Global freedoms and viral harms: The controversy around governance of speech and social media

PhD in Sociology

April 2021

Chiara Poletti

1453250

Supervisors: William Housley and Adam Edwards

Abstract

In the study I address the controversy surrounding the governance of speech and social media communications. In less than 15 years, the regulation of content on social media platforms has increasingly taken over public discussions all over the globe. Social media's charming narrative of 'liberation technology' and space of free speech, has progressively switched into the frightening character of 'threat to democracy' and space of hate speech and fake information. Whichever idea one might be leaning on, the diffusion and entanglement of social media platforms with every aspect of our society has made content regulation on social media a global public issue.

Scholars have stressed how governance of speech has been in the hand of a plurality of actors, in a plurality of settings. In the lack of a single decision-making process, governance initiatives emerge as a reaction to public shocks. In this study, I investigate how public shocks have contributed to regulation initiatives. Using theoretical concepts from Actor-Network Theory (ANT) and critical data studies and the methodological tools from controversy mapping, I have analysed narratives about free speech, technology and governance models on websites and in the UK press from 2015 until 2018. The analysis reveals public bodies have increasingly assigned public policy responsibilities to social media and their technology (algorithms and A.I.). However, they miss considerations about the social implication of this type of governance of speech, which reinforces the structure of organisation of platform economy and algorithmic management of social life. With this study, I hope to contribute to the empirical study of governance of speech as well as presenting a normative reflection on the type of governance. I also include a meta-reflection on the role of researchers, and in particular on how this methodology and theory can expose the paradoxes hidden in the black boxes of technology.

Acknowledgements

The road goes ever on. As much as it might sounds cliché, I am deeply convinced that every PhD is a journey. Down from the door where it began, it takes you to people, places and events you would never expect. As in the most memorable journeys, not everything goes as planned. Sometimes you might get lost in the woods of literature, other times you might hit a dead end with the data. Sometimes you enjoy the journey so much that a part of you would not want it to end. Other times, you might want to leave it, as life happens and reclaims its space. In my PhD journey, I passed all these places. Coming to an end, I clearly see that I would never had been able to accomplish what I have done without the help and support of a vast number of people, to whom goes my deepest gratitude. A huge thank you goes to the ERSC Wales DTP and the School of Social Sciences in Cardiff University and to my supervisors William Housley and Adam Edwards, for believing in my project and sustaining me throughout this journey. Thank you to Cardiff University I have met great fellow PhD students and academics who also happen to be fantastic friends and human beings. I have a huge debt of gratitude to all of them and in particular to every member of the Curemede, and the T-Grains teams who taught me so much about the love and passion of doing research as a job. I had the joy and privilege of sharing part of my journey with fellow adventures in the world of academia who inspired me and motivated me. To all of you my deepest gratitude and admiration. My journey would have never started without the help and inspiration from Claudia Padovani, Francesca Musiani, Stefania Milan, Julia Pohle and Arne Hintz. I am very grateful to all the people who patiently answered my questions and doubts: thank you to Andrea Calderaro, Mathieu Jacomy and the members of the MediaLab in Paris, Christian Katzenbach and the great groups of research at the HIIG in Berlin, Noortje Marres at the University of Warwick and Stefania Milan and the team in Amsterdam. I am deeply grateful to Davide Crivelli and Allison McKechnie for their help and to Niels ten Oever for his support and unconditional cheerleading. I would like to express my gratitude to all of those who have accompanied my journey also from outside the world of academia, supporting me with love and patience. My chosen family, my friends of a lifetime Bomb, Kerol, Annuzza. My wonderful and precious Birbe e portagioielli. Marcorasao Cathy and Dan. The Italian Connection in Wales and now conquering London. My sunshine, Leon. You are my daily reminder that the world is full of blessing. Finally, the deepest thanks to my big fat Italian family. The Clan, you are the pillars of everything. I am very proud of you, and I love you all, *anca se no capí un casso e gavé da convertirve*.

Contents

Abstract i

Acknowledgements ii

List of Tables ix

List of Figures ix

1. Introduction 1

1.1 Social media, free expression, and speech regulation 1

1.2 Governance of speech, social media and academic research 2

1.3 Contribution to the field and main findings 5

1.4 Structure of the thesis 7

2. Literature review 9

2.1 Introduction 9

2.2 Narratives about technology and regulations of speech 10

2.2.1 Optimistic narratives and low regulation of speech 11

2.2.2 Pessimistic narratives and regulatory backlash 16

2.2.3 Regulatory policies governing mass media before the Internet 19

2.3 Governance of speech: overview 20

2.3.1 Plurality of actors 22

2.3.2 Plurality of settings and means used to influence policy used by single actors 23

2.3.3 Plurality of settings – mixed-actors initiatives 29

2.4 Main issues from the literature 31

2.4.1 Unbalanced distribution of power 32

2.4.2 Platformisation 33

2.4.3 Threats to free speech 35

2.4.4 Societal interests 36

2.5 Governance of speech as object of research: gaps, and how to fill them 37

2.6 Conclusion 41

3. Theoretical framework.. 42

3.1 Introduction 42

3.2 Theoretical framework.. 43

3.2.1 Material semiotics and digital society: an introduction.. 43

3.2.2 Material semiotics.. 45

3.2.3 Critical data studies.. 46

3.3 Using STS, ANT and material semiotics to close the research gaps........................ 48

3.3.1 Previous use of theories.. 50

3.4 Key concepts.. 50

3.4.1 Actors and associations 50

3.4.2. Translation 51

3.4.3 Roles.. 52

3.4.4 Controversies.. 54

3.4.5 Public and tools for measurement...55

3.4.6 Datafication... 56

3.5 ANT and critical data studies concepts in the literature 58

3.6 Key concepts and governance of free speech online 59

3.7 Drawbacks and limitations of material semiotics 61

3.8 Conclusion... 65

4 Methodology 66

4.1 Introduction 66

4.2 Methodology and research aims 67

4.3 Suitability of controversy mapping for research on governance of speech................. 70

4.4 Epistemological and ontological positions.. 72

4.5 Data collection tools and reason for use 74

4.5.1 Digital tools for collection of statements.. 74

4.5.2 Implications of using digital tools for data collection... 76

4.5 3 Keyword selection 79

4.5.4 Google's ranking algorithm 80

4.7 Sources of data collection 82

4.8 Data analysis tools and justification 84

4.8.1 Qualitative content and discourse analysis 84

4.8.2 Computational tools for text analysis 86

4.8.3 Samples used in analysis 92

4.9 Limitations of controversy mapping with web pages 94

4.9.1 Mixing methods: centrality of researcher and quantitative approach 95

4.10 Limitations of newspapers 99

4.11 Ethical considerations 99

4.11.1 Public vs private space 100

4.11.2 Subject vs author 101

4.12 Conclusion 102

5. Mapping controversies using web pages 104

5.1 Introduction 104

5.2 Construction of corpus 104

5.3 Presentation of findings 106

5.3.1 Identification of actors 106

5.3.2 Overview of actors 111

5.4 Types of documents 118

5.5 Study of hyperlink relationships 119

5.5.1 Interpreting the networks 120

5.5.2 Analysis of measures of in-degree/out-degree 123

5.5.3 Social media platforms 131

5.5.4 News media 133

5.5.5 Public bodies and international organisations 134

5.5.6 Civil society: NGOs 135

5.5.7 Academia and think tanks 137

5.6 Other findings from the analysis of structure of hyperlinks 139

5.6.1 Technological dynamics 140

5.7 Identification of shocks and regulation initiatives 141

5.8 Emblematic episodes or recurring topics 142

5.9 Key issues emerging from qualitative analysis 152

5.10 Groups of actors' positions on issues 156

5.11 Narratives about free speech 158

5.12 Narratives about governance 164

5.12.1 Private companies 169

5.12.2 The position of other international bodies 169

5.13 Narratives about technology 171

5.14 Findings and literature 174

5.15 Conclusions 177

6. Mapping controversies using newspaper articles 179

6.1 Introduction 179

6.2 Data collection 180

6.2.1 Construction of corpus 180

6.3 Presentation of findings 182

6.3.1 Identification of actors 182

6.3.2 Identification of shocks and issues they summarise 184

6.3.3 Narratives about free speech 201

6.3.4 Narratives about governance 206

6.3.5 Narratives about technology 213

6.3.6 The temporal dimension 215

6.3.7 The shape of the controversy 217

6.4 Conclusion ... 220

7. Discussion ... 223

7.1 Introduction ... 223

7.2 Morphology of the controversy ... 224

7.2.1 Actors' composition ... 225

7.2.2 Public shocks and regulations ... 228

7.3 Dynamics connecting actors ... 231

7.4 Detection of roles ... 232

7.4.1 Spokespersons and inscriptions ... 232

7.4.2 Intermediaries and Mediators ... 234

7.4.3 Tools for measurement ... 235

7.5 Translation processes ... 236

7.5.1 Problematisation 1 – terrorism and national security ... 238

7.5.2 Problematisation 2 – hate speech ... 241

7.5.3 Problematisation 3 – harassment ... 242

7.5.4 Problematisation 4 – fake news and misinformation ... 243

7.6 Translation as a tale of contrasting views ... 244

7.6.1 Free speech ... 245

7.6.2 Governance ... 246

7.6.3 Technology ... 248

7.7 Macro perspective ... 249

7.8 Technological dynamics ... 254

7.9 Role of researcher ... 255

7.10 Conclusions ... 256

8. Conclusion ... 258

8.1 Research questions and objectives ... 259

8.2 Contribution and significance of the findings ... 261

8.3 Limitations of the study, opportunity for further research 263

References 268

References to web documents (chapter 5) 294

References to newspapers' articles (chapter 6) 298

Appendices 1

Appendix: List of tables 1

Appendix: List of figures 1

Appendix. 1 1

1) Web pages dataset 1

2) Data from Qualitative analysis of texts 4

3) Data from quantitative analysis of texts 8

4) Data from SNA - Graphs visualisations 9

Appendix 2 18

1) Newspaper's dataset 18

2) Data from quantitative analysis of texts 20

4) Newspapers: Semantic relations topics-keywords 32

5) Time plot of the topics 33

6) Newspapers sampling 46

List of Tables

Table 5. 1 - Summary of URLs collected after removal of duplicates 105

Table 5. 2 - Expansion of list of actors after the crawl 105

Table 5. 3 - Example of identification of actor from URLs.. 107

Table 5. 4 - Groups of actors explained 108

Table 5. 5 - List of actors from list 1 – scrape of Google.co.uk 109

Table 5. 6 - Expanded list of actors 111

Table 5. 7 - News media publications that appear most often, across the years (first list)......... 112

Table 5. 8 - Summary of NGOs with more than 2 pages collected in the dataset 113

Table 5. 9 - Most frequent academia and think tank actors 115

Table 5. 10 - Most frequent public bodies actors.. 116

Table 5. 11 - Most frequent private companies 117

Table 5. 12 – Most frequent International organisations or European institutions.................... 118

Table 5. 13 - From URLs to networks 122

Table 5. 14 - Exemplary cases extracted from statements 143

Table 6. 1 - Summary of articles collection and coding per year 181

Table 6. 2 - Publications and political orientation 183

Table 6. 3 - Cortext LDA topics output and exemplary cases from qualitative analysis 185

Table 7. 1 - Overview of actors' roles in the controversy 236

List of Figures

Figure 4. 1 - top 100 results from Google's ranking algorithm 82

Figure 4. 2 - Coherence score output produced for newspaper topic analysis in LDAvis, Cortext. 89

Figure 4. 3 LDAvis output. Visual exploration of relationship between terms-topics-corpus..... 90

Figure 4. 4 LDAvis output. Visual exploration of relationship between terms-topics-corpus. 91

Figure 5. 1 - Colours assigned to each group.. 121

Figure 5. 2 - Example of overview of a network 122
Figure 5. 3 - Nodes with higher in-degree in 2015, coloured by groups 124
Figure 5. 4 - Nodes with higher in-degree in 2016, coloured by groups 125
Figure 5. 5 - Nodes with higher in-degree in 2017, coloured by groups 126
Figure 5. 6 - Nodes with higher in-degree in 2018, coloured by groups 127
Figure 5. 7 - Nodes with higher out-degree in 2015, coloured by groups 128
Figure 5. 8 - Nodes with higher out-degree in 2016, coloured by groups 129
Figure 5. 9 - Nodes with higher out-degree in 2017, coloured by groups 130
Figure 5. 10 - Nodes with higher out-degree in 2018, coloured by groups 131
Figure 5. 11 - Detail of the network of URLs in 2018.. 133
Figure 5. 12 - Overview of hyperlinks connections in 2018.. 134
Figure 5. 13 – Detail of nodes with higher indegree in network in 2015 137
Figure 5. 14 – Detail of network with higher outdegree in 2015... 138
Figure 5. 15 - Detail of network with higher outdegree in 2018 139
Figure 5. 16- Paul Chambers' tweet. Source: Daily Mail (2012) 148
Figure 5. 17 - Actors' statements on issues 157
Figure 5. 18 – Narrative about freedom of expression divided by group of actors 158
Figure 5. 19 – Example of contents in the blog Fortress of Faith.. 160
Figure 5. 20 – Narratives about Governance models by groups of actors................................. 165
Figure 5. 21 - Most mentioned technological aspects in the texts from web pages 171

Figure 6. 1 - Percentage of articles in the dataset per publication 183
Figure 6. 2 - Semantic connection between Topic 5 and 2 (yellow links) and most relevant terms 188
Figure 6. 3 – Semantic connection between Topic 3 and 10 (both in blue links) and most relevant terms.. 191
Figure 6. 4 – Semantic connection between Topic 4 and most relevant words (red links) 193
Figure 6. 5 – Semantic connection between Topic 6 and 9 (green and dark green) and most relevant words.. 197
Figure 6. 6 – Semantic connection between Topic 7 (yellow) and 4 (red) and most relevant words 199

Figure 6. 7 - Nick Ut's photograph of Kim Phúc 200
Figure 6. 8 – Semantic connection between Topics 6 and 9 (green and dark green) and topic 10 (blue) and most relevant words 201
Figure 6. 9 – Semantic connection between Topics 1-8 (yellow) and most relevant words 205
Figure 6. 10a Temporal evolution of the topics: colours associated with topics 215
Figure 6. 10b Temporal evolution of the topics. 216
Figure 6. 11 Racist tweet (partially represented here) included in the article in the Daily Mail (2016-07-19) 218
Figure 6. 12 Racist tweet (partially represented here) included in the article in the Independent (2016-07-19) 219

1. Introduction

1.1 Social media, free expression, and speech regulation

In recent years, speech on social media (SM) platforms has become a controversial global policy issue. Terrorist attacks, disinformation and fake news on the occasion of high-stakes political events, together with racist and misogynistic abuse and harassment, have underlined how social media can exacerbate and contribute to the spread of harmful content, with massive social and political repercussions.

SM platforms were previously regarded as critical spaces for the protection of free expression (Dutton, 2009; Diamond, 2010; Diamond and Platter, 2012) and benefited from a non-regulatory regime. However, their technology, internal rules, and business models are increasingly viewed as sources of public risk and subject to national state regulation initiatives (Cusumano et al., 2021). In recent years, SM companies have updated their internal policies several times (Belli and Venturini, 2020), enhancing their content moderation systems by introducing algorithms for automatically detecting content and hiring thousands of human moderators. However, such initiatives have yet to be considered adequate to face the threats and harm deriving from free speech on their platforms (Cusumano et al., 2021) and SM companies' role, powers, and responsibilities in the global governance of speech have become a significant issue for national and international policymakers (Gillespie et al., 2020; Edwards et al., 2021).

Political institutions at all levels have been demanding that SM companies put in place more effective forms of content regulation and have issued legislation to increase state powers in the field of content monitoring and regulation (EU Commission, 2016; Shields, 2017). Only recently a large number of policies and *ad hoc* services for monitoring and policing content on SM companies have been created at the national level, such as the French *Vigipirate* plan and 'state of emergency law' established in 2015, as well as the 'famous' German 'NetzDG' law in 2017, or the UK Investigatory Powers Bill and Digital Economy Act, and more recently, the UK White Paper on Online Harms (2019). Regulation initiatives have taken place also at the supra-national level, as in the EU Internet Referral Unit (EU IRU), established in 2015 and the EU Code of Conduct on Countering Illegal Hate Speech Online (2016).

The combination of the regulatory trend initiated by Western states and the increasing number of internal policy changes adopted by private corporations in the area of blocking, filtering, and removal of content, has started to raise a concern about implications for human rights (EDRi, 2016), especially in terms of freedom of expression and access to information, as highlighted by Council of Europe (CoE) Secretary-General Thorbjørn Jagland (2016) and UN Special Rapporteur on freedom of expression, David Kaye (2016, 2017). Without a clear definition of threats, measures, and remedies in national, local, and supranational legislation, all efforts to control content and data on SM are at risk of being disproportionate and illegitimate measures (Kaye, 2016, 2017). Furthermore, shifting the responsibility of governance of fundamental freedoms and rights to private companies and their technologies increases the risk of illegitimate and opaque influence on public life, especially considering that SM companies' 'for-profit nature' does not resonate well with democratic ideals (Deibert, 2003; Mueller, 2010; De Nardis, 2012; MacKinnon, 2012; Morozov, 2010; Fuchs, 2014; Pickard, 2014; De Nardis and Hackl, 2015; Ippolita, 2015; O'Neil, 2016; Hintz, 2016).

Scholars, noticing how governance of content and the current online governance regime interests human rights and fundamental freedoms, have started to revisit difficult questions about how speech online is structured and how governance of speech can be theorised and researched (van Dijck, 2013, 2014; Wagner, 2013, 2016; DeNardis and Hackl, 2015; MacKinnon et al., 2014; Gillespie, 2015, 2018, Ananny and Gillespie, 2016; Balkin 2017; Gillespie et al. 2020).

1.2 Governance of speech, social media and academic research

Defining the governance of speech online as a research object presents several theoretical and methodological challenges. The concept covers political, economic and technological aspects. Firstly, it refers to a variety of processes, involving a large number of heterogeneous actors (i.e. private corporations, public institutions and bodies, state governments, but also representatives of organised civil society such as academics and NGOs, as well as less structured networks of individuals and users) in a transnational ecosystem. In addition, technology itself has a crucial role in the *mise en place* of this governance regime, which is, for many, a sociotechnical regime. Among these complexities, scholars tend to agree on some main features: online forms of governance take place as a performative, emerging order rather than a planned action, often as a reaction to public shocks or controversial situations (Hofmann et al., 2016; Ananny and Gillespie, 2016). Furthermore, regulatory powers do not belong to a single actor, and different groups of actors can use various tools and means to influence governance in various settings with varying degrees of formality and hierarchical organisation (Wagner, 2016; Hintz, 2016; Balkin, 2017; Gorwa, 2019). In particular, authors studying governance processes online have stressed the importance of discourses and discursive tools in the norm and subsequent policy creation (Padovani and Santaniello, 2018; Radu, 2019).

As I will discuss in greater depth in chapter 2 (literature review), historically, research on speech online has focused on a restricted number of spaces, mainly the more prominent SM platforms (Twitter, Facebook, and Youtube). Moreover, research on governance tends to focus on specific regulatory initiatives, either at the national or international level, rather than considering transversal and informal relationships connecting actors. Recently, scholars have started to stress the limitations of such approaches and recommended studying with more attention to the role of less-institutionalised elements in the study of speech governance regimes, such as civil society and 'traditional media, as well as the role of technology itself (in particular algorithms) as a tool for regulation. Scholars have also called for more research focusing on the transversal and informal relationships which contribute to orientate governance in such complex environments (see, among others, van Eeten and Mueller, 2012; Gillespie et al., 2020).

Based on these considerations, I ask in this study: how can we study the governance of speech online as an emerging phenomenon without focusing on a single actor or a single regulatory setting? How do governance initiatives 'initiate' and take form? Moreover, what does it mean for the broader governance of freedom of expression and democracy? By answering these questions, I aim to contribute to the understanding of the governance of speech online by adopting an empirical rather than theoretical approach to identifying actors, narratives, and material elements attached to technology and the power dynamics that link them. In order to achieve my aim, my objectives include:

- Using the tools of controversy mapping, *empirically* scope the actors and elements involved in the controversy surrounding the governance of speech online, identifying the actors animating the controversy online and in the UK press.
- Assess the role of institutional (e.g. states and businesses) and less-institutionalised actors (e.g. civil society, the press) as well as non-institutionalised actors (such as technology) in

the governance system using the theoretical tools provided by the sociology of associations (i.e. Actor-Network Theory (ANT)).

- Explore the material implication of this type of governance regime using concepts from critical data studies.

To answer these questions and achieve my objectives, I combined two theoretical approaches in this study: one derived from material semiotics and the other belonging to the tradition of critical studies. First, considering platforms as one of the exemplary sites of socio-technological controversies, I used controversy mapping to define the limits of my object of study and define a sample of actors. Second, in line with the inductive approach prescribed in controversy mapping, I have observed the actors' agendas and alliances through their documentary productions available online (i.e. websites) and the description given by the public press (i.e. UK newspapers). I have thus adopted a more deductive form of analysis, using discourse and content analysis to identify elements characterising the issues at stake and the narratives about free speech, governance, and associated technology. I subsequently interpreted the findings in two steps. First, I gave an interpretation of the controversy by linking the findings of the mapping exercise with the theoretical elements of ANT's sociology of translation (Callon 1986a). Secondly, I interpreted the findings in the light of the theoretical elements typical of critical data studies focusing on the power dynamics embedded in the socio-technical assemblage of SM users' speech and data.

The study focuses on the controversy around social media governance and freedom of expression, using as privileged point of observation the UK context. The choice to approach the study of governance of speech from the UK context was motivated by the fact that the UK has been one of the European countries most active in the debate on the roles and responsibilities of SM platforms. It is a stable democracy whose history renders it a 'bridge' between continental European approaches and US positions. Moreover, English is one of the most employed languages to produce documents online. However, in order to understand where the controversy takes place and to include as many actors as possible, the study considers two different public spaces: the internet and traditional media, using as sources of data web pages collected from Google.co.uk, and newspaper articles from LexisNexis as critical sources of data.

1.3 Contribution to the field and main findings

With this thesis, I aim to fill a gap in research by questioning how the emergent governance of speech takes place, focusing on the controversial events and the narrative attached to them rather than on a specific actor or regulatory process. I have investigated the dynamics of the creation of governance of speech by studying controversial elements involving SM platforms and broader society, using an ANT-informed theoretical and methodological approach.

In this study, I present a new approach to the study of governance regimes online, based on the theorisation of governance of freedom of speech on SM as a 'controversy' and the empirical definition of actors involved in the governance system, using ANT's concepts to identify the elements that make governance 'visible'. This choice displaces the focus of the research from the ontology of governance of freedom of speech online towards the practical ways in which governance of freedom of speech online is performed. Using a combination of controversy mapping, the ANT theoretical framework, and critical data studies, I highlight how heterogeneous actors concur in creating the governance regime of freedom of expression. This approach fits with the literature on governance as emerging and related to moments or episodes of shocks or emblematic issues (Pohle, 2016a, 2016b; Hofmann, 2016; Hofmann et al., 2016; Ananny and Gillespie, 2017). However, instead of adopting an approach privileging one specific actor or regulatory tool, I performed the study of governance with an initial agnosticism about the roles and hierarchies, leaving it to the actors to indicate who or what does or does not take part in the governance system.

I contribute to the understanding of the governance of speech on SM platforms by collecting empirical data supporting the idea that governance is the outcome of a reaction to public shocks and, on the other hand, presenting a more extensive range of actors and means to influence speech governance. In particular, I was able to highlight the strategic role of technology narratives and the media (as the disseminators of narratives) in the development and orientation of regulatory measures. Moreover, using a critical data approach, I placed this empirical evidence within the framework of the specific power structure that links speech online to the economy and politics of digital data.

My main contributions to the field can be summarised in four layers of findings, some more related to the 'substance' of the controversy and others pointing towards the 'technological dynamics' that make the controversy visible (Marres 2015).

