follower graph" (ibid, 2). As a criterion for diffusion of information, they used the parameter *retweeting*, which shows re-posted status updates and thereby reaching broader audience, than the one of the original sender. After that, the researchers explored this indicator in relation to users' number of followers (also known as *in-degree*), friends, and tweets. As a result of their study, they concluded that the most influential users in their definition were not the ones with the highest values of these parameters, but instead, the ones which scored average numbers of tweets, followers and friends. Furthermore, Anger and Kittle (2011) ranked Austrian *Twitter* users based on two criteria – the amount of followers that they had and the number of interactions that they evoked. Cha et al. (2010) on the other hand, suggested three types of influence, based on users' number of followers, number of retweets and number of mentions.

In contrast to the above mentioned approaches, few studies adopt methods, which combine the *provided indicators* and a *new type of quantitative indicator*. Here, we can mention the studies by Kwak and colleagues (2010) and Weng et al. (2010) who used the *Ragerank* algorithm, introduced by the founders of Google for defining the most relevant search results. The first applied this algorithm to a network of followings and followers and the latter used it to measure the topic-sensitive influence on *Twitter*. As a result of this meta-analysis, we can argue that the indicators, provided by *Twitter* are used as a main *data source* by academics in defining the particular question of *Twitter* influence. This review showed that researches tend to use the provided types of data in certain combinations and thereby, formulate different rankings. The lack of guidelines from *Twitter* also invites researchers to experiment and search for one common criterion that can be used as an overarching method. Going beyond the measurements, suggested by the platform, and applying

algorithms which are external to *Twitter*, is still rare among researchers but it shows a kind of urgency for more independence from the platform's measurements.

As concluding remarks of this chapter, I would like to emphasise on the importance of a detailed study of the database of *Twitter*, as a major part of the research process. As I suggested, data which is extracted via the API of the microblogging site comes as information, categorised and ordered in a very particular platform-specific fashion. This segmentation of the data on the one hand, might require researchers to use additional tools for data filtering and on the other hand, might invite scholars to use these indicators in their research and thereby making the research highly dependable by the platform. Thus, the database is an actor that can change the preliminary selected methods and mobilise more actors in the process of knowledge-making.

2.4. The role of software tools for data manipulation and visualisation

> Like alphabet, mathematics, printing press, combustion engine, electricity, and integrated circuits, software re-adjusts and re-shapes everything it is applied to – or at least, it has a potential to do this.
>
> — Manovich 2008, 14

The final steps of this academic study of *Twitter* are the visualisation and analysis of the researched data. The use of visualisation algorithms is necessary due to the scale of data that can be obtained through an API of a social platform. When exploring the scraped data, simple observation through the interface of the database does not afford an understanding about the relationships between the entities. Our human information processing abilities require the data to be translated into a visual format so that we can detect patterns, connections or causalities in the dataset. Thus, the use of a software that can aggregate, calculate and present data in a graph is a *prerequisite* for the actual analysis of the data.

A quick look at the corpus of research papers on *Twitter* data, mentioned in the previous chapter, shows the ubiquitous role of visualisations in academic articles as all but one scholarly work included visual imageries as part of the analysis of the data. The majority of the articles included a combination of charts, tables, scatter plot graphs and network

graphs. Lev Manovich defines these types of imagery with the term *information visualization* (Manovich 2010, 4). According to him, information visualisation has risen due to the availability of the big data sets and the programming languages which were suited to produce such graphics. He argues that these are two-dimensional images, which are used to "discover the structure of a (typically large) data set" (ibid, 4). Analysing the visual artefacts, used by academics of *Twitter*, might provide us with insights about the cognitive styles of the different visual formats, but would not reveal anything about the processes involved into the creation of these images, which is a crucial element in the production of knowledge. As Latour argues, "[d]iagrams, lists, formulae, archives, engineering drawings, files, equations, dictionaries, collections and so on, depending on the way they are put into focus, may explain almost everything or almost nothing" (Latour 1986, 4).

In ANT terms, visualisations are not just single actors, but a network of heterogeneous associations. Sociologist John Law defines this the term "heterogeneous engineering", where "bits and pieces from the social, the technical, the conceptual and the textual are fitted together, and so converted (or "translated") into a set of equally heterogeneous scientific products"(Law 1992, 2). According to him, when networks act as a coherent object, the object can become *punctualized*, simplifying the network into a single entity. For example, it is easier to perceive a laptop as a technical object which performs variety of functions and help us in our daily work. However, if the battery ceases to charge, the laptop becomes a network, made up technical components and human expertise. Thereby, in order to understand how *Twitter* visualisations are constructed, we need to explore the *actors* and the *process* of their creation with visualisations tools. My perspective to software and software-based artefacts is also based the notion, developed by ANT scholars, which states that non-humans are embedded with goals and functions and they prescribe actions, roles and skills back to human actors (Latour 1994, 31).

As an object of study, software can be formally defined as a set of "instructions and associated data that directs the computer to accomplish a task" (Shaffer 2010, 26). New media scholar Mirko Tobias Schäfer (2011) argues that software is "in-material" (ibid, 64), located between the real and the symbolic, being both the formulation in a programming language and the execution of the formulated action. In a Latourian sense, it is standing between word an action (ibid). Software programmes have a performative nature – they are carriers of scripts that enable users to perform activities, accomplish tasks and produce new

artefacts. In this perspective, tools for data visualisation can be perceived as objects that are embedded with conventions, deriving from the academic discipline of the creators of the tools. We can then expect that the visual artefacts, created by the tools will also follow these principles and thereby translate the embedded meanings into the new object.

2.4.1. Case study: Visualisations made with Gephi

As a particular object of study I would like to present an explorative research of a network visualisation, created with the software *Gephi*[13]. I chose this software tool because of its growing popularity for visual analysis of *Twitter* networks by academics and because it is freely available and accessible to researchers of all levels of experience.

