LSE Media and Communications

Media@LSE Working Paper Series

Editors: Bart Cammaerts, Nick Anstead and Richard Stupart

In Search for 'Liveliness'

Experimenting with Co-Occurrence Analysis Using #GDPR on Twitter

Sameeh Selim

Published by Media@LSE, London School of Economics and Political Science ("LSE"), Houghton Street, London WC2A 2AE. The LSE is a School of the University of London. It is a Charity and is incorporated in England as a company limited by guarantee under the Companies Act (Reg number 70527).

ABSTRACT

This study seeks to contribute to recent debates concerning computational social science by experimenting with 'co-occurrence analysis' on a Twitter dataset relating to the subject of the recently introduced General Data Protection Regulation (GDPR). In so doing, the study implements the method in the context of hashtags associated with the subject, in an attempt to test the method's capacities and limitations to detect liveliness - dynamism and engagement with a given topic - on Twitter. The study highlights the significance of a qualitative assessment of tweets in combination with network mapping. It further suggests that the ability of hashtag analysis to detect liveliness of issues on Twitter is influenced by the type of actors involved, insofar as the ubiquitous use of multiple hashtags in the same tweet as observed in this study poses serious limitations to hashtag analysis. In terms of specific hashtag analyses, the research stresses the importance of a comprehensive exploration of the various entities associated with the hashtag, including other hashtags, users, hyperlinks, and tweet source, as well as how they shift over time, if a true understanding the hashtag's 'life' is to be obtained.

1 INTRODUCTION

Datafication can be defined as the process by which different aspects of our lives are continuously translated into data (Cukier & Mayer-Schoenberger, 2013). Smartphones in our pockets monitor our daily movements through their connections to GPS satellites, credit card transactions record our purchasing behavior, while social networking sites capture people's interactions, political views, and mundane everyday thoughts. This dissertation is specifically concerned with the latter example of datafication and its implications on knowledge production and social research.

Scholarly debates about the extent to which the ongoing datafication of our social lives enables new ways of studying society have been gaining pace over the last 20 years. The rise of what is referred to as Web 2.0 at the turn of the century changed the nature of the Internet from a source of information to a space of interaction and participation leaving behind an ever-growing footprint of digital traces. Unsurprisingly, social scientists are interested in not only the social, economic, and political implications of these developments, but also, in what the increasingly important role 'the digital' plays in people's lives can mean for social research. Terms like digital humanities (Schreibmann et al., 2008; Kirschenbaum, 2010), digital sociology (Wynn, 2009; Lupton, 2014; Marres, 2017), and computational social science (Lazer et al., 2009) all revolve around what can be referred to as the computational turn - that is, the growing interest in technology and computational research approaches across the humanities and social sciences (Berry, 2011).

While the continuous hype around 'big data' and its potential to revolutionize research in both the natural and social sciences has rightfully lead a growing number of scholars to critically evaluate the implications of this development from an ethics and privacy perspective (see boyd & Crawford, 2012; Schroeder, 2014), Jose Van Dijck (2017) stresses that in addition to tackling these challenges, the shift to a 'datafied society' requires scholars and students to work on expanding their research skills and 'to become critical data practitioners who are both capable of working with data and of critically questioning the big data myths that frame the datafied society (p.12). The use of new computational methods to analyze large amounts of online data is celebrated by many, while others are rather skeptical. Perhaps Chris Anderson's (2008) suggestion that 'the data deluge makes the scientific method obsolete' is an example of an extreme statement challenged by many academics. While Anderson's suggestion has been discredited (see Kitchin, 2014a), it remains important for researchers in the social sciences to examine the value of various 'digital methods' which are becoming an 'integral part of the social research toolkit' (Snee, Hine, Morey, Roberts, & Watson, 2016). Exploring innovative methods in empirical research while critically reflecting on the epistemological implications of this new media environment is therefore crucial (Kubitschko & Kaun, 2016). That being said, the use of digital methods seems not to have yet been accepted as mainstream within the social sciences (Snee et al., 2016).

With this as a backdrop, this dissertation aims to contribute to debates around digital methods, by applying a specific method- co-occurrence analysis -on a Twitter dataset relating to the subject of the General Data Protection Regulation (GDPR). In doing so, the study follows Noortje Marres and Carolin Gerlitz's call for 'experimental inquiry into what makes [digital methods'] deployment productive for social inquiry' (Marres & Gerlitz, 2016, p.23), and aims to test the method's ability to detect 'liveliness' of issues on Twitter in a relevant and timely subject. To be clear, the topic in question (GDPR) is taken as an *occasion* to study the method. The choice was made following Tommaso Venturini's (2010) advice of what constitutes 'a good controversy' to study, which will be elaborated on below.

2 THEORETICAL CHAPTER

2.1 Literature Review

2.1.1 The Rise of Internet Research

With the ubiquitous rise of Internet use in the 1990s, the interest in investigating its implications gradually grew and scholars began examining the dynamics of what used to be referred to as the virtual world. At the time, Internet research was primarily focused on studying 'virtual communities'. Wellman & Gulia (1999) for example ask questions such as whether or not strong and intimate relationships are possible on the Internet and the extent to which such virtual communities affect 'real life communities', while Howard Rheingold (2000) explores the dynamics and implications of virtual communities in light of his own experience as an active member in an online community named the WELL. In 'Doing Internet Research', Jones (1998) collaborated with a number of scholars to examine and critique different methods of studying the Internet, exploring the use of various quantitative and qualitative methodologies and how they can be adapted in the study of the Internet. Similarly, Hine (2005) presents a number of case studies that explore different methodological solutions to the increased use of established social science methods such as surveys, interviews, and ethnographic studies via the Internet.

2.1.2 The Internet as a Source of Insight: On Big Data and Digital Methods

As such, in the late 1990s and the early years of the 21st century the focus was on doing both research *about* the Internet as a new media technology worthy of studying, and on adapting existing social science methodologies to the new reality of computer-mediated communications. In 'Digital Methods' Richard Rogers (2013) emphasizes the distinction between the 'digitized' and the 'natively digital'. That is, the importation of traditional research methods in the social sciences (e.g. surveys) into the web on the one hand, and the utilization of what he refers to as 'methods of the medium' (p.1) on the other. With methods of the medium Rogers is referring to digital acts like crawling, scraping, crowdsourcing, etc. He poses the questions of how such digital acts together with digital objects like hashtags, tweets, hyperlinks, mentions, etc.., can be utilized for social and cultural research purposes and to what extent cultural and societal claims can be grounded in web data, a concept he terms 'online groundedness' (p.23).

While the idea of using 'big data' from social media and other online platforms to make claims about the social in new ways has been discussed by many (Cukier & Schoenberger, 2012; Kitchin, 2014a), the 'digital methods' proposal presented by Rogers differs on key aspects. The 'big' in big data reveals a clear preoccupation with the size of the data set: big data is 'huge in volume' and seeks to be 'exhaustive in scope' (Kitchin, 2014a, p.1). The large size and supposed exhaustive nature of the data leads the big data argument to be primarily concerned with detecting correlations and recognizing patterns. As Cukier & Schoenberger (2012) put it: 'Correlations are useful in a small-data world, but in the context of big data they really shine' (p.124). In contrast, Rogers' proposal is to examine the web as a source of data regardless of the size of the data set (Rogers, 2013). He proposes a type of digital social research that considers the effects of the device from which the data is extracted. His proposition is for researchers to 'follow the methods of the medium as they evolve, learn from how dominant devices treat natively digital objects, and think along with those object treatments and devices so as to recombine or build on top of them' (Rogers, 2013, p. 5). As such, for this kind of 'device-driven' digital research what is significant is how digital media enable new and distinctive modes of analysis (Weltevrede, 2016).

Although established methods that are 'tried and tested' clearly dominate the social sciences and humanities, the use of innovative digital methods is gradually picking up. Yet the definition of what qualifies as innovation in the context of social research methods is not something that scholars seem to agree on. Wiles, Howell, Crow, and Nind (2013) explain that while some regard innovation in methods as applying only to the development of new methods, others' definition of innovation is inclusive of advances in established research methods (Taylor & Coffey, 2008). For Snee et al. (2016) innovation considers how digital methods supplement established social science research (p.222). In the context of media and communications research specifically, Kubitschko & Kaun (2016) suggest that the term innovation should reflect 'the lively and productive qualities of emerging methods' (p.3), as opposed to anything that is new. They recommend 'the widening and rethinking [of] research methods to further understand the role that media technologies and infrastructure play in society' (p.4).

2.1.3 Actor-Network Theory and Digital Controversy Analysis

The potential of digital media to enable new forms of social research is discussed by Latour et. al. (2012) suggesting that the increased presence of digital traces allows for social research which questions classical theories about social order. The authors demonstrate that newly available digital tools for network analysis and visualization enable the practice of actor-network theory (ANT) in new and practical ways.

It is worthwhile at this point to briefly elaborate on the arguments of ANT, as it forms an important basis for the method employed in this research. ANT was first developed in the 1980s by Bruno Latour and Michel Callon in the field of science and technology studies (STS) which is essentially interested in the interplay between science and society and has ever since been applied in various other disciplines within the social sciences and humanities. ANT is both a theory and a practical method of conducting social research. It is a theory insofar as it challenges established notions in social theory. In Reassembling the Social, Latour (2005) proposes a type of sociology he refers to as 'sociology of association'. He suggests that there is no such thing as a 'social explanation' of a phenomenon and instead of looking for one, the sociologist needs to be tracing associations. Whereas sociologists would typically be concerned with the social context within which different non-social activities take place, Latour refutes the presence of a 'social force' that 'explain[s] the residual features other domains cannot account for' (p.4). He redefines the social as a movement rather than a structure and calls for researchers 'to follow the actors themselves' (Latour, 1987). Perhaps what is most controversial about actor-network theory is its inclusion of non-humans as actors (Sayes, 2014). Latour argues that '*any thing* that does modify a state of affairs by making a difference is an actor [...]' (Latour, 2005, p.71). As such, in ANT humans and non-humans are placed at par as actors with agency. In order to explain 'the social', Latour suggests that researchers 'feed off controversies'. Controversies simply put are situations of disagreement among actors. They begin with actors' realization that they cannot ignore one another and end when a compromise is reached (Venturini, 2010). Actor-network theory sees an opportunity in controversies and 'claims that it is possible to trace more sturdy relations and discover more revealing patterns by finding a way to register the links between unstable and shifting frames of reference rather than by trying to keep one frame stable' (Latour, 2005, p.24).

That being said, the question then is what are the practical implications of actor-network theory? To translate ANT into practice, Rogers, Sánches, and Kil (2015) suggest that digital methods can act as a bridge. Concepts such as associations and traces, which are central to ANT are translated into more tangible forms like mentions and links. Furthermore, the use of software tools makes possible the processing and visualization of digital data 'to deploy complex issue networks in order to tell stories with maps' (Rogers et al., 2015, p.29). As such, digital methods are seen as tools that enable the mapping and analysis of controversies.

Marres (2015) notes that while the analysis of controversies as a research method was used primarily within STS, its implementation in digital settings can be described as an interdisciplinary effort that involves a number of contributing disciplines in addition to STS, such as media studies, communication, computer science, and policy analysis. She further explains that while different disciplines may follow different approaches to digital controversy analysis, they all aim to 'render legible disputes about public issues' (Marres, 2015, p.658). Thus, central to controversy analysis is the ability to map the controversy, which today is facilitated by the proliferation of digital tools for analysis and visualization. Marres & Moats (2015) highlight how prior to the rise in popularity of social media platforms, social scientists from various disciplines had already begun engaging in digital controversy analysis, for example by analyzing networks of hyperlinks (Roger and Marres, 2000; Scharnhorst & Wouters, 2006), or conducting textual analysis to examine controversies as they unfold in blogs (Foot & Schneider, 2004). The prominence of social media platforms in recent years has unsurprisingly created a lot of excitement about its potential for social research in general and controversy analysis in specific. Important are the ways in which platforms like Twitter and Facebook direct and limit the users' possibilities of action and expression (Rieder, 2013). Acts like 'sharing', tweeting, and 'liking', present researchers with more structured data than what is otherwise available on the web (Marres & Moats, 2015). Taking advantage of these affordances, Marres & Moats (2015) go on to stress how scholars have engaged with controversy analysis using various social media platforms with Twitter likely being the most popular one being studied by academics, especially in relation to political controversies and debates.