The first layer relates to the actors involved in the governance process, their normative interpretation of free speech and governance, and the power dynamics in which they have been involved. Using ANT translation theory in the analysis, I show the politicisation of the prominent positions concerning free speech on the axis left-right, and, at the level of governance and how public bodies in European states have pushed for the development of a co-regulatory framework, placing SM companies in the strategic role of enforcers of regulation of speech online through the application of their technological tools.

The second level of findings adopts a macro perspective and connects the previous layer to further significant societal implications. In this case, the data shows how public bodies are helping to legitimise the role of private companies and the use of technology in the governance of speech as part of a broader trend toward platformisation of public policies and datafication of life. It also stresses the social and material implications of such a trend, recommending further reflection on the implications of a co-regulatory framework of governance grounded in AI and machine learning technology as a tool for policy.

The third layer of findings verges on the technological dynamics that 'publicise' contemporary controversies (Marres, 2017): here the data shows that newspapers have increasingly been configured and limited by SM platform architecture, with articles taking on issues generated by the platforms without providing a critical contextualisation for the readers. The analysis of exemplary cases and storylines highlights how most controversial cases used as storylines (as we will see there are several, as in the case of the' feud' between Leslie Jones and Milo Yiannopoulos in chapter 6) acquired their exemplarity because of the divisive contents and virality provided by SM platforms' architecture. In this thesis, I argue that reproducing such content, on the one hand, creates a storyline and helps the public identify issues and take a position. However, at the same time, it reproduces rather than critically assesses the biases present in the platforms, as the virality of examples and cases is not critically assessed and put into the larger perspective of technological dynamics governing the platforms. This consideration connects the findings to the larger context of the attention economy and the platform economy in general.

The last type of finding relates to how sociology approaches digital materials and artefacts and can be considered a retake on the 'social life of methods and data' (Savage, 2013; Law and Ruppert, 2013). In this part, I discuss the role of researchers working with and on digital data, and how through the labour of doing digital social research (often not documented in academia), it is possible to open black-boxes and demystify narratives about technology and public discourse.

1.4 Structure of the thesis

In the next chapter (chapter 2), I introduce the topic by giving an overview of the historical changes in the narratives about freedom of expression associated with internet and social media technology, followed by a literature review on the topic of governance of freedom of expression online. Based on the literature I identify gaps and I situate my research questions. In chapter 3 I present my theoretical background and describe the theoretical concepts that I derive from material semiotics approaches, such as Actor-Network Theory (ANT) and those concepts from critical data studies that I use. I introduce the importance of studies of controversies as a way to investigate public shocks leading to regulatory initiatives.

In chapter 4 I describe the methodology, derived from the ANT-informed approach. I explain how I have performed a study of freedom of expression on SM platforms as a controversy. I discuss the issue of digital bias and the issues related to the data collection and analysis. Based on these considerations, in the last part of the chapter I explain the selection of my case study and sources for my data collection, i.e. websites and articles in the British press. I conclude with an overview of the workflow and ethical considerations.

In chapters 5 and 6 I present the findings for the data collection and mapping exercise using two different sets of data. I start with how I have identified actors and issues animating the controversy around social media platforms and freedom of expression in the public space constituted by web pages (chapter 5) and I follow with the study of the public and the issues that emerge from the British general press (chapter 6). I used qualitative discourse and quantitative content analysis to identify actors, issues and their relationships.

In the discussion chapter (chapter 7), I apply theoretical concepts described in chapter 3 to the findings of the mapping exercise. I interpret the data according to the roles described in ‘ANT’ sociology of translation (i.e. the spokespersons, mediators and intermediaries taking part in the controversy). I use critical data studies to interpret the societal implications of the roles that emerge in the translation processes and the other findings from the controversy mapping.

In the conclusion, I discuss the various implications of my findings, starting from the technological dynamics behind the morphology of the controversies concerning free speech and social media platforms and the dynamics of transmission of content/activation of controversies. I then move towards the macro perspective and consider the relationship between the dynamics highlighted above, and the general structure of organisation of contemporary society, i.e. capitalist system of production applied to information technologies. Finally, I include a meta-reflection on the role of researchers and the methodology used, discussing the limitations and possible avenues for future research.

2. Literature review

2.1 Introduction

In Chapter 1, I presented the governance of freedom of expression online as the core of one of the most prominent contemporary public controversies. In this chapter, I discuss in greater detail some of the key conceptual, methodological and empirical challenges encountered when designing a study of the governance of speech in contemporary online communications, which are dominated by SM platforms.

In the first part of this chapter, I review historical and recent literature produced on online content regulation. I highlight how academic scholarship and national policies about free expression online have developed together with specific narratives about technology, pointing out that an optimistic narrative of internet and SM technology corresponded to a period of low regulation of speech. In contrast, a pessimistic narrative associated with technology corresponds to increased calls for regulation of expression. In this part, I argue that even though similar paths of regulatory policy have appeared before in pre-internet companies (Cusumano et al., 2021) and other mass media (Wu, 2010; Pickard, 2015), the commercial expansion of the internet in the 1990s, and the global socio-technical system that ensued, has had socio and political repercussions of a magnitude never experienced before (Radu, 2019). Consequently, it requires a different way of studying how governance of speech is now performed.

In the second part of the chapter, I focus on recent literature exploring how the governance of speech on SM can be conceptualised as an object of research. In this regard, I review the problematic elements represented by SM, being both regulators (or governors as in Klonick, 2017) and objects of regulation (Gillespie, 2018; De Nardis and Hackl, 2015). Drawing from the literature, I argue that governance of speech is a more fitting concept than regulation since free speech and its limitations on SM platforms are determined by a plurality of actors in a plurality of settings, both formal and informal, and implemented by the different actors involved in the process of governance using a heterogeneous set of tools.

Drawing on the work of Gillespie, Wagner, Balkin, Gorwa, and other scholars of internet governance and international relations, I present an overview of the actors traditionally considered in the study of governance online (i.e., states, private companies and civil society). I then review their main means of engaging in the debate and influencing policy (e.g., states' regulatory frameworks, civil society and social movements' different type of actions, companies' private ordering and self-regulation tools). As a result of this overview, I argue that there is a necessity to expand the range of actors in the study of the governance system and recognise the political role of technology as a tool for content regulation (Bucher, 2018; Kaye, 2018; Gillespie et al., 2020; Sinnreich, 2020). In particular, I propose to fill this gap in the literature about the governance of speech using an empirical rather than theoretical method to identify the range of actors involved in governance and to consider, among the means to influence regulation, the performative power of narratives associated with technology and the public discourse created by the media. Finally, I conclude the chapter with a list of research questions developed from the review discussed and briefly introduce my theoretical framework.

2.2 Narratives about technology and regulations of speech

Regulation of speech online is a much debated area that in the last twenty years has attracted scholars from different disciplines and has increasingly gained space in the public debate, especially in terms of dealing with harms and unforeseen consequences of emergent technologies (Deibert, 2003; Heimlick, 2008; Peng Hwa, 2008; Mueller, 2010; De Nardis, 2012; Edwards et al., 2013; van Dijck, 2013; Fuchs, 2014; Isin and Ruppert, 2015; De Nardis and Hackl, 2015; Wagner, 2016; Gillespie, 2015, 2017, 2018; Zubiaga et al., 2016; Procter et al., 2019; Gillespie et al., 2020; Gorwa, 2019a, 2019b, 2020, Gorwa et al. 2020; Edwards et al., 2021).

Drawing from cognitive studies and psychology, scholars have stressed on several occasions that legislation or regulations for new technologies are often developed based on analogies, metaphors and narratives attached to pre-existing technologies or social situations (Gore, 2003; Luokkanen et al., 2014; Suzor, 2019). Scholarship has found evidence that this is particularly common with digital technologies (Mueller, 2010; Couldry, 2012; Mezei and Verteș-Olteanu, 2020; De Filippi et al., 2020). Scholars are increasingly interested in the social implications of narratives associated with technology and, in particular, the ethical implications that metaphors of technology can have for research and politics, as, for instance, the values attached to the idea of ‘Big Data’ (Mayer-Schonberger and Cukier, 2013; Puschmann and Burgess, 2014).

As mentioned in Chapter 1, public discussions are dominated by calls for the regulation of speech on SM. SM companies are now considered primarily responsible for the toxic nature of the public sphere, the spread of misinformation and disinformation, the polarisation of opinions, and even for mental health issues created by addictive models on their platforms (Owen, 2019). However, it is essential to recall that, for years, scholarship, governments and the general public have been influenced by different narratives and metaphors of information and communications technology (ICT), for instance that of cyberspace. In the following paragraphs, I provide a historical overview of the narrative associated with ICT to highlight how different narratives correspond to different attitudes towards the regulation of speech online.

2.2.1 Optimistic narratives and low regulation of speech

The early years of the internet corresponded to relatively low regulation of communication technology and especially low control of expression. Very few people were using the network, and most of them were in academia and research centres. In case of necessity, it was quite easy for states to resort to existing legislation and resources to regulate (Peng Hwa, 2008). For years, the government of the United States, where the majority of the digital communication network was created, followed a so-called ‘hands-off’ approach with minimal interference. These first years were characterised by a specific narrative associated with technology, i.e., the values and principles developed within the hacker counterculture and cyberlibertarian culture of the 1970s and 1980s. In this narrative, cyberspace represented a different and more unrestricted space, opposed to the control of governments (Mezei and Vertes-Olteanu, 2020:5). Computers, digital technologies and communication networks were tools of personal liberation (Malcomson, 2016). In a highly cited document, John Perry Barlow's 'A Declaration of the Independence of Cyberspace' (1996), Barlow summarises this position, stressing the exceptional importance of the internet for individual freedoms and the fundamental mistrust of any form of authority of laws and states:

> Governments of the Corporate World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather.

The same idea can be found in Steven Levy’s book *Hackers: Heroes of the Computer Revolution* (1984):

> All information should be free. Mistrust authority–promote decentralization. Hackers should be judged by their hacking. You can create art and beauty on a computer. Computers can change your life for the better (Levy, 1994:33-36).

Digital communication technologies were associated with ideas of egalitarianism and community, and a fundamental trust in the positive effect of technological development (van Dijck, 2013).

> We will create a civilization of the Mind in Cyberspace. May it be more humane and fair than the world your governments have made before (Barlow, 1996).

Digital communication technologies were associated with ideas of egalitarianism and community and a fundamental trust in the positive effect of technological development (van Dijck, 2013).

As several academics have stressed, this narrative was mainly diffused across programmers and technologists on the West Coast (especially around Silicon Valley) who took part in the development of the infrastructure of digital communications as we know it today (Mueller, 2002; Malcomson, 2016; Mezei and Vertes-Olteanu, 2020; De Filippi et al., 2020). In this spirit of community and human progress, at the beginning of the 1990s, Tim Berners-Lee shared the Hypertext Transfer Protocol (HTTP) technology and opened the World Wide Web (or the internet) to become a mass communication technology. The change in people's lives was of a magnitude never before experienced, and academic literature was attracted to the positive effects and opportunities generated for communication by this new medium. It was:

> [A] [L]awless frontier immune to regulation and control by governments. Libertarian by nature, open in its architecture, the Internet was seen by many as encouraging democracy, freedom, and liberty around the world (Deibert, 2003:504).

The internet's decentralised architecture fascinated advocates of freedom of expression and global democracy. The new horizontal, many-to-many network of communication was considered revolutionary in the creation of a global public sphere, overcoming dispersion and individualization (Castells, 1996, 2008, 2013; Dutton, 2009; Diamond, 2010; Lotan et al., 2011; Diamond and Plattner, 2012; Shirky, 2008; Gerbaudo, 2015; Maireder and Ausserhofer, 2014, Margetts et al., 2015). Deliberative public sphere scholars saw in the internet a structure in which debate on common problems could develop between citizens at a global level and create a critically informed public opinion capable of guiding decision-making processes (Bohman, 2004; Froomkin, 2004; Dahlberg, 2001).

Manuel Castells was among the first scholars to focus on the potential for digital spaces to create a global 'networked public sphere' where public opinion can easily be gathered and communicated to decision-makers, collectively constructing political meaning (Castells, 1996, 2008, 2013). Drawing from his ideas, many other scholars adopted the concept of the internet as the global public sphere. For instance, Dutton (2009) saw in the global communication system created by digital ICTs and SM the foundation of a new 'Fifth Estate' enabling 'networked individuals' to challenge the boundaries of existing institutions, and increasing the accountability of politicians, the press and experts and of power and influence in general (Dutton, 2009:2).

However, after the global opening of the network in the mid-1990s, the worldwide spread of the technology made it clear that the 'Internet was outgrowing its research and education roots', becoming an open platform for global data networking (Mueller, 2002:2). The internet's global expansion corresponded with an increase in problematic aspects of content accessible worldwide. Other national governments began to express uneasiness with unilateral US control of such an essential part of the global communication infrastructure and requested to be part of the regulation (Mueller, 2002). In this changing environment, two US laws played a fundamental role in paving the way for today's internet governance of speech.

The first and most significant regulation aimed at regulating expression on the internet dates back to 1996. The Communications Decency Act (CDA) was created to regulate pedo-pornographic content, introducing restrictions on communication online. It also included a section (now called

Section 230 CDA) that became fundamental for developing the governance ecosystem as we know it today. This section established that companies owning infrastructure for digital communication (intermediaries providing access to the internet or other network services) could not be considered responsible for crimes occurring on their infrastructure. It also stated that intermediaries could police content without losing their 'safe harbour' status and without the need to meet a standard of effective policing. Over the years, most parts of the CDA have been declared unconstitutional (Mueller, 2015; Gillespie, 2018), but Section 230 is still in place, and it plays a crucial role in the regulatory discourse over SM responsibility for user-produced content. Since then, intermediaries providing access to the internet or other network services have not been judged as 'publishers' and cannot be held legally liable for their users' speech. As Gillespie (2018) recalls, even though Section 230 was not enacted with SM platforms in mind, it created one of the broadest immunity regimes for platforms, with the complication that platforms as intermediaries are active not only in the US but also in countries where legislation entails stricter rules for intermediaries, such as in the EU.

The US government passed a second crucial act for the current governance of freedom of speech online in 1997[1]: the Framework for Global Electronic Commerce, aimed at solving the tension created by emerging requests for control of the communication network from other states. In that act, the Clinton administration set a non-regulatory, market-oriented approach to the internet, opening the incorporation of the web to private developers such as Google, AOL and Amazon. The act was a crucial moment in history that marked the beginning of today's internet ecosystem, also known as the commercial internet (Radu and Chenou, 2015).

The creation of the commercial internet opened the way to the development of privately owned online sites and services. The ensuing financial success rapidly led to a bubble in the telecommunication market that burst at the beginning of the 2000s. A different version of the web, called Web 2.0 (or social web), ensued and opened the market to new online sites and services

[1] Also, a third act could be included; in 1998 the US government published a policy document officially titled 'Management of Internet Names and Addresses', asking international support for a new, not-for-profit corporation formed by private sector Internet stakeholders (Mueller, 2002:4) since the intellectual, commercial, and political climate surrounding the internet disliked any form of state action or intergovernmental organisations. 'The ultimate result, for better or worse, was the Internet Corporation for Assigned Names and Numbers' (ICANN) (Mueller, 2002:4). The organisation created the first multi-stakeholder form of representation at the international level, and deeply influenced the development of subsequent initiatives aimed at governing internet-related issues.

such as blogs, wikis, web applications and social networking/media. The product known as 'social media' developed incredibly quickly in different shapes, depending on the aims and main functions available. For example, in 2002, Friendster was created, followed in 2003 by LinkedIn and Myspace, and Facebook in 2004. Twitter followed in 2006, and Google+ in 2007. Others such as Instagram, Snapchat, Foursquare, Tinder, Grindr, and TikTok were created in only a few more years. These sites differed from the previous generation because of their architecture based on user-generated content, which allows interactivity among users as well as with the content, together with network connections within users (De Nardis and Hackl, 2015:762). The real innovation, though, was in the economic system intertwined with the architecture. Social media not only host, organise, and circulate public expression without producing or commissioning (the bulk of) that content, but also beneath that circulation of information – and differently from other media – they hold an infrastructure for processing data for customer service, advertising, and profit (Gillespie, 2018), which represents the core of their business model (see Helmond, 2015).

Just as the internet had previously been associated with lawless frontiers of free expression, Web 2.0 was also initially described using positive metaphors and images about freedom of expression. In particular, Web 2.0 was associated with the idea of 'user centred' and 'collaborative' systems. Thanks to the potential for the user to 'talk back' and send messages instantly, the design of Web 2.0 was described as *participatory* and an extension of democracy online. Exploiting the metaphor of platform in framing the technology as open to all, SM presented themselves as unprecedented tools for empowerment and online self-communication (van Dijck, 2013), facilitators of public expression, and impartial hosts protecting freedom of speech and information (Gregory, 2016, 2018; Casilli, 2017; Gillespie, 2018).

In academia, scholars (Diamond, 2010; Tucker et al., 2017) were fascinated by SM's potential for widening the public sphere and creating a more pluralistic arena for the creation and exchange of information. Moreover, when used as instruments for transparency and accountability, for documenting and deterring abuses of human rights and democratic procedures, SM could even become 'liberation technologies', as in the case of the uprisings that took place during the Arab Spring (2009–2011) and the so-called 'Twitter revolutions' (Diamond, 2010:71).

The role of the internet and SM in creating a public sphere for free speech has been recognised as fundamental at the level of international political organizations too (De Nardis and Hackl, 2015; United Nations and Kaye, 2016, 2018). However, the public space that has been developing since is quite different from the ideal of the public sphere envisaged by the idealists of deliberative democracy.

2.2.2 Pessimistic narratives and regulatory backlash

Following two decades of 'techno-optimism', we are now experiencing a much more pessimistic approach towards these technologies and the companies that own them. As a result, a different narrative has arisen, challenging the idea that digital communication and SM are beneficial for the 'public sphere'. This narrative focuses on the challenges to free expression and the public sphere presented by both state and private surveillance and control and the degraded nature of the public discourse that takes place on SM services.

As far as state surveillance is concerned, scholars have stressed that since 11 September 2001, terrorism and national interest have been used to justify surveillance and restrictions on citizens' freedom of expression by states ranging from the most democratic to the least democratic. Moreover, as a reaction to the terrorist attacks in Europe in 2015, Western states have increasingly put pressure on SM corporations to police content and users (in France see McGoogan, 2015; and in the UK see Carey, 2016). Similarly, the liberation role played by SM on the occasion of the uprisings that took place during the Arab Spring (2009–2011) and the so-called 'Twitter revolutions' has been progressively re-evaluated critically. Gladwell (2010) and Morozov (2009, 2011) disputed the idea that SM is beneficial for democracy and free expression, focusing on the use of their infrastructures made by states for state surveillance and repression, as well as their weakening effect on global activist networks with the rise of 'slacktivism' with no political or social impact (Morozov, 2010).

Scholars have also developed concerns about the nature of the sphere of communication created by the internet, stressing how it is a private rather than public sphere. A darker narrative (Pasquale, 2016a, 2016b) has begun to be attached to platforms, leading to requests for regulation of the corporations owning the platforms. Concerns have started to emerge about the power of private companies that own SM platforms, which manage the data of billions of people and exert an increasingly strong influence over public life. A few corporations have been particularly successful and managed to occupy strategic positions in the digital market, creating a situation of oligopoly and concentration of power, such as the platforms owned by Google (Alphabet), Amazon, Facebook, Apple, and Microsoft (GAFAM) (Van Dijck et al., 2019). The size of the corporations behind SM platforms is seen as a threat to the idea of independent public space (De Nardis and Hackl, 2015), contributing to the idea of a fragmented public sphere (Papacharissi, 2002; Malcomson, 2016). Scholars stress that SM are capitalistic enterprises and not a neutral public space for discussion. They have their own agenda, values and private norms, which influence their policies on content in the case of mass social events. Fuchs (2014) underlined how companies like Twitter could not be considered a 'public sphere' as they can limit the free development of social interaction, censoring or blocking users in line with regulations or the interests of the platform.

Among the most discussed elements in the negative narrative of platform technology is the toxic nature of communication created by SM. Scholars in the last 15 years have underlined how SM have offered global visibility to instances of racism, misogyny and abuse and contributed to the spread of dangerous and fake information, with massive social and political repercussions.
Initially, the technological aspects of 'collapsed audiences' and virality were called into question. SM platforms collapse different audiences into a unified context, with no distinction between public and private or different spheres of social relations (Boyd et al., 2007; van Dijck, 2013; Schmidt, 2014). Any casual speech in SM platforms 'is turned into formalised inscriptions, which, once embedded in the larger economy of wider publics, take on a different value' and 'can have far-reaching and long-lasting effects' (van Dijck, 2013:7).
Social media are specifically designed so that any form of content, even the most abusive, can potentially reach a critical mass and have consequences for single individuals, groups of people (e.g., encouragement of misogyny or racism) and states (e.g., online recruitment of terrorists). Several authors have studied how SM have contributed to the diffusion of negative and harmful content created by users (Cammaerts, 2009; Özarslan, 2014; Awan, 2014, 2016; Daniels, 2008, 2013) and the possible negative offline implications of hyperconnectivity (Webb et al., 2015). Daniels (2008, 2013) highlighted unintended consequences of digital media for racism, civil rights, and hate speech. Freiburger and Crane (2008) have applied the theory of online social learning to

study the diffusion of extremism online, showing how terrorists can use ICTs to spread ideology and propaganda. Awan (2014) has focused on how Islamophobia was expressed on Twitter after the Woolwich attack in 2013 and on Facebook pages (2016).

Similarly, Özarslan (2014) found that Twitter had been used to diffuse hateful content towards Kurds in the aftermath of the 2011 earthquake in Turkey. Cammaerts (2009) and Bostdorff (2004) stress how extremism can be facilitated by the very same liberal principles of free speech built into the design of the internet. In particular, Bostdorff (2004) stressed how anonymity and separation facilitate aggressive speech online (as in the case of the Ku Klux Klan) while disavowing responsibility for the consequences of their messages. Similarly, Meddaugh and Kay (2009) and Perry and Olson (2009) stressed how the internet contributes to spreading racist and white supremacist ideology.

The link between content produced on SM and episodes of disinformation and fake news has been demonstrated on several occasions, including the manipulation of information in respect of the Brexit referendum and the 2016 US elections and, more recently, in January 2021 the storming of Capitol Hill by Donald Trump's supporters, convinced that the elections were invalid, and the consequent SM platforms’ decision to block Trump's SM accounts (Zubiaga et al. 2016; Procter et al. 2019; Mezei and Verteș-Olteanu, 2020).

Moreover, more recently, studies have moved on to investigate SM influence on the norms around content or speech (Klonick, 2017) and how SM companies moderate content, taking important decisions about the users and types of speech that they permit (Roberts, 2018). The role that platforms played during the Arab Spring, the #MeToo and #BlackLivesMatter campaigns, and Twitter and Facebook's choices with regard to Donald Trump and fake news (Heldt, 2018) are examples of the ambiguous role that platforms play in the management of content (Mezei and Vertes-Olteanu, 2020:7). Scholarly discussions about legal liability for SM have surged (Napoli and Caplan, 2017), especially in relation to the role of algorithms in selecting and presenting information (Bucher, 2018).

2.2.3 Regulatory policies governing mass media before the Internet

Regulatory policy development for media companies is not a new phenomenon, and it has been studied, especially in antitrust and competition law. For example, in 2010 Tim Wu identified similar paths in other information–communication industry giants (such as AT&T). According to Wu, in 'the Cycle', companies in their emergent phases seemed limitless and open; however, the more they consolidated, the more they became objects of state regulation. Also, Cusumano et al. (2021) put the current situation of increased regulation of speech on SM in a historical perspective, stressing how, in the past, companies in the business of producing movies, video games, and television shows and commercials have all faced issues around the appropriateness of content in a way that resembles today's SM platforms (Cusumano et al., 2021).