Gephi is an open-source platform, specialised in the visualisation of networks and complex systems. It is designed in modules with allow researchers to import large data sets, to visualise them in a network form, spatialize, filter, manipulate and export the network image for further presentation (Bastian et al. 2009). The creators of the tool define is as: "[l]ike Photoshop but for data, the user interacts with the representation, manipulate the structures, shapes and colors to reveal hidden properties" (Gephi 2013). The networks consists of nodes (dots) connected with edges (lines) and different layouts give the shape of the graph. For example the *Force Atlas 2* layout (Figure 9) arranges the nodes in a linear fashion, based on the attraction and repulsion proportional to distance between nodes (Gephi 2011). Simply put, the most connected nodes will be pushed towards each other, while the least connected will be pulled away and thereby clusters of users will be formed. It is used to visualise small and middle sized graphs and is specially developed for qualitative interpretation of graphs, where the visual emphasis is put on entities which are complementarities (Gephi 2011b). In contrast, the *Circular* layout (Figure 10) distributes the nodes in a circle and orders them by criteria, chosen by the user such as ID, in-degree, attributes' type. This kind of topology is recommended when the goal of the visualisation is to highlight the ranking of notes. Furthermore, the software provides filtering, ranking and statistical metrics that can be applied to the visualised data. Ranking relates to the size of the notes by a selected criterion and statistical metrics, based on a parameter. For instance, a researcher might rank the nodes by the number of links they receive (in-degree) so that the size of a node will correspond to this value. Parallel to this, users can perform the metric *Modularity*, which measures how well a network decomposes into modular communities.

[13] Official website: <https://gephi.org>.

Figure 9: Force Atlas 2 (Gephi 2011a).

Figure 10: Circular layout (Gephi 2012).

To research the properties of Gephi in relation to the process of research, I visualised a network of *Twitter* replies, which I extracted from the database that I analysed during my internship. In my earlier study, I did not perform a network analysis of the data and in this respect my current work allowed me to add a new dimension to my previous work. The database consists of users who have used the word *flight ticket* in a conversation with at least one more user. In terms of size, the network has 1675 unique users (nodes) and 931 connections (edges).

In my research, I defined the following research questions:

a) Which are the most connected users?

b) Which users receive the most tweets?

c) Are there some distinct communities of users that communicate about the topic of flight tickets?

To explore these questions, I used the algorithm *Force Atlas 2* and employed three parameters: *between centrality*, *in-degree* and *modularity*. Structurally, this algorithm shows that the nodes in the analysed network are distributed almost evenly between each other, with the majority of the nodes concentrated closely together and few which are positioned outside of the group. The distance between the nodes is a parameter, called *average path length* and in the case of this data, it is rather low.

The first graph (Figure 11) presents the nodes, ranked by the parameter *between centrality*. This metric is defined by sociologists Linton Freeman as "the degree to which a point falls

on the shortest path between others and therefore has a potential for control of the communication" (Freeman 1977, 35). In *Gephi* this parameter has been discussed by computer scientist Patrick McSweeney who defines it in a similar fashion, simply as the algorithm that calculates "the shortest path between two nodes in the network" (Patrick 2009, 3). In this case, the bigger and darker nodes are the most central nodes in the graph. A zoom into the particular actors shows that these are users KLM and vacantieboeken4, followed by vliegticketsnl and TunafishNL. This graph served as a basis for the next visualisations, in which I added a second layer of calculation on top of it.

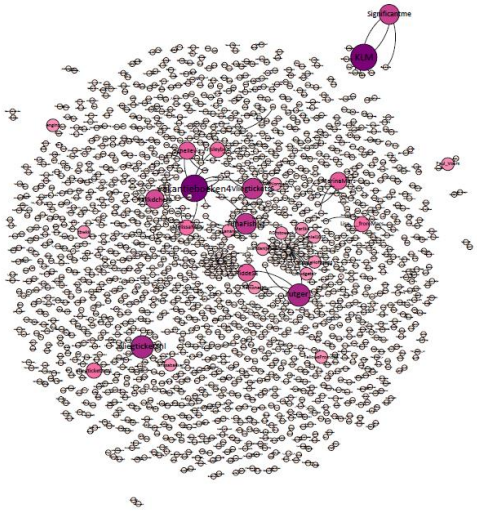


Figure 11: Nodes, ranked by the parameter "Between centrality".

In the second graph (Figure 12), the size of nodes is defined by the calculation *between centrality*, but the colour is based on the parameter *in-degree*, or the number of times a user has received tweets from other users. Thereby, the darker the nodes, the more attention these users have received. At first glance, the first two graphs look almost the same, but a closer look shows a shift in the main actors. In the second visualisation, the most popular actors are users KLM and Rutger_ followed by Schellevis, vliegticketsnl and MarinaMars. Thus, based on these two calculations, we can conclude that users who receive the most tweets are not the ones who are the most central in the network. This statement brings the issue of measuring influence, mentioned in the previous part of this study. It raises the question about how we should evaluate influence of Twitter users – based on the attention they receive or based on their relative position within the network of users? Furthermore, we can also see that both graphs display highly steep power law graphs so that very few users receive most of the attention while more than 90% of the users receive much less than the average.

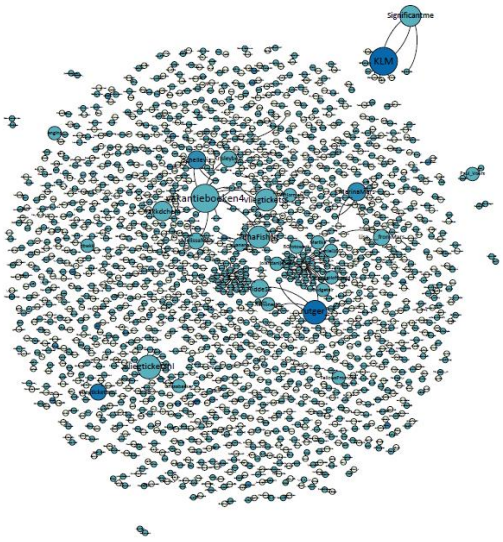


Figure 12: The size of nodes is defined by the calculation "Between centrality", and colour is based on the parameter "In-degree".