2.1.4 Critical Reflection: Epistemic Problems

A number of issues relating to the nature of knowledge produced using digital methods are being increasingly addressed by scholars. Among the main supposed strengths of big data analysis is the notion of objectivity, the idea that the utilization of big data analytics eliminates human bias. Gitelman & Jackson (2013) note that while it is easy to think of data as 'before the fact', '[d]ata need to be imagined *as* data to exist and function as such, and the imagination of data entails an interpretive base' (p.3). Kitchin (2014b) also stresses that such empiricist thinking is flawed. He argues that data are not 'abstracted from the world in neutral and objective ways (p.175). Similarly, Rieder & Röhle (2012) warn against the notion that computational tools provide 'a higher epistemological status of the results' (p.73). They argue instead that machines create complications rather than solutions on an epistemological level: issues such as how algorithmic procedures are determined and decisions as to how to present the results raise important questions of bias.

Furthermore, the so called black-boxed nature of algorithms constitutes another challenge to the increasing use of social media data and computational methods in the social sciences. The notion of 'black boxes' is usually used as a metaphor to refer to the inaccessible nature of digital platforms' algorithms. Since algorithms determine what information is deemed relevant (Gillespie, 2014), it becomes crucial for researchers to be able to scrutinize the way they actually work. Yet in most cases access to the source code is blocked for proprietary reasons, preventing researchers from understanding the algorithm's underlying assumptions. This has caused a number of scholars in recent years to discuss and debate the epistemological problems algorithms create (Gillespie, 2014; Pasquale, 2015).

Apart from the significance of algorithms running digital media platform, Rieder & Röhle (2012) point our attention to the technological blackboxing of the analysis tools that process and analyze data. They show how graph layout algorithms used in network analysis tend to valorize certain features of the network thereby affecting the interpretations of the results. While many tools are open source with easy access to the source code, Rieder & Röhle note that even when the source codes are available, 'code literacy' remains an issue (p.76). In other words, the extent to which the social researcher possesses the necessary skills to assess and critique the workings of the algorithms adds an additional challenge to researchers engaging with computational methods of analysis. In 'The Datafied Society', Rieder & Röhle (2016) address this challenge again and situate it within the concept of digital *Bildung* (Berry, 2011), which emphasizes the importance of coding and computational techniques. They argue against the idea of training social science researchers to code and to write their own computational methods, since 'the practice of programming and software development requires far-reaching acculturation and many, many hours of practice' (Rieder & Röhle, 2016, p. 115). Their notion of digital *Bildung* is one that stresses the *concepts* based on which the software is built. For example, just as the user of SPSS must have a thorough understanding of statistical theories as opposed to an understanding of Java programming language, social researchers using network visualization tool Gephi need not, they argue, understand the codes with which graph layout algorithms are built. Instead, what is important is an understanding of the underlying concepts - in this case the study of graph theory and social network analysis concepts.

Another significant issue is that of data collection. The two main techniques used for the extraction of digital data are scraping and the use of the platforms' application programming interfaces (APIs). Web scraping is a technique for collecting digital data, whereby a number of steps are followed to extract formatted data from the web (Marres & Weltevrede, 2013). In recent years, digital platforms have been introducing APIs, in order to regulate the use of the platforms' data as a result of the increasing popularity of scrapers (Marres & Weltevrede, 2013). While APIs are much easier to use, especially for the less technical user, they pose notable challenges for digital social researchers. Bucher (2013) stresses that APIs are by no means neutral. For example, while Twitter offers two APIs to access and retrieve data, each has a number of limitations. The REST API allows the collection of a number of data points yet has a low rate limit (Puschmann & Gafney, 2014). The streaming API enables the collection of real-time tweets with a higher rate limit yet does not allow for the collection of any historical tweets.

Concurrently, the question of how representative the data set is, remains of capital importance. Whereas Puschmann & Gaffney (2014) note that with the streaming API, 'an acceptable degree of randomness' (p.58) is generally agreed on, boyd & Crawford (2012) stress the difficulty of making claims about the quality of the analysis if one does not know how the API actually works. As such, research that relies on APIs is seen by some as 'platform-dependent' (Marres & Weltevrede, 2013). That being said, in facing these problems Weltevrede (2016) calls for a response which foregrounds what she calls the 'research affordances' that both scrapers and APIs offer. She argues that data features like date-stamps and location, as well as attributes, such as freshness and connectedness enabled by APIs provide key indicators for 'device-driven' research and goes on to stress that the pre-formatting of data allows 'digital research to derive its analytic capacities in part from these effects' (p.43). Whereas many scholars emphasize problems of bias associated with pre-formatted data of this sort, Weltevrede suggests an attitude that embraces such issues and 'render[s] these problems researchable, to make their effects visible and reportable for practical purposes' (p.45).

A tricky question often debated as a result of digital bias issues is the extent to which one is actually studying the social phenomenon of interest, as opposed to the social media platform itself (Marres & Weltevrede, 2013; Roger, 2013, Marres & Moats, 2015; Weltevrede, 2016; Marres, 2017). In other words, is the researcher using the platform as a tool to analyze a controversy within and beyond the digital setting, or is the issue being used as an occasion to investigate the platform itself (Marres & Moats, 2015)? To elaborate on the practical implications of this problem, Marres (2017) gives an example of a research project in which

Moats (2015) studies a Twitter data set related to debates about the Hinkley Point - a nuclear power plant to be built in the UK. In the data set the researcher was faced with a hashtag (#FF, which stands for #FollowFriday) that seemed completely unrelated to the issue at hand. The question is then whether such a hashtag should be regarded as noise and thus excluded from the data, or whether it should instead be seen in a positive light. Marres (2017) goes on to lay out three different tactics in response to issues of bias: critical extraction, performative deployment, and radical empiricism. Critical extraction is a precautionary approach which takes a negative view of digital bias and seeks to eliminate it. For instance, when faced with bots, the instinct would be to immediately remove them. Critical extraction however serves also a practical purpose- that of data reduction. For example, in analyzing Twitter data, the exclusion of less used hashtags helps to render the data less cluttered and thus analyzable. In contrast, the performative deployment approach has 'a more positive appreciation of the role of digital devices in social life' (Marres, 2017, p.136). It pays special attention to the dynamics of the platform and how they act as indicators of issue activity (Marres & Moats, 2015). Finally, in radical empiricism the question of whether the social issue or the platform is the object of study becomes itself the empirical question. It adds to the performative approach the awareness that what is under study is variable, in the sense that while researchers may be concerned with studying the social issue via social media, they might end up studying the platform dynamics (Marres, 2017). As such, Marres (2017) calls for the researcher 'not [to] give up on the critical task of specifying and re-specifying, our objects of enquiry [...]' (p.138).

2.1.5 From 'Liveness' to 'Liveliness'

A closely related topic of inquiry here is the differences between social science methods and methods embedded in online media. For example, Marres & Gerlitz (2016) draw our attention to the fact that tools like MentionMapp[1] focus on capturing popularity in Twitter, while content analysis methods are more focused on detecting emerging issues (Callon et al., 1983). Social media platforms generally tend to valorize 'trending' and 'most recent' topics, whereas sociologists might be rather interested in historical trends. Emma Uprichard (2012) argues that 'the strength of focusing on the 'now' is simultaneously this genre's ultimate weakness' (p.128). Her argument is based on C Wright Mills' (1959) notion of 'sociological imagination'

[1] A tool for Twitter network visualization. For more details, see www. mentionmapp.com

which among other things stresses the importance of 'historically oriented work'. Uprichard criticizes the relatively narrow time span of digital social data, which she argues is unable to provide any true insights into historical trends. This real-time or 'live' nature that characterizes today's digital data environment has lead scholars to propose new ways of conducting research, so called 'live methods' (Lury, 2012; Back & Puwar, 2012). For Back & Puwar (2012) live methods include tools that both enable the investigation of 'real-time' data, but also put the data in its historical context (Lupton, 2014).

That said, Marres & Gerlitz (2016) argue that by emphasizing the differences between the methods, we risk ignoring 'the appropriability and instability of boundaries between digital and social research methods (p.26). With 'instability and indeterminacy' of digital research methods, the authors highlight the lack of clarity as to which purposes and research objectives digital tools serve, suggesting that such tools 'enable a variety of different agencies of research in different settings [...]' (p.26). Consequently, instead foregrounding the differences between digital and social methods, it is more constructive to focus on the programmability of digital methods and regard the dynamic and unstable nature of digital data and digital tools 'as an enabling condition for social enquiry' (Marres, 2017, p.107). To understand how the tools at hand can be adapted to serve social research purposes, Marres & Gerlitz propose the adoption of 'interface methods', which they define as methods that social and cultural researchers cannot claim as their own, 'but which resonate sufficiently with our interests and familiar approaches to offer a productive site of empirical engagement with wider research contexts, practices, and apparatuses' (Marres & Gerlitz, 2016, p.27). They call for an experimental approach with digitally-native methods with the aim of detecting 'liveliness' instead of 'liveness': Whereas liveness is concerned with popularity at a given moment, liveliness asks 'which entities are the most happening: which terms, sources, actors are the most *active*, which fluctuate interestingly over a certain period' (Marres & Weltervelde, 2013, p.327). As such, an actor-network approach, whereby shifts in associations define the entity in question is central to the detection of liveliness. Marres & Gerlitz (2016) demonstrate this by studying Twitter hashtags in relation to climate change. They carry out co-occurrence analysis by analyzing the hashtag profile, actor profile and user profile of certain hashtags. That is, which other hashtags co-occur with the hashtag in question, which URLs are connected to it, what kind of user engages with the hashtag, and how do these change and fluctuate over time? The analysis of the profile of a hashtag as such enables the detection of emerging issues in relation to a topic and allows for tracing liveliness.

2.1.6 Conceptual Framework and Research Questions

In the previous section I discussed debates on digital methods, their limitations and introduced actor-network theory. Further, discussions as to how digital methods can help with the practical implementation of ANT for controversy analysis were presented and finally the detection of 'liveliness' was introduced as a challenge in the use of digital methods. While I will not go into these topics in this section, I will briefly elaborate on co-occurrence analysis as it is the method I experiment with in this research.

In the previous example, Marres & Gerlitz (2016) demonstrate the similarities between co-occurrence as a widely used measure in digital tools and an established social science method: co-word analysis[2]. Co-occurrence can be measured by tools such as Twitter Streamgraph, which captures 'words [that] prominently occur together...and show[s] how these word relations change over time' (p.28). While co-word analysis also detects which words occur together, it gives different values to word associations depending on their proximity in the document. As such, it does not just measure the frequency of co-occurrence, but also the strength of relations (Marres & Gerlitz, 2016). The purpose behind co-word analysis in science and technology studies was to identify emerging topics that other forms of content analysis were unable to detect (Marres, 2017), in contrast to co-occurrence as a measure that highlights popularity.

The possibility of using co-occurrence analysis to detect liveliness instead of liveness - the emergence and change in relevance as opposed to popularity and hype - is therefore tested by Marres & Gerlitz (2016). In their study, Marres & Gerlitz (2016) carry out co-occurrence analysis of hashtags to see how relations change over time by comparing top hashtags (frequency) in the dataset with most connected hashtags (co-occurrence). Their analysis suggests that top hashtags (in terms of frequency of mentions) point towards the specific dynamics of Twitter[3] while relational measures 'can help foreground more substantive dynamics' (p.34) by providing more issue-specific hashtags. However, their study offers only 'initial empirical support' (p.34) for this claim thus leaving room for further investigation.

The overall aim of this dissertation is to contribute to the understanding of the dynamics and inner workings of Twitter--as one of the most popular social media platforms--in its potential

[2] Co-word analysis was developed by Michel Callon with the purpose of detecting and visualizing what he calls 'problematic networks' in science and technology documents. For more information, see Callon et al. (1983)

[3] Their analysis shows that top hashtags by frequency point attention toward 'Twitter specific' hashtags like #qanda (question and answer)

for social science research. More specifically, the study aims to test the relevance of 'co-occurrence' analysis methods to measure liveliness on Twitter and to contribute to the nascent but growing body of literature on this topic. The dissertation will attempt to do so by focusing on the GDPR Twitter dataset and will test for liveliness by examining the profiles of prominent hashtags in the dataset. That said, in addition to the actor, hashtag, and user profiles, it will also attempt to trace liveliness by exploring the source profile (the client software from which the tweet is sent), which Gerlitz & Rieder (2018) argue enables an understanding of 'the potentially heterogeneous contexts, practices, and cultures that feed into platform data' (p.530). Examining the source profile of a hashtag can reveal the extent to which it is driven by cross-platform syndication (tweets occuring as a result of actions on other platforms), various degrees of automation, or the Twitter interface. Sources they argue may 'offer traces of particular practices or user groups, adding to our interpretative arsenal' (p.543). As such, the source variable is a relevant way of detecting liveliness.

While this study does not propose an altogether *new* method to the digital methods debate, it contributes to the topic by experimenting with some of the recent developments in digital methods and by applying co-occurrence analysis on hashtags in multiple ways thereby allowing for more grounded and solid findings. Furthermore, the fact that social media research using such tools remains at its infancy, there is a need and a duty to continuously build on the available findings in order to widen our understanding of the medium given its constantly changing nature by looking out for topics and issues to put such tools to the test. I argue that analysing a Twitter dataset relating to the subject of 'GDPR' has the potential to offer insights into the use and usefulness of these tools and thereby presents a relevant occasion for such an analysis.