Cusumano et al. (2021) present examples of pre-internet communication industries that 'successfully' avoided state regulation by putting in place self-regulation for their content, as for example, the US National Association of Radio and Television Broadcasters and the ban on cigarette advertisement. The authors want to show that increased state regulation is not a new phenomenon; instead, it occurs every time large companies are perceived as threats to national security or public health. They justify the current situation of increased 'state' pressure by pointing out that SM companies have 'failed' to set up self-regulation initiatives to deal with the 'threats' and consequently have brought national and supranational rulemaking upon themselves. Cusumano et al. (2021) highlight a major issue with these companies, recognizing that companies tend not to pursue self-regulation when the perceived short-term costs are high. The authors stress that this is

> 'a problem for SM platforms in particular because fake news stories and damaging videos, and reports of spectacular conspiracy theories, are more frequently read and forwarded than real news items, and they generate more activity, stronger network effects, and more advertising revenues (Aral, 2020)' (in Cusumano et al., 2021:20).

Radu (2019) explains that the commercialization of the internet in the 1990s and the associated liberal narratives and ideology (which brought forward Section 230 in the CDA) contributed towards creating a system where private companies yield a power never seen before. As a result, SM companies have grown within a regulatory system considered strictly related to the market and contracts. However, the development of internet applications and an economy based on data has proved that governance of the internet and its related aspects directly impinge upon political, societal issues and human rights.

Between self-regulation and heavy-handed state intervention, normative discussions about internet policy have emerged at different levels. Radu et al. (2021), Padovani and Santaniello (2021) and Palladino (2021) present these discussions as an alternative idea of governance, labelled 'digital constitutionalism', in competition with one more focused on state sovereignty and private companies' self-regulation. The expansion of the internet has given rise to concerns that have more political salience, such as human rights, algorithms, big data, privacy, and surveillance. It has opened the door to engagement from civil society seeking to articulate political rights, governance norms, and limitations on the exercise of power on the internet (Palladino, 2021).

Scholars have increasingly started to focus on investigating regulation of speech online as an object of research, exploring the features of the regulation of content. In the following paragraphs, I will provide an overview of the main features of the current governance of free speech that emerge from the literature.

2.3 Governance of speech: overview

When focusing on the ways in which speech is regulated online, scholars recognise a complex ecosystem. Scholars have long stressed that societies organised around ICT have pluricentric regimes of power and different sites and tools of ordering (Barry, 2001). The traditional decision-making authority, the nation-state, is no longer understood to be the 'control centre of society' (Mayntz, 2003:29), and is rather one actor among others in an increasingly heterogeneous system. In the governance of freedom of expression on SM platforms, new structures, processes and actors replace the traditional concept of command and control (Hofman, 2016; Hofman et al., 2017). For this reason, the idea of governance – i.e., the process of coordinating multiple actors in order to work towards a shared goal (Rhodes, 1996) – fits the digital environment better than the idea of rules and regulations. Governance focuses on how forms of 'order' are created in a context in which there is no steering, governing body, and on how coordination has an emergent quality rather than being imposed from above (Levi-Faur, 2012).

Several studies focusing on the nature of governance have developed in the field of internet governance in recent years (Epstein, 2013; Flyverbom, 2011; Hofmann, 2016; Hofmann et al., 2017). They have contributed to stabilizing the idea that online governance is a performative and relational form of order, involving social and technological elements emerging into self-organised networks contingently and continuously (De Nardis, 2014; Epstein, 2013; Montenegro Mayer and Bulgakov, 2014; Musiani, 2015; Pohle, 2016b, 2016c). Moreover, the literature shows that forms of order are not achieved through a single formal or institutionalised planning order – rather they 'happen' as the product of a 'reflexive action' of coordination of the different actors involved, as a response to critical moments (Hofmann et al., 2017). Napoli and Caplan (2017) and Ananny and Gillespie (2017) have observed how critical moments, or 'public shocks', have been fundamental in driving the regulation of these platforms. Public shocks have been defined as:

> public moments that interrupt the functioning and governance of these ostensibly private platforms, by suddenly highlighting a platform's infrastructural qualities and call it to account for its public implications. These shocks sometimes give rise to a cycle of public indignation and regulatory pushback that produces critical – but often unsatisfying and insufficient – exceptions made by the platform (Ananny and Gillespie, 2017: 2–3).

When it comes to platform governance, and specifically the governance of admissible speech, scholars tend to agree on the following main features defining the ecosystem:

1) The control of expression does not belong to a single actor;
2) Governance takes place in a plurality of governance settings through a variety of regulatory solutions with different degrees of formality and no hierarchical organization.

The following paragraphs will provide an overview of the range of actors contributing to the governance debate and how they engage with the processes and influence governance.

2.3.1 Plurality of actors

Scholars of internet governance agree that governance of speech online does not belong to a single actor and does not follow the traditional dyadic model of states regulating the speech of private parties (Balkin, 2017). Wagner (2016), Balkin (2009, 2017), Helberger et al. (2018), Gillespie (2018) as well as Gorwa (2019a, 2019b, 2020) describe in different ways a pluralist model with responsibility divided between *three* main groups: states or governments, civil society and private companies.

States or governments include individual governments and supranational forms of government (e.g., the European Union, the United Nations). They produce statutory regulations (i.e., laws issued by parliaments) which are geographically bounded but with powers to regulate speech directly, without direct collaboration from other stakeholders (Gorwa, 2019a, 2019b). They can also coerce or co-opt companies into regulation (Balkin, 2017). These regulatory forms can be defined as co-regulation and represent a self-regulatory system supported by legislation (Tambini et al., 2008; Article 19, 2018).

Civil society includes organised groups such as advocacy groups, international NGOs, scholars, journalists, activists, hackers and hacktivists, and less institutionalised networks of individuals and users (Bennet and Segerberg, 2012). Scholarship highlights the variety within this group, which in turn corresponds to different ways of influencing governance. For example, some studies focus on how organised civil society influences governance by playing an accountability function with regard to state and companies (Balkin, 2017; Gorwa 2019a, 2019b). Others have highlighted how civil society can ‘change the regulatory narrative’ by introducing new norms within larger institutional settings (Milan and ten Oever, 2017). While others, concentrating on grassroots movements in internet governance, show how other types of action focus on developing alternative infrastructure and technical ‘bypasses’ around rules and regulations (Hintz and Milan, 2009). Finally, other studies have focused on users in different online communities and their self-regulation through digital interactions (e.g., counter-speech) or using the reporting tools and protocols provided by the platforms (Zubiaga et al., 2016).

Among the traditionally explored actors involved in governance online are also platform companies, data brokers, advertisers, and developers. Scholars have investigated the extent of their regulatory power and the features of private ordering (Hintz, 2016; Hustzi-Orban, 2018). They have increasingly stressed how companies are acting as private governors (Klonick, 2017), creating and enforcing rules and norms on the communities they govern (Balkin, 2017; Fuchs, 2014). As explained in the overview of the narratives associated with technology, SM platforms have been self-regulating for several years, using their internal policies, such as statutes, codes of conduct, and terms of services as an internal regulatory system (Tambini et al., 2008). Central to SM self-regulation have been users, who are the primary elements that start a moderation process through their reporting action.

As explained above, governance of speech involves the actors described above, in a multitude of settings with varying degrees of formality 'where a hierarchy is impossible to discern' (Dingwerth and Pattberg, 2006:192). Similarly, the tools and means that different actors have to influence the 'shape' of governance can vary greatly, ranging from more institutionalised and binding regulations (usually in the hands of states) to less binding initiatives taken at the level of international organization (soft law), to private companies' internal self-regulation as well as influence from civil society advocacy and challenges to the system (infrastructures). Such a variety of settings opens up the opportunities and means by which actors can engage with the governance of speech and influence policy. In the coming paragraphs, I will provide an overview of the main settings and the regulatory tools and other means used by actors to influence the shape of the governance of speech, starting with those created by single actors and moving on to those that involve a combination of actors.

2.3.2 Plurality of settings and means used to influence policy used by single actors

According to Gorwa (2019a, 2019b, 2020) regulatory initiatives tend to involve just one type of actor or dyad, and only rarely involve decision-making distributed across all the actors (in a model called multi-stakeholders). While Wagner (2016) emphasises how the vast majority of content regulation is 'based on private norms and practices' and 'the Internet is governed by numerous informal power relations and agreements, predominantly between private actors, or self-regulatory bodies or quasi-public NGOs' (Wagner, 2016:122), which have the technical means to enforce their decisions. In this system, regulations are initiatives of 'communities of practice', regulating speech according to their own logic of appropriateness (Wagner, 2016).

With respect to the means to influence governance, from the larger field of internet governance, Padovani and Santaniello (2018) recall several examples in the scholarship that acknowledge the role of discourse in shaping political reality online. Similarly, drawing on social constructivism in international relations, Radu et al. (2021) and other scholars stress how, in such a complex environment, the role of norms and 'norm entrepreneurs' is fundamental in steering governance and policies online. Radu et al. (2021:2) define 'normfare' as

> 'the assiduous development of norms of very different character (public and private, formal and informal, technically mediated and directly implemented) by different actors (platforms, standard-setters, states) as an answer to the wide range of challenges facing internet governance'.

In this perspective, the means and the opportunities for actors to influence governance move beyond the 'restricted' settings of national legislation or international treaties to include regulations developed from discursive norm creation, as for instance from non-binding instruments such as statements and declarations, such as guiding principles, charters, codes of conduct or operative guidance tools, such as recommendations, guidelines. (Radu, 2019).

States and international organizations

When considering individual actors, governments are key regulators. State-based content regulation still provides the baseline for building other governance initiatives (Gorwa, 2020). States tend to regulate based on national security and the protection of citizens. For instance, since 2015, as a reaction to the wave of terrorist attacks in France and Belgium, France has introduced a 'state of emergency law' that augmented controls on SM for security reasons (McGoogan, 2015). In November 2015, the UK government introduced the Investigatory Powers Bill to define more clearly surveillance powers and reform oversight. The bill included obligations for communications companies to collect and record users and allow investigators to gain access to data stored in personal devices (Burgess, 2016). In 2017, as a reaction against cases of abusive speech against refugees online, the German *Netzwerkdurchsetzungsgesetz* (NetzDG) was approved. This law established fines for social networks failing to remove illegal content (according to German law) within 24 hours. A similar approach was taken in 2019 by the UK government. In the *Online Harms White Paper*, the UK government established for the first time that companies have a 'duty of care and responsibility for the safety of users' (Woods, 2019).

At the supranational level, in the EU, the most relevant regulatory frameworks that apply to SM companies derive from consumer law, competition law, antitrust law, and privacy law, focusing on competition and consumer welfare. Examples are the E-Commerce Directive, which has established baseline provisions for intermediary liability, the EU Audiovisual Media Service Directive (AVMSD), the EU Copyright Directive and the General Data Protection Directive (GDPR).

Other international organizations do not retain the same power as sovereign states; however, they can work similarly to civil society by issuing soft law instruments. These non-binding instruments can change online governance by instituting norms that end up transposed to policymaking (for instance, as in the case of the Council of Europe) (Marzuki, 2019). Indeed, for a long time, authors studying internet governance have stressed how in the governance of the internet and its related aspects, the presence of states and other formal institutions has been inflated (van Eeten and Mueller, 2012) and that:

> "in most areas, governance of the Internet takes place under very different conditions: low formalization, heterogeneous organizational forms and technological architectures, large numbers of actors and massively distributed authority and decision-making power" (van Eeten and Mueller, 2012:730).

Civil society

While lacking binding instruments with which to 'impose' regulation, civil society does have other means to influence the governance of speech. Hintz (2016) provides an overview of four main ways in which civil society can influence media policies. Some strategies are based on actions 'inside' the policy-making system (i.e., in conversation with policymakers and private companies), while others include actions performed 'outside' or in opposition to the 'institutional system' (i.e., in the case of protest); other actions, according to Hintz, go 'beyond' the policy system (i.e., in the case of grassroots organizations building infrastructural ‘alternatives’) and other are policy hacking actions (i.e., actions that are originated outside the policy-making system to reach changes at the level of legislation).

Within the ‘inside’ actions, scholars have studied how civil society has taken part in the creation of norms and regulatory discourses or has acted as a state and corporate accountability watchdog (Gorwa, 2019a, 2019b) as well as working at the level of infrastructure (Hintz and Milan, 2009; Milan and ten Oever, 2017). In particular, free speech NGOs and advocacies have engaged with state and private companies on several occasions, expressing concern over initiatives such as the NetzDG in Germany, where private actors are called to enforce the law and other national legal provisions under short deadlines and the threat of hefty fines (EDRi 2017, Article 19). In terms of other cases of ‘dialogue’ with other actors, Milan and ten Oever (2017) studied how human rights advocates operated as a critical community advancing discursive tactics and creating socio-technical imaginaries within one of the technical bodies, the NCUC.

Examples of actions ‘outside’ include protests, and social mobilisation, as in the case of the protest started against the Stop Online Piracy Act (SOPA). While actions ‘beyond’ include the development of technological alternatives, such as Tor for encryption (Hintz, 2016). In the case of actions beyond, rather than participating in policy debate with governments and the corporate sector, movements value developing alternative infrastructure and technical ‘bypasses’ around rules and regulations (Hintz and Milan 2009; Hintz, 2016).

Examples of policy hacking actions include creating principles that can craft policies based on human rights to be endorsed by states and private companies. Examples are the ‘Manila Principles on Intermediary Liability’, presented by NGOs to address states’ requirements of SM companies, stressing the need for transparency and due process and at the same time preserving non-liability of companies for content produced by third parties. On the other hand, in 2018, the ‘Santa Clara Principles for Content Moderation’ (SCPs) provided recommendations for private companies. They stressed the need to introduce changes in their internal moderation processes, making them more transparent and inclusive of the right to appeal and the need for notice to be given to users whose content is undergoing a moderation process (Gorwa, 2020).

Standard-setting attempts have continued more recently. In 2018, the NGO Global Partners Digital published the white paper ‘A Rights Respecting Model of Online Content Regulation by Platforms’ proposing a model of online content regulation by platforms in line with international human rights law and standards (Bradley and Wingfield, 2018). In 2019, Article 19 launched the Social Media Council (SMC) initiative as an open and voluntary accountability system to apply human rights-based principles to the review of content moderation decisions made by SM platforms (Article 19, 2019). The initiative was based on the UN Special Rapporteur request to develop an industry-wide accountability mechanism (UN and Kaye, 2018). Gorwa (2019a, 2019b) stresses how civil society is mainly left to itself in these initiatives, with very low involvement of states and industry representatives. However, these initiatives have value from the point of view of studies on norm creation and diffusion in international settings (Radu et al., 2021). In her historical overview of the development of regulatory instruments and institutions in internet governance, Radu (2019) stresses the role of discursive agreements and non-binding regulation in the development of the current internet governance ecosystem. Similarly, Milan and ten Oever (2017) showed how the inclusion of advocacy members within technical communities brought new ways of framing issues and created emerging 'ordering narratives' from the bottom up. Together with other scholars, they show the importance of constructing narratives and shared discourse as a preliminary phase of norm creation (Palladino, 2021; Padovani and Santaniello, 2018).

Private companies

If states set the baseline for regulation, SM companies are the only entities with enforcement power over their infrastructure. For a long time, platforms have managed speech using self-regulation, i.e., their internal rules (e.g., community standards, or terms of service), to set the limits of acceptable speech. According to Gorwa (2019a), this strategy allowed them to 'improve their bargaining position with other actors, to win public relations points, and to evade more costly regulation' (p.9).

However, as already mentioned above, scholars have stressed that private companies have been acquiring law-making and law enforcement powers, playing an increasing role in enforcing regulations, setting new rules and providing significant resources for surveillance and information control. The main concern of this private ordering is that companies are not subject to constitutional constraints and procedures, which raises relevant concerns about their legitimacy and accountability (DeNardis, 2014; DeNardis and Musiani, 2016; Hintz, 2016:119).

Content regulation on their infrastructures is referred to as ‘moderation’, and it includes the screening, evaluation, categorization, approval or removal of online content according to the platform's internal rules and relevant communication policies. Each platform has its own content moderation policies, and they differ by sise, reach, language, technical design, genre, corporate ethos, business model, and stated purpose (Gillespie et al., 2020:5). Internal policies can vary significantly according to the different companies, and some companies are notoriously more liberal (e.g., Reddit or Gab) than others (e.g., Facebook). In addition, platforms' content moderation systems differ in terms of the presence of institutional mechanisms of adjudication, enforcement and the appeal of decisions (Belli et al., 2017; Gillespie et al., 2020:2). All forms of content moderation on platforms are enacted through a mixture of human screening and automatic recognition of speech through artificial intelligence (AI) and machine learning.

As the first step in content moderation, all platforms rely on their community of users to report offensive or abusive or fake content (i.e., to ‘flag’). Human moderators then review the content and decide whether the specific content represents acceptable speech according to the platforms' internal rules and definitions (for instance, using a list of protected categories) (Ulmann and Tomalin, 2020). However, this system has been criticised on several occasions either for its failure to prevent harm or hateful speech or for creating an opaque system in which decisions to remove content are obscure (Crawford and Gillespie, 2016; Ulmann and Tomalin, 2020; Edwards et al., 2021). In order to intervene before the occurrence of harm, SM platforms have also been developing pre-emptive forms of content moderation based on automatic recognition. However, the use of AI in content moderation is a debated issue. Civil society and international organizations have expressed concern about automated content moderation and algorithmic decision-making processes. These practices offer very little transparency and virtually no remedy to individual users when their content is taken down or demoted (Article 19 2018, UN; Kaye, 2018). Scholars also stress how automated moderation worsens the already lacking communication with those users who are the objects of speech restriction (Suzor et al., 2019). Recent reports have also expressed concern about the failure of platforms to self-regulate with algorithms, as in the case of SM self-regulation initiatives produced to curb the spread of fake news (Hoffman et al., 2019). Scholars have focused on the human costs hidden in algorithmic content moderation and have highlighted the presence of biases as well as well-being and mental health costs for moderators (Casilli, 2017; Dencik et al., 2018a, 2018b; Carmi, 2019; Caplan, 2019; Roberts, 2019). The free speech costs of automated moderation are a significant source of concern, and research has found that in a trade-off between the protection of vulnerable groups and free speech, experts tend to express preferences for solutions that maintain free speech (Edwards et al., 2021). Some scholars are trying to develop automated systems that are able to proactively intervene in cases of harmful or abusive speech, minimizing the costs of over-policing free speech. Ulmann and Tomalin (2020) have worked on the combination of automated content recognition leading to a quarantine period for offensive posts rather than automatic deletion (Ulmann and Tomalin, 2020). However, other studies (Copland, 2020) have found evidence that similar changes in the platform's policies (e.g., Reddit) result in a migration of users towards less regulated platforms. Other scholars have stressed the potential of counter-speech, and online communities' self-regulation in situations of abusive speech or fake news, as the best way to develop pre-emptive regulatory solutions without the censorship burden of introducing a filter or deletion of content (Bartlett and Reinolds, 2015; Huey, 2015; Housley et al., 2018; Procter et al., 2013b; Procter et al., 2019).

2.3.3 Plurality of settings – mixed-actors initiatives

Together with single actor initiatives, governance of speech has been enacted by creating several regulatory initiatives involving more than one actor, either as public body–private company frameworks of co-regulation or as forms of self-regulation based on the collaboration between private sector companies and civil society. A particular form of mixed actor regulatory initiative is represented by multistakeholder forums, which involve representatives of all actors. The level of institutionalization and binding nature vary a lot across these initiatives.

Public–private initiatives

In 2014, the European Commission started the 'EU Internet Forum', involving the EU and private companies, which led in May 2016 to the launch of the Code of Conduct on countering illegal online hate speech. The code of conduct includes a series of commitments to fight the spread of illegal hate speech online in Europe, including the removal of illegal hate speech within 24 hours (EU Commission, 2016).

Civil society associations such as the European Digital Rights initiative (EDRi) and Access Now criticised the agreement and expressed apprehension about the chilling effect on free speech of having technology companies policing illegal content. They also condemned the systematic exclusion of civil society organizations from the dialogue and announced their withdrawal from future discussions (EDRi, 2016). Similar criticisms came on 1 June 2016, from the Council of Europe (CoE) Secretary-General Thorbjørn Jagland, who urged European governments to ensure that their legal frameworks and procedures in the area of blocking, filtering and removing internet content are transparent and incorporate adequate safeguards for freedom of expression and access to information in compliance with Article 10 of the European Convention on Human Rights (CoE, 2016). June 2016 also saw the publication of the report of the Special Rapporteur on freedom of expression, David Kaye, to the United Nations Human Rights Council. In that report, and on several subsequent occasions, the Special Rapporteur expressed warnings concerning the fact that online expression is increasingly eroded by new forms of state regulation and mediated through private networks and platforms created, maintained and operated by companies in the ICT sector (United Nations and Kaye, 2016, 2017, 2018a, 2018b).

Private companies–civil society

In response to the request for increased transparency, companies have also developed *ad hoc* self-regulatory instruments with the involvement of civil society. In 2016, Twitter created the 'Trust and Safety Council', which involves over fifty groups, including advocates, academics, researchers, grassroots advocacy organizations, and community groups. The members are activists campaigning to prevent abuse, harassment, bullying and suicide, and safeguard mental health (Twitter, 2016). However, scholars see some limitations in the initiative, which lacks transparency and does not have any specific governance responsibility (Gorwa, 2019a, 2019b). As stated on the Twitter page: 'membership is voluntary and does not imply endorsement of any decisions we make. Members also don't speak on Twitter's behalf. A small number of organizations on the Council requested not to be named' (Twitter, 2019).

Similarly, Facebook has adopted some initiatives in cooperation with NGOs, for example the Online Civil Courage Initiative launched in 2016 to address the issue of hate speech (Facebook, 2020). In response to increasing requests for transparency, the Facebook Oversight Body was appointed in May 2020 to provide oversight or input into Facebook's content policy process. The outcome of public consultations and workshops with experts, institutions, and people worldwide that began in 2018, it comprises 20 experts, including lawyers, academics, journalists and ex-politicians (Facebook, 2020). It was immediately criticised for lack of independence and, in response, a ‘Real Facebook Oversight Board’ not sponsored by Facebook was announced on 25 September 2020, claiming to have better oversight over Facebook (among the members is Shoshana Zuboff, author of *The Age of Surveillance Capitalism* (2019)) (Butcher 2020). Scholars find that similar initiatives are still under-studied and should be comprehensively examined by future research (Gorwa, 2019a, 2020).

Multi-stakeholder initiatives

Multi-stakeholder initiatives represent other forms of ‘mixed’ arrangements. These initiatives involve actors from at least two groups: states, non-governmental organizations (including civil society, researchers, and other parties), companies, and international organizations. This form of governance was created with the foundation of ICANN in 1998, and reproduced at the 2003 World Summit on the Information Society (WSIS) and the subsequent United Nations Working Group on Internet Governance (WGIG), which proposed the creation of an Internet Governance Forum (IGF). The IGF, however, has been criticised as not being representative of the whole internet governance (van Eeten and Mueller, 2012) and for a lack of inclusivity, especially concerning the involvement of the Global South (Mc Laughlin and Pickard, 2005; Radu, 2019).

As far as the governance of speech is concerned, the Global Network Initiative (GNI) was launched in 2008 and has developed principles and guidelines to help governments, companies and civil society to respond to demands from governments around the world that could restrict users’ freedom of speech and privacy (GNI, 2015). However, Gorwa (2019a, 2019b) presents criticism of this initiative too, as he states that it is an insular organization, not known to the public, where participants sign non-disclosure clauses promising not to disclose classified information raised at board sessions (similar to what happens in Twitter's Trust and Safety Council).

2.4 Main issues from the literature

Even in this summary, it is possible to get a sense of the wide range of players and settings involved in the governance of online expression. In this complex system of different kinds of regulatory means, scholarship has identified issues that are currently posing challenges to the online expression governance system. One of the issues that emerges from the literature is that even though governance of speech is the result of the initiative of different actors, they do not share the same powers to orientate the course of regulation.