In the third set of graphs (Figures 13 and 14), I applied the metric *modularity* on top of the nodes, ranked by the factor *between centrality*. Here, the great variety of colours shows the overall *lack of communities* that can be detected within this particular network. After filtering the data (Figure 14) we can distinguish seven small clusters of users. Not surprisingly, the most connected nodes in the network seem to be the core, uniting factor of these groups of users. Of a particular interest for me was the fact one community is gathered among a node of the user "fflekkerweg", which was neither the most centred actor, nor the most popular by other users. In order to examine the factors, behind this peculiar grouping I explored the database and discovered that this was the user that I previously had detected as a non-human agent (bot) which had sent a large number of personal messages (factor called *out-degree*) and thus the *Gephi* algorithm arranged the recipients closer around this particular user.

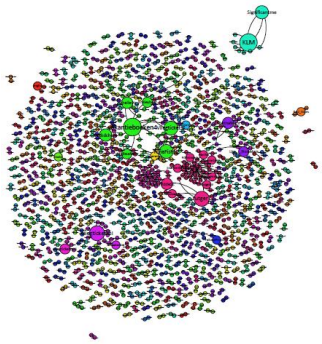


Figure 13: Nodes, divided by the parameter "Modularity".

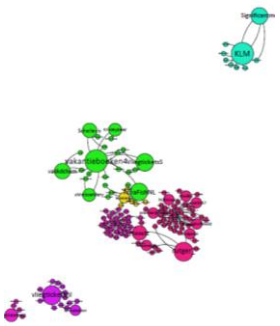


Figure 14: "Modularity", filtered.

As a result of this empirical experiment, I can conclude that the algorithm of *Gephi* allowed me to perceive the *Twitter* data as a network visualisation and interactively explore the structure and properties of users and their connections. The combination of algorithms *Between centrality* and *In-degree* was highly beneficial as it not only revealed the actors that I had not focused on in my previous study, but they raised the important questions regarding the methods of measuring user influence. However, the use of the *Modularity* algorithm was problematic. Although it did reveal few little clusters and insights, this particular dataset was too homogeneous and thus, the use of an algorithm that highlights community structures brought too much noise in the analysis.

2.4.2. Reflection on the software's influence on the process of research

In the final part of this chapter I would like to reflect on my experience with the tool with regards to the methods that I used and the decisions that I made within the research process.

First of all, in order to be able to use the software of *Gephi*, I first had to gain *theoretical knowledge* and *practical experience* about the properties of the tool. Uploading the data into the software required the database to be structured and labelled in a particular way, which in the case of the database that I was working with, involved the development of two new databases. After I upload the data, I decided to focus on analysing the structure of the data, by applying the *Force Atlas 2* algorithm. I chose this particular layout, after reading about its benefits to produce clear and readable spatialization of relational data (Jacomy et al. 2011) and that it has been used in studies with Twitter data (Bruns 2013; Paßmann et al. (Forthcoming)).

Getting acquainted with the features of the software was an experience in which I learned not only how to operate the tool, but also major principles and approaches, deriving from the field of network science, visual statistics and mathematics. The main methods of analysis on which *Gephi* is embedded with are the principles of graph analysis, introduced by Jacob Moreno (1953) and sociologist Linton Freeman (1977). Moreno is considered the 'father' of the sociogram (Freeman 2000) and for him the visualisation of networks was the starting point of the analysis of a social group:

> "We have first to visualize . . . A process of charting has been devised by the sociometrists, the sociogram, which is more than merely a method of presentation. It is first of all a method of exploration. (Moreno 1953, 95-96)

Coming from the field of new media, the process of understanding the properties of *Gephi* was also a process of acquiring *skills and knowledge* about the main principles of graph theory and social network analysis.

When approaching the tool with the database of my internship, my first intention was to focus on examining the structure of the network and identifying the most important nodes. But as soon as I started exploring the different algorithms, I found that the tool provided a much greater scope of calculations than I had first planned to apply. The quick and easy interface allowed me to explore the possibilities of the tool – the immediate result which appeared as a visualisation and the element of exploration made the experience feel more like a game, than a research process. I found myself in constant process of *negotiation* between my initial goals and the possibilities of the software. As a result, I reconfigured my research questions with parameters which seemed more beneficial for the research.

Lastly, my research method was influenced by the visual form of particular visualisation algorithms, applied to the database that I was working with. For instance, when I applied the ranking of the nodes based on the factor *In-degree,* the visualisation was too homogenous (Figure 15) and thus the navigation between the nodes was highly problematic. Here, I based my criteria for clarity of the network image on the principles, defined by statistician and information design theorist Edward Tufte (2001). For Tufte, proper statistical visualisations must present complex data with clarity, avoid distortion of the data and reveal the data at several levels of detail (ibid, 13). The lack of these factors in the visualisation with ranking *In-degree* made me *rethink my method* and try different criteria, despite the fact that first calculation was mathematically accurate. As a result, I kept the ranking of the nodes on the calculated centrality between the nodes and the colouring was based on the factor on the metric *In-degree*.