The overall research question can be framed as follows: what does the specific application of co-occurrence analysis on a Twitter dataset related to GDPR tell us about the ability of hashtags to divulge the liveliness of issues?

More concretely:

- To what extent does co-hashtag analysis on Twitter offer insights into issue dynamics?
- To what extent does the choice of measure (frequency vs. co-word) drive analysis towards medium vs. issue dynamics?
- How does the analysis of a hashtag's 'profile' in terms of its association to other entities (other hashtags, URLs, users, and sources) offer insights into the liveliness of the issue?

3 RESEARCH DESIGN AND METHODOLOGY

This chapter starts by discussing the research strategy and the rationale behind the choice of 'GDPR' as a relevant topic that can provide insights into the possibility of capturing liveliness of issues through Twitter. It then proceeds with a discussion of the specific processes followed and tools used for data capture and analysis and finally details the ethical considerations applied in the study.

3.1 Why GDPR?

In the previous chapter I elaborated on the 'climate change' study by Marres & Gerlitz (2016) which is considered to be a controversial topic with a large number of actors involved. Venturini (2010) takes the topic of climate change/global warming as a reference to offer some advice for those seeking to find a 'good controversy' for analysis. His recommendations were used as guidelines in the choice of GDPR as a topic. The subject of data privacy has lately been widely discussed especially in light of news concerning Cambridge Analytica's unauthorized use of the personal Facebook data of millions. The General Data Protection Regulation (GDPR) which aims 'to protect and empower all EU citizens data privacy' (eugdpr.org, n.d.), has been described as 'an ambitious, complicated, [and] contested law' (Selbst & Powels, 2017, p.233). Particularly interesting is the lack of clarity about its practical implications on different actors given its overall complex nature. This is a point stressed by Venturini (2010): 'when you look for controversies, search where collective life gets most complex' (p.262). He further warns against 'boundless controversies': '...the more a controversy is restricted to a specific subject, the easier will be its analysis' (p.264) While the overall topic in question is that of 'data privacy', I opted instead to focus on GDPR particularly to avoid 'boundless controversies'. A simple search for the keywords 'data privacy' reveals endless and to some extent unrelated issues, with actors from all over the world. While also a complex and controversial subject, GDPR is characterized by 'issue specificity' (Marres, 2015), rendering its mapping less impractical. Lastly, Venturini (2010) advocates controversies that are related to technical or scientific issues, where 'the border between science and politics, culture and technology, morals and economy [is blurred]' (p.265). While the spirit of the regulation is the protection of citizens' privacy, questions about its economic effects and how innovation may be impacted remain. As such, debates pertaining to morals vs economy are central to the issue.

3.2 Access and Capture of Dataset

Collecting Twitter data requires access to the platform's APIs. Twitter offers two APIs: the REST and the streaming API. As explained above, each of the two comes with certain limitations. That said, one of the main advantages of the streaming API is the supposed random sampling of up to 1% of all tweets it allows for (Twitter, 2018), which makes it the preferred API for most empirical research (Janetzko, 2017). Furthermore, the streaming API enables an almost real-time and continuous collection of data, provided a connection to the stream of data is maintained (Gaffney & Puschmann, 2014). This is important for the purposes of this study since the detection of liveliness requires an analysis of trends over time as opposed to a snapshot view. The streaming API was therefore chosen for this study. I used the Twitter Capture and Analysis Toolset (DMI-TCAT) (Borra & Rieder, 2014) which was developed by the Digital Methods Initiative[4] to connect to the API, as this tool was also used subsequently for the processing and analysis of the retrieved data. Determining the time frame of the study was the next challenge. Venturini (2010) emphasizes the importance of capturing the controversy when it is 'hot'. While discussions about GDPR have been circulating for months, the regulation went into force on May 25th, 2018. It would perhaps make sense to assume that the debates would be at their most intense in the period close to this date. However, one of the limitations of the API is its inability to access historical tweets. To access tweets older than a week, one has no option but to buy the data from Twitter (Twitter, 2018). That being said, others argue that 'a time lag can be beneficial to capture engagement by a broader public instead of limiting the investigation to the initial reaction' (Yang, Quan-Haase, & Rannenberg, 2017, p.1987). As such, from the period of 02.07.2018 until 02.08.2018 a connection to the API was created and tweets including the keyword GDPR or the hashtag #GDPR were collected. This resulted in a total of 179,201 tweets with 82,106 unique users.

3.3 Data Processing & Analysis

The analysis of the captured data primarily relied on DMI-TCAT but also in combination with Gephi[5]. DMI-TCAT is an open-source platform for the capture and analysis of Twitter data. It was chosen for its unique ability to address the research questions posed in this study. The tool offers three main types of analyses: Tweet statistics, network analysis, and content analysis (Borra & Rieder, 2014). Since the aim of the study is to account for liveliness of GDPR by analyzing the behavior of hashtags associated with it, network mapping helps to make visible the association between hashtags and other entities, including other hashtags and

[4] For more information about the Digital Methods Initiative see www.wiki.digitalmethods.net

[5] Gephi is a free software for graph visualization and network analysis. For more details, see www.gephi.org

users. To do so, TCAT enables the creation of co-hashtag networks, as well as a number of bipartite networks (two-mode networks), such as a hashtag-user network. The tool however only provides the network data which then needs to be visualized in other software programs. In this case Gephi was used.

The analysis can be divided into the following two main phases:

3.3.1 Composition of issue and fluctuations over time

To get an overview of the composition of GDPR on Twitter, a co-hashtag network - i.e. a network where two hashtags are linked if they appear in the same tweet - is created. Gephi's modularity algorithm is used to detect clusters of closely connected nodes and so identify sub-issues composing the topic. The OpenOrd layout algorithm is used to spatialize the network, as it helps make visible clusters in large networks (Martin et al., 2011). Further, to evaluate how issue composition fluctuates over time, the 32 day dataset is divided into four intervals of eight days each. For each interval I created a co-hashtag graph. Here, the Force Atlas 2 layout is used as it is suitable for relatively smaller networks (Jacomy, Venturini, & Bastian, 2014). The hashtag #GDPR is eliminated from the graph as it does not add any insight to the analysis. The co-hashtag graph helps give an idea about the extent to which the vocabulary associated with GDPR (in terms of hashtags) changes over time. While dividing the dataset into four intervals clearly reduces the size of the data, each interval remained too large for analysis, ranging from 6,500 to 7,500 nodes per interval. To examine co-hashtag fluctuations, it would be impossible to decide which ones to focus on within this dataset. As such, following a 'precautionary mode' (Marres & Moats, 2015) the data is further 'cleaned' for practical purposes. To do so, I calculated the weighted degree of each nodes, which can easily be done in Gephi. Degree refers to the number of links (or edges) of any point (or node) in the network (Scott, 2017). The weighted degree takes into account not only the number of links of a given a node (in this case the number of other hashtags a certain hashtag is linked to), but also the number of times these two hashtags co-occur, and as such is seen as a suitable measure to explore which hashtags are most connected in the dataset (Marres & Gerlitz, 2016). I then included in the network only hashtags that have a minimum weighted degree of 50. This reduced the size of each of the four interval graphs to anywhere between 300 and 400 nodes.

3.3.2 Hashtag profiles and shifts in associations

While assessing the fluctuations in issue composition enables the detection of liveliness of a topic (Marres & Weltevrede, 2013; Weltevrede, 2016), liveliness can also be assessed by studying how certain hashtags relate to other entities and how these relations shift from one interval to another (Marres & Gerlitz, 2016; Marres, 2017). Therefore, two of the 'top' hashtags in the dataset (determined by weighted degree) are chosen and the profile of each is explored. The following analyses are performed:

- **Hashtag Profile:** Here I look into which other hashtags co-occur with each of the two hashtags across the four intervals. This is visualized using an experimental module offered by the TCAT by the name of the 'Associational Profile' which explores shifts in hashtag associations over time.

- **Actor-Profile:** An understanding of the links that are associated with a specific hashtag helps evaluate how diverse the hashtag is. This section looks into and categorizes the top 10 domains associated with the two hashtags and explores how they change over time. This is made possible by the TCAT's ability to produce a spreadsheet that includes the number of times a host co-occurs with a given hashtag.

- **User & Source Profiles:** Finally, associations between the hashtags and the users are considered. As mentioned above, TCAT allows for the creation of a number of bi-partite networks, one of which is a graph that links hashtags to the users employing them. This is used to make visible the relationship between users and each of the two hashtags. The top 10 most active users, in terms of how often they tweet using the hashtags are further studied in terms of the client used to send the tweets thereby allowing for an understanding of the level of automation in the production of the tweets (Mowbray, 2014; Gerlitz & Rieder, 2018).

3.4 Ethical Considerations

The ethical implications of conducting research using Twitter data differ from those of other traditional methods such as surveys and interviews, seeing as 'tweets are inherently public and readable, when posted to a public account […]' (Thelwall, 2014, p.85). That said, scholars do warn against the sharing of personal and sensitive information posted on Twitter (Zimmer & Proferes, 2014). This study follows the guidelines presented by Townsend & Wallace (2016) in which the authors address four areas of concern: 'private vs. public', 'informed consent', 'Anonymity' and 'risk of harm'. The first point relates to the second in the sense that it helps assess 'the extent to which [one] is ethically bound to seek informed consent from social media users' (Townsend & Wallace, 2016, p.5). In this research, all tweets are publicly available and therefore informed consent as per Twitter's terms of services (Twitter, 2018) is given.

Furthermore, Townsend and Wallace suggest that risk of harm is not present in cases where the data is public and when ‘the social media user is aiming for broad readership’ (p.8), as is the case with Twitter. As such, it is determined that no risk of harm is present in this research, particularly considering the public nature of the subject in question. Ethical approval was received from the academic supervisor and permission to conduct the study was granted.

4 MAIN FINDINGS AND DISCUSSION

4.1 Main Findings

4.1.1 Composition of Issue and Fluctuations Over Time

In order to be able to illustrate the sub-issues associated with GDPR, Gephi’s modularity class algorithm is used to detect clusters of closely connected hashtags in the co-hashtag graph as mentioned above. The co-hashtag graph reveals a number of prominent (strongly connected) hashtags. The main cluster (Figure 2) includes the following main hashtags: #privacy, #cybersecurity, #security, and #dataprotection. Other strongly connected hashtags (in terms of weighted degree) such as #blockchain, #data, and #ai are also notable. In addition, a number of industry related terms including #digitalmarketing, #marketing, and #business appear throughout. Examining the top hashtags in more detail reveals a dominating presence of business and technology news sources, as well as strong presence of organizations trying to market services related to GDPR compliance.

While the modularity algorithm detected more than 60 communities, the following clusters represent the main sub-issues identified.

Cluster 2: online marketing related hashtags

Cluster 3: security related hashtags

Cluster 4: education related hashtags

Cluster 5: social media/personal data related hashtags

Figure 1: co-hashtag network of entire dataset. OpenOrd layout is used. Colors represent clusters identified by the modularity algorithm

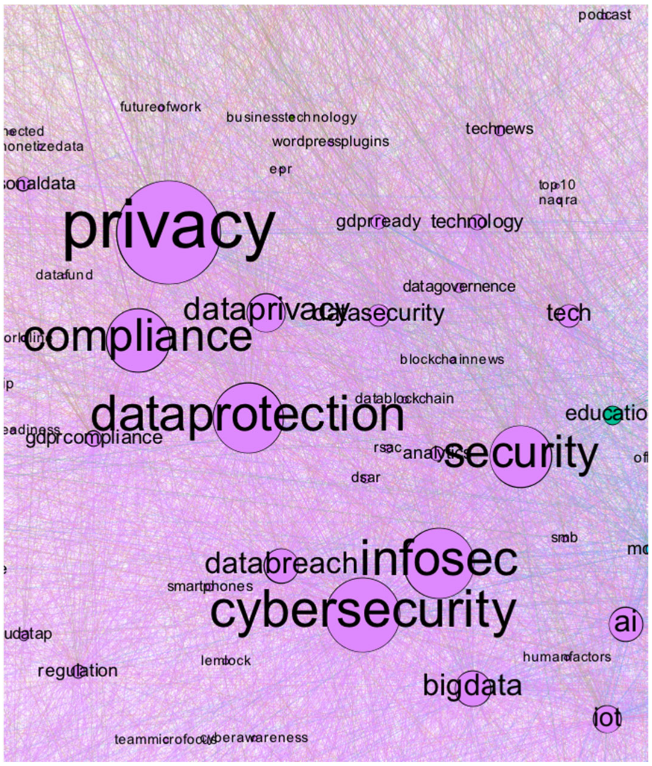

Figure 2: Main cluster of densely connected hashtags.