2.4.1 Unbalanced distribution of power

Although governance of speech does not happen in any one of these specific sites, some authors still believe that power relations are not ‘equal’. Most scholars accept that civil society actors are underrepresented in the ‘governance triangle’ involving corporations, governments and civil society (Gorwa, 2019a), and that civil society merely helps to legitimise the mechanism for other, more influential actors. Others contend that the current state of ‘platform governance’ is increasingly shifting away from companies' self-regulation and toward greater government participation (Helberger et al., 2018). Others, such as Hintz (2016), Balkin (2017) and Zuboff (2019), are concerned about the increased power of platforms to govern speech and consider that platforms are seizing governance by exploiting neo-liberal loopholes in governing institutions (Zuboff, 2019). The increasing prevalence of forms of private ordering, together with state demands for more regulation of content, signal a move toward outsourcing public responsibilities to private actors and, as a result, the increased privatization of policies (Hintz, 2016). SM companies are now recognised as political institutions that make major political decisions when constructing the global governance framework of free speech (Gillespie, 2018). Governance of speech is at the crossroads of the political effects of their platforms (governance *by* platforms) (Gillespie, 2018) and the local, national, and supranational mechanisms of governance that constrain SM platforms (governance *of* platforms) (Gillespie, 2018).

Kate Klonick (2017) famously defined platforms as the ‘new governors’ due to their power to enforce policies on their technological interfaces. She explained how SM platforms have moulded their self-regulatory system on the juridical tradition and interpretations of free speech intended in the First Amendment in the Unites States Constitution. This is conventionally seen as a juridical tradition more protective of free speech than in its European counterparts (Tambini et al., 2008). However, this self-regulatory approach has been challenged by the increase in demand for regulation included in recent statutory initiatives. Douek (2021) observes that SM companies have been required to change their internal rules and adapt to a different tradition of interpretation and adjudication of freedom of expression on their platforms. According to Douek (2021), SM platforms are progressively adopting decisions based on proportionality and probability. Proportionality means recognizing that the right to free speech has to be weighed against other societal interests, more in line with the European tradition. Probability means that content moderation at the scale of SM platforms will always involve some form of error and that a balance has to be struck on reasonable error rates and which kinds of errors are acted upon (Douek, 2021).

However, this change of paradigm raises several issues concerning the capacity and legitimacy of SM companies to operate this systemic balancing, especially through their automated content moderation systems. The scholarship is calling for more research on this topic (see Gillespie et al., 2020). The implications of content regulation by and on platforms are highly political, and the governance structures that arise will shape the future of online expression and public discourse (Douek, 2021). Considering the form of future governance structure, scholars have explored the movement towards a progressive 'platformisation' of society, i.e., the gradual movement of the economic and social system towards a platform organization (Fuchs, 2014; Casilli, 2017; Srnicek, 2017).

2.4.2 Platformisation

Tarleton Gillespie's works on platforms (2010, 2015, 2018), Frank Pasquale's *Black Box Society* (2016) and Cathy O'Neil's *Weapons of Math Destruction* (2017) are all examples of very successful studies trying to unveil how platforms and algorithms have become the allocation or ordering system actively structuring society. Global companies that run SM sites have been embroiled in nearly every area of everyday life, from politics (Gillespie, 2018) and labour relations (Srnicek, 2016; Van Doorn, 2017; Casilli, 2017) to cultural development and consumption (Poell et al., 2017). ‘Platform society’ (van Dijk, 2014) and ‘platform capitalism’ (Srnicek, 2017) are two of the terms used to highlight this specific ordering system, created around the capitalist production of profit and based on the extraction of value from digital social data (Casilli, 2017). Similar concepts are ‘surveillance capitalism’ or ‘dataveillance’ (Lyon, 2014; Andrejevic, 2011; Fuchs; 2012; Zuboff, 2019) or ‘data colonialism’ (Couldry and Mejias, 2019), all terms coined to describe a society structured around the production, aggregation, quantification, and profiling of people's data through constant monitoring and tracking of (meta)data produced via platforms. Some, such as Shoshana Zuboff (2019), are concerned with how digital surveillance (i.e., exploitation of digital social data) is used in this capitalist system to predict and influence emotions in order to sell products. Similarly, Bunz and Meikle (2018) and Lupton (2016) also show how discursive elements developed together with technical affordances have the power to shape behaviour and bodies. In this way, they show how fitness tracking apps such as Fitbit stimulate users using competition and challenges and create ‘playful surveillance’. The system exploits the neo-liberal ideas of individuality to create profit.

Scholars have been exploring this order's theoretical, methodological, and ethical implications, stressing how the narratives and metaphors associated with platforms can be misleading. Platform capitalism captures and extracts value from users' data. On SM platforms, users' data produce commodities in information, social relationships and social networks. SM platforms create profit from these activities by selling advertisement space and through targeted advertising. Users enable this through the visibility and engagement that their interactions generate and by being the recipients of targeted advertising (Poletti and Gray, 2019). This creates issues of commodification and exploitation (Casilli, 2017). However, on SM, this method of value extraction is depicted as an improvement in the procurement of goods and services, whether public or private, and it often uses terms such as ‘sharing’, ‘participation’, and ‘collaboration’ (Gregory, 2017). The narrative behind the concept of ‘participatory culture’, such as entrepreneurialism and free choice, hide forms of exploitation on which the system is based (Gregory, 2017). Platform companies are co-opting key terms typical of human rights activism and changing their meaning, as for instance, privileging the idea of *openness* to the one of *transparency* (Milan, 2015). In this way, industries embrace human rights labels but not their ethical commitments. Scholars have used critical discourse analysis to explore the narrative elements associated with platforms to stress how these technologies are connected to a specific vision of the world, a neo-liberal conception of society developed in Silicon Valley. For years, a similar vision has also dominated the development of media policies in the US. Victor Pickard (2013) highlights two overlapping discourses regulating the US communication policy: corporate libertarianism (i.e., the idea that economy and society benefits more from being free from government intervention) and market fundamentalism (i.e., the idea that the market is the most effective and therefore beneficial method of allocating resources).

These have been, in his view, the dominating ideas that have oriented communication policy in the US since the 1940s. However, in his historical overview, he shows that traditional media are, in fact, in a situation of market failure, since the market (and selling advertisement) on its own is incapable of supporting media (which are, however, a public good for democracy). Hence, a crisis increased by SM and the possibility to sell advertisements online. The only way to protect information and communications in a democratic system is to protect them from commercial pressures (Pickard, 2013).

2.4.3 Threats to free speech

Hintz (2016) stresses the consequence of the increasing role of private intermediaries in formulating, implementing and enforcing regulatory mechanisms of speech:

> Social media and other digital platforms have provided an important means of activist and dissident communication, but they are also key sites where the tension between free communication and the emerging reality of restriction and censorship is played out (Hintz, 2016:336).

As stressed above, SM platforms' primary content moderation tool – algorithms and automated speech recognition – presents fundamental problems for freedom of expression. The influence of the 'politics of algorithms' on freedom of expression is hotly debated in academic and popular publications (Ziewitz, 2016). Even the best automated moderation systems make mistakes: analysing meaning and context of expression has always been a challenge and, even with an improved machine learning system, there are limitations based on the type of language for instance (Hustzi-Orban, 2017). The automated system might end up under- or over-censoring speech and, at the scale of giant platforms such as Facebook, this could mean hundreds of thousands of wrong decisions every day (Hustzi-Orban, 2017:235). Moreover, studies have stressed how automated technology applied to speech online (such as algorithms and bots) contributes to reproducing biases in society, as in, for instance, search engines which tend to retrieve defamatory content for women and BAME groups (Ziewitz, 2016), or reproduce sexist or misogynistic biases (Gerrard and Thornham, 2020).

Automated moderation adds to the opacity of SM companies' internal procedures (Crawford and Gillespie, 2016; Hustzi-Orban, 2017; Ulmann and Tomalin, 2020; Edwards et al., 2021), complicates questions of equity and justice in large-scale socio-technical processes, and re-obscures the inherently political essence of speech decisions made at scale (Sinnreich, 2018). According to Sinnreich (2018), there are three main threats to democratic norms and institutions arising from algorithms:

1) Quantization of culture. Delegating complex and contextualised cultural decision-making (and sense-making) mechanisms to an algorithm risks making choices inconsistent with cultural norms and reifying algorithmic reasoning as the arbiter of meaning and validity.
2) Institutional convergence. The division of powers is necessary for a democratic government. By delegating regulatory, judicial, and executive duties to a centralised, unaccountable, privately-owned body, platform content moderation contradicts this concept. When people recognise automated oversight as an acceptable form of governance, they are also more likely to accept autocracy as an acceptable form of government.
3) Expansion of scale. Platforms' decisions on freedom of expression, taken via algorithms, are imposed upon national sovereignty and self-determination. However, corporations, unlike states, have no obligation to maintain democratic principles.

2.4.4 Societal interests

Scholars increasingly seem to agree that to meet these rising challenges and demands from nation-states and end-users alike, digital infrastructure firms will have to take on additional social responsibilities or should be recognised as public utilities (Rahman, 2018; Balkin, 2017). The more corporations serve as governors (Klonick, 2017), the more they are expected to follow governors' responsibilities to the people they rule and to introduce procedural guarantees, due process, transparency, and fair rights (Balkin, 2017). These responsibilities require procedural justice, accountability, and adherence to the companies' publicly defined standards and policies (Balkin, 2017). In light of the role of platforms in the system, according to van Dijck et al. (2019), the power of platforms should no longer be assessed merely in terms of economic markets and consumer welfare (which has been the position of the EU for a long time). Since people and institutions have become reliant on networks for their social and political well-being, platform control also needs to be rethought in the light of evolving platform ecosystems (van Dijck et al., 2019).

The growing expectations of the accountability of digital platforms to citizens expressed through demands for legislation and regulation at a national and regional level mean that global digital platform companies are expected to be accountable for the content available from their sites. Companies are increasingly called on to be responsible for managing the content on their sites in ways that both meet public interest concerns and are practised openly and transparently.

2.5 Governance of speech as object of research: gaps, and how to fill them

Although publications on the topic have multiplied in the last few years, the review of relevant literature shows that research in the field is still developing, and scholars agree that expanding the scope and range of research on free speech and SM is critical. More normative questions can be asked about the current governance system and how speech on platforms should be governed in the future. Scholars invite society to engage in evaluations of our current governance system and the political perspectives it lays out, thinking about alternative arrangements (van Dijck and Reider, 2019:4) and bearing in mind that the ultimate goal is to achieve a governance policy grounded in human rights and open societies (Gillespie et al., 2020:4).

The review of recent publications on the governance of speech on SM has shown how research has focused on three specific actors or instruments of governance (i.e., state laws, civil society advocacy or self-regulation by private companies) (Wagner, 2016; Gorwa, 2019). Studies have also focused on regulation as a reaction to highly visible individual cases, the so-called public shocks (e.g., terrorist attacks, US presidential elections), or on high-impact themes such as pornography or hate speech (Gillespie, 2018; Cusumano et al., 2021). However, these approaches tell only a partial story about the complex ecology of governance of speech. Gillespie et al. (2020) call for an approach that moves beyond the macro perspective on the regulation of misinformation, or hate, or pornography.

Scholars have shown that in internet-related governance issues, often ‘soft’ and less institutionalised means have succeeded in creating norms that have been picked up at a higher level of policy-making (Radu et al., 2021). This line of research focused on the mundane practices and discourses that contribute to the creation of governance. In this line, together with studies on large regulatory instruments, further research is needed, focusing on the day-to-day construction of narratives about technology and free speech, which contribute to the development of norms that will influence governance.

In the same way, the scholarship recognises a gap in the study of the relationships and dynamics connecting different actors, for instance in terms of how companies receive pressure from policymakers and how platforms exert influence politically (Gorwa, 2020), as well as the strategies that civil society uses to push content regulation by private companies in a direction respectful of human rights (Jørgensen and Zuleta, 2020). Studies on large-scale approaches to the governance of speech (as in Gorwa, 2019, 2020; Balkin, 2017) tend to privilege 'traditional' or more institutionalised actors and settings (states, private companies, advocacy bodies and NGOs in civil society within multi-actors initiatives) and do not say much about smaller elements of the system, for instance individuals who still have a significant role as both citizens of democratic countries and users of the sites.

As future lines for development of the scholarship, Gillespie et al. (2020) suggest considering the plurality of actors in different geographical locations and not just the big US-based platforms. As much as big platforms create the most visible cases, they do not represent SM ecology. Smaller networks, or platforms outside the US, with very different visions and cultures than Facebook or Google, may develop innovative moderation techniques (Gillespie et al., 2020:3). This approach also emphasises the policies pursued by countries such as those in the Global South and ensures equal justice for their citizens (Couldry and Mirthas, 2019).

However, the studies on the governance of speech mentioned above focus on traditional 'social actors' such as states, or companies or civil society and tend to underestimate one of the central tenets of internet governance studies, i.e., the idea that the governance of speech takes place in a socio-technical system in which technological infrastructure, although not always visible, is pervasive and has agency. Some other studies of governance of speech on SM have highlighted the importance of technology and algorithms and the implications of such tools in terms of democratic legitimacy and social justice (Sinnreich, 2020). Dencik et al. (2018a, 2018b) call for more research on the debates on how technology is incorporated in governance practices. Scholars call for more studies on the ethical problems raised by algorithmic content moderation activities, as well as the threats associated with delegating social policy regulation to artificial intelligence and other non-human processes (Sinnreich, 2020).

A further gap in the literature concerns the use of narratives in the study of the regulation of governance of speech online. In the first part of this chapter, I discussed the relationship between narratives and regulations, stressing how often it is possible to find a correspondence between narratives and metaphors about technology and changes in the regulation of speech. Previously in this chapter, I mentioned the scholarship that recognises the importance of norms and narrative elements as complementary aspects to be considered in the study of governance online (Radu et al., 2021).

It is important to recall this now and to relate it to the other major findings from the literature described above. Firstly, the literature review has defined the governance of speech as an emergent process, a 'reflexive act of coordination', stimulated by major public shocks related to speech and SM technology. Secondly, scholarship has also stressed the importance of norms developed 'outside' the more institutionalised arenas.

These definitions imply a fundamental role for public discourse and, within this, the role of media as the main tools used in the creation and narration of 'public shocks' and the reinforcement or challenge of existing norms. As seen above, societies marked by technological innovation tend to continuously generate new types of public shock that break established routines (Dewey, 1927) and all media (not only SM) are fundamental in mobilizing public discussion about technology, citizens and public opinion (Barry, 2001). Media are the instruments that represent what is relevant for the collective (Couldry, 2012). As Couldry (2012:35) put it, 'representations are a material site for the exercise of, and struggle over power'. The studies presented above leave a gap in the consideration of the material effects of narratives on the creation of 'public shocks' intended as a spark for regulation initiatives and the role of those who create these narratives, i.e., media (not necessarily SM).

Including the role of narratives and, in particular, the role of media in the study of governance opens interesting opportunities to address the recommendations for further research presented above and, in particular, creating the opportunity to expand the range of controversial cases and include more of the less powerful or represented actors. Moreover, the analysis of images and metaphors mobilised by the media in the construction of public discourse highlights fundamental principles and global visions about technologies. It provides the key to interpreting the initiatives composing our current governance of speech online. At the same time, it provides fascinating insights to assess the current initiatives against the normative ideals of the type of governance of speech that we would like to have in democratic regimes respectful of human rights.

Based on the critical assessment of the literature and the recommendations for future research presented in the different studies, in this study I have chosen not to focus on one specific actor (either states, private companies or civil society) or one specific highly controversial event. In order to consider both 'non-traditional' or institutional actors as well as fewer famous cases, I have chosen to adopt a different approach. Since scholars (Barry, 2001 and others) have stressed the important role that media play in the shaping of 'public shocks', I will be studying the narratives emerging from media and public discourse concerning speech regulation and SM, and how they contribute to creating contentious phenomena (public shocks) able to initiate regulation processes (Barry, 2001; Gillespie et al., 2018). As stressed in the literature, the political and social (and economic) implications of governance of speech on SM platforms are huge. For this reason, in this study, I intend to contribute to the literature on the implications that technology can have for materially shaping governance.

Considering the gaps and opportunities for further research described in the literature, (i.e., not many studies on the relationships connecting actors; the importance of algorithms; the importance of narratives and media in creating those narratives), in this study, I ask: how can we study governance of speech online as an emerging phenomenon and without focusing on one actor or one specific setting? How do governance initiatives 'initiate' and take form? Moreover, what does it mean for the broader governance of freedom of expression and democracy?

I have focused my research on the following research questions:

- What are the 'public shocks' (not necessarily major) that contributed to breaking routine or pre-existing forms of decision-making concerning public expression on SM?
- What types of norms and governance model are taking place due to public shocks in the last few years?
- What actors and dynamics of power are revealed when using an approach that does not simply focus on a single platform or actor and includes technology in the study of regulation initiatives?

- How do ‘public shocks’' relate to the narratives associated with free speech and technology? And what is the role of media in reproducing narratives and shocks?

In answering these questions, I aim to contribute to understanding the governance of speech on SM by adopting an empirical rather than theoretical approach to identifying actors, narratives, and material elements attached to technology and the power dynamics that link them.

2.6 Conclusion

Based on this critical assessment of the literature, in the next chapter, I explain why, in my view, material semiotics (and in particular Actor-Network Theory and its empirical applications: controversy mapping), combined with critical data studies, are the most suitable theoretical and methodological framework to study the governance of speech on SM. I will explain how the specific ontology of material-semiotic approaches, considering both the material (technology) and semiotic (narrative) elements, can prove extremely useful in understanding the development of the current governance of speech online. However, since controversy mapping is fundamentally a bottom-up methodology, I argue that it lacks the depth to provide a robust normative interpretation of the results (and answer the questions I posed above). For this reason, I have chosen to put the data empirically collected through controversy mapping in dialogue with the emergent field of critical data studies and to interpret the results in the light of the analysis of power relations highlighted in the platform society (van Dijk, 2018).

3. Theoretical framework

3.1 Introduction

In the previous chapter I gave an overview of the literature on the topic of governance of speech online. The scholarship agrees on the idea that speech governance online takes place as the emerging result of a series of initiatives by a plurality of actors in different settings (from the most to the least institutionalised). The literature also agrees that such initiatives to regulate speech are the reaction to a 'public shock' or contentious events related to social media's platform technology (Ananny and Gillespie, 2017). I have also explained how narratives and media play a fundamental role in creating these public shocks that initiate regulatory initiatives. In the final part, considering the gaps and opportunities for further research described in the literature, I briefly introduced the research questions and my main goals: how can we study governance of speech online as an emerging phenomenon and without focusing on one actor or one specific setting? How do governance initiatives 'initiate' and take form? And what does it mean for the wider governance of freedom of expression and democracy?

In this chapter, I present my theoretical framework: a combination of material semiotics (in particular Actor-Network Theory (ANT)) and critical data studies. I will explain how these two approaches are useful for filling the gaps found in the literature review, namely furthering research on platform governance that avoids focusing just on one single platform or event, recognising a fundamental role for the narratives of technologies presented in the public discourse by the media and at the same time providing a critical assessment of the material implications that technologies like algorithms can have on social life.

In the first part, I introduce material semiotics as both an ontological and epistemological framework able to describe what happens during public shocks concerning freedom of expression online. I then focus on the specific theoretical and methodological tools developed within ANT. I argue that ANT seems particularly fitting to fill the gap found in the literature, as it empirically defines significant actors and the main 'public shocks', without focusing on a specific actor or event. Moreover, it includes a specific interest in technology, and the narrative elements associated with it, offering interesting tools to identify and study its role within regulation initiatives concerning freedom of expression or content regulation (as in the case of state regulation vs. self-regulation by private companies). Critical data studies is a younger field, still strongly related to the ontological background of material semiotics, but at the same time interested in how digital materialities produce a hierarchy of power, highlighting the power structures embedded in technology, and especially in the management of digital social data.

Below I present the main concepts that I use in this study and how these relate to the findings in the literature review. Borrowing from ANT terminology, I introduce the concept of actors and associations, translation, and other terms used to isolate the role of different elements in the governance ecosystem. I also introduce the concepts of 'public' and 'controversy', terms that are developed further in the methodology (chapter 4), but that find their origin in the ANT interest in the study of public shocks, when the routine and the taken for granted are interrupted. Drawing from critical data studies I also introduce the concept of datafication to critically assess the material impact of technologies. I then explain how, combining ANT with critical data studies, I aim to overcome the main limitation which concerns me, i.e. the ontological lack of interest in any interpretation of social dynamics in terms of larger concepts or structures. I argue that by combining the 'ethnographic' approach of ANT with critical data studies I provide a theoretical framework able to answer both my descriptive and my normative questions about the type of model of governance of speech that we are experiencing and that might develop further, such as how the lack of critical analysis of digital social data contributed to legitimise a specific ordering power in contemporary social life (i.e. platform economy, algorithmic management of social life). I conclude the chapter with a list of operational research questions, developed from the theoretical framework discussed, and aimed at orienting the methodology and the data collection chapters.

3.2 Theoretical framework

3.2.1 Material semiotics and digital society: an introduction

In recent years, a vast body of scholarship interested in digital society and technologies has adopted theoretical frameworks compatible with – if not directly developed from – the Science and Technology Studies (STS) and Actor-Network Theory (ANT) traditions. For instance, many of the studies discussed in the literature review (DeNardis, 2014; De Nardis and Hackl, 2015; Musiani, 2015; Ananny and Gillespie, 2016; Epstein et al. 2016; Hofman et al. 2016; Pohle et al. 2016) routinely use concepts and frameworks developed from these two original fields. STS has become a quite common approach for scholars active in the field of internet governance (IG), and concepts such as the ‘socio-technological system’ are increasingly adopted to define the internet and related digital communications technology (Musiani, 2015; Pohle, 2016a, 2016b). The main reason why STS was adopted for this type of study in the early 2000s relates to the same limitations in the study of governance in the digital ecosystem that I discussed in the literature: the lack of a specific institutional framework or legitimate procedure, the decision-making power diffused across actors of different types (states, private companies), the role of civil society and technology. Approaching IG through an STS lens, these authors could address issues that political and legal sciences were incapable of addressing up until then, and yet are crucial for the understanding current governance of the internet (Musiani, 2015).

Even if STS and ANT are not the same, as they have developed from different fields and they bring forward the heritage of different scholarly communities (in this regard see one of the many articles that have been published on the difference between STS and sociology, for instance by John Law, 2008), in these studies they are often used interchangeably (De Nardis, 2014; Musiani, 2015; Pohle, 2016a, 2016b, Epstein et al. 2016). I am not interested in discussing the differences between the two approaches here; what I am interested in is discussing how these traditions together have contributed to shape studies of digital society towards a *materialist ontology, the interpretation of social order as emergent from hybrid elements* and a *relational idea of power*. Most of all, STS and ANT share a fundamental interest in demystifying realities taken for granted, such as scientific knowledge, showing how they are in fact the product of interactions among hybrid social elements (Cloatre and Pickersgill, 2015).

As seen in the chapter 2, several studies investigating freedom of speech online have adopted a materialist ontology, as an approach able to interrogate the place that technology occupies in today’s society (Gillespie et al., 2020; Sinnreich, 2018, 2020). The literature also showed how order and governance of speech online emerge as an effect or a reaction rather than planned government (Wagner, 2016; Gorwa, 2019), some of them stressing the performative aspect of this type of governance (Musiani 2015; Hofman, 2016; Hofman et al. 2016; Epstein et al. 2016). In the lack of an institutionalised hierarchy, power too emerges from the relations across the elements in the ecosystem, where boundaries between public and private are increasingly blurred, and holders of traditional forms of power (such as states) are challenged by other strong actors (such as social media or financial companies) and technological infrastructure (De Nardis, 2014; Gillespie 2018).

Below I introduce the main concepts that I use in this study, borrowed from STS and ANT theoretical traditions, and from the larger perspective of material semiotics. As I explain further in the second part of the study, these are not the only concepts that I use, since I intend to expand the framework with insights derived from critical data studies.