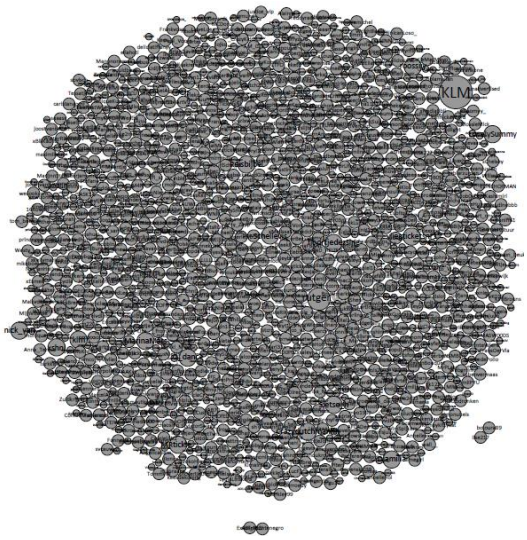


Figure 15: Ranking the nodes based on the metric "In-degree".

To conclude, in this chapter I argued that software tools are embedded with properties and functions which invite scholars to *rethink* their *research goals* and *methods of study*. In order to explore the extent of this effect I presented a research done with the visualisation software *Gephi*. By using a combination of algorithms, I produced several visualisations which allowed me to add a new dimension to the analysis of the data and to identify actors which I would not otherwise be able to distinguish as important. Based on this experience I stated that the properties of the software of *Gephi* influenced my method of research in three main directions. First, my research question was changed as a result of the *knowledge and skills* I acquired while learning the properties of the tool. Second, the easy to use interface and fast calculations, invited me to explore much more algorithms and layouts which led to a *change in my initial method* of research. Third, my choices of visual representation were to a great extent influenced by the quality of the visual form and the final visualisations were a result of a *negotiation* between my goals and the algorithms which would afford clarity and precision of the visual analysis.

The analysis of the role of the visualisation software *Gephi,* in the methods of research is the last element of the thick description, representing a holistic approach to the study of *Twitter*. Reflecting on this experience, I would like to state that my research question, goals and methods were in *constant flux,* spilling within the whole process and being transformed by the properties of each actor. I summarise the steps that I took in this research journey in the graph below (Figure 16). I would like to emphasise that this is a simplified version of a

very particular research project. However, I hope that it can be a useful model which sheds light on some of the possible stages of a *Twitter* study.

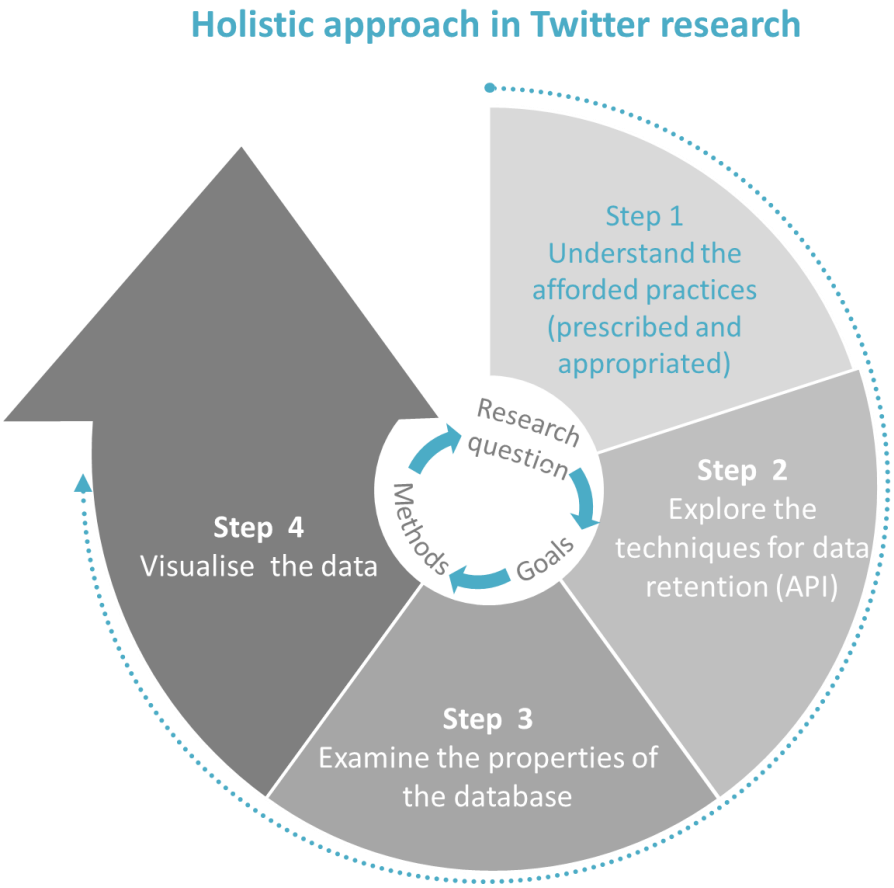


Figure 16: The steps, taken in the process of *Twitter* research (simplified).

Conclusion

Online social networks have already been established as part of the daily life of online users worldwide. Whether we support their practices or not, their ubiquitous presence will continue to trigger the interest of academics to explore users' relations, communication and patterns that occur in these environments. With this, the question behind research methods for scholarly exploration becomes a main concern and increasingly, the debate regarding techniques that can be applied is emerging. This issue is particularly important for scholars in the fields of humanities and social sciences, because in online environments their main objectives cannot be reached using traditional methods and scholars are forced to develop new research techniques.

I began this thesis by outlining the theoretical discussion about the notion of online research, employed by academics in ethnography, sociology, digital humanities, computational social sciences, digital sociology and digital methods. This overview has shed light on two contradicting perspectives about the approach of academics to Internet exploration – one, which advocates for adapting existing methods and techniques to digital environments, and the other, which argues for the adoption of brand new research methods. While the first have been argued to be out-dated and limiting the research potential of online methods, I stated that the latter view does not necessarily provide a comprehensive solution and proper guidance about the kind of questions that scholars should ask before conducting online research. I have criticised the recently developed computational methods, arguing that they reduce the analysis to a particular tool or an algorithm, encouraging scholars to think first about the availability of software tools, instead defining their research question and goals. Furthermore, I opposed current views on the notion of online methods which use data as a single object, and do not recognise it as a result of processes, which take place within a particular context. In order to bridge this widening gap in the understanding of methods within academia, I have suggested the concept of a *holistic approach* to online research. Based on the principles of Actor-Network Theory, this approach proposes that both the research objects and research methods should be perceived as socially constructed, where both human and non-human actors form a network of associations.