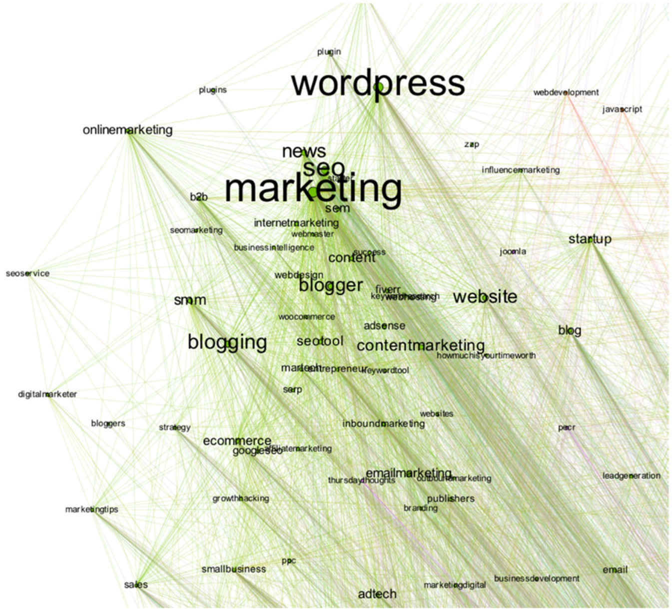

Figure 3: cluster 2: 'online marketing' hashtags

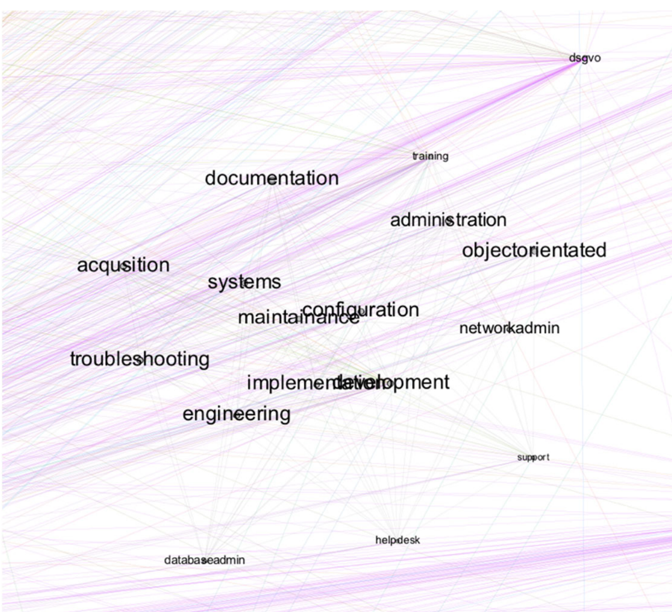

Figure 4: cluster 3: 'security' hashtags

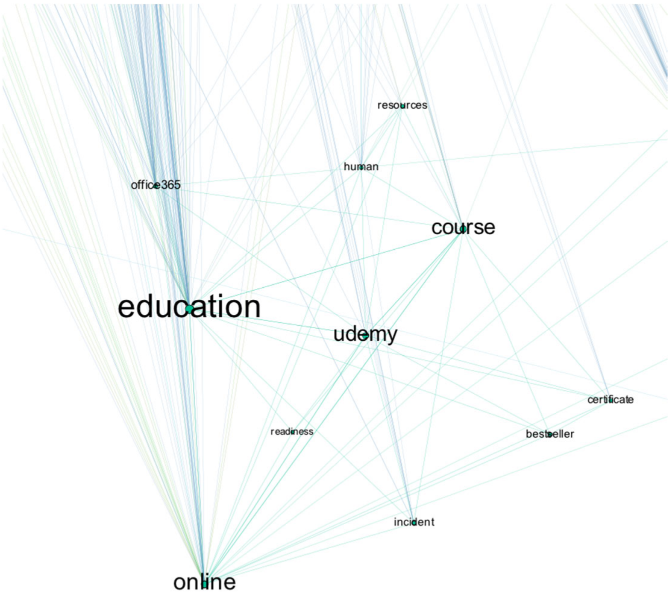

Figure 5: cluster 4: 'education' hashtags

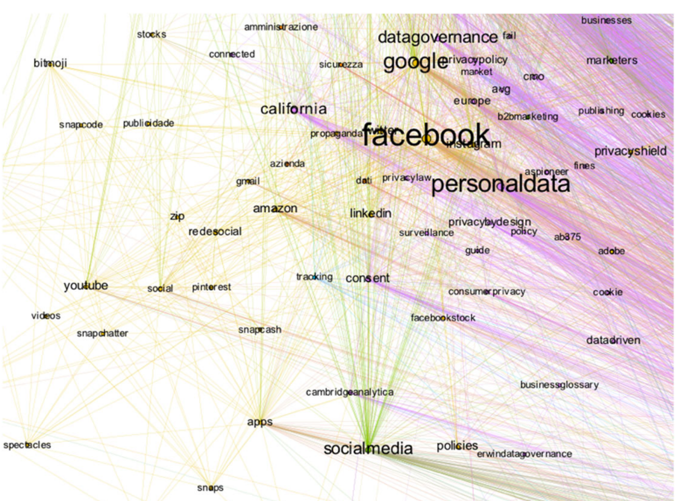

Figure 6: cluster 5: 'social media' hashtags

While this analysis provides a general impression of the composition of GDPR on Twitter, in order to establish how lively the issue is on this platform, a closer look at the fluctuations over time is necessary. What is clear is that throughout the four intervals the hashtags comprising the main cluster appear across the four intervals. The purpose is then to identify the more dynamic hashtags: that is, the ones that appear and disappear across the four intervals in relation to the set of stable hashtags. To do so, a co-hashtag network is created for each interval, in order to evaluate the evolution of issue composition.

In the first interval a cluster of education related hashtags is clearly visible. This includes #learning, #schools, #highered, and #classrooms. The content associated with the hashtags seems generally concerned with the impact of GDPR on educational institutional, with users discussing and sharing different sources about the potential impacts. Also visible are organizations offering their services to educational institutions looking to be GDPR compliant. These hashtags seemed to appear and disappear throughout the time frame of analysis. Specifically, they were more visible in the first and fourth interval of analysis as illustrated in the below graph.

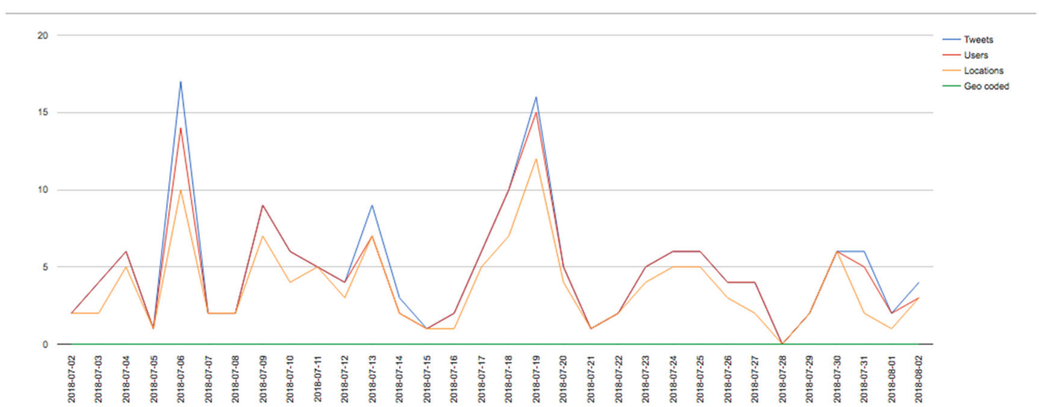

Figure 7: Fluctuations of #learning, #schools, #highered, and #classrooms.

A notable fluctuation is witnessed in regards to hashtags related to ethics, specifically the hashtag: #aiethics. The hashtag is used in relation to discussions about the ethical implications of GDPR in general, as well as specifically on artificial intelligence (AI). This hashtag appears strongly in the second interval (particularly on July 10th) only to disappear after that. The below graph demonstrates its fluctuation throughout the 32 days.

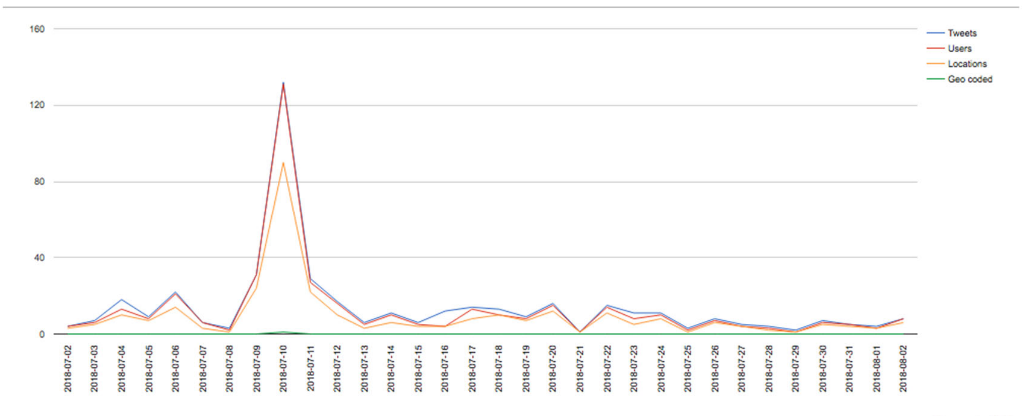

Figure 8: Fluctuations of #aiethics.

A striking cluster of five strongly connected hashtags emerges in the fourth interval. The hashtags in question are #economy, #gdp, #jobs, #donaldtump, and #obama. Although the emergence of hashtags such as #economy and #gdp (gross domestic product) in relation to GDPR initially suggests a discussion around the macro-economic impacts of the regulation, the presence of hashtags #donaldtrump and #obama in the same cluster seem rather curious, seeing as GDPR is a European regulation. Indeed, a closer look at the tweets associated with this cluster highlights that it is entirely composed of retweets of single a tweet by a Trump supporter. The hashtag #GDPR thus seems to merely serve the goal of attracting a larger audience.

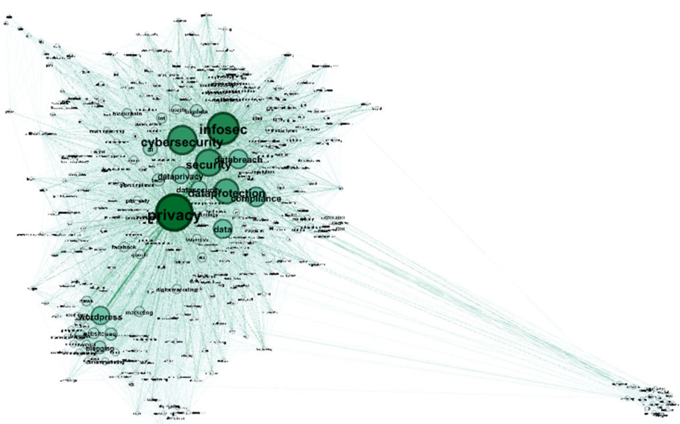

Figure 9: Co-hashtag network of Interval 1. Node size: weighted degree. Node color (from light green to dark green): weighted degree

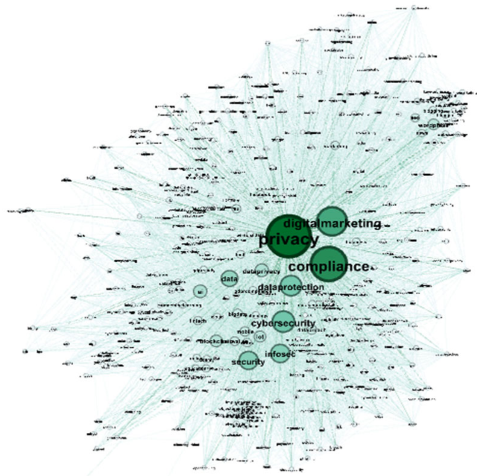

Figure 10: Co-hashtag network of Interval 2. Node size: weighted degree. Node color (from light green to dark green): weighted degree

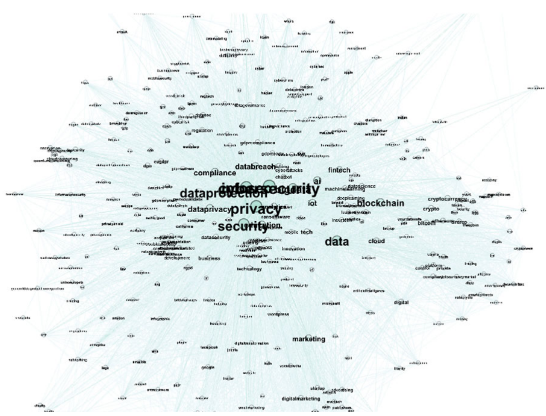

Figure 11: Co-hashtag network of Interval 3. Node size: weighted degree. Node color (from light green to dark green): weighted degree

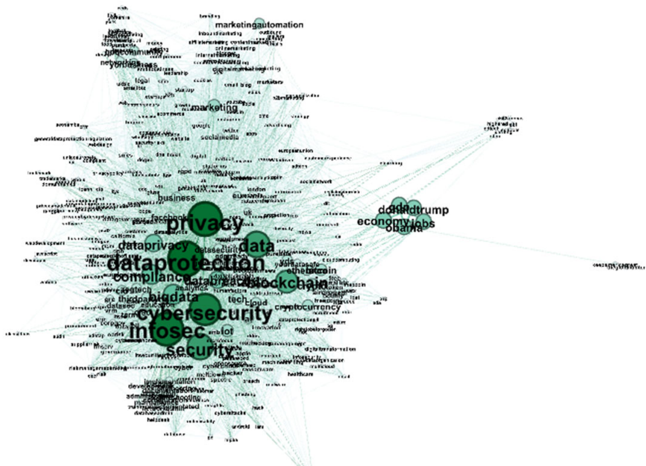

Figure 12: Co-hashtag network of Interval 4. Node size: weighted degree. Node color (from light green to dark green): weighted degree

To address the second research question, a comparison is made between the top hashtags in terms of their co-occurrence with other hashtags and top hashtags in terms of the frequency of their mention. The results reveal very few differences in the top hashtags identified as shown in the below table across the four intervals.