3.2.2 Material semiotics

Material semiotics describes a particular theoretical position in social sciences, which has been built on the works of philosophers, semioticians and sociologists such as Gilles Deleuze and Félix Guattari, Michel Serres, Alfred Whitehead, Isabelle Stengers, Algirdas Julien Greimas, Gabriel Tarde and Michel Foucault. It gathers different authors: representatives of ANT such as Bruno Latour, John Law and Michel Callon, but also scholars in feminist studies such as Donna Haraway and Annemarie Mol and Marilyn Strathern, among others.

Material semiotics' particular ontology is highly indebted to the work of Gilles Deleuze and Félix Guattari, and the concept of *agencement* (translated as 'assemblage' in English). According to this view, social reality is made of multiple, heterogeneous elements such as human and non-human bodies, discourses, practices, artefacts and technologies. Agency or materiality is not a quality of the subject; it emerges as the result or effect of the association between these heterogeneous elements (Beetz, 2016; Müeller, 2016). Based on this concept, Bruno Latour (2005b) in *Reassembling the Social* defines sociology as the 'science of living together'. ANT sociology sees the 'social' as the result of continuous movements of association and dissociation between the hybrid elements (e.g. humans and non-humans). What appear as fixed social objects are de facto *assemblages* of different elements, which become 'detectable' only because they successfully 'freeze' an otherwise dynamic situation, 'masking' the negotiations and potentially competing strategies that separate the different elements. ANT is also particularly indebted to Greimas's theory of narrative structure and Garfinkel's ethnomethodology and their focus on uncovering the process through which different social elements make sense of social reality adopting their narrative and perspective.

Scholars in Science and Technology Studies have also underlined this performative aspect of social reality, and the fact that what we take for granted is an effect rather than an intrinsic quality of objects, in particular considering knowledge and technology. Both STS and ANT share the goal of ‘demystifying technology’, also called the opening of the ‘black-box’ of what is taken for granted. STS scholar Harry Collins’s (1975, 1981, 1985) famous metaphor of the miniature ship in the bottle has also been used several times by ANT academics (Venturini, 2010) to explain how scientific knowledge is created and transformed in shared reality. As the miniature ship in a bottle looks as if it has always been there, so scientific claims that become accepted knowledge appear as if they have always existed. However, a closer investigation in the world of bottled ships reveals that different procedures might have been used to put a ship in a bottle. Similarly, in the case of knowledge creation, STS scholars underline that different ideas and social influences are negotiated in order to generate consensus around any form of knowledge (Venturini, 2010).

The unboxing of social reality is also a fundamental interest for ANT, which has developed a whole specific toolbox to empirically identify the elements that are involved in these types of negotiations, also called ‘translation’ (see Callon, 1986a, 1986b), and I will discuss this in greater detail in the next paragraphs. The interest in the demystification of technology, and in particular the unveiling of the different elements and power relations underlying the production of Big Data science, is also the interest of critical data studies (Kitchin 2014; Kitchin and Lauriault, 2014; Iliadis and Russo, 2016), which I introduce below.

3.2.3 Critical data studies

Critical data studies are very much indebted to material semiotics approaches (in particular to the concept of assemblage). From a theoretical point of view, scholars in this field share the idea of distributed agency and the prominent role of technology in society. However, they are distinguishable for the specific interest in digital data’s social construction and materiality and because they use critical concepts and methods to theorise digital data in the context of domination in society. A growing number of academic works have started to give a critical account of the *agency or effects* of technological objects as part of the social world, theorising data and ICTs as sociomaterial (or socio-technical) objects (Lupton, 2015, 2016). The subjects of critical data studies are the socio-technical ‘data assemblages’, i.e.

> the combination of narrative elements such as systems of thought, forms of knowledge, finance, economy, governmentalities and legalities, as well as materialities and infrastructures, practices, organisations and institutions, subjectivities and communities,

> places, and the marketplace where data are constituted (Kitchin and Lauriault, 2014 in Iliadis and Russo, 2016:2).

In the sociomaterial perspective, data about humans and humans are always part of each other and emerge together (Lupton, 2018:5). The key purpose of these studies is to reframe the questions that inform epistemological systems surrounding data-related social issues (Iliadis and Russo, 2016). This line of inquiry started with boyd and Crawford (2012), who proposed a key set of critical questions for Big Data. Critical data studies today criticise research on 'potentially depoliticised data science' and suggest the need to 'track the ways in which data are generated and curated, and how they permeate and exert power on all manner of forms of life' (Iliadis and Russo, 2016:2). To provide a critical interpretation of technology, critical data studies stress the role of the context in which data and digital technologies in general are produced. Data are a form of power (Iliadis and Russo, 2016) produced through commercial platforms, aimed at exploiting their quantifiable materiality to monetise trends in society (Savage and Burrows, 2007). Companies process data and have the ability to influence society for profit (Gillespie, 2015; Casilli, 2017a; Zuboff, 2018). As noted in the literature review (chapter 2), scholars have identified platforms and algorithms as the fundamental ordering system structuring this society (calling it either the platform society (van Dijk, 2014) or platform capitalism (Srnicek, 2017)). Critical data studies highlight the implications, biases, risks and inequalities, as well as the counter-potential of digital data, which are increasingly used to inform decision-making processes, in economic, political and legal systems as well as the social justice system (Beer, 2016, 2017; Dencik et al., 2016, 2018a, 2018b; Richterich, 2018; Redden, 2018) creating a real form of algorithmic management of social life.

The extent to which digital data are influencing civic rights and personal autonomy (Fuchs, 2015) has fostered scholarly discussion in critical data studies about the ethical implications of research *on* and *with* big digital social data. They include discussions about the nature of digital metrics, their innate biases and politics, and their implications when used for research or to inform decision-making processes. For instance, by raising the risks behind 'dataism' as the widespread ideology of Big Data's desirability and unquestioned superiority (van Dijk, 2014: 198; boyd and Crawford, 2012) they question the fundamental principle behind forms of decision making based on algorithms.These studies call for the deconstruction of the narrative surrounding data as neutral, objective, independent, raw representations of the world, and expose the extent of its embeddedness in society (Kitchin and Lauriault, 2014). This process starts from research which needs to break the idea of the objectivity of data in favour of more qualitative, empirical approaches (Lupton, 2015; Metcalf and Crawford, 2016).

This brief review shows how these theoretical approaches can provide interesting insights to approach the governance of speech on social media. However, there are specific reasons why I have chosen this approach, based on my research goals and the gaps identified in the literature review.

3.3 Using STS, ANT and material semiotics to close the research gaps

The overview of the main literature on the topic of governance of speech highlighted gaps and possible directions for future research. In particular, Gorwa (2019a, 2019b, 2020) stresses the importance of deepening the understanding of the associations and power relations linking the actors taking part in governance initiatives concerning freedom of expression online. The literature review also revealed potential in the development of studies that consider the role of narratives, i.e. the way that technology is presented and mediated through the media, and how they contribute to build the public shocks that initiate governance initiatives. Moreover, as stressed by Gillespie and others, scholars are now recommending that the focus should not be only on specific actors or events, so that they are not used as explanations for the whole governance system.

As mentioned above, ANT and STS can bring interesting insights in the study of processes of ordering construction, as has been fruitfully discussed in relation to governance theories (Katzenbach, 2012, 2013, Epstein et al. 2016, Musiani 2015). As presented in the brief introduction above, STS, material semiotics and in particular ANT have already been adopted in studies of digital society, as they make sense of situations where heterogeneous elements (such as states, private companies and technologies) are engaged in ordering processes. The emphasis on associations in ANT is useful for distinguishing groups or networks of actors that cross conventional social categories and dichotomies (e.g. nation states, or public/private, local/global, and formal/informal). Similarly, the relational idea of power highlights how formal holders of power are not necessarily the same as the effective holders of power (Callon, 1986a, 1986b; Schouten, 2014). The inclusion of non-human elements in the socio-technical system allows us to consider the role of technology such as platforms' technical architecture, software, filters and algorithms for content recognition, private companies' Terms of Services/Community standards, managerial strategies, public/private agreements (European code of conduct), European/national/city-region legislation, law enforcement bodies, users, and terrorist groups, in the enactment of the social reality (Poletti and Michieli, 2018).

Considering the opportunity for further research described in the literature review, ANT offers a framework to study and analyse the role of actors who do not fall in the traditional 'governance triangle' of states, private companies and civil society (Gorwa, 2019a, 2019b, 2020).
Following the ANT perspective, social media platforms can be theorised as sociomaterial objects: socio-technical assemblages, entangling humans, capital, social actions and technological elements such as algorithms, data, servers, work flows, business plans, communications protocols and trackers (Gillespie 2010, 2018; Kitchin and Lauriault (2014); Marres, 2017). Similarly, the governance of freedom of expression on social media can be theorised as a highly complex assemblage composed of 'coordinated but dispersed regulations, calculative arrangements, infrastructures and technical procedures' that render online content governable (adapted from Schouten, 2014). It is composed of 'socio-technical' arrangements that mediate relations and interactions, black-boxing some concerns and threats while foregrounding others.

Moreover, as I describe in greater detail below, ANT is interested in combining narratives and material elements, and assigns a fundamental place to the study of different narratives and technological artefacts present in a controversy about technology and how media contribute to reproduce or challenge them. It does this without assuming the predominance of one actor over the others (ANT considers all the elements as isomorphs) and with complete 'agnosticism' towards the chances of an actor influencing the final outcomes. This approach translates into a set of methodological tools that allow for the collection of data about governance of speech and social media, without starting with or fixing the focus on specific actors or events (as recommended by Gillespie et al., 2020). However, as much as this empirical and flat ontology can provide a valuable direction to approach the study of speech governance, it might turn out to be unfit for the development of interpretations in terms of normative models of governance. For this reason, I integrate this perspective with the one from critical data studies.

3.3.1 Previous use of theories

STS has been used a lot in research of internet governance in general and of multi-stakeholder organisations, but not many authors have applied the ANT theory of translation to the analysis of governance processes on the internet. However, Julia Pohle (2016a, 2016b) provides a good example of how sociology of translation can be used to study discursive production in internet governance. In her study she analysed the deliberations within the UN Working Group on Enhanced Cooperation (WGEC), a multi-stakeholder group created to overcome controversies on the role of governments in internet governance, which have persisted since the World Summit on the Information Society (WSIS) (Pohle, 2016a:2). In particular, she combined translation theory with different types of discourse analysis (e.g. Interpretive Policy Analysis (IPA) (Pohle, 2016a) and Argumentative Discourse Analysis (ADA) (Pohle, 2016b), to explore how actors translated ideas, shaped meaning and competed over the inscription of discourse into policy outcome. Even though she focused on different types of documents from this study (i.e. she was considering policy documents, while this study includes a more varied group of documents), many of the concepts and choices of analysis presented in her work can provide useful guidance for the analysis in this study: in particular the concept of exemplary cases and storylines. In this study I have used Pohle’s work to orientate the detection of narrative structures and for the detection and interpretation of actors and roles in the light of the sociology of translations.

In the next paragraphs I will describe the main concepts from ANT and critical discourse analysis that I used in this study.

3.4 Key concepts

3.4.1 Actors and associations

ANT is also called the sociology of associations (Latour, 2005b). According to Latour, there are no social aggregates behind social activities, just as there is no difference in sise or ‘nature’ among the different elements. ANT’s generalised symmetry principle (Callon and Latour, 1981) insists on the fact that all elements should be treated as isomorphic, and that there is no difference between human and non-human elements. This implies a specific choice of language, derived from narrative theories (i.e. in particular from Greimas' structural linguistics), which describe agency without deciding beforehand who is allowed to make a difference in the story. This allows non-human entities to be considered as subjects, as well as humans. In ANT, agents acquire their role of 'actors' and are included in the description only when they become a subject of matter, i.e. when they are 'detectable'; because they have an effect, they 'appear' in the story.

3.4.2. Translation

In the ANT perspective, associations are created through a process of 'translation' of interest, which includes:

> [A]ll negotiations, intrigues, calculations, acts of persuasion and violence, thanks to which an actor or force takes, or causes to be conferred on itself, authority to speak on behalf of another actor or force (Callon and Latour, 1981: 279).

Translation revolves around power relationships. While creating associations/attachments or alliances with others, actors dissociate these entities from their previous relations (i.e. 'disentanglement') and simultaneously oblige them to remain faithful to their alliances (Callon, 1986:19). An actor that succeeds in translating other entities' interests becomes the spokesperson for the whole network and the representation of the network itself (hence the term 'actor-network'). In the process, certain relations are put in 'black boxes', i.e. become things that no longer need to be reconsidered. From the moment that a set of associations is 'black-boxed' or simplified (Callon, 1986a) it can act as a single entity, creating other associations and building new networks (Micheal, 2017). An actor-network is thus a 'network of simplified entities which in turn are other networks' (Callon, 1986a:32).

Callon describes the different phases in the process of translation in his study of the scallops of St Brieuc Bay (1986a): during the first stage of *problematisation*, the primary actors identify a problem, and formulate a claim about the issue (i.e. take a position which requires support from other actors within the network). In the second phase, the *interessement*, negotiations may take place with other actors about the roles they will perform within the network. *Interessement* is the formation of a network of 'alliances' in order to reach an understanding among the different actors about their respective interests and how they can be aligned with those of the primary actor's (Alcouffe et al., 2008). *Enrolment* corresponds to the strengthening of links between the various interests of the actors and the stabilisation of the shape that the ties will take. Once enrolled, actors accept the roles they have been given. *Mobilisation* happens as those outside the network (allies) join in to sustain the network's interest monitoring so that it remains stable (Alcouffe et al., 2008). Describing the process of translation, Callon (1986a, 1986b) stresses that if, on the one hand, certain elements succeed in playing the key role for the assemblages they constituted (i.e. successfully occupying the prominent position and enrolling and assigning specific roles to others), on the other hand elements can resist this process, refusing roles or trying to become the spokesperson too. This changes the focus on the definition of the problem. For example, adopting this perspective in this study I can theorise code, algorithms, users, and social media companies as actors all enrolled in the socio-technical assemblage that we call 'social media platforms'.

As I will discuss in later chapters, the different roles and positions of actors might seem 'stable': for instance, SM platform companies develop codes, and algorithms, that create the interface of platforms, which becomes the horizon of possibility for users. They present their technology as essential for freedom of expression, dipping into the traditional 'cyber libertarian' narrative that has for a long time characterised the development of the internet. In this situation, the different roles might seem 'unquestioned', and the translation achieved. However, as we will see, certain elements might refuse to 'follow the rules' – exposing the network of assigned roles to uncertainty. This might be the case with users posting hate speech, combined with technology such as algorithms pushing for virality of certain content.

3.4.3 Roles

As mentioned above, ANT has derived several of the concepts used from narrative theories (i.e. from Greimas's theory of narrative structures). Narrative structuralist theories postulate that every narrative, either micro or macro, is based on a constant structural relation between the narrative's characters (the roles of characters in a story, called narrative structure) and the themes (discursive structures) (Beetz, 2016). In the study of translation processes, ANT draws on this interest for the structural role of actors and focuses on identifying both the role of the characters in a story and the main discursive elements that are developed. Specifically, in ANT the attention is placed on those elements that become involved in association processes (Latour, 1986; Callon, 1986a, 1986b). The more powerful ones succeed in creating the occasion for these associations, translating the interests of all the others within the network and engaging the different actors to pursue such interests. They occupy the role of spokespersons for the others, giving voice and general interpretation for the other actors in the controversy (Latour, 2005). Spokespersons are identifiable by their position in the network, as 'obligatory passage points (OPP)' (Callon, 1986a). As OPP, either individual or collective actors become indispensable for all others to achieve their goals.

> Spokespersons are accompanied by 'intermediaries', which 'transport… meaning or force without transformation' (Latour, 2005:39), contributing to the diffusion of an 'hegemonic' vision in the associated group of actors. By contrast, spokespersons are challenged by mediators, which 'transform, translate, distort, and modify the meaning or the elements they are supposed to carry' (Latour, 2005:39).

Mediators, in ANT's vocabulary, are those actors that 'transform, translate, distort, and modify the meaning or the elements they are supposed to carry', unlike the intermediaries that 'transport… meaning or force without transformation' (Latour 2005:39). In Callon's view, mediators are material, tangible objects that can break through the boundaries of fixed associations and opening up the frame to new actors, involving new actors that originally were not part of the 'system'. Because of the existence of mediators, no translation should be taken for granted, or be considered to take place without resistance. Elements might always resist the endeavour of enrolment, and try to create their own network (Callon, 1986a). Callon underlines that any procedure of disentanglement produces new attachments. Any attempt at framing, i.e. creating clear and precise boundaries, results in externalities. It is impossible to achieve a total framing, as every framing carries with it other associations.

According to ANT's relational interpretation of power, the value and influence of actors are determined by their role in the network and their ability to persuade other actors to share their interests and behave accordingly. To translate other actors' interests and vision of the world into their own agenda or 'programme of action' (Latour 2005), actors use certain resources and strategies. One strategy is to inscribe forms of order in durable materials (Callon, 1986a; Latour, 2005). For instance, when a specific interpretation of the problem developed within a group is adopted by different actors it can be 'fixed' into written text or even lead to changes in the organisational setting, for example through the creation of new procedures or institutions (Pohle, 2016a, 2016b).

Another strategy is to mobilise more elements (in terms of money, sub-actors, etc.) using imaginaries and narratives as tools to problematise and enrol other actors. The struggle between associations, coalitions, alliances of human and non-human elements mobilised around matters of concern corresponds to a struggle between different realities/imaginaries performed through the associations (Law, 2007; Müeller, 2016). Being an agent in this or that arrangement or assemblage corresponds to a specific knowledge about the social reality, or a distinct 'cosmology' (Stengers, 2005). Reality in different associations coheres in fundamentally distinct ways (Wardle and Shaffner, 2017). Different enactment from different associations creates different realities; a competition between different ontologies. In the enactment of different forms of knowledge, non-human elements – for instance technological innovations – occupy a vital role, reconfiguring boundaries, making connections and creating interoperability where previously there was none (Barry, 2001). This is the idea behind ontological politics (Mol, 1999), or cosmopolitics (Stengers, 2005), as ways to bring together the different practitioners and practices that contribute to the making/emergence of the issue at stake. In this instance, the media can be used as a 'tool' to enrol citizens. Newspapers can be instruments to identify externalities, enabling the different actors to spread their vision and define their interests.

3.4.4 Controversies

Even though relations need to be repeatedly performed to make the network appear as a whole (Latour, 1986, 2005), once the associations composing society and technology are 'established', it is very difficult to observe their internal dynamics of negotiation among elements. They become visible only in the moment when an agreement has not yet been reached, i.e. in case of socio-technological or scientific controversies (Collins, 1975; Venturini et al., 2015). Therefore Latour (2005) argues that socio-technical debates are extremely interesting topics of research, as they represent the places where society exists at its 'magmatic stage' (Venturini, 2010, 2012). In socio-technical controversies, society lacks the stable associations between the different elements of which it is composed. They have not yet been 'taken for granted'. Socio-technical controversies are particularly privileged spaces from which to observe how social reality is constructed.

In a similar way as with the tradition of STS (Collins, 1981, 1985), in ANT controversies represent the unfolding moments when the black-box, the 'routine' of what is taken for granted, is suspended. This is the truly political moment when artefacts, activities or practices become objects of contestation, matters of concern (or 'hot situations' in Callon, 1998). Controversies are moments when the process of translation is 'broken', when something does not work in the way that the assigned roles would. Then the black boxes start to reveal new actors (Callon, 1986a). It is possible to observe how the diverse heterogeneous elements struggle to impose their 'wills' or framing of what has to be taken into account (matter of concern) and what has to be ignored, until the moment when one element will succeed in speaking on behalf of the other elements. In this sense, controversies around socio-technological assemblages as platforms are reminiscent of Callon' s hybrid forums (Callon, 1998): 'highly confused situations' where 'facts and values have become entangled to such an extent that it is no longer possible to distinguish between two successive stages: first, the production and dissemination of information or knowledge, and second, the decision-making process itself.' The actual list of actors, as well as their identities, will fluctuate in the course of the controversy itself, and they will put forward mutually incompatible descriptions of future world states (Callon, 1998:11).

3.4.5 Public and tools for measurement

The concept of public and the public sphere is central in the idea of socio-technical controversy. Political and media scholars like Nortjie Marres and Richard Rogers found a prolific field merging STS and ANT with political theories of public participation in the democratic process, drawing especially from US pragmatist philosophers Walter Lippmann and John Dewey. The writings of Dewey and Lippmann develop a particular conception of the public as organised by material means, and suggest that the public is best understood as a group created around problems (Marres, 2012). In this idea the public is created because it is affected by issues; however, it does not take part in formal decision-making processes, and lacks the connections, skills and vocabulary required to address these issues. American pragmatists state that non-experts and interested viewers will challenge and manipulate expert statements (Barry, 2013; Marres, 2005), and their voices – as expressed in the media – therefore have to be mapped as part of the controversy (Klauser 2009). The attention to the role of media in staging controversy is stressed also in Barry, who sees the public sphere as a set of spaces mediated through technology (Barry, 2001, 2013). He argues that newspapers and websites are technological environments where the role of technology in society is discussed. In controversies some actors do work as tools to enable the public to measure their interests. In this sense, as mentioned by Callon (1998), newspapers can play the role of instruments for citizens to 1) realise that they are involved in the system of framing and 2) provide a tool to understand (i.e., measure) their interests in view of the negotiations taking place during translation processes. Media are usually seen as one way to map non-expert voices in the controversy, i.e. acting as spokesperson for non-expert groups (Schouten, 2014; Barry, 2013; Marres, 2005). As stressed above, media are the instruments for representing what is relevant for the collective (Couldry, 2012). Drawing on Foucault, these authors acknowledge the fundamental role of communications technology, since it is the tool through which power is executed and its affordances and practices contribute to creating the feeling that particular realities are more 'natural' than other alternative ones (Couldry, 2012). Media contribute to shape the controversies, by producing and reproducing the different narratives and world-views of the actors. By attaching governance of free speech to everyday concern, they contribute to problematise governance and to create matter of concern. At the same time, authors recognise that media are also 'publicity devices' (Marres, 2005), where content and information are strictly related to advertisements. For this reason their influence on staging the controversy has to be studied both from the point of view of the substance of the controversy as well as from the point of view of the their specific technological dynamics.

The concepts described above orientated my data collection and analysis. As methodology I used controversy mapping: a specific empirical application of the theoretical concepts of controversy and public. In the analysis I have identified roles from translation processes. To answer my normative question about the shape that speech governance should take in the future, I preferred to rely on consideration of power developed in the context of critical data studies.

3.4.6 Datafication

With the development of ICT, society has relied increasingly on technology for daily actions and interactions, producing a massive amount of digital social data, i.e. socio-technical phenomena associated with terms such as 'Big Data' and the 'data deluge' (Anderson, 2008). Digital social data have become the backbone of the economic system and an opportunity for forms of social research produced by ‘new’ research bodies (Savage and Burrows, 2007; Law et al., 2011).

The ‘big data deluge’ (Anderson, 2008) has been accompanied by an initial enthusiasm for the apparently endless potential for empirical uses of digital social data. In this regard, Chris Anderson’s editorial piece in *Wired* (2008) entitled ‘The end of theory: the data deluge makes the scientific method obsolete’ is exemplary. Influenced and attracted by the new frontiers in Big Data analytics, studies on ICTs and society from the early 2000s have generally favoured (i.e. funded) studies with empirical approaches with low reference to theory (Fuchs, 2017). Twitter metrics such as in-degree (followers) and out-degree (following) connections have been used to define influence/popularity (Cha et al., 2010; Kwak et al., 2010; Marwick and boyd, 2011a, 2011b; Sun and Ng, 2012). Twitter metrics have, for instance, been employed in studies to isolate the elements composing different flows of communication (see for instance, Bruns and Stieglitz, 2014; Weller et al., 2014). A wide production includes conducting sentiment analysis for quantities of Twitter data (see Thelwall, 2014) or computer-assisted content analysis to forecast events (Burnap et al., 2014; Einspänner-Pflock et al., 2014).