In my empirical part, I have presented an explorative research on the microblogging platform *Twitter*, employing the holistic approach towards the actors and reflecting on the

practices involved in this exploration. I first elaborated on the platform's characteristics and the way in which they shape user behaviour. By investigating the type of interactions enabled by *Twitter* I have argue that the platform affords practices which are *implicit*, *obscure* and *appropriated* by *Twitter* users. Thus, in order to understand these dynamics, I have stated that scholars need to develop tactics, which would allow them to look *beyond the logic of the database*. In this regards, I have shown that traditional *ethnographical* methods can be employed as a possible technique in detecting practices of user appropriation.

Furthermore, I presented the role of the provided APIs for data retention, stating that they confine the research within particular rules. The rules then *limit* the scope of actors that can be explored and problematize the analytical process. Next, I investigated the properties of the data, extracted through the API of *Twitter*. This study showed that data comes in an already formatted fashion that invites scholars to use the provided categories as part of their studies and thereby creating a research, which is highly *dependent* on the platform. Finally, I visualised the data with the software tool *Gephi* and reflectively analysed how it affected my research method. As a conclusion of this visualisation process, I stated that the properties of the software not only translated particular skills but they also altered my initial research question, goals and methods, making them much more fluid and unstable.

While the use of algorithms and tools is expanding in academia, scholars in the fields of humanities and social sciences will continue to adopt more software applications as part of their research practice. Working with massive amounts of data, deriving from online environments will continue to proliferate and transform the existing practices of academic research. With the use of a *holistic approach*, presented in this thesis, I showed that academics can employ traditional as well as new methods in their social media studies and thereby, bridging the gap between existing academic principles of research and the recently developed digital techniques. In this thesis, I have proven that social media exploration takes place on several levels and on each stage, the research practice is shaped by the actors. Thus, the concept of research method cannot be confined within the traditional frames, but instead, it should be perceived as an evolving process, being in a state of constant flux, where heterogeneous actors influence the research decisions, mobilise traditional and new methods and negotiate the research choices of scientific exploration. Instead of black-boxing these processes, I suggest that we need to deconstruct them, to record the steps we take and investigate how scholars make knowledge with social media

data. In this exploration, we might need to create more thick descriptions, conduct laboratory studies or employ ethnographic methods that will shed light on the context in which knowledge is being created. It is up to future academic studies to develop a praxeology of social media research, which can establish this type of exploration not only as a technological accomplishment, but as complex holistic process.

Bibliography

Akrich, Madeleine. 1992. "The De-Scription of Technical Objects", In W.E. Bijker & J. Law (eds), *Shaping Technology/Building Society*. Cambridge, MA: MIT Press: 205—224.

Anderson, Chris. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete". *Wired magazine website*. <http://www.wired.com/science/discoveries/magazine/16-07/pb_theory>. (accessed on February 23 2013).

Andrews, Marc, Bootsman, Joris, Borra, Erik, Dresscher, Paulien, Rek, Declan, van Sprang, Maarten and Matthijs Volkers. 2008. "PostDemographics". *Digital Methods Initiative Website* <https://www.digitalmethods.net/Digitalmethods/PostDemographics>. (accessed April 11 2013).

Anger, Isabel and Christian Kittle. 2011. "Measuring Influence on Twitter". Proceedings of the *11th International Conference on Knowledge Management and Knowledge Technologies* 31. ACM New York.

Appadurai, Arjun. 2005. *Modernity At Large: Cultural Dimensions of Globalization*. Minneapolis: University of Minnesota Press.

Barabàsi, Albert-László. 2003. *Linked: How Everything Is Connected to Everything Else and What It Means*. Plume.

Barocas, Solon, Hood, Sophie and Malte Ziewitz. 2013. Governing Algorithms – A Provocation Piece. Paper, presented as Conference on Computation, Automation, and Control at New York University, May 16-17, 2013. <http://governingalgorithms.org/resources/provocation-piece>.

Bashky, Eytan, Janke M. Hofman, Winter A. Manson, and Duncan J. Watts. 2011. "Everyone's an Influencer: Quantifying Influence on Twitter". Paper presented at the *WSDM '11: Fourth ACM international conference on Web search and data mining*. New York, NY, USA.< http://dl.acm.org/citation.cfm?id=1935845>.

Bastian, Mathieu, Heymann, Sebastien and Mathieu Jacomy. 2009. "Gephi: an open source software for exploring and manipulating networks". Paper, presented at *International AAAI Conference on Weblogs and Social Media.* <https://gephi.org/users/publications>.

Beevolve. 2012. "An Exhaustive Study of Twitter Users Across the World". *Beevolve Website.* <http://www.beevolve.com/twitter-statistics/#c1>. (accessed May 6 2013).

Bell, Gavin. 2009. *Building Social Web Applications*. Sebastopol: O'Reilly Media. <https://www.mgov2.org/documentos/social_web_applications_oreilly.pdf>.

Berry, David. (ed). 2012. *Understanding Digital Humanities*. Palgrave Macmillan.

---. 2011. The Computational Turn: Thinking About Digital Humanities. In *Digital Humanities: Beyond Computing* 12. <http://www.culturemachine.net/index.php/cm/issue/view/23>.

boyd, danah. 2006. Friends, Friendsters, and MySpace Top 8: Writing Community into Being on Social. Network Sites. *First Monday* 11(12). <http://www.firstmonday.org/ojs/index.php/fm/article/view/1418>.

boyd, danah and Nicole Ellison. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication* 13(1), article 11. <http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>.

boyd, danah. 2010. Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. In *Networked Self: Identity, Community, and Culture on Social Network Sites* (ed. Zizi Papacharissi). London: Routledge: 39-58.

boyd, danah and Kate Crawford. 2011. Six Provocations for Big Data. Paper presented at *Oxford Internet Institute's – A Decade in Internet Time: Symposium, on the Dynamics of the Internet and Society*. University of Oxford.