Interval 1	
Top hashtags by frequency of mention	**Top hashtags by frequency of co-hashtag connections**
Privacy	Privacy
dataprotection	infosec
cybersecurity	cybersecurity
Data	dataprotection
infosec	Security
Compliance	Compliance
Security	Data
dataprivacy	databreach
WordPress	dataprivacy
AI	WordPress

Table 1: Frequency of hashtag mention vs frequency of co-hashtag connections in interval 1

Interval 2	
Top hashtags by frequency of mention	**Top hashtags by frequency of co-hashtag connections**
privacy	privacy
compliance	compliance
digitalmarketing	digitalmarketing
dataprotection	cybersecurity
cybersecurity	dataprotection
Data	infosec
security	security
infosec	Data
dataprivacy	dataprivacy
AI	AI

Table 2: Frequency of hashtag mention vs frequency of co-hashtag connections in interval 2

Interval 3	
Top hashtags by frequency of mention	**Top hashtags by frequency of co-hashtag connections**
Privacy	CyberSecurity
dataprotection	Privacy
CyberSecurity	infosec
data	security
infosec	dataprotection
security	data
dataprivacy	BigData
compliance	Blockchain
BigData	ai
Blockchain	dataprivacy

Table 3: Frequency of hashtag mention vs frequency of co-hashtag connections in interval 3

Interval 4	
Top hashtags by frequency of mention	**Top hashtags by frequency of co-hashtag connections**
Privacy	DataProtection
DataProtection	Privacy
Data	infosec
CyberSecurity	CyberSecurity
infosec	Data
Security	Security
blockchain	blockchain
Compliance	bigdata
dataprivacy	Compliance
Marketing	dataprivacy

Table 3: Frequency of hashtag mention vs frequency of co-hashtag connections in interval 3

4.1.2 Profiles of key hashtags

This section looks into the profiles of two specific hashtags-#privacy and #ai-in more detail, examining their association to other entities and exploring how relations change across the four intervals. Determining which hashtags to further study is a challenge seeing as there is a huge number of hashtags co-occurring with other hashtags and fluctuating over time. Following in the footsteps of previous studies (see Marres & Weltevrede, 2013; Marres & Gerlitz, 2016), I focus on two of the key hashtags in the dataset in terms of their frequency of co-occurrence - calculated by weighted degree. Still, this leaves a large number of hashtags to choose from. The hashtag #privacy was chosen for further examination, since unlike other prominent hashtags (security, infosec, cybersecurity), it seems less specific to businesses and therefore has the potential to reveal a more diverse profile. In choosing the second hashtag I opted for #ai, not only for its significance in the dataset, but to have the research study both abstract notions such as privacy and more issue-specific concepts[6].

[6] To determine which hashtags to further investigate, a subjective decision needs to be made. Similarly, Marres & Gerlitz (2016) determined which hashtags to further analyze based on what were deemed more 'polarizing' hashtags.

I start by examining the hashtag profile of the two hashtags in question, that is exploring what other hashtags these two hashtags co-occur with, and how these associations shift from one interval to the next.

4.2 Hashtag Profile of #ai:

In the first interval #ai is mostly connected to the rather static hashtags in the dataset, such as #privacy and #cybersecurity. Qualitatively examining the tweets reveals that the space is mainly dominated by consultancies offering cyber-security and risk management services to organizations looking to comply with the GDPR. In the second interval #machinelearning becomes the top hashtag associated with #ai, surpassing #privacy and #cybersecurity. The tweets associated with #machinelearning are mainly concerned with the impact of privacy regulations on innovation. Interestingly, new hashtags also emerge as densely associated with #ai. These are #abdsc, #dataethics, #xai and #aiethics. On closer inspection, #abdsc turns out to refer to an online data science community and #xai stands for 'Explainable AI'. The rise of these hashtags in the second interval is primarily the result of significant retweeting activity of a tweet by a prominent data scientist, addressing the need for the 'demystification' of artificial intelligence. In the third interval the top hashtags are once again #privacy and #cybersecurity - with consultancies dominating the conversation. That said, two new hashtags that were previously absent from the dataset emerge in this interval: #chatbot and #phishing. Their prominence is the result of significant retweets of a single tweet sharing a video calling for victims of email scammers to forward the scam email to a 'smart' chatbot email which in turn would start a conversation with the scammers to waste their time. A strong overall drop in associations with hashtag #ai is noticed in the fourth interval, where #cybersecurity becomes the most connected hashtag once again, albeit to a much lesser degree.

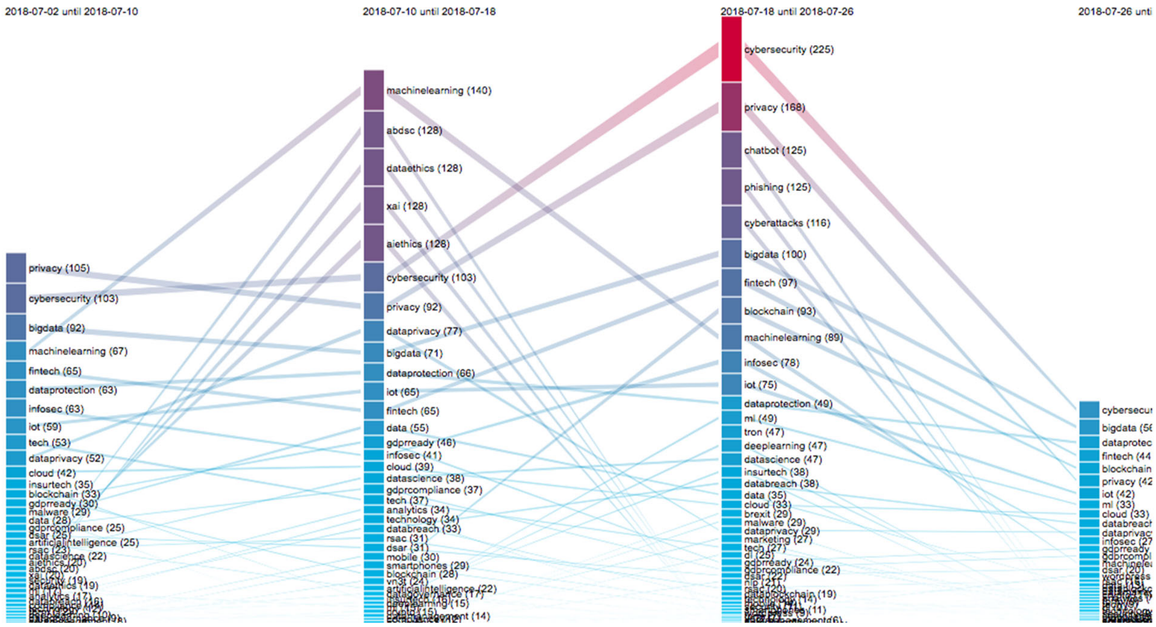

Figure 13: Hashtag profile of hashtag #ai. Produced by DMI-TCAT's 'Associational Profile' module, showing hashtag associations per interval.

4.3 Hashtag Profile of #Privacy

In regards to hashtag #privacy, strong retweeting activity of a single tweet leads to the notable rise of hashtags #compliance and #digitalmarketing in the second interval and seems to be responsible for the peaking of #privacy across the timeframe of analysis. The tweet in question is posted by a marketing consultant discussing the future of digital marketing in view of the introduction of the GDPR.

Furthermore, the profile of #privacy demonstrates a clear prominence of security related hashtags, including #infosec, #security, #cybersecurity, and #dataprotection. These four hashtags appear relatively stable throughout the four intervals. Unsurprisingly, when co-occurring with #privacy, these hashtags are used in a similar context to that of #ai, in that organizations offering advice and promoting their products are dominating the space. Interestingly, the hashtag #dataprotection also appears in relation to privacy activism with a significant presence of tweets raising awareness about data privacy and calling on people to support a campaign for the protection of privacy.

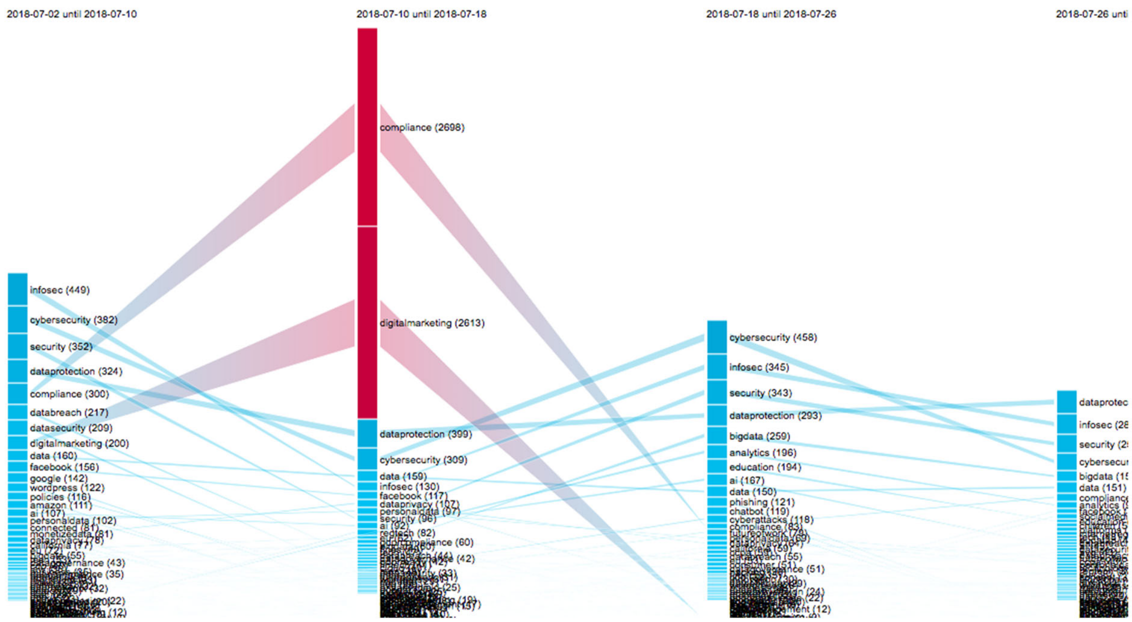

Figure 14: Hashtag profile of hashtag #privacy. Produced by DMI-TCAT’s ‘Associational Profile’ module, showing hashtag associations per interval.

4.4 Actor Profiles

In this section I analyze the links (URLs) that the two hashtags are associated with in the four intervals. To be able to make sense of the large number of URLs, I manually categorize them based on the type of domain they relate to[7]. The following categories are identified:

Blogs: links to personal or community blogs

Companies: links to corporate websites

Media: links to specialized media outlets, publishing technology or GDPR specific news

[7] Manual categorization of URLs was followed by the ‘Climate Change on Twitter: Issue Lifelines’ study. For details on the project, see: http://blogs.cim.warwick.ac.uk/issuemapping/cases/issue-lifelines/

News: links to general news websites

Social media: links to social media websites

Organizations: links to other not for profit organizations

4.5 Actor Profile of #ai

The most prominent URL associated with #ai across the four intervals is that of a data science blogging community appearing 153 times in tweets associated with #ai. Apart from this specific URL, it is clear that the links most associated with the hashtag are either corporate websites or specialized media outlets. The below word cloud offers an overview on the the most prominent domains associated with #ai.