However, in academia, a more cautious position has developed alongside a more enthusiastic approach. Several authors have described works based on digital social data analytics as ‘positivist’ (Fuchs, 2016; Mosco, 2015) and expressed concern about the uncritical acceptance of the process of ‘datafication’, i.e. the conversion of social action into online quantified data that can be tracked and analysed in real time (Mayer-Schoenberger and Cukier, 2013; van Dijk, 2014). The problem of datafication is that it completely covers all aspects of contemporary society. As discussed also in the literature review, scholars have explored the movement towards a progressive ‘platformisation’ of society, i.e. the gradual movement of companies towards a form of capitalist economy based on extraction of value from data created within platform organisations (Andrejevic 2011, Fuchs, 2014; Pasquale 2015; O’Neill 2016; Casilli, 2017; Srnicek, 2017). In this perspective, critical data studies call for research focusing on the implication of the system of production and allocation of resources (e.g. data and algorithms) for social life. The concept of datafication allows to rethink definition of power relations and adapt traditional theoretical concept such as Marx’s definition of *exploitation* and *subordination* (Casilli, 2017) or Foucault’s *governmentality* (Lupton 2017) to a system where data generated by users create at the same time value and create collective meanings. In particular, critical data studies’ perspective helps theorising power in an algorithmic and technological society as transversal, diffuse and yet able to influence the materiality of individuals up to their behaviours and bodies.

3.5 ANT and critical data studies concepts in the literature

This approach displaces the focus from the ontology of the governance of freedom of speech online (and on SM) towards the ways in which governance of freedom of speech online is performed, and ‘assembled’, and thus from a theoretical enquiry about the definition of governance to analysing governance as the product of the association and the resources mobilised to assemble and stabilise the various actors involved.

As seen in the literature, this approach fits with the theorisation of governance as emerging and related to moments or episodes of shocks or emblematic issues (Pohle, 2016a, 2016b; Hofmann, 2016; Hofmann et al., 2016; Ananny and Gillespie, 2017). Considering regulation initiatives as a reaction to episodes that break the routine fits the ANT idea of controversies as the place where the social can be identified and analysed. The sociology of translation reminds us that forms of order are the accomplishments (rather than the origin) of associations between elements, which produce their positions in a controversy through competition with others to better articulate the problem in ways that interest others, enrol and mobilise them.

I thus see the potential for theorising the governance of freedom of speech on SM as a controversy and use ANT’s concepts to point at the elements that make governance ‘visible’. What counts in this method is articulating the position of the various elements involved in controversies. However, instead of adopting an approach privileging one specific actor or event, using ANT the researcher approaches the actors with an initial agnosticism on the roles and hierarchies, leaving the actors themselves to be the ones that provide the indication of who or what does or does not matter.

This theoretical structure draws on the literature's finding (Wagner, 2013; Balkin, 2016; Gorwa, 2019a, 2019b, 2020) that governance is carried out by networks of actors that cross public/private, local/global, and formal/informal divides, mixing forms of statutory, co-regulation and self-regulation of SM platforms and users.

In sum, these insights from ANT lead to an understanding of governance of speech on SM in which, on the one hand, the ontology of digital content governance is the outcome of reaction to public shocks, and, on the other hand, the spokespersons for digital content governance – be they policy documents or public statements – are only part of the ‘actors’ making up the ordering system. The translation theoretical concepts highlight the actors' argumentation techniques and the ways in which actors find compromises or persuade other actors, not only to collaborate for the same aims and interests, but also to adopt the same view of an issue (Pohle 2016a). This theoretical approach considers the strategic role of technology narratives and the role of media (as the disseminators of narratives) in the development and orientation of regulatory measures.

The general critical approach to digital social data reminds us that digital social data – including content and data from social media users – is the result of associations of hybrid elements and it carries a specific power structure with economic and political meaning. Hence, the ways in which this technological artefact are described (or hidden) in the main narratives acquire a specific political implication.

3.6 Key concepts and governance of free speech online

Adopting ANT in the study of the controversy around social media and freedom of expression, it is possible to retrace actors and the issues around which they mobilise; studying how certain concerns and elements have been taken for granted (e.g. black-boxed) and while others have not. Rather than using a priori assumptions about the ontology of the phenomenon (e.g. starting from the assumption that a clear system of global communication governance exists), I investigate the processes through which a form of governance is emerging as an outcome of the associations of all the different elements around matters of concern.

My interest is in the many associated and heterogeneous elements (combining scientific, political and economic elements) composing the actor-world/network surrounding the technical object ‘social media’ (Callon, 1986a). These could be: private companies, users, governments, but also code, algorithms, specific cultural traits developed within SM as the ‘connectivity culture’ (van Dijck, 2013), regulations, specific forms of communication (e.g. tweets, posts, video), and economic system (e.g. platform society). I am interested in understanding what element is succeeding in the process of translation: namely enrolling and assigning specific roles to the different elements, playing the spokesperson of the entities they constituted. At the same time, I also consider the ways in which different elements obey or resist this process, refusing roles or trying to become the spokesperson. In order to understand who is at the centre of translation the study aims to identify the elements that serve as obligatory passage points, i.e. through which all other entities must pass and how, which strategies and problems are used by the actors to ‘convince’ the others to follow their vision, and the different types of displacements, e.g. movements of materials that stabilise the translation. I am particularly interested in understanding what types of actors emerge, whether they can be associated with traditional social categories or whether new ones can be added (i.e. technological actors and semantic elements), and what discursive and material strategies they put into place to succeed in the process of translation.

In the first part of the empirical study, I use the ‘expedient’ of isomorphism and the ‘descriptive imperative’ typical of ANT to try to empirically, rather than theoretically, retrieve the different elements (material and semantic) that associate around an issue. In the second part, to ‘interpret’ the position and effort of translation I rely also on concepts ‘alien’ to ANT, such as pre-existing circumstances or other concepts developed in critical discourse and critical digital studies. As in an investigative search, the aim is to understand who benefits from framing the issue in a certain way. I try to critically analyse the imaginaries mobilised and, using critical digital data, focus on what model of society actors envisage by opening the black-box of SM and platform data technology. In this regard, I am interested in observing the ways in which different socio-technological imaginaries are employed as enrolment strategies. I want to consider which ontological perspective and world interpretation they belong to.

I use ANT analytical vocabulary to try to answer the research questions of this study, by focusing on the actors’ narratives emerging from controversial issues related to social media platforms and freedom of expression. Actors’ roles (e.g., OPP, mediators) and tools (e.g. inscriptions) and different strategies employed to create order as well as the role of ‘technological objects’ in facilitating or hindering the results.

In this study, I consider two different public spaces to understand how material and semiotic objects are created through the associations of heterogeneous elements via practices and discourses. Firstly, I consider the showcase offered by Google, in its collection of websites, as a case study of how the public emerge on the internet. Secondly, I consider newspapers as another environment where the ‘public’ (as in the ‘*group concerné’*) and the controversy can be detected. Drawing from Barry (2001, 2013) I treat the two environments as technologically mediated examples of public spaces, also relying on Dewey’s (1927) definition of public (as in the group concerné) (Marres 2005).

3.7 Drawbacks and limitations of material semiotics

The ANT approach comes with some limitations and criticisms. The classical criticism concerns the symmetrical treatment of humans and non-humans and the accusation of positivism advanced within Science and Technology Studies (Collins, 1981, 1985). However, as discussed in the overview of the theory, it is exactly because of this symmetrical treatment that ANT and material semiotics seem so appropriate for interpreting today's technology.

Other aspects are, in my view, more contentious: in particular, ANT's generalised symmetry and the principle of isomorphism among elements. As there is no difference between human and non-human elements, the difference between micro and macro elements depends only on the resilience of assemblies/associations (and not on intrinsic characteristics of the actors) (Callon and Latour, 1981). The ANT preference for description, and refusal of any forms of reification (e.g. society, capitalism, etc.) or dichotomous categories as the explanandum for associations, can create limitations in the interpretation of the data. If actors are all perceived to be equal, social constructions and configurations, as well as 'exogenous contingencies' such as economic crises, the market, economics, organisations, management or culture may be under-evaluated (Bloomfield and Vurdubakis, 1999).

ANT's flat ontology is in contradiction to conventional social theory, and as Couldry and Hepp (2018) point out, it risks undervaluing capital, sense-making mechanisms, and other systemic characteristics. Furthermore, ANT, as non-representational theory in general, makes no effort to clarify how representational contents and representations become rooted in the universe (Couldry, 2012). In *Unscrewing the Big Leviathan* (1981), Callon and Latour try to explain the concept of isomorphy of all elements, and of how micro actors become macro actors. According to these authors, more durable, solid materials create more stable actors. It is possible to recognise macro actors as they are the ones that tell others what they want, what they will be able to do in the future and in what order.

In Callon and Latour's case study, the story of the competition between the French electrical company Electricité de France (EDF) and the car factory Renault are presented as examples of macro actors in the controversy that concerned the introduction of electric cars in France in the 1970's. As in other writings (Callon, 1986b), the authors use the failure of the EDF programme to stress that the process of translation is constantly undergoing a process of resistance from other elements. The case is exemplary as it shows how EDF almost succeeded at translating the interests of the whole system of production of cars, switching to electric cars: assigning to the CGE (Compagnie Generale d'Electricité) the role of developing the electric motor, and demoting Renault to the manufacturing of the car bodies. However, the programme failed as the technological elements (in particular the catalyst in the electrical engine) did not work as planned, and this led to the collapse of the shared 'vision' of the world, and the ultimate abandonment of the project (Callon 1986b).

In this sense, the authors want to prove that the strength of an actor depends on the power to break off and bind together, rather than on the 'sise' of the actor, as in their example EDF which is a state-subsidised company competing with a 'smaller' competitor (i.e. Renault).

In *Unscrewing the Big Leviathan*, the authors argue that macro actors are those that can enrol more durable elements and consequently can stabilise associations more easily. This entails that macro actors have more elements that are affected and can help the effort of translation. Going back to the above example, if we compare enterprises such as Renault with a startup producing cars using 3D printing, we can identify similar elements composing the respective 'actor-networks': workers, CEO, machines (3D printer), investors, consumers, selling strategy. However, the number of single elements that compose the actor-network Renault is necessarily bigger, as would be their effect in the event of the 'malfunctioning' of their programme of action. An economic loss for a large enterprise can affect other macro actor-networks, for instance the state. A government in a democratic state has an interest in and responsibility for the economic performance of the whole actor-network state. Such economic and political consequences would not result from the closure of a startup (it would if there were vast numbers of start-ups or small/medium sise enterprises).

So even though all actor-networks are made of elements (analogous to atomic particles) of the same sise, some actors include a high number of associations, or mobilise more resources (e.g. money), increasing the potential for externalities – influences beyond the original circuit of actors. The amount of money/economic power it can mobilise gives the actor strength also in ANT terms: it gives more power to break off or bind together. The 'sise' of an actor potentially results in its actions leading to 'externalities' (see Callon 1998). Economic power attracts more shareholders, potentially mobilising other actors and aligning them to the main interest: i.e. growth and survival of the platform. Using an ANT type of example, consider the case of *Phytophthora infestans*, the parasite that destroyed the potato crop in Ireland in the nineteenth century, killing almost half of the population and leading to one of the biggest migrations from Europe to North America. Would it have had the same effect if the entire Irish economy was not built on English occupation and potato monoculture? As much as an actor-network is composed of isomorphic elements, some actor-networks manage to enrol more durable elements and have a larger potential to affect the environment. To translate other actors into their programme of action, actors have to enrol other actors and use resources and strategies. The more elements they can mobilise (in terms of money, sub-actors, etc) the bigger they become. For instance, the media can be used as a 'platform' to enrol citizens. If the actor owns the platform, and at the same time owns the technology that regulates the flux of information, it can mobilise a potentially enormous number of users and influence public opinion activating different imaginaries as strategies to problematise and enrol other actors. For instance, unlike ANT, Edwards (2016) recognises asymmetry of relations between the different elements of the system. Edwards accepts the idea of distributed agency, and the idea that all powers are dependent on the other actors (power-dependence), but he stresses how certain asymmetries imply different capacities of different actors to project their power (more in line with the material realist perspective and Bourdieu's concept of pre-existing conditions). This connects to Callon's idea of externalities (1998), as something that cannot be avoided. In Callon, every attempt at framing results in externalities and overflows. Callon (1998) uses the example of the chemical plant that pollutes the rivers with toxic wastes to show that actors' boundaries are both a technical and a political matter. The chemical plant is made of a clear and distinguishable set of elements/actors; however, the boundaries are challenged and once that pollution spills to the rivers, since this opens the network to the involvement of other actors. Framing an actor means, in other words setting its boundaries; clearly distinguishing with which actors it is associated, and this requires investments (economic, physical and technical).

> [T]his work of cleansing, of disconnection, in short, of framing, is never over and that in reality it is impossible to take it to a conclusion. There are always relations which defy framing (Callon 1998:189).

In this sense all actors (actor-network) are always contained by other networks; arrangements that can cross different domains (Barry, 2001).

More in line with critical theorists, I consider SM characteristics, as their business model and the forms of exploitation this entails, as fundamental elements to understand the dynamics of power and specific interpretations of the world (imaginaries) behind platform technology, and that contribute to create more equal alternatives (Fuchs, 2014; Casilli, 2017, Gregory 2017). For this reason critical data studies are important, since they consider the context where data are produced and the domination that the narrative surrounding data creates.

Another criticism is that the excessive reliance on the configuration of the actor-networks to explain why and how some actors are more empowered, while others are disempowered, could seem ‘incomplete’ (McLean and Hassard, 2004; Boullier, 2018). This criticism applies particularly to the methodology section and I will discuss it in more detail in the next chapter. However, Boullier (2018) warns against the excessive reliance on the description of the structure, rather than on the ‘meanings’; semantic messages that create/divide associations between actors. In this regard, Couldry (2008) underlines that ANT does not explain how and why a certain actor has taken these beliefs for granted and how they have shaped the actor’s interests or what comes after the establishment of networks, once that order has been obtained. In particular, ANT does not say much about how actors’ ideas, programmes of action or ontologies are shaped by the underlying features of the networks in which they are situated.

Some criticism concerns the interpretation of power relations but also the role of the affective dimension and importance given to the neglected elements (Latimer and Munro, 2006; Puig de la Bellacasa, 2011). In this sense, every effort of translation and mediation ‘enacts’ the separation with other forms of association. It is a case of matter of care, vs matter of concern. According to Annemarie Mol (2008) the logic of care modulates the logic of ‘rational’ choice. Matters of facts are always matters of concern. Some matters of concern are also matters of care. It is important when studying controversies to pay attention to the neglected things. In Callon’s (1986a) example of the scallops, it would be the fisherman's children, expecting presents for Christmas, who are not included in the network. Or in the case of an SUV, it is the flowers, the trees, babies that suffer from pollution. The choice of including or excluding certain aspects has a highly political consequence (who is in? who is out?) which has important effects on the direction and legitimacy of an ordering system. In this regard, feminist data studies scholars, such as Kate O’Riordan (2016) suggest beginning social enquiry by asking what is not included – opening the black-box of what data and data visualisation does not show.

3.8 Conclusion

In this chapter I have described the theoretical framework that I use, ANT and critical data study, providing arguments in support of this choice. In particular, I argue that the combination of these two perspectives can be very interesting to orientate my investigation on the governance of freedom of speech on social media. The ontological and epistemological assumptions are in my view particularly fitting to answer the research questions and aims that I have developed through the literature review. In particular, theorising the governance of freedom of speech on social media as a controversy, I can contribute to the literature of freedom of speech governance as an emerging effect of public shocks, without a specific hierarchy of actors and formal decision-making process. The hybrid ontological conception of the social includes different types of institutional (states) and non-institutional actors (private companies, civil society, as well as technology) in the group of elements that have to be considered as the origin of regulatory initiatives. In particular, it is inclusive, as it refrains from focusing on a single actor or single environment, and at the same time, critical, since interpreting digital data and technologies as social products highlights the political implications of technology.

As a result, not only exploratory but also normative research questions can be answered, for instance what governance initiatives related to freedom of speech on social media emerge? What actors are involved? What kinds of power relationships are they establishing? What are the dominant narratives on free expression and technology? What does all of this teach us about society's broader socio-technical dynamics? In the next chapter I describe the methodology I have used to answer these questions: controversy mapping, and how I have used it for the purposes of my study.

4 Methodology

4.1 Introduction

In the previous chapter, I argued that ANT combined with critical data studies provides a ‘complete’ theoretical framework, suited to the findings and recommendations from the literature about the governance of freedom of expression online, and my research questions and aims. I explained how the governance of freedom of expression can be theorised as the outcome of regulatory initiatives prompted by socio-technical controversies surrounding social media platforms. I explained how this theorisation fits with the idea of governance as reaction to ‘public shocks’ that break the routine or pre-existing forms of decision making concerning public expression on social media highlighted in the literature and how this contributes to answer my research question about the controversial issues that have the power to mobilise different elements involved. I explained how ANT and critical data studies can provide an answer to my question about the competing narratives presented in the media, and the type of governance that is emerging from the interaction of all these elements.

As anticipated in the presentation of the theory, ANT not only provides concepts that are useful for the understanding of the roles and dynamics linking different actors; it also comes with a set of methodological tools for the empirical collection of data about controversies. This methodological toolkit fits the purpose of empirical exploration of the actors and emblematic issues (Pohle, 2016a, 2016b) and narratives present in a controversy. It also fits the study of governance of freedom of speech on social media, where, as the literature has shown, the actors involved and their roles and means to influence outcomes are not clear or institutionalised. In the chapter 3 I explained the fundamental role of the media and the public of non-experts in performing controversies. Bearing this in mind, I explain in this chapter how I focus my data collection on statements about freedom of expression and social media produced by the public at large in two different media environments: web pages and newspapers (in the UK). I start the chapter with a description of my research strategy, which is a combination of inductive and deductive moments, and in general an interpretivist and realist approach. I present the main aspects of the methodology, as developed within Sciences Po (the Paris Institute of Political Studies), University of Amsterdam and Warwick University. I discuss limitation of controversy mapping's methodology, focusing in particular on the problem of digital bias and the role of researchers in this methodology. In the last part of the chapter, I explain the selection of my case study and sources for my data collection: i.e. websites and articles in the British press. I conclude with an overview of different steps of my data collection and analysis and ethical considerations.

4.2 Methodology and research aims

In order to achieve the objectives of my study, I have applied the theoretical framework of ANT in one of its empirical applications: controversy mapping. ANT has indeed often been described as a set of methodological tools, rather than purely a theory (Law, 2007). Controversy mapping (or issue mapping) is a methodological toolkit developed by scholars in STS and ANT studies to train students in the observation and description of socio-technical controversies. There is no real difference in the theoretical background and practical application of either one or the other, even though they describe slightly different things. Marres (2015) prefers to use the term 'issue mapping' rather than controversy mapping, when, in order to detect actor relations, researchers start with a specific topic and from there they move to detect whether there are emerging formations of issues. In this study I use 'controversy mapping' when I want to refer to the larger socio-technical controversy and 'issue mapping' when I am referring to constituent elements of the larger controversy.[2]

Scholars at different universities have developed different protocols or 'recipes' to help make sense of the different elements involved in a socio-technical controversy. The mapping exercise in this study relies greatly on Venturini (2010, 2012, Venturini et al.,2015) and Rogers (2009, 2013a,

[2] The methodology for mapping social controversy developed over the years as the result of the work of an expanding international community of researchers, a non-exhaustive list includes the MACOSPOL (Mapping Controversies on Science for Politics) project at SciencesPo MediaLab and partners worked in direct contact with Bruno Latour on the empirical application of ANT. Scholars Nortjie Marres and Richard Rogers merged STS and ANT with political theories of public discourse (such as Walter Lippmann 1927 and Dewey), and applied the theory and methodology in research centres as the Digital Methods Initiative (University of Amsterdam), and the Public Data Lab and the Centre for Interdisciplinary Methodologies at the University of Warwick.
Scholars empirically applying ANT have produced a large literature and case studies on socio-technical controversies, massively contributing to develop new tools for the gathering, analysis and visualisation of data. These methodological toolkits are the practical application of the theoretical study of controversies, respecting the imperative t0 'follow the actors', i.e. understand their role in the staging a controversy in the public realm (Latour, 2005).

Rogers et al., 2015), as well as on Mathieu Jacomy and Noortje Marres' advice on practical workshops[3]at Sciences Po, Amsterdam and Warwick University.

In this study I followed the recommendations presented in the work *Issue mapping for an ageing Europe* by Rogers et al. (2015), where the authors use Latour's (2005) theory to orient the exercise of mapping controversies. Firstly, they recommend considering actors in groups only as 'group-like' formations whose boundaries have to be constantly redefined (Venturini, 2010; Rogers et al., 2015). Secondly, they recommend focusing on what causes the 'translation' of interest between elements in a network and how some actors manage to engage others on their side to do something collectively (Rogers et al., 2015: 17). Thirdly, they stress the importance of mapping not only "human-to-human connections or object-to-object ones, but the zigzag from one to the other" (Rogers et al., 2015:17). The fourth instruction concerns the focus on how facts become issues (or in ANT terms matters of concern) in the public life (Rogers et al., 2015:17). This corresponds to what Callon describes as 'problematisation', in his example of the scallops of St Brieuc Bay (1986) discussed in the previous chapter. Finally, the last instruction concerns the effort that the researcher should make in order to give an account of actors' different positions on the issues while trying to maintain as much agnosticism in the interpretation of positions, i.e. what is also called the "the second-degree objectivity" of the researcher (Rogers et al., 2015:17).

Based on these recommendations, I have followed what Venturini calls the 'pathway' process for the practical mapping exercise (Venturini 2010, 2012, Venturini et al. 2015). Following ANT's interest in tracing the flows and the movements in the network of actors, Venturini identifies five different steps that can be used to uncover different layers of a controversy and links between actors (Venturini, 2010: 265). These steps have become a sort of 'procedure', a standard way to perform controversy mapping in the context of the MédiaLab in Paris, but also in several other research centres, including the University of Amsterdam (the 'pathway' is also followed in Rogers et al., 2015) and Warwick. Adapting Venturini's (2010, 2012) pathway to mapping controversies, in my methodology I have included three phases. The first one is an empirical detection of actors,

[3] Workshop "How to map issues? Mixing methods for the study of topical affairs" - September 2016 - University of Warwick. Workshop: 3rd FORCCAST Summer School on 'Controversies and Conspiracies. Conceptual boundaries and empirical practices' – September 2016 – Sciences Po MédiaLab Paris. Workshop: Digital Methods Summer School 'Only Connect? A Critical Appraisal of Connecting Practices in the Age of Social Media' – July 2016 – Digital Methods Initiative (DMI) and DATACTIVE – University of Amsterdam

based on the retrieval of statements, either textual or in any other form (such as slogans, phrases, keywords, terms, etc.) that represent examples of opposite views or controversial issues where the controversy takes place (step 1 in Venturini's pathway).

In this phase, following the theoretical background, the methodology assumes that actors involved in the controversy can be detected from the statements that they have left in the public sphere (according to the definition of 'public' described in the previous chapter). Also, it presumes that controversies are never a binary opposition between two arguments, but rather that they are composed by several different positions (Venturini, 2010, 2012). The identification of actors provides an idea of who wants to 'animate' to create the controversy and who is 'left out'. It is thus an interesting point of departure for the study of the 'neglected' elements (O'Riordan, 2016).

In the second phase I used the statements collected to identify a list of the actors who enunciated them. This phase included an exploration of the identity of the actors, and the eventual gathering in group-like formation (step 2 in Venturini's pathway). It also included an exploration of the relationships of the statements with each other (either through citations as in academic publications, or mentions in articles or hyperlinks, etc.) (step 3 in Venturini's pathway).One of the most employed and taught methods in controversy mapping is to retrieve statements using search engines online, using the web as a source and URLs as proxy for authors, and subsequently to explore the relationships between actors following the hyperlinks present in the pages (I describe this method in greater detail below).