Bollen, Johan, Mao, Huina, Xiao-Jun Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2(1): 1-8.

Bruns, Axel and Jean Burgess. 2011. The Use of Twitter Hashtags in the Formation of Ad Hoc publics. In *6th European Consortium for Political Research General Conference*, 25 - 27 August 2011, University of Iceland, Reykjavik. <http://eprints.qut.edu.au/46515>.

Bruns, Axel. 2013. "Visualising Twitter Dynamics in Gephi, Part 2". *Mapping Online Publics Website*. <http://mappingonlinepublics.net/2010/12/30/visualising-twitter-dynamics-in-gephi-part-2>. (accessed May 21 2013).

Burger, John, Henderson, John, Kim, George and Guido Zarrella. 2011. "Discriminating Gender on Twitter". Proceedings of the *Conference on Empirical Methods in Natural Language Processing*. ACM New York: 1301-1309.

Callon, Michel. 1986. "The Sociology of an Actor-Network: The Case of the Electric Vehicle." In M*apping the Dynamics of Science and Technology: Sociology of Science in the Real World*, edited by M. Callon, J. Law, and A. Rip. Houndmills, UK: Macmillan.

Cha, Meeyoung, Hamed Haddadi, Hamed Benevenuto and Krishna P. Gummadi. 2010. "Measuring User Influence in Twitter: The Million Follower Fallacy". Proceedings of International AAAI Conference on Weblogs and Social Media (ICWSM), May 2010. <http://twitter.mpi-sws.org/icwsm2010_fallacy.pdf>.

Chen, Curtis. 2010. "Twitter API Adds Retweet Count and More". *Programmable Web*. <http://blog.programmableweb.com/2010/08/30/twitter-api-adds-retweet-count-and-more>. (accessed May 11 2013).

Collier, Nigel, Truong Son, Nguyen and Ngoc Mai Nguyen. 2011. OMG U got flu? Analysis of Shared Health Messages for Bio-Surveillance. *Journal of Biomedical Semantics*, 2(5): S9. <http://www.jbiomedsem.com/content/2/S5/S9>.

Durkheim, Emile. 1982. *Rules of Sociological Method*. Free Press.

Farrel, Dan and Peter Petersen. 2010. The Growth of Internet Research Methods and the Reluctant Sociologist. *Sociological Inquiry* 80 (1): 114–125.

Freeman, Linton. 1977. A Set of Measures of Centrality Based on Betweenness. *Sociometry* 40 (1): 35-41.

---. 2000. Visualizing Social Networks. *Journal of Social Structure 1 (1)*. <http://www.cmu.edu/joss/content/articles/volume1/Freeman.html>.

Foucault, Michel. 2002. *The Archaeology of Knowledge*. London: Routledge.

Fuller, Matthew. 2008. *Software Studies. A Lexicon*. Cambridge, MA: The MIT Press.

Galloway, Alexander. R. 2004. *Protocol: How Control Exists After Decentralization.* Cambridge, MA: MIT press.

Garton, Laura, Haythornthwaite, Caroline and Barry Wellman. 2006. Studying Online Social Networks. *Journal of Computer-Mediated Communication 3 (1),* June 1997.

Gephi. 2013. "Features". *Gephi Official Website*. <https://gephi.org/features>. (accessed on May 18 2013).

---. 2012. "Circular Layout". Gephi Marketplace Website. <https://marketplace.gephi.org/plugin/circular-layout>. (accessed on 20 May 2013. accessed on May 20 2013).

---. 2011a. "Force Atlas 2, the New Version of a Home Brew Layout". *Gephi Official Website*..<http://gephi.org/2011/forceatlas2-the-new-version-of-our-home-brew-layout>. (accessed on May 20 2013).

---. 2011b. "New Tutorial Layouts in Gephi". Gephi Official Website. <https://gephi.org/2011/new-tutorial-layouts-in-gephi>. (accessed on May 20 2013).

Gibson, James Jerome. 1986. *The Ecological Approach to Visual Perception*. Psychology Press: New Ed edition.

Halevi, Gali, & Henk F. Moed. 2012. Research Trends. *Special Issue on Big Data*, issue 30, September 2012. <http://www.researchtrends.com/wp-content/uploads/2012/09/Research_Trends_Issue30.pdf>.

Hayles, N. Katherine. 2012. "How We Think: Transforming Power and Digital Technologies". In *Understanding the Digital Humanities*, D. M. Berry (ed.). London: Palgrave Macmillan.

Helmond, Anne. 2010. On the Evolution of Methods: Banditry and the Volatility of Methods. *Personal blog,* <http://www.annehelmond.nl/2010/05/17/on-the-evolution-of-methods-banditry-and-the-volatility-of-methods>. (accessed on May 23 2013).

Hecht, B.; Hong, L.; Suh, B.; and Chi, E. H. 2011. "Tweets from Justin Bieber's heart: The dynamics of the location field in user profiles". In *Proceedings of the 2011 annual conference on Human factors in computing systems*, *CHI '11*. ACM New York: 237–246.

Hine, Christine. 1994. "Virtual ethnography". Public Communication on Science and Technology. <http://www.cirst.uqam.ca/PCST3/PDF/Communications/HINE.PDF>.

---. 2000. *Virtual Ethnography*. London: Sage.

---. 2005. *Virtual Methods: Issues in Social Research on the Internet*. Oxford:Berg.