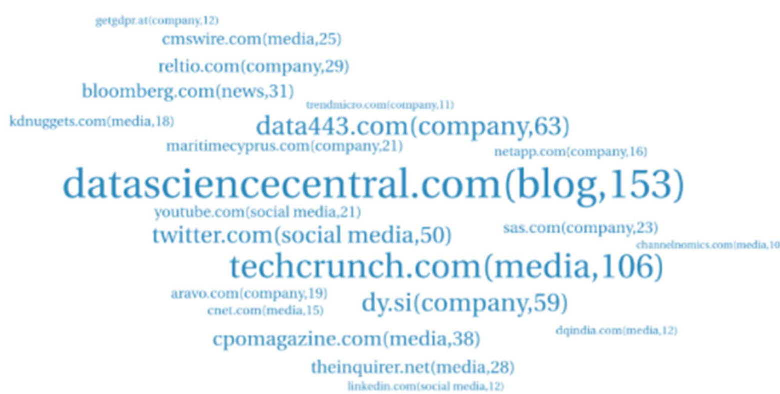

Figure 15: Word cloud of 20 most prominent domains associated with #ai across the four intervals. Classification and frequency of co-occurrence with the hashtag included in brackets. Word cloud created using wordart.com.

The figure below demonstrates the shift of different domain categories from one interval to another based on the 10 most prominent domains that appear in each interval[8]. In the first interval, links to specialized media outlets are the most frequent. These include techcrunch.com, an online publisher specialized in technology news, followed by

[8] Details of the 10 most linked domains per interval are available in Appendix C

cpomagazine.com, a website dedicated to data protection and security related news. The high frequency of media URLs in the first interval is primarily driven by retweets of a techcrunch article. Links to corporate websites also figure strongly. While the second interval is also characterized by dominance of corporate and specialized media outlet links, the most frequently linked URL is that of the data science blogger community mentioned earlier. Its prominence is the result of strong retweeting activity during this interval. The third and fourth interval are also dominated by corporate websites followed by media outlets. Overall it seems that links to corporate websites are the most stable throughout the four time periods.

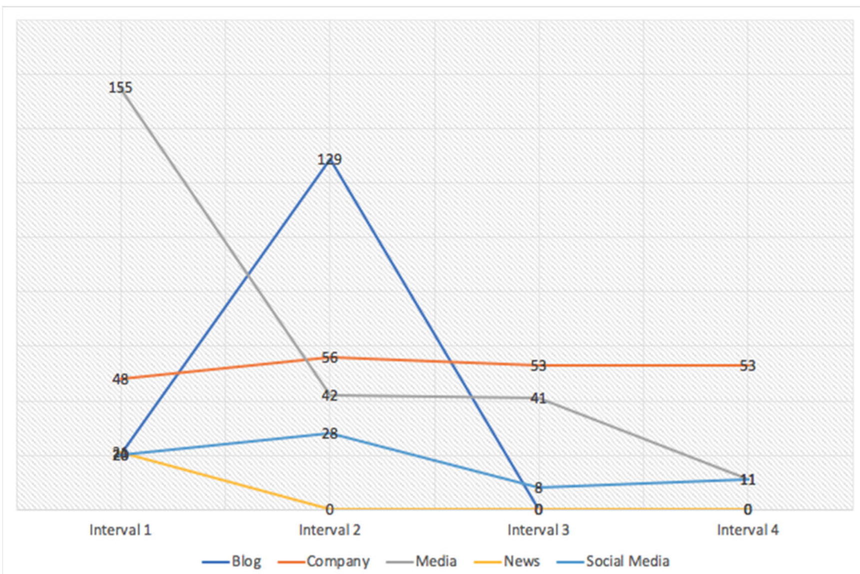

Figure 16: Ten most prominent domains associated with #ai, classified by type

4.6 Actor Profile of #privacy

While similar to #ai in the strong visibility of companies and specialized media links, the hashtag #privacy seems more diverse: much stronger presence of general news websites and social media is seen. In addition, more blogs feature within the top 10 domains compared to #ai.

Figure 17: Word cloud of 20 most prominent domains associated with #privacy across the four intervals. Classification and frequency of co-occurrence with the hashtag included in brackets. Word cloud created using wordart.com.

Figure 18 illustrates the most prominent domain categories in each interval. The second and fourth intervals are characterized by very high frequencies of corporate and media domains respectively. The burst of company URLs in the second interval is clearly the result of retweets. Specifically, retweets of a single tweet posted by a marketing consultant, which links to his own website. Similarly, the rise of the media category in the fourth interval is mainly the result of heavy retweeting of a tweet containing a link to cpomagazine.com. The magnitude of these two bursts makes it rather challenging to assess how the hashtag relates to other categories of domains and how these change over time. Thus, figure 19 excludes the two bursts from the graph. The new graph brings into view a more diverse hashtag in terms of its association to different categories. Links to news websites are especially prominent in the first interval, slowly decreasing in the second and third intervals then picking up again in the fourth. Social media figure throughout the time frame of analysis, peaking in the second interval, while links to blogs appear both in the second and third intervals.

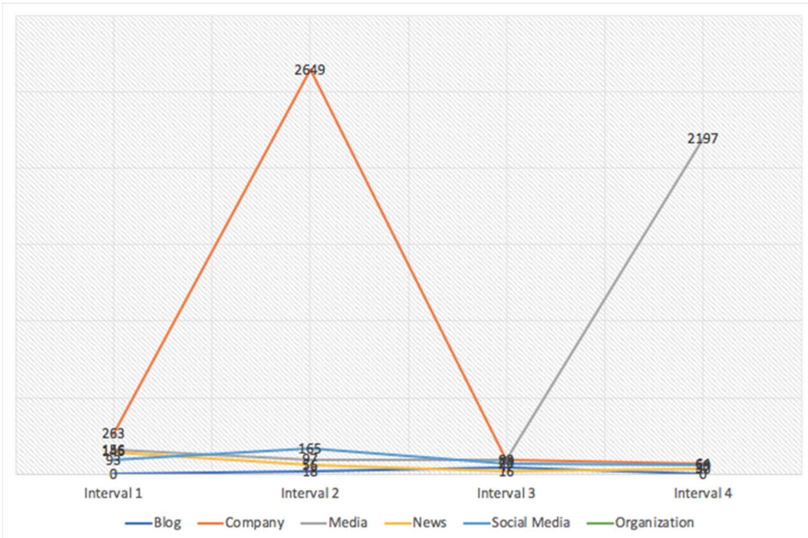

Figure 18: Ten most prominent domains associated with #privacy, classified by type

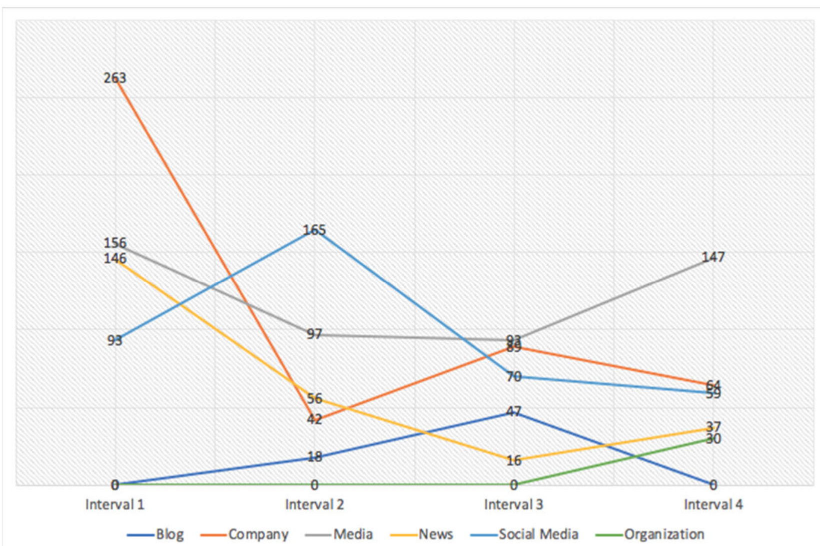

Figure 19: Ten most prominent domains classified by type, excluding retweet bursts in the first and fourth intervals.

4.7 User & Source Profiles

The following section examines the user profiles of the two hashtags. In this sense, it looks into the production side of the hashtags to explore the type of users engaging with the hashtags. One interesting question this analysis attempts to answer is the extent to which the production of the hashtags in question is automated. Mowbray (2014) notes that there are a number of measures enabling the detection of automated tweets, including an assessment of the type of client through which the tweet is sent: humans are more likely to use web and mobile clients. As such, the most noteworthy users in terms of the frequency of their deployment of the hashtag are identified together with the source of their tweets.

4.8 User Profile of #ai

To visualize the association between the hashtag and the users, a bi-partite graph is created, whereby hashtags and users are connected by an edge if the user employs the hashtag. The more the hashtag is employed by the user, the thicker the edge. Figure 20 depicts the hashtag-user network created from all tweets that include the hashtag #ai. In order to get a better view of the 'top users' in terms of the frequency with which they deploy hashtag #ai, figure 21 illustrates the same network after excluding all other hashtags. This allows for better visibility of the edges, the thickness of which indicates the most active users.

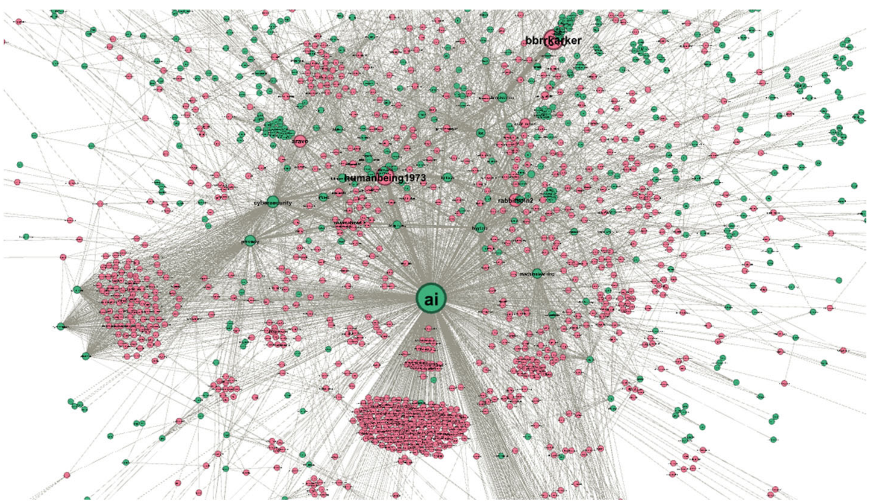

Figure 20: Hashtag-user network within all tweets that include the hashtag #ai. Hashtag #gdpr is deleted

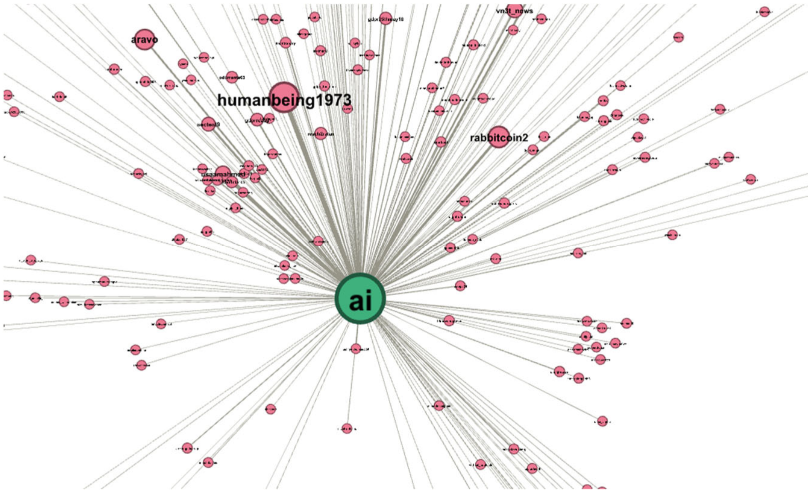

Figure 21: Hashtag-user network within all tweets that include the hashtag #ai. All other hashtags except #ai are filtered out.