The next phase involves the exploration of narratives and what values and ideas they represent (phase 3) and how actors relate to them or compete against them. In this part I moved from the actual terms and statements to the narratives that emerged from the statements, and I analysed what socio-political stances they were taking (i.e. absolute or relative conception of freedom of expression, etc). From there I tried to observe how different groups of actors relate to the larger 'ideas' behind the specific topics and how these narratives have changed through time (steps 4 and 5 in Venturini's pathway). In this part of the interpretation I have relied particularly on critical data study, demystifying the discourse of neutrality of digital data (Iliadis and Russo, 2016).

This pathway for controversy mapping protocol has become quite standard in several workshops and trainings. However, it is not without criticism. Another researcher from Sciences Po, Dominique Boullier (2018) criticises the excessive reliance on the structural properties of networks of actors built with hyperlinks, which do not say anything about the content of the statements. In particular, he criticises the shift from the identification of actors to the identification of their values using this pathway. Especially in the case of data from web pages, the construction of a network based on hyperlinks does not necessarily give information about the positions of the actors concerning the issue. In Boullier's view, it is by engaging with the content, rather than looking at the 'position' of the actors, that processes of translation can be identified.

In this study, I firstly developed a list of search terms taking into consideration the affordances of my research tools (a procedure described below). Using an extremely broad definition of the basic elements of the controversy I let the data show how this 'issue' is staged/framed according to different groups (in a sort of 'follow the actor' position). In order to identify the actors I analysed the data looking for 'authors' of statements on websites and newspapers. I had to use different ways to identify actors, according to the type of document that I was considering (i.e. web pages or articles from newspapers). However, in order to interpret the data in the light of sociology of translation I combined the elements underlined in Venturini's pathway with the analysis of contents and the identification of 'key issues' with the power to mobilise actors, more in line with the focus on the 'contents' that circulated the most stressed in Boullier (2018).

4.3 Suitability of controversy mapping for research on governance of speech

As seen in the literature review (chapter 1), in the last years increasingly more studies on governance online have been adopting ontological and epistemological approaches in line with STS or ANT. The majority of studies converge on the definition of a plural, 'complex' and 'heterogeneous' social space, with a tendency towards a materialistic approach able to account for the role of non-human entities in the 'social', and an interpretation of power as diffuse and emergent. Adopting an ANT-informed approach to the study of SM platforms, I started to approach my object of study (i.e. freedom of speech regulation on SM) as a controversy. Through this lens, my interest is to study how governance of free speech on 'social media platforms' moves from being a matter of fact or a matter of indifference to become a matter of concern, consequently unveiling the 'whole machinery behind the stage' of actors and dynamics involved in the definition (Latour, 2008). Bearing this in mind, I opted to follow an empirical application of the theory and use controversy mapping. Not only is this methodology in line with my theoretical framework (both ANT and critical data studies), but it also fits well with the specific interest in the role of narratives about technology.

In the previous chapter I explained the fundamental role of the media and the public of non-experts in performing controversies. Controversies are made by the group of actors that produce arguments about them. The engagement with the issue is only detectable from the moment that an actor 'publishes' or makes public statements about the controversy. By doing so they contribute to the creation of a collection of statements initiating or connecting to pre-existing narratives about technology. Similarly, theory shows how media play the fundamental role of producing and reproducing (or silencing) narratives. In the study of controversy, it is essential to understand what is the 'public' and what the media say about the narratives raised by the public. I thus decided to use controversy mapping tools to collect the statements about freedom of expression and social media produced by the public engaged in two different media environments: web pages and newspapers (in the UK).

The idea of using web pages is directly connected to the major application of controversy mapping, and it rests on the idea of traceability of materials online. Digital data are used to develop accounts of social processes (Ruppert et al., 2013), taking advantage of the 'social traceability' created by digitisation, and the masses of data created by social media platforms and search engines (Beer and Burrows, 2007; Marres and Gerlitz, 2016). As I describe in greater detail later in this chapter, controversy mapping online uses web pages to understand the position of actors engaged on a divisive topic. It is based on the collection of statements which can 'show' the values and attitudes of the main actors in the controversy. However, this method comes with a number of limitations, and in order to gain a better understanding of narratives and shocks that stimulate the reactions online, I integrated my observation of websites with the study of articles from British newspapers. The choice of using newspaper articles is based on the interest in media as a fundamental tool in the production of the controversy (Barry, 2001; Couldry, 2012). Newspapers do not produce original statements on the controversial issues, but create narratives, focusing on storylines and exemplary cases that become the embodiment of specific problematic issues. Moreover, newspapers provide a stage for the controversy and important insight on what elements are object of disagreement and what are taken for granted.

4.4 Epistemological and ontological positions

As discussed in the previous chapters (chapter 2 and 3), the study of regulation of freedom of speech on social media platforms raises several ontological and epistemological questions. The first type of question concerns the elements that have to be considered as part of the research problem: who are the actors that create the governance? How do I decide which actors to include and to exclude? how do I avoid focusing always on the 'usual suspects' (i.e. big companies, states, etc.)? The second type of question concerns the way in which it is possible to study and understand this social space, given the ontological assumptions. How do I find these actors, and how can I understand the type of relationships and dynamics that are taking place among them? Controversy mapping as a method is an attempt to create a procedure to answer these questions.

As noted in the literature review and theory (chapters 2 and 3), material semiotics' ontological perspective involves a suspension of the traditional theoretical dichotomies such as subject/object, or agent/structure, as it prefers to focus on the process through which entities acquire attributes and agency as the effect of semiotic and material associations among the elements. Similarly, critical data studies recognise the social constructedness and the materiality of data (called socio-technical assemblages). According to the theoretical framework, the actors that compose a controversy are the ones that have taken a position or are interested by a matter of concern; in other words, only those that are actively involved in shaping the issue, the so-called '*group concerné'*. They do not represent the public in general or a form of rational public debate as in the 'Habermasian sense' (Habermas, 1989), but in the sense understood by Dewey (1927), for whom every issue assembles a specific public which is made up of the network of actors involved through negative externalities (Dewey, 1927).

It is difficult to define controversy mapping and the material semiotics approach in terms of traditional categories of research design. Muniesa (2015) defines ANT as a distinctively materialist, radically constructivist approach to social theory and to empirical research. This is because the whole theoretical and methodological framework is based on the concept of a hybrid social system, and as such it is interested in the materiality (i.e. realist approach) as much as the symbolic, semiotic elements (i.e. constructivist approach) of both objects and language.

With regard to the specific method, controversy mapping, it is in one sense an inductive method, since authors in ANT stress the importance of refraining from imposing the researcher's view on the issue (as in the second-degree objectivity mentioned above); rather, they invite the researcher to discover processes of translations by following the actors, and recognising issues as defined by the actors themselves. Considering traditional techniques of social research, controversy mapping resembles (digital) ethnography and ethnomethodology in certain respects, as it requires the researcher to 'follow' the actors and observe the meanings and relationships that develop within the community or situation they are researching. An inductive approach fits the object of this study, as it adapts to the lack of a clear structure of reference of social actors in the context of content regulation. In this context, and in the light of the gaps in research highlighted in the literature, an empirical observation of actors appears to be the most 'reliable' form for identifying the group of statements and elements involved in the governance initiatives.

ANT is open to mixing quantitative and qualitative research methods in the process of data collection and analysis, as it recognises the existence of a material aspect of reality, objects which can be 'quantified'. At the same time, in all these studies quantitative data are always interpreted from the 'qualitative' point of view, as is often the case for networks of actors reconstructed via controversy mapping (see Jacomy et al., 2016; Boullier, 2018; Ooghe-Tabanou et al., 2018). All forms of quantification of data are aimed at representing the different perspectives/points of view that are included in the process of translation, in the possible venues/arenas/spheres where the issue is brought to life (Venturini, 2012). This might lead to an overrepresentation of minorities/extreme positions so that focusing only on 'directly' involved actors means some elements are necessarily neglected.

As I introduced in the description of the theoretical framework, in this study I intend to interpret the results in the light of concepts developed within sociology of translation, as well as within the approach of critical data studies. In the interpretation part, I will then switch to a more deductive approach, where I will discuss the implications of the findings emerging from the inductive exercise, in the light of theoretical concepts described in the literature review and theoretical chapter (such as datafication and algorithmic governance).

4.5 Data collection tools and reason for use

As mentioned above, the main sources of data for this study are written statements – inscriptions – left by actors involved in the discussion of controversial aspects of freedom of expression and social media (in the years 2015–2018). In order to collect this type of data I relied on digital tools of data collection (for the collection of statements from web pages), and archives for articles. Of these different types of data collection, digital tools in particular have been the object of scholarly debates.

4.5.1 Digital tools for collection of statements

In controversy mapping, digital tools are particularly appreciated by scholars as instruments to identify the relations between key actors active in a controversy. As much as ANT has found a practical application in controversy mapping (Venturini, 2010, 2012), controversy mapping has in some way found a 'spontaneous' application in digital methods/tools of data collection (Rogers, 2009; Rogers et al., 2015).

> Using digital methods in intersection with mapping theories is currently of relevance as the web and the ubiquity of digital technologies are affecting how a social issue is communicated, and staged (Rogers et al., 2015:29).

Digital data are often organised and arranged in ways that make it ideal for controversy study, such as using network and textual analysis and visualisation to track the evolution of controversies across many platforms and over time (Marres, 2015). The concept of online groundedness states that traces of the social can be found in the form of content and metadata, relationships and interactions, as well as links, shared vocabularies and keywords (Rogers et al., 2015:24, 44). On this basis, digital data are used to develop accounts of social processes (Ruppert et al., 2013), taking advantage of the 'social traceability' created by digitisation, and the masses of data created by social media platforms and search engines (Beer and Burrows, 2007; Marres and Gerlitz, 2016).

In recent years, a large number of digital tools for data analysis and visualisation have been developed, either by repurposing tools originally created by tech companies for research aims (e.g. the use of search engines as tools to mine the internet as an archive), or creating new 'custom-made' tools or what Rogers (2009) calls digital native methods, i.e. developing new tools specifically tailored to the needs of researchers. A quick look at several research centres' websites reveals an increasing variety of both types of tools. In Paris, the MédiaLab has tools for the analysis of academic references links (i.e. scientometrics), for the construction of hyperlink networks (i.e. Gephi and Hyphe), and for the semantic analysis of texts (a tool called ANT). The Digital Media Initiative (DMI) in Amsterdam has developed tools for the scraping of websites (i.e. Google Scraper), the construction of networks based on hyperlinks (i.e. Issue crawler), the gathering of data from Twitter (i.e. T-CAT), and many more. Many of these materials are open access and available in the repository of the websites of these projects.[4]

Marres and Rogers (2005) have developed a method to delineate controversies about techno-scientific issues on the Web, following hyperlinks among pages dealing with a given issue. They use the term 'issue-network' to describe a heterogeneous set of entities (actors, documents, slogans, imagery) that have configured into a hyperlink network around a common problematic, summarised in a set of keywords. Issue-networks are not public debate, but as in Dewey (1927) are networks of actors assembled around an issue. The research interest is thus how these networks involve affected actors in the articulation of the issue, if the issue-definitions capture the ways in which actors perceive the controversy, and lastly whether the articulation of the issue, and the organisation of a public in the issue-network, contribute to the issue being addressed (Marres and Rogers, 2005).

In this study, drawing from Marres and Rogers (2005) as well as Jacomy et al. (2016) I delineate the controversy around techno-scientific issues using the Web as one of the main sources, following hyperlinks among pages mentioning the issues relating to freedom of expression and social media platforms. Among all the tools for mapping controversies online I employed a tool for the gathering of data from search engines called Google Scraper and a tool to expand the original list of URLs called Hyphe.

- The Search Engine Scraper (former Google Scraper) was developed by the Digital Methods Initiative laboratory at the University of Amsterdam. The Search Engine Scraper searches

[4] http://blogs.cim.warwick.ac.uk/issuemapping/
http://mappingonlinepublics.net/resources/
https://wiki.digitalmethods.net/Dmi/DmiAbout
http://densitydesign.org/course_projects/climate-change-controversy-report/
www.mappingcontroversies.net/Home/PlatformOverview
https://www.digitalmethods.net/Dmi/ThingsInternetResearchersShouldKnowAboutGoogle
https://www.digitalmethods.net/Dmi/DmiProtocols#Issue_Networks

data with a particular search query and outputs a list of the results retrieved by the search engine. When I did the first data collection the scraper was only available for Google. Today, researchers can select which search engine to scrape, allowing them to compare the results returned by different engines with the same query.

- Hyphe is a crawler which reconstructs direct and indirect relationships among web pages, following links between URLs and adding new URLs to the original list. In this study, the crawl was done starting from the list of URLs that I had collected using the list of keywords in Google, and setting the crawl at 'depth 1'. This means that the crawler gathers hyperlinks to web pages and documents only connected directly to my original list. Eventual indirect connections might appear, but only if they are the combination of direct links between two of the original URLs. This crawler retrieves two levels of connections. The first one is at the level of relationships between domain names (e.g. the website that can be considered the origin of the URLs collected). The second is at the level of specific URLs or pages, which means that it retrieves only the connections with the specific pages (e.g. documents) that were collected with the queries. If the first level can show the relative positioning of large actors in the internet environment, the second one is useful to give an idea of how a specific discourse/issue is articulated online, and is the modality usually employed for issue-networks, as in Marres and Rogers (2004). For newspaper articles I used the LexisNexis repository (available through Cardiff University library).

In the empirical chapters I describe in greater detail the construction of the sample of the documents collected (i.e. the corpus).

4.5.2 Implications of using digital tools for data collection

Of all the tools for data collection that I have used, digital tools for data collection are the most debated and the ones that present most complications. As discussed in the theoretical chapter (chapter 3), scholars generally recognise the validity and the interest in using digital social data; nevertheless, there is growing awareness of problems and bias created by the 'overreliance' on digital networked data and methods (Marres, 2017:311). Scholars have long discussed how digital social data can contribute positively to traditional research and what the main limitations are (boyd and Crawford, 2012; Edwards et al. 2013; Housley et al. 2014; Tufekci, 2014).

Concerning the use of digital tools for social research, some issues particularly attract the attention of sociologists. These are issues related to the ontology of digital social data and biases linked to the use of digital research tools and the epistemological bias that derives from these biases. Sociologists are aware that every tool we use in research, as well as our own presence as researchers, is de facto contributing to define the 'object' of our own study. However, the problem concerning the ontology of data in the digital space adds new layers of complexity. Mapping social dynamics using digital data requires specific tools whose own features need to be considered.

In the current situation of the internet, the only way for research tools to access digital data is by gaining access through interfaces provided by commercial companies. Whether it involves scraping web pages via search engines, or collecting SM platforms' data through public API, the main points of access to digital social data are controlled by ICT companies. As discussed in the chapter 3, scholars in the social sciences have been concerned by the ways in which digitisation and private companies producing digital social data are changing traditional social research (Savage and Burrows, 2007). Commercial data collection and research become increasingly entangled, and dependent on influencing norms and values relevant to scientific knowledge production (Richterich, 2018b).

In the case of controversy mapping, the problem is even deeper: the digital is one of the domains where controversy and public participation takes place. SM platforms and websites are media technologies: in so far as they are based on the public, but at the same time build the public (Gillespie, 2010). The bias associated with digital data collection methods is at risk of weakening controversy research because it makes it difficult to understand whether we are observing the controversies or rather the digital environments that make them observable (Venturini and Guido, 2012). Researchers have to deal with the results of the ever-evolving algorithms developed by private companies, such as Google's web search, as well as the algorithms behind the functioning of social media platforms.

It seems that scholars have progressively abandoned the idea of minimising the effect of tools (Madsen, 2012), moving instead towards an affirmative approach, able to highlight the role of digital devices in the dynamics of social, political and public life as organisers of relevant socio-technical formations (Gillespie 2010).

> [I]n relation to digital devices, then, we need to get our hands dirty and explore their affordances (Ruppert et al., 2013:32).

In recognition of the embeddedness of the digital in every aspect of the social world, it is impossible to use digital tools without reflexively considering how:

> [D]igital devices themselves are materially implicated in the production and performance of contemporary social life (Ruppert et al., 2013:22).

Some authors suggest that researchers learn from the tools and repurpose methods in order to ‘exploit’ the specificity of the medium on which they depend (Rogers et al., 2015). Others, like Marres, argue in favour of adopting an empirical approach, i.e. including the role of technology in the definition of the object of study (Marres, 2015). The empirical approach requires one to consider the devices and the technologies that produce the data as part of the inquiry itself. Treating the ‘ambiguity of online issue formations as a topic of critical inquiry’(Marres, 2015:673) means accepting the inherent ambiguity of the empirical object of research, and that every issue is formed by both substantive and technological dynamics. ‘Substantive’ dynamics are the ones related to substantial aspects of the controversy (e.g. conflicts around the definition of freedom of expression or responsibilities of actors), while technological dynamics are the ones related to the media and technological environments where the topic is debated (e.g. web pages, SM platforms or newspapers, television) (see Marres, 2015). This means including in the study a critical assessment of how digital affordances (e.g. social media metrics, retweets, likes, hashtag and any other measure of ‘trend’, as well as the structure of a web page and its hyperlinks, and the ranking algorithm behind any search engine) have an influence on shaping the issues (Graves and Anderson, 2020). Taking the structural aspect into account, Marres is optimistic about the capacity of researchers to disentangle, and rebalance, the power of machines and algorithms. Rather than dismissing digital methods, she promotes a constant reconfiguration of the research, i.e. based on openness and the flexibility to discuss either of the aspects, depending on the highlights from the data. For this study I adopted an affirmative approach by critically inquiring how Google’s search engine works, from the selection of keywords to the ranking orders of the web pages. Below I describe the process I followed to select the keywords to use as starting points for my queries on Google, and the considerations about Google’s ranking algorithm.

4.5 3 Keyword selection

Studies often report that their initial query is based on a list of keywords or URLs compiled by experts or defined through search engines (Rogers, 2013b; Marres, 2015; Gray, 2017), even though the process that led to the selection of keywords is not discussed in detail. This initial step can open criticisms on the account of the controversy as an artefact of the arbitrary concepts you have used. However, scholars in controversy mapping have considered possible procedures to answer this criticism. According to Rogers and Zelman (2002), search engines were not considered the best starting point for selection, since "The top ten returns may not lead directly to an issue network" (Rogers and Zelman, 2002:15). However, in a more recent study, Rogers et al. (2015) did rely on search engine queries to create a map of ageing in Europe, using broad terms of research in the local version of Google. In the study they analysed the results (i.e. a list of websites URLs) accepting the local version of Google ranking as ordering factor and they explored the websites and collected the issues that were derived via qualitative inspection (Rogers et al., 2015:106).

In line with the scholars' recommendation to adopt an empirical and reflexive approach to the use of digital tools (Rupert et al., 2013; Marres, 2015), in this study I have considered the role of the medium in the staging of the controversy. For the collection of data from websites, I verified the ways in which the issue was defined in Google.co.uk, using Google's trend vocabulary. The aim was to determine the terminology used to describe the main 'concepts' composing the issue. Since in controversy mapping it is necessary to make sure that the keywords used will make it possible to find who is actively 'talking' or producing 'traces' (Latour, 2005b) about the issue in that specific sphere of public, and what are the different framings that are adopted to create (or minimise) the issue.

The output was a list of synonyms and related concepts for 'social media' and 'freedom of expression' to use for the query in the search engines: 'Freedom of speech' OR 'freedom of expression' AND 'social media' OR 'social network*' OR 'social media site*' OR 'social media platform*' OR 'business social media' OR 'new social media' OR 'social media content' OR 'social media governance'.

I acknowledge that in this way I have a 'skewed' list of results, or anyway that the list already frames the issue in a specific way: i.e. as connected to freedom of expression. However, I justified this selection on the basis that in this way I was able to reduce the number of results to the ones more related to the object of my study without imposing a framework too strict. By using the same keywords at different points in time I was able to monitor the variation in the amount of results and evaluate whether to introduce changes in the list to make closer to the voices of the actors (e.g. an eventual reduction would have signalled a change in the controversy and the necessity to change group of words).

4.5.4 Google's ranking algorithm

Another consideration concerns the influence of the ranking principle in search engines. In this regard, it is an assumption of the methodology that the Web decides upon relevance i.e., "that 'the Web', one way or the other, is the judge" (Rogers and Zelman, 2002:11). Researchers can exert a minimal control by selecting to use structured or unstructured queries. An unstructured query is open ended, and it is adapted to situations when researchers do not know exactly what they are after. In this type of search, researchers rely completely on Google and the PageRank algorithm's ability to provide significant results, as PageRank will present first in the list those sites which received the most links from the most influential sites or reputable sources. In a structured search, researchers are interested in specific terms and for this reason they are queried using 'quotation marks', instructing Google to return only sources which use those specific terms. In this way, 'equivalent' terms are excluded from the results. The list of web pages was ordered by Google based on relevance. The DMI research centre in Amsterdam suggests to use the SEO consultancy MOZ to keep track of the changes algorithm (which can happen as much as 500-600 times per year). While the bulk of these improvements are small, others are notable in terms of their effects on research (from the DMI website).

While building the corpus I came across another important limitation of performing controversy mapping using web pages as data source: i.e. the internet does not work well as an archive of events that happened in the past. URLs and web pages can be updated several times and change their content (Rogers talks about "unstable media" and "ephemerality of content"; 2015:31). In order to get a coherent set of web pages linked to the controversy, I had to perform different data collections from December 2015, checking their publication and update dates. For the final corpus, I only kept the URLs lists collected in December 2015 and 2016, September 2017, and April 2018 (last data collection) as they cover the largest time span in the most 'regular' way (see Table 5.1

in chapter 5). During the development of the analysis, I had to check the availability of the web pages, and exclude the ones that are no longer available (not even as archived form in the internet Wayback machine[5]) from the analysis. Moreover, search engines only retrieve 'public' pages; i.e. they do not permit access to materials which are beyond a paywall (such as the full text of journal articles, or materials protected with passwords such as social media posts protected with privacy settings).

Adopting an empirical approach and taking into consideration the affordances of digital tools (such as the influence of keywords and ranking algorithms in the final output of search engines) does not solve another issue related to the use of digital tools for data collection. This relates more to the issues presented in Savage and Burrows in the "Coming crisis of empirical sociology" (2007), and the realisation that all data collected from the digital are created in a system, where private companies are competing with academia for the creation of validated knowledge of and expertise about the social. In practical terms, it is impossible to collect data from the internet which have not been created and shaped with the purpose of commercial companies. As discussed in the following chapters, the structure of web pages is clustered with commercial links for advertisements. This is even more true in the case of social media platforms' infrastructures, where the visible content and associated metrics are the result of commercial profiling. Rather than giving us an idea of society, content and the data of users describe to us, the system through which the companies survive, and the metrics and content available should be subject to a critical approach both in terms of reliability and more generally as a product of a system of exploitation (see Poletti and Gray, 2019).

In the study, I checked which categories of actors occupy higher positions in the different data collection (using the ID assigned by the crawler to each URL at the moment of the data collection). In general, web pages from civil society and NGOs, as well as academia and think tanks, appear among the highest positions in the ranking presented by Google.co.uk. In Figure 4.1, in the first 100 results presented from Google, academia and NGOs (yellow and red bars) account for the majority of results and the news media (blue) are the largest category in terms of number of records. However, when checking the proportion of actors, news media come in a lower position

[5] The Internet Archive, a nonprofit library located in San Francisco, created the Wayback Machine as a digital archive of the World Wide Web.

in the relevance ranking presented in Google; i.e. news media produce many pages but they appear lower in the results presented by Google.