Jacomy, Mathieu, Heymann, Sebastien and Tommaso Venturini. 2011. "ForceAtlas2, A Graph Layout Algorithm for Handy Network Visualization". *WebAtlas Website.* <http://webatlas.fr/tempshare/ForceAtlas2_Paper.pdf>. (accessed June 18).

Java, Akshay, Song, Xiaodan, Finin, Tim and Belle Tseng. 2007. "Why We Twitter: Understanding Microblogging Usage and Communities". Proceedings of the *Joint 9th WEBKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis.* ACM New York: 56-65.

Jones, Steve (ed). 1998. *Doing Internet Research. Critical Issues and Methods for Exploring the Internet*. London: Sage.

Kim, Dongwoo, Jo, Yohan, Moon, Il-Chul, and Oh, Alice. 2010. "Analysis of Twitter Lists as a Potential Source for Discovering Latent Characteristics of Users". Conference paper at the *ACM Conference on Human Factors in Computer Systems*. <http://people.csail.mit.edu/aoh/papers/chi2010microblogging.pdf>.

Kinash, Shelley. 2012. "Paradigms, Methodology & Methods". *Bond University Website,* <http://www.bond.edu.au/prod_ext/groups/public/@pub-tls-gen/documents/genericwebdocument/bd3_012336.pdf>. (accessed May 22 2013).

Köhler, Wolfgang. 1970[1929]. *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*. New York: Liveright.

Koffka, Kurt. 1922. Perception: And Introduction to the Gestalt-theorie. *Psychological Bulletin* 19: 531-585.

Knorr-Cetina, Karin. 1981. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Pergamon Press.

Kwak, Haewoon, Lee, Changhyun, Park, Hosung and Sue Moon. 2010. "What is Twitter, a Social Network or a News Media". Proceedings of the *19th International World Wide Web (WWW) Conference*, Raleigh NC.

Law, John.1992. Notes on the Theory of the Actor Network: Ordering, Strategy and Heterogeneity *Systems Practice* 5 *(4)*. <http://www.lancs.ac.uk/fass/sociology/papers/law-notes-on-ant.pdf>.

---. 2004. *After Method. Mess in Social Science Research*. London/New York: Routledge.

Latour, Bruno. 1986. Visualisation and Cognition: Drawing Things Together. In H. Kuklick (editor). K*nowledge and Society Studies in the Sociology of Culture Past and Present* 6. Jai Press: 1-40.

---. 1987. Science in Action. Harvard University Press.

---. 1994. On Technical Mediation-Philosophy, Sociology, Genealogy. *Common Knowledge 3 (2):* 29-64.

---. 1997. Stengers'Shibbolet. Foreword to *Power and Invention: Situating Science*, by Isabelle Stengers. University of Minnesota Press: xiii–xiv.

---. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. and Van Alstyne, M. 2009. Computational Social Science. *Science* 323(5915): 721–723.

Lessig, Lawrence. 2006. *Code: And Other Laws of Cyberspace, Version 2.0*. Basic Books.

Madge, Clare and Henrietta O'Connor. 2004. Exploring the Internet as a medium for research: web-based questionnaires and online synchronous interviews. *ESRC Research Methods*. <http://www.ccsr.ac.uk/methods/publications/documents/WorkingPaper9>.

Manovich, Lev. 2008a. Software Takes Command. <http://black2.fri.uni-lj.si/humbug/files/doktorat-vaupotic/zotero/storage/D22GEWS3/manovich_softbook_11_20_2008.pdf>. (accessed on May 20 2013).

---. 2008b. "Cultural Analytics". *Software Studies Initiative Website.* <http://lab.softwarestudies.com/2008/09/cultural-analytics.html>. (accessed on June 12 2013).

---. 2012a. "How to follow software users". *Software Studies Initiative Website.* <http://lab.softwarestudies.com/2012/04/new-article-lev-manovich-how-to-follow.html>.

---. 2012b. "Media Visualization: Visual Techniques for Exploring Large Media Collections". In *Media Studies Futures*, ed. Kelly Gates. Blackwell Publishers.

---. 2010. What is Visualization?. In *Visual Studies* 26 (1): 36-49.

---.2011. "Trending: The Promises and the Challenges of Big Social Data". *Debates in the Digital Humanities,* ed. M.K.Gold. The University of Minnesota Press, Minneapolis, MN. <http://www.manovich.net/DOCS/Manovich_trending_paper.pdf>.

Markoff, John. 2011. "Government Aims to Build a "Data Eye in the Sky". *The New York Time*s < https://www.nytimes.com/2011/10/11/science/11predict.html>. (accessed on June 18).

Marin, Alexandra and Barry Wellman. 2011. Social Network Analysis: An Introduction. *The SAGE Handbook of Social Network Analysis*, Ed. Peter Carrington and John Scott. London: Sage.

Marres, Noortje. 2012. The Redistribution of Methods: On Intervention in Digital Social Research, Broadly Conceived. The *Sociological Review* 60: S1, pp. 139–165.

Marres, Noortje and Esther Weltevrede. 2013. Scraping the Social?. *Journal of Cultural Economy*. <http://dx.doi.org/10.1080/17530350.2013.772070>.

McMillan, Robert. 2013. "The Big-Data Interview: Making Sense of the New World Order". *Wired Magazine.* < http://www.wired.com/wiredenterprise/2013/03/big-data>. (accessed June 18).

Mitchell, Melanie. 2006. Complex systems: Network thinking. *Artificial Intelligence 170*: 1194–1212.

Moreno, Jacob. L. 1953. *Who Shall Survive?* .Beacon, N.Y.: Beacon House Inc.

Newman, M., Barabási, A. and Watts, D. 2007. *The Structure and Dynamics of Networks*. Princeton, NJ: Princeton University Press.

Norman, Donald. 2002[1988]. *The Design of Everyday Things*. Cambridge, MA: MIT Press.