The following users are identified as the ones with the highest edge weight:

User	Source of Tweets
Humanbeing1973	Twitter Web Client
RabbitCoin2	Twitter Web Client
Energiatutka	Twitter for Android
1davidclarke	Smartqueue
Aravo	Hubspot
usaamahmed	IFTTT
gdpr25thmay	GDPR Funnel (bot)
Vn3t_news	Twitter Web Client
gdprready	GDPR Bot 250218
itknowingness	ttools it knowingness

Table 5: Top users employing hashtag #ai

Among the top users engaging with the hashtag a strong presence of automated accounts is noted. These include bots such as gdpr25thmay and gdprready, which retweet GDPR related tweets, and an organization utilizing a marketing automation software (Hubspot) to post tweets promoting its products. Surprisingly, some of the nodes with quite thick edges which one assumes ought to be bots (thick edges indicate more use of the hashtag) turn out to be sent from the Twitter web client. Also interestingly, while one would assume that automated tweeting would be associated with non-personal accounts such as the above mentioned bots and with organizations, a number of seemingly 'human' profiles are in fact automated through clients such as IFTTT (If This Then That), which is mainly focused on cross-platform syndication, whereby a tweet is sent whenever an action is taken on another platform, while another user described in his profile as a data protection expert, sends his tweets using a scheduling tool by the name of 'Smarterqueue'.

4.9 User Profile of #privacy

The hashtag #privacy is dominated by companies operating in the fields of security and data governance, including consultancies such as 'Datastreams_io', 'Sectest9', and 'Lemlock_app'. Other notable users include 'Privacystrategy', an account advocating for privacy and raising awareness about privacy related matters primarily by retweeting news from various sources. The level of automation of the top users also stretches from fully automated bots (gdpr25thmay), to companies utilizing social media management tools like Hoosuite and users engaging directly through Twitter such as a data ethics and privacy speaker tweeting through the Twitter for Android application.

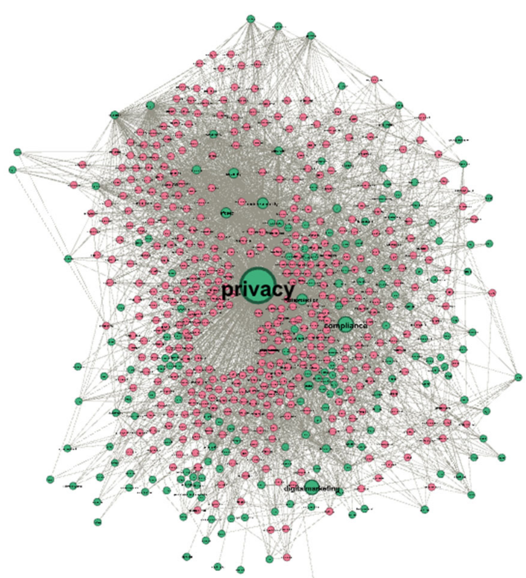

Figure 22: Hashtag-user network within all tweets that include the hashtag #privacy. Hashtag #gdpr is deleted

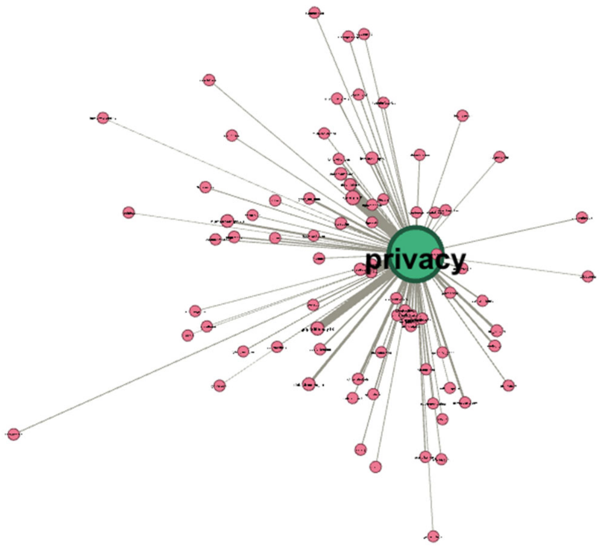

Figure 23: Hashtag-user network within all tweets that include the hashtag #privacy. All other hashtags except #privacy are filtered out.

User	Source of Tweets
gdpr25thmay	GDPR Funnel (bot)
Datastreams_io	Hootsuite
Sectest9	"Auto is the only way it can be"
newsgdpr	Twitter Web Client
usaamahmed	IFTTT
privacystrategy	Roundteam
Eti_entein	Twitter Web Client
privasense	Twitter for Android
Humanbeing1973	Twitter Web Client
Lemlock_app	Hootsuite

Table 6: Top users employing hashtag #privacy

4.10 Discussion

In exploring the thematic changes across the four intervals in relation to the overall topic, a preliminary idea of the liveliness of GDPR on Twitter is made possible. First, the co-hashtag networks reveal a set of relatively stable hashtags that are present in all intervals, while a number of other hashtags appear only in certain intervals. In contrast to Marres and Weltevrede (2013) whose research on the hashtags associated with the keywords 'crisis' and 'austerity' on Twitter demonstrates that 'newsy' economic and political hashtags such as 'euro'', 'debt' and 'syria' are the main stable ones, in this study stable hashtags across the four intervals appear to be substantive in nature, including hashtags like #privacy, #security and #dataprotection. This however is not driven by social conversations but is primarily a result of corporates continuously using such hashtags to market their services to other businesses. They specifically seem to be utilizing a large number of hashtags in order to draw more attention to their tweets - a reflection of the medium itself, since Twitter like other social media platforms valorizes popularity, and hashtags are therefore often utilized for marketing campaigns (Bruns & Stieglitz, 2012). In the aforementioned study a clear differentiation between static and dynamic hashtags is made, whereby the stable set of hashtags are more 'newsy' and the dynamic ones appeared more social (e.g. #screwyouassad, #solidaritywithgreece..etc). This research however does not indicate any substantial distinction between static and dynamic hashtags, though it is perhaps fair to say that the set of stable hashtags are broader in nature (privacy, security, compliance, etc..), while the more dynamic ones are somewhat more specific. Some variability around the subject of GDPR however seems to be present, for example with the hashtag #aiethics and education hashtags (#learning, #schools, #highered) appearing strongly only in certain intervals.

The study of Marres and Weltervrede (2013) describes the fluctuating hashtags as evidence of the liveliness of the issues they are studying. Looking more closely at #aiethics however, its sudden prominence is mainly the result of very strong retweets of one single tweet linking to a blog post on AI with no relation to ethics implications[9]. A close look at the tweets proves necessary: the presence of a hashtag is not necessarily reflective of the context of the tweet. Similarly, examining the cluster composed of #gpp #economy #donaldrump and #obama reinforces the importance of a qualitative assessment of the tweet to exclude irrelevant tweets, showing that solely relying on network mapping can easily lead to misinterpretation (Niederer, 2016).

The above hashtag analysis is carried out based on top hashtags in terms of their connectedeness to other hashtags. The question to be asked is whether or not the results would be different had the top hashtags been determined by the frequency of their mention. The comparison demonstrates that across the four intervals the top hashtags determined by both

[9] Link to the blog post: http://nirvacana.com/thoughts/2017/12/27/demystifying-artificial-intelligence/

measures are almost identical, notwithstanding one or two differences. This is in contrast to Marres and Gerlitz (2016) whose study of climate change on Twitter reveals strong differences between frequency of hashtags and co-hashtag occurrences. The results of their research indicate that top hashtags by frequency of mention are less stable, in the sense that they appear and disappear quickly. They are also more specific to Twitter, such as the case with the hashtag #qanda (question and answer). This study exhibits no such distinction, with both measures producing more or less the same top hashtags. The primary reason behind this seems to be the overwhelming presence of business accounts within the dataset whose purpose is to promote their data management and compliance services in light of the introduction of the GDPR and as such, tend to utilize a large number of hashtags in the same tweet to gain more visibility. While the deployment of co-hashtag analysis as a form co-word analysis is proposed by Marres (2012) as an alternative measure 'to dominant popularity metrics' (157), this research suggests that the utilization of co-hashtag analysis within tweets has its limitations. The original purpose behind co-word analysis in science and technology studies is 'to identify problematic networks and study their evolution on the basis of the analysis of documents' (Callon et al., 1983, p. 196), which is achieved by looking for associations among keywords and dispensing with words randomly associated with others, as frequently as they may appear (Marres, 2012). Although a tweet may 'provide a workable data unit within which to detect co-occurrence relations' (Marres & Gerlitz, 2016, p. 32), the decision to investigate hashtag co-occurrences as opposed to simply keyword co-occurrences brought into play the dynamics of the medium: In the GDPR dataset the hashtag as a digital object seems to be predominantly used to gain more audience, more so than as a way 'for coordinating conversations around themes..,'(Bruns & Stieglitz, 2012, p.164). That is not to imply that co-occurrence of hashtags is not a meaningful measure, for as Marres and Weltevrede (2013) put it: 'the difference between scraping the medium and scraping the social is probably best understood as a difference in degree: in some cases online devices play an ostensibly large role in the structuration of data [...]' (p.329). In the case of this research, the medium dynamics play a crucial role in the overall structure of the data. As such, this study suggests that the claim that 'proportional measures (frequency) are more likely to direct our attention to medium-specific dynamics (bursting; hyping), while relational measures (connectedness) can help to foreground more substantie dynamics' (Marres & Gerlitz, 2016, p. 34) in the context of hashtag analysis on Twitter is not generalizable, and likely depends on the dynamics of the issue under investigation.

Examining the profile of certain hashtags as a form of actor-network analysis, where the entities associated with the hashtag define it (Latour, 2005), helps provide more insights into the liveliness of issues on Twitter (Marres & Gerlitz, 2016). For example, looking at the changes in associations between the hashtags in question (#ai and #privacy) 'give[s] us a way to narrate the 'life' or 'liveliness' of an issue term' (Marres & Gerlitz, 2016, p.38). Indeed, it is through detailed analysis of the two hashtags that some liveliness is detected. The visualization produced by the associational profiler module allowed for the detection of new hashtag relations, for example with the rise of hashtags #chatbot and #phishing in the case of #ai for example. These two hashtags related to tweets concerning an 'intelligent' chatbot that starts endless conversations with email spammers. The hashtag #ai as such is at some point used in an entirely different context: not in relation to the impact of GDPR on artificial intelligence applications, but in the use of AI in a 'funny' way to combat privacy invasions by spammers. This echoes with the notion that 'engagement on a platform like Twitter does not just involve 'substantive position taking' [...] but relies on a wide variety of information formats' (Marres & Moats, 2015, p.9) and therefore, 'topics of concern' articulated in different registers including humor should be analyzed (Marres & Moats, 2015).

Analyzing the association between the hashtags and the links they co-occur with offers a different lens to evaluate the diversity of a hashtag. The analysis reveals that both hashtags (#ai and #privacy) are mainly associated with company websites and specialized media outlets, though it also suggests that #privacy is somewhat more 'lively' in the sense of being more diverse, with stronger presence of social media, general news websites, and blogs than #ai. However, as explained in the findings, the liveliness of #privacy was only then uncovered when two bursts of retweets (linking to a company website and a news article) were removed from the analysis. In taking on a 'performative deployment approach' (Marres, 2017), where a medium-specific activity such as retweets is not excluded from the analysis as a way to avoid duplicates (Pearce et al., 2014), the analysis risked overlooking interesting dynamics: it was only through the elimination of these two bursts that links to more 'social' domains such as blogs and social media stood out. As such, while '[i]n some cases, applying a measure like 'removing all duplicates' would mean distorting the empirical object' (Marres, 2017, p.136), in this research returning to a precautionary mode proves necessary in certain situations.

Finally, the analysis of the users employing the two hashtags confirms the strong presence of accounts representing companies. Examining the source of the tweets associated with the users reveals a large degree of automation. Gerlitz and Rieder (2018) reject the notion of a 'binary opposition between 'human' and 'bot' (p.538), suggesting instead that '[automations is] a fine-grained nuanced continuum that is organized around the automation of specific functions' (p.538). The results of the analysis display this 'continuum', since the accounts deploying the hashtags include users interacting directly with the platform, bots retweeting certain tweets, but also a number of professional sources including Hootsuite and Hubspot, which are primarily used for scheduling and managing posts on various social media platforms. Thus, one can perhaps suggest that knowing the source of the tweets brings into view the level of engagement of the user employing the hashtag in question. As such, exploring the source variable of the tweets allows for an understanding of the different ways of 'being on Twitter' (Gerlitz & Rieder, 2018).

Put together, the analysis of the profile of specific hashtags proves fruitful for the detection of liveliness. That said, it is only through a comprehensive exploration of the various entities associated with the hashtag that a true understanding of the 'life' of the hashtag is made possible. For example, analyzing the hashtag profile of a hashtag like #privacy demonstrates the various other hashtags it is associated with and how they change and fluctuate over time. Yet these other hashtags are also used in different contexts: the hashtag #dataprotection for instance is used by corporates and privacy activists. It is then only through exploring other entities associated with #privacy like its actor and user profiles that the hashtag's true diversity is revealed.