Oppenheim, Paul and Hilary Putnam. 1958. "The unity of science as a working hypothesis", in H. Feigl et al., eds., *Minnesota Studies in the Philosophy of Science* 2. Minneapolis: Minnesota University Press.

Page, Larry., Brin, Sergey., Motwani, R Rajeev and Terry Winograd, 1999 "The PageRank Citation Ranking: bringing order to the Web. Technical Report". *Stanford InfoLab*. <http://ilpubs.stanford.edu:8090/422>.

Papacharissi, Zizi. 2009. The Virtual Geographies of Social Networks: A comparative analysis of Facebook, LinkedIn and ASmallWorld. *New Media & Society* 11. London: SAGE: 199-220.

Paßmann, Johannes., Boeschoten, Thomas. and Schäfer, Mirko Tobias. (Forthcoming). "The Gift of the Gab: Retweet cartels and gift economies on Twitter", in *Twitter and Society*, Weller, K., Bruns, A., Burgess, J., Mahrt, M., and Puschmann, C. (eds.), New York: Peter Lang.

Patrick, McSweeney J. 2009. "Gephi Network Statistics". Paper proposed for *Google Summer of Code 2009*. <http://web.ecs.syr.edu/~pjmcswee/gephi.pdf>.

Polycom. 2013. Policom Website, <http://www.polycom.co.jp/global/documents/services/professional_services/datasheets/api-software-services-faq.pdf>. (accessed 10 May 2013).

Popper, Karl. 2002. *The Logic of Scientific Discovery*. London: Routledge.

Procter, Rob, Vis, Farida and Alex Voss. 2011. "Reading the riots. Investigating England's summer disorder". *The Guardian*.

<http://www.guardian.co.uk/uk/interactive/2011/dec/07/london-riots-twitter>. (accessed March 25 2013).

Rogers, Richard. 2009. *The End of the Virtual: Digital Methods*. Amsterdam: Amsterdam University Press. <http://www.govcom.org/publications/full_list/oratie_Rogers_2009_preprint.pdf>.

Saukko, Paula. 2003. *Doing Research in Cultural Studies: An Introduction to Classical and New Methodological Approaches*. London: SAGE.

Schäfer, Mirko Tobias. 2011. *Bastard Culture! How User Participation Transforms Cultural Production*. Amsterdam: Amsterdam University Press.

Schnapp, Jeffrey and Todd Presner. 2009. "Digital Humanities Manifesto 2.0". *Humanities Blast Website.* <http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf>. (accessed April 11 2013).

Shaffer, Ann, Patrick Carey at al. 2010. *New Perspectives on Microsoft Office 2010, First Course*. Boston. MA: Course Technology.

Shapin, Steven and Simon Schaffer. 1989. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton University Press.

Skågeby, Jörgen. 2010. "Online Ethnographic Methods: Towards a Qualitative Understanding of Virtual Community Practices". In *Handbook of Research on Methods and Techniques for Studying Virtual Communities: Paradigms and Phenomena*, ed. Ben Kei Daniel. IGI Global; 1 edition: 410-428.

Tergan, Sigmar-Olaf and Tanya Keller (Eds.). 2005. *Knowledge and Information Visualization: Searching for Synergies (Lecture Notes in Computer Science / Information Systems and Applications, incl. Internet/Web, and HCI)*. Springer-Verlag Berlin Heidelberg.

Tufte, Edward. 2011. *The Visual Display of Quantitative Information*. Graphics Press.

Twitter. 2009. "The Streaming APIs".*Dev.Twitter.com.* <https://dev.twitter.com/docs/streaming-apis>. (accessed May 8).

---. 2012a. "Things Every Developer Should Know". *Dev.Twitter.com.* <https://dev.twitter.com/docs/things-every-developer-should-know>. (accessed May 7 2013).

---. 2012b. "Developer Rules to the Road". *Dev.Twitter.com.* <https://dev.twitter.com/terms/api-terms>. (accessed May 9 2013).

---. 2013a. "About Twitter". Official blog. <https://twitter.com/about>. (accessed April 19 2013)

---. 2013b. "Twitter Certified Products". *Dev.Twitter.com.* <https://dev.twitter.com/programs/twitter-certified-products/products>. (accessed April 19 2013).

---. 2013c. "Platform Objects". *Dev.Twitter.com* . <https://dev.twitter.com/docs/platform-objects>. (accessed May 11 2013).

---. 2013d. "What are Hashtags Symbols". *Support.Twitter.com* . <https://support.twitter.com/entries/49309-what-are-hashtags-symbols>. (accessed May 11 2013).

Watts, Duncan. 2007. A Twenty-First Century Science. *Nature*. 445 (7127): 489-489.

Weng, J.; Lim, E.-P.; Jiang, J.; and He, Q. 2010. "TwitterRank: Finding Topic-Sensitive Influential Twitterers". In: Proceedings of the *Third ACM International Conference on Web Search and Data Mining*. ACM New York: 261-270.

Wertheimer, Max. 1912. Experimentelle Studien über das Sehen Bewegung. *Zeitschrift für Psychologie*, 61: 247-250.

Wertheimer, Max. 1938 [1925]. "Gestalt theory". In W. D. Ellis (Ed. & Trans.), *A source Book of Gestalt Psychology*. London: Routledge & Kegan Paul.

Willig, Carla 2008. *Introducing Qualitative Research in Psychology*. Open University Press: 1s-11.

Word of Mouth Marketing Association (WOMMA). 2010. "An Introduction to WOM Marketing with Definitions". *WOMMA Website.* <http://www.womma.org/wom101> . (accessed on March 6 2012).

Zhao, Dejin, and Mary B. Rosson. 2009. "How and Why People Twitter: The Role that Micro-Blogging Plays in Informal Communication at Work". Proceedings of *GROUP 2009*. <http://cscl.ist.psu.edu/public/users/mrosson/Working/GROUP09-Zhao.pdf>.