5 CONCLUSION

Applying co-occurrence analysis on a one-month Twitter dataset about GDPR offers a broader understanding of the capacities and limitations of the method. In this study co-occurrence analysis is used to assess the extent to which insights into the liveliness of issues can be drawn by investigating the 'behavior' of hashtags. The findings of the study suggest that while fluctuations in hashtags associated with the topic provide insights into the life of the issue, a qualitative assessment of the tweets proves crucial, insofar as the content of the tweet is often unrelated to the hashtag. As such, future work that aims to understand the inner workings of Twitter hashtags must combine multiple tools to ascertain 'behaviour' and liveliness. This contrasts with purely quantitative, correlation-based 'big data' research. Moreover, the type of issue under investigation seems to play a significant role in the 'research affordances' (Weltevrede, 2016) that the hashtag as a digital object offers. The topic of GDPR on Twitter is dominated by various businesses whose primary goal is to gain more audiences to market their products and hashtags act as enablers for more visibility. Consequently, the context in which hashtags are used is often unrelated to the substance of the hashtag. This also has another implication on the distinction between frequency and relational measures (frequency of hashtag mention vs frequency of hashtag co-occurrences). The extensive use of multiple hashtags in the same tweet renders the results of the two measures very similar, in contrast to Marres and Gerlitz (2016) whose analysis demonstrates notable differences.

The results of this study further support the notion that an analysis of the profile of specific hashtags in the dataset is especially fruitful for the detection of liveliness. At the same time, the study strongly suggests that a deeper engagement with the dynamics of the hashtag provides more meaningful insights and highlights important shortcomings that must be taken into account in future research. Finally, it is worth noting that since the investigation is based on hashtag analysis, the 'story' of GDPR on Twitter had keywords in the tweets been investigated instead remains a mystery. Further investigation in this area may help us uncover more insights into the usefulness and limitations of hashtag analysis for the detection of liveliness of issues on Twitter.

REFERENCES

Anderson, C. (2008). The end of theory: the data deluge makes the scientific method obsolete. Wired 16 (7).

Back, L., & Puwar, N. (2012). A manifesto for live methods: provocations and capacities. *The sociological review*, *60*(1_suppl), 6-17.

Berry, D. M. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine*, *12*.

Berry, D. (2011). The philosophy of software: Code and mediation in the digital age. Palgrave Macmillan.

Borra, E., & Rieder, B. (2014). Programmed method: developing a toolset for capturing and analyzing tweets. *Aslib Journal of Information Management*, *66*(3), 262-278.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, *15*(5), 662-679.

Bruns, A., & Stieglitz, S. (2012). Quantitative approaches to comparing communication patterns on Twitter. *Journal of Technology in Human Services*, *30*(3-4), 160-185.

Bucher, T. (2013). Objects of intense feeling: The case of the Twitter APIs. *Computational Culture*, *3*.

Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Information (International Social Science Council)*, *22*(2), 191-235.

Cukier, K., & Mayer-Schoenberger, V. (2013). The rise of big data: How it's changing the way we think about the world. *Foreign Aff.*, *92*, 28.

Dijck, J. Van (2017). Foreword. In Schäfer, M. T., & Van Es, K. (Eds.), *The datafied society: Studying culture through data (pp*. 11-12). Amsterdam University Press.

EU GDPR Information Portal. (n.d.). Retrieved July 20, 2018, from https://www.eugdpr.org/

Gaffney, D., & Puschmann, C. (2014). Data collection on Twitter. In Weller, K., Bruns, A., Burgess, J., Mahrt, M., & Puschmann, C. (Eds.), *Twitter and society*, (pp. 55-67). Peter Lang.

Gerlitz, C., & Rieder, B. (2018). Tweets are not created equal. *International Journal of Communication*, *12*, 528-547.

Gillespie, T. (2014). The relevance of algorithms. Media technologies: Essays on communication, materiality, and society, 167.

Gitelman, L. & Jackson, V. (2013). Introduction. In Gitelman, L. (Ed.), *"Raw Data" is an Oxymoron (pp.* 1-14). MIT Press.

Foot, K. A., & Schneider, S. M. (2004). Online structure for civic engagement in the September 11 web sphere. *Electronic J of Communication*, *14*, 3-4.

Hine, C. (Ed.). (2005). Virtual methods: Issues in social research on the Internet. Berg Publishers.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one*, *9*(6), e98679.

Janetzko, D. (2017). The Role of APIs in Data Sampling from Social Media. *The SAGE Handbook of Social Media Research Methods*, 146.

Jones, S. (Ed.). (1998). Doing Internet research: Critical issues and methods for examining the Net. Sage Publications.

Kirschenbaum, M. (2012). What is digital humanities and what's it doing in English departments. *Debates in the digital humanities*, *3*.

Kitchin, R. (2014a). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 2053951714528481.

Kitchin, R. (2014b). The data revolution: Big data, open data, data infrastructures and their consequences. Sage.

Kubitschko, S., & Kaun, A. (2016). An Introduction to Innovative Methods in Media and Communication Research. In *Innovative Methods in Media and Communication Research*(pp. 1-12). Palgrave Macmillan, Cham.

Latour, B. (1987). Science in action: How to follow scientists and engineers through society. Harvard university press.

Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford university press.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... & Jebara, T. (2009). Life in the network: the coming age of computational social science. *Science (New York, NY)*, *323*(5915), 721.

Lupton, D. (2014). *Digital sociology*. Routledge.

Lury, C. (2012). Going live: towards an amphibious sociology. *The Sociological Review*, *60*, 184-197.

Marres, N. (2012). The redistribution of methods: on intervention in digital social research, broadly conceived. *The sociological review*, *60*(1_suppl), 139-165.

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

Marres, N., & Moats, D. (2015). Mapping controversies with social media: The case for symmetry. *Social Media+ Society*, *1*(2), 2056305115604176.

Marres, N. (2015). Why map issues? On controversy analysis as a digital method. *Science, Technology, & Human Values*, *40*(5), 655-686.

Marres, N., & Gerlitz, C. (2016). Interface methods: renegotiating relations between digital social research, STS and sociology. *The Sociological Review*, *64*(1), 21-46.

Marres, N. (2017). Digital sociology: The reinvention of social research. John Wiley & Sons.

Martin, S., Brown, W. M., Klavans, R., & Boyack, K. W. (2011, January). OpenOrd: an open-source toolbox for large graph layout. In *Visualization and Data Analysis 2011* (Vol. 7868, p. 786806). International Society for Optics and Photonics.

Mayer-Schönberger, V., & Cukier, K. (2012). Big Data: A revolution that transforms how we work, live, and think.

Mills, C. W. (1959). *The sociological imagination*. New York: Oxford University Press.

Moats, D. (2015). 'Mapping Controversies with Wikipedia: The Case of Fukushima. *Doctoral diss., University of London, Goldsmiths, UK*.

Mowbray, M. (2014). Automated Twitter Accounts. In Weller, K., Bruns, A., Burgess, J., Mahrt, M., & Puschmann, C. (Eds.), *Twitter and society*, (pp. 183-194). Peter Lang.

Niederer, S. M. C. (2016). *Networked Content Analysis: The case of climate change* (Doctoral dissertation, Doctoral dissertation, University of Amsterdam).

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

Pearce, W., Holmberg, K., Hellsten, I., & Nerlich, B. (2014). Climate change on Twitter: Topics, communities and conversations about the 2013 IPCC Working Group 1 report. *PloS one*, *9*(4), e94785.

Rheingold, H. (2000). The virtual community: Homesteading on the electronic frontier. MIT press.

Rieder, B. (2013). Studying Facebook via data extraction: the Netvizz application. In *Proceedings of the 5th annual ACM web science conference* (pp. 346-355). ACM.

Rieder, B. & Röhle, T. (2012). Digital methods: Five challenges. In Berry, D.M. (Ed.), *Understanding digital humanities* (pp. 67-84). Palgrave Macmillan, London.

Rieder, B., & Röhle, T. (2017). Digital methods: From challenges to Bildung. In Schäfer, M. T., & Van Es, K. (Eds.), *The datafied society: Studying culture through data (pp*. 109-124). Amsterdam University Press.

Rogers, R., & Marres, N. (2000). Landscaping climate change: A mapping technique for understanding science and technology debates on the World Wide Web. *Public Understanding of Science*.

Rogers, R. (2013). *Digital methods*. MIT press.

Rogers, R., Sánchez-Querubín, N., & Kil, A. (2015). *Issue mapping for an ageing Europe*. Amsterdam University Press.

Sayes, E. (2014). Actor–Network Theory and methodology: Just what does it mean to say that nonhumans have agency?. *Social Studies of Science*, *44*(1), 134-149.

Scott, J. (2017). *Social network analysis*. Sage.

Selbst, A. D., & Powles, J. (2017). Meaningful information and the right to explanation. *International Data Privacy Law*, *7*(4), 233-242.

Scharnhorst, A., & Wouters, P. (2006). Webindicators: a new generation of S&T indicators. *Cybermetrics*, *10*.

Schreibman, S., Siemens, R., & Unsworth, J. (Eds.). (2008). *A companion to digital humanities*. John Wiley & Sons.

Schroeder, R. (2014). Big Data and the brave new world of social media research. *Big Data & Society*, *1*(2), 2053951714563194.

Snee, H., Hine, C., Morey, Y., Roberts, S., & Watson, H. (2016). Digital methods as mainstream methodology: An introduction. In *Digital Methods for Social Science* (pp. 1-11). Palgrave Macmillan, London.

Taylor, C., & Coffey, A. (2008). Innovation in qualitative research methods: Possibilities and challenges.

Thelwall, M. (2014). Sentiment Analysis and Time Series with Twitter. In Weller, K., Bruns, A., Burgess, J., Mahrt, M., & Puschmann, C. (Eds.), *Twitter and society*, (pp. 83-95). Peter Lang.

Townsend, L., & Wallace, C. (2016). Social media research: A guide to ethics. University of Aberdeen.

Twitter. (n.d.) Retrieved August 1, 2018, from https://developer.twitter.com/en/docs/basics/getting-started

Uprihard, E. (2012). Being stuck in (live) time: The sticky sociological imagination. *The Sociological Review*, *60*(1_suppl), 124-138.

Venturini, T. (2010). Diving in magma: how to explore controversies with actor-network theory. *Public understanding of science*, *19*(3), 258-273.

Wellman, B., & Gulia, M. (2018). Net-surfers don't ride alone: Virtual communities as communities. In *Networks in the global village* (pp. 331-366). Routledge.

Weltevrede, E. J. T. (2016). *Repurposing digital methods: The research affordances of platforms and engines* (Doctoral dissertation, Universiteit van Amsterdam [Host]).

Wiles, R., Bengry-Howell, A., Crow, G., & Nind, M. (2013). But is it innovation?: The development of novel methodological approaches in qualitative research. *Methodological Innovations Online*, *8*(1), 18-33.

Wynn, J. R. (2009, June). Digital sociology: emergent technologies in the field and the classroom. In *Sociological Forum* (Vol. 24, No. 2, pp. 448-456). Oxford, UK: Blackwell Publishing Ltd.

Yang, S., Quan-Haase, A., & Rannenberg, K. (2017). The changing public sphere on Twitter: Network structure, elites and topics of the #righttobeforgotten. *new media & society*, *19*(12), 1983-2002.

Media@LSE Working Paper Series

Media@LSE Working Paper Series will:

Present high quality research and writing (including research in-progress) to a wide audience of academics, policy-makers and commercial/media organisations.

Set the agenda in the broad field of media and communication studies.

Stimulate and inform debate and policy. All papers will be published electronically as PDF files, subject to review and approval by the Editors and will be given an ISSN.

An advantage of the series is a quick turnaround between submission and publication. Authors retain copyright, and publication here does not preclude the subsequent development of the paper for publication elsewhere.

Notes for contributors:

Contributors are encouraged to submit papers that address the social, political, economic and cultural context of the media and communication, including their forms, institutions, audiences and experiences, and their global, national, regional and local development. Papers addressing any of the themes mentioned below are welcome, but other themes related to media and communication are also acceptable:

Communication and Difference

Globalisation and Comparative Studies

Innovation, Governance and Policy

Democracy, Politics and Journalism Ethics

Mediation and Resistance

Media and Identity

Media and New Media Literacies

The Cultural Economy

Contributions are welcomed from academics and PhD students. In the Autumn Term we also invite selected Master's students from the preceding year to submit their dissertations which will be hosted in a separate part of this site as 'dissertations' rather than as Working Papers. Contributors should bear in mind when they are preparing their paper that it will be read online.

Papers should conform to the following format:

6,000-10,000 words (excluding bibliography, including footnotes)

150-200 word abstract

Headings and sub-headings are encouraged

The Harvard system of referencing should be used

Papers should be prepared as a Word file

Graphs, pictures and tables should be included as appropriate in the same file as the paper

The paper should be sent by email to Bart Cammaerts (b.cammaerts@lse.ac.uk), the editor of the Media@LSE Working Paper Series

ISSN: 1474-1938/1946