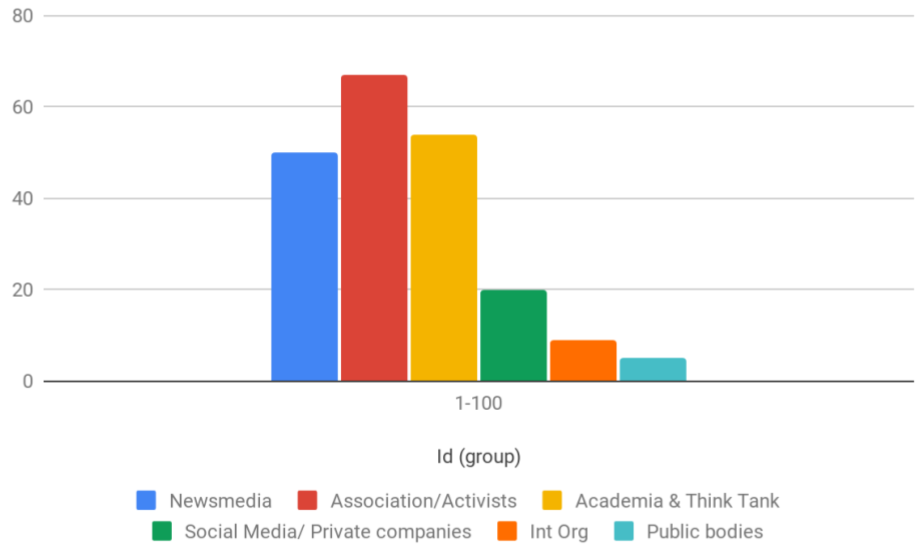


Figure 4. 1 - top 100 results from Google's ranking algorithm

Results presented grouped by different groups

Also, in the light of all the limitations presented above, I felt the need to integrate the data collected from the websites with a second source. In the light of the literature, I focused on other media, since they occupy a special place in controversies as they are the tool through which certain actors' narratives become more visible and established than others (Barry, 2001; Couldry, 2012). I decided to integrate the results from web pages with the analysis of statements collected from newspapers in the UK using an archive available to Cardiff University: LexisNexis. Performing controversy mapping using newspapers is a chance to study how ideas about freedom of expression and social media are distributed. Moreover, newspapers do not suffer the 'ephemerality' of web pages, and work very well as an archive. This means that as a data source they can provide the extra support of data which can be verified again in the future when web pages might not always be available.

4.7 Sources of data collection

Controversy mapping is extremely time-consuming. Potentially a mapping exercise could develop over years (Venturini, 2010, 2012). However, it is acknowledged that researchers have boundaries given by time and resources that can be used to limit the width of the mapping exercise (Venturini, 2010, 2012). Some studies (see Rogers et al., 2015) limited the study to a specific time frame, established by the use of Google Insights, a tool that displays the lifespan of a topic in the online sphere. In this study, I set boundaries to the collection of data by focusing on the cases of the UK, in the period January 2015 to April 2018 (date of last collection). The choice of the timeframe is linked to the length of the PhD programme.

The choice of focusing on the UK as geographical unit of analysis is motivated by the fact that the UK has been one of the European countries most active in the debate about roles and responsibility of SM platforms. It is a longstanding democracy, whose particular history renders it a sort of 'bridge' between continental European approaches and US positions. Moreover, English is one of the most employed languages for the production of documents online.

In order to understand where the controversy takes place and include as many actors as possible, the study considers two different spaces, internet public, and traditional media and as sources of data Google.co.uk, and UK national newspapers.

Google is the most used search engine in the world. Over 90% of internet searches in the UK take place through this medium (BBC, 2013). However, as big as Google has become, using it as the only source to describe a controversy would pose several problems. As stated by Venturini:

> approaching the digital realm must be done carefully, for (1) search engines are not the web; (2) The web is not the Internet; (3) the Internet is not the digital; (4) the digital is not the world (Venturini, 2012, p. 803).

In particular, as seen above, even controlling for the keywords, and taking into account algorithmic ranking, some limitations remain as web pages are subject to frequent updates and do not work as archives, and search engines do not retrieve materials which are beyond paywall, or that require passwords. The idea of combining web pages with statements from newspapers represents also the choice to mitigate the risk of neglecting elements implicit in the bias of digital tools. In line with the empirical approach proposed by Marres and in recognition of the technological dynamics of media that might influence the shape of the controversy, I have considered the influence of technological dynamics (Marres, 2005) also in the analysis of data.

4.8 Data analysis tools and justification

As highlighted in the literature review and theory (chapters 2 and 3), in this study, I am trying to contribute to the knowledge on processes of governance of speech online without focusing on a specific actor or decision-making arena. For this reason, I have adopted an 'inductive' approach, using a theory, ANT, that can embrace simultaneously many actors of different kinds. My methodological approach was developed in the same direction and adopted controversy mapping to identify the actors. Rather than testing hypotheses concerning the main actors or issues composing the controversy, I let the actors emerge from the observation of publications on the controversy 'largely' defined online and in newspapers. I observed the actors' agendas and alliances through their documentary productions available online (websites) and the description given by the public press (in this specific case, UK newspapers). I adopted forms of discourse and content analysis to identify features that could describe the different elements of sociology of translation. To study the interessement and problematisation phases (i.e. identification of the main issues mobilising the public), I performed a qualitative and quantitative analysis of texts from a sample of the web pages and newspapers articles collected.

4.8.1 Qualitative content and discourse analysis

As explained in the literature review and the theoretical chapters (chapters 2 and 3), in this study, I am interested in how governance of speech emerges as a reaction to public shocks (Ananny and Gillespie, 2016) where publics made of experts and non-experts discuss existing forms of order. I have also argued that these types of situations can be studied through the lens of sociology of translation and that other studies have adopted a similar methodology and merged it with discourse analysis (Pohle, 2016a, 2016b). As underlined in the section on theory (chapter 3), the terminology and several aspects of controversy mapping and the translation process are indebted to narratives theory (Greimas 1971; Latour 2005b) and narrative tools of analysis. Controversy mapping is a way to detect stories, actors, narratives, settings and the culmination of a story.

Drawing on Pohle's work described in the previous chapter (2016a, 2016b), I have found helpful the use of two concepts from discourse analysis applied to policy documents: exemplary cases and storylines. As a way to orient the analysis, I relied on the indications left by Pohle (2016a, 2016b) on her interpretation of the process of translation taking place within the deliberations within the

UN Working Group on Enhanced Cooperation (WGEC). Pohle identifies translation processes where a specific discourse coalition becomes dominant by persuading key players to follow its worldview or version of the narrative about the issue. To orient the analysis and identify successful translation processes, I looked for situations in which one narrative about speech and technology developed within a specific group begins to be adopted by a plurality of other actors.

Similarly, to identify mediators and intermediaries, and in the light of the literature interest for public shocks, I followed Pohle's approach and looked for 'emblematic issues' and 'storylines' (2016a, 2016b). Emblematic issues are cases used to simplify complex issues, becoming symbolic for understanding the problem. On the other hand, storylines are stories that are used to summarise complex narratives and assist the actors in conveying facts and data. Actors are encouraged to share opinions and create suggestions when exploring iconic topics and storylines, preferably addressing the issues in their entirety Pohle (2016a, 2016b). They can be used to summarise the main controversial elements that constitute the public shocks (Ananny and Gillespie, 2016) and can also be intended as a signal that some aspects of the routine management of freedom of speech have started to become 'matters of concern'.

Using the tools for analysis developed within narrative discourse analysis and Pohle's work, I built my initial analysis on two questions: is there an entity, problem, event, series of episodes that are taken as an example of a more significant issue by a group of actors? What aspects are they summarising?

Following these two questions, I developed a code, which included the identification of:

1) Emblematic episodes or stories that are recurring across the texts

2) The type of issues they summarise

3) The type of narrative (worldview) of freedom of expression which is intended

4) The type of narrative (worldview) of technology that is intended

5) The type of narrative (worldview) of governance that is intended

6) To what type of regulation initiatives do they relate?

To identify recurring episodes and stories, I relied on qualitative analysis of the texts and quantitative (computational) tools for text analysis, such as the Cortext Platform.

I did not use specific tools for the qualitative analysis of texts. I started initially using Nvivo 11, but the incompatibility between different software versions made me drop the choice. So instead, I created a dataset with all my texts in excel, which I found comfortable enough, given the possibility to add codes to the texts adding columns and colours.

4.8.2 Computational tools for text analysis

The field of corpus linguistics has developed several applications to analyze a large amount of textual data and summarise information. Several studies have used these tools in recent years, mainly term extraction and topic analysis with Latent Dirichlet Allocation (LDA) to analyze documents. For example, Farrel (2015) used a type of LDA to study how political and financial actors influence ideological polarization in the climate change debate. In his work, Farrel argues that LDA provides a credible content analysis of extensive text collections that would be too complicated to analyze manually and that, even if unsupervised, LDA can outperform human coders on analysis of documents. Similarly, Dobša et al. (2020) analyzed texts from Web of Science using the LDA extraction of topics, provided a valid alternative to pre-determined categories, and was able to identify interdisciplinary fields directly from the textual content of paper titles, abstracts, or keywords.

The methods are still quite debated, and the level of performance can vary, especially depending on the size of the corpus, the length of the documents and the language used. These methods tend to work better on longer documents and large English texts (Seemab Latif et al., 2021). However, scholars are developing tools for the analysis of shorter documents as well as for less commonly spoken languages (Seemab Latif et al., 2021).

Drawing on these and other similar studies, I chose to use this type of tool to help me cope with a large amount of text (for a single coder) and as an alternative with which to compare my qualitative coding. Cortext is one of these tools. It is a project launched and sustained by IFRIS and INRA (Cortext website, 2021) and its web application provides open-access tools for the computational analysis of texts, such as word frequencies, topic analysis (for instance, using the LDA system), as well as semantic network mapping of documents and temporal analysis (called ‘demographic analysis’) (Cortext documentation, 2021a).

Terms extraction

Cortext provides a term extraction tool based on Natural Language Processing (NLP) to identify the frequency and relevance of simple terms and multi-terms. In the study, I used measures of TF-IDF and co-occurrence measures. These measures can provide an overview of the relevant terms by considering the totality of articles in the corpus. They calculate the average frequency of terms in documents and correct the results based on the likelihood of very frequent terms appearing several times in the same content. In the study, I calculated the frequency of terms at the document level, as I was interested in observing the term frequencies computed according to the number of different documents in which they appear. The other option would have been to calculate the frequency at sentence level (which would be the default choice) which considers repetitions of a term across sentences within a document (Cortext documentation, 2021b). I used the Cortext terms extraction tool to list the 300 most relevant terms in the corpus of newspaper articles and websites (i.e., based on quantitative techniques for weighting terms in documents TF-IDF). To begin with, I extracted a list of the 1,000 most relevant terms. However, I realised that excessive granularity invalidates the purpose of isolating the actors since it introduces too many items (e.g., terms that do not have much meaning). For this reason, I chose to focus on a shorter list (i.e., the 300 most relevant words).

Topic modelling

Cortext topic modelling produces a topic representation of a corpus's textual field using the LDA model. LDA is an unsupervised generative probabilistic method for extracting topics from a corpus of documents. It assesses the co-occurrence patterns of terms within individual texts and throughout the entire corpus, assuming that each document can be represented as a probabilistic distribution over latent topics (Blei et al., 2003).

Identification of number of topics: Conventionally in LDA, the variable referring to the number of topics is called *k*. However, *k* is not necessarily pre-programmed into the system, and can arise from the patterns underlying a set of texts (i.e., leaving the algorithm ‘unsupervised’). The optimal value of *k* indicates a number of topics able to be semantically interpretable (i.e., coherent) rather than artefacts of the statistical inference not interpretable by humans (Sievert and Shirley, 2014).

An insufficient number of topics could render an LDA model too coarse to identify accurate classifiers. On the other hand, an excessive number of topics could result in an overly complex model, making interpretation and subjective validation difficult (Zhao et al., 2015).

Scholarship is divided on how to select the value of *k* (i.e., the optimal number of topics in a model). Different 'versions' of LDA have developed calculations to identify coherence or classify the accuracy of topics. They often create systems to attribute scores based on the degree of semantic similarity between words (see Arun et al., 2010; Newman et al., 2010; Taddy, 2011; Bischof and Airoldi, 2012, cited in Sievert and Shirley, 2014; Zhao et al., 2015). However, no method has prevailed (see Zhao et al., 2015). The most commonly employed rule to identify *k* still relies on some form of iterative approach (i.e., trial and error) and the use of human rankings as the gold standard for coherence evaluation.

In Cortext, the tool for topic analysis adopts Sievert and Shirley's (2014) interpretation of LDA and visualization, called Davis. In LDAvis, the device identifies the *k* number of topics in an unsupervised way. It bases its optimization on the score of relevance for each topic, calculated as a weighted average of the logarithms of a term's probability and its lift (i.e., the ratio of a term's probability within a topic to its marginal probability across the corpus). The authors explain that the weight is grounded empirically on a user study involving 13,695 documents and 29 human coders (Sievert and Shirley, 2014). The result is a script set to optimise the number of topics (i.e., the default setting in Cortext is 0 for automatic search, min. 10 max. 40 – see Appendix p.21-A for the full parameter lists).

By letting the script be 'unsupervised' (i.e., typing 0), the optimal number of topics is identified using the method described above, optimizing over the number of topics produced by the model with the highest topic coherence possible (Cortext documentation, 2021).

The system reports the overview of different coherence scores attributed to various topics (see Figure 4.2). I used this output to study and compare the differences in the coherence score produced by changing the settings versus adopting the unsupervised script. I was able to assess that the unsupervised script returned the highest coherence scores. However, high coherence scores do not say much about the 'meaning' of the topics. For instance, a high coherence score would still be enough if the topic identified is not 'understandable' to a human reader. Moreover, relying only on scores might risk falling into the 'black box' situation where the researcher can only 'accept or reject' the machine's decision.

For this reason, I found LDAvis particularly suitable for social research. On top of the calculation of coherence score, LDAvis provides a visual interface that allows the researcher to interactively explore the list of terms included in the topics and the relationships between terms, topics and the corpus of documents. In this way, it is possible to visually see whether and how particular terms appear in more than one topic and how they compare to the distribution of terms in the whole corpus (Sievert and Shirley, 2014).

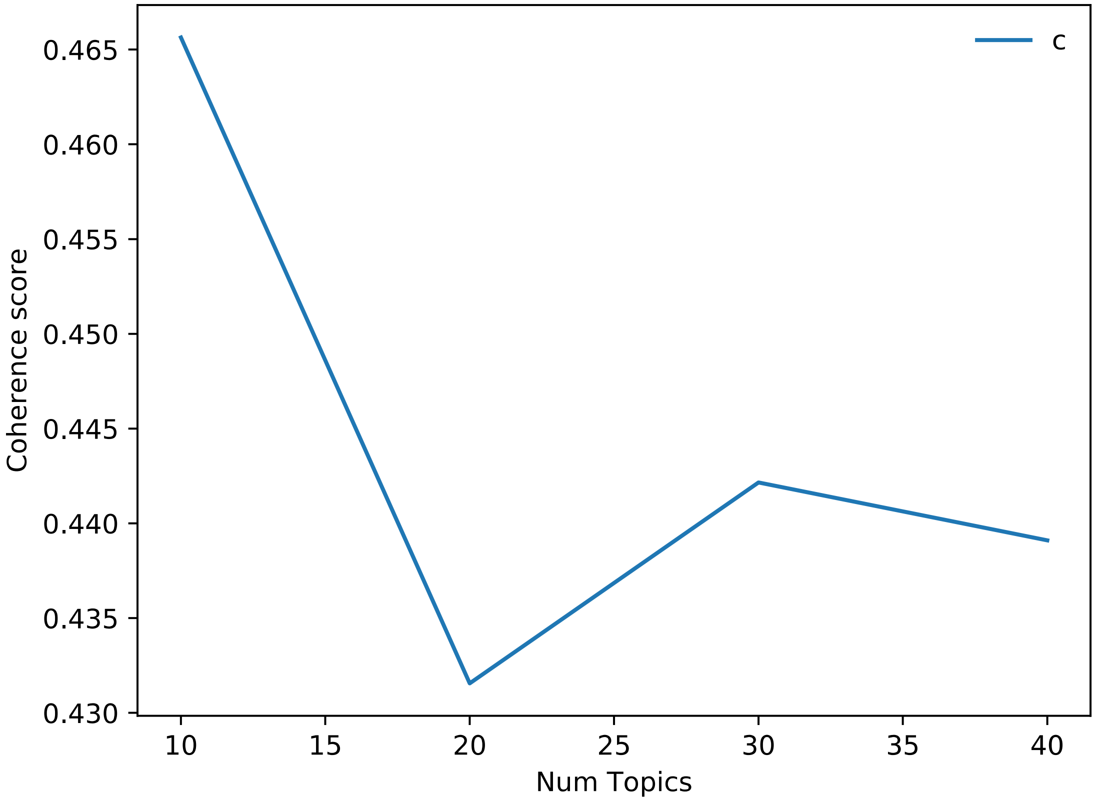
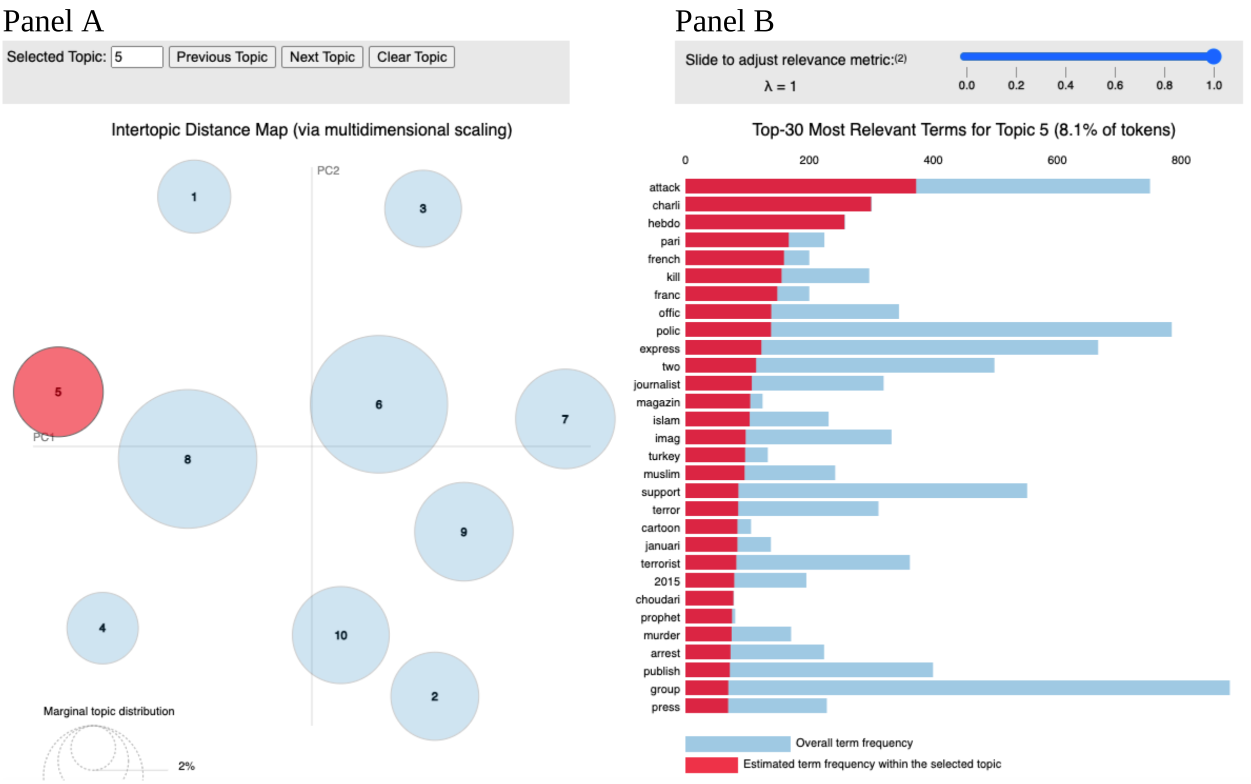


Figure 4. 2 - Coherence score output produced for newspaper topic analysis in LDAvis, Cortext.

This graph shows that 10 topics have the highest coherence score, and it is considered the 'optimal' number of topics.

The typical visualization (shown in Figure 4.3) consists of two panels. Panel A on the left provides a global perspective of the topics. In this panel, the areas shown in the circles are proportional to the relative prevalence of the topics in the corpus, and the distance between circles reflects the inter-topic differences.


Panel A

Panel B

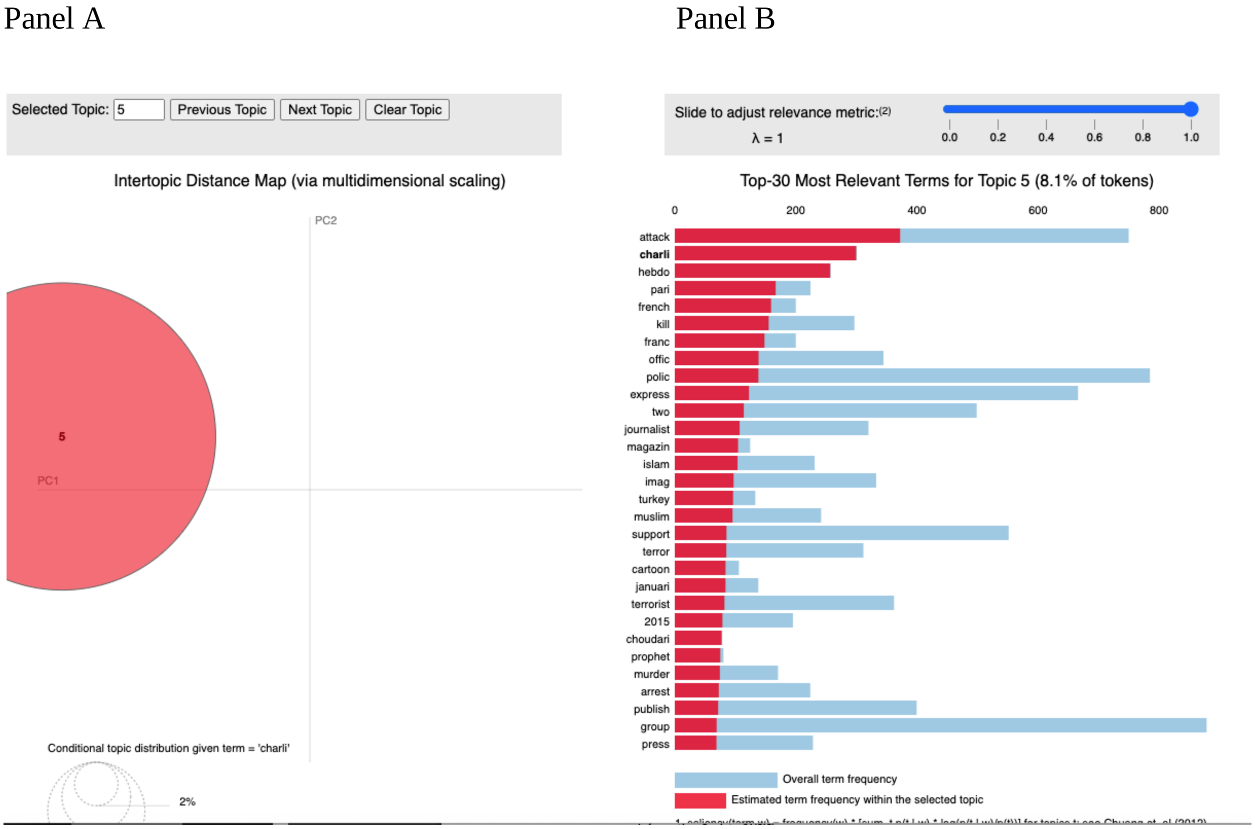
Selected Topic: 5

Slide to adjust relevance metric:(2) λ = 1

Figure 4. 3 LDAvis output. Visual exploration of relationship between terms-topics-corpus

In Panel A the topics identified by the system are visible. In Panel B, it is possible to explore the most relevant terms associated with the topics. The red bar indicates their prevalence within the topic, while the blue bar indicates their overall presence in the corpus.

Panel B on the right shows how the terms are distributed over topics and across the corpus. The terms are ranked from the most to the least relevant (i.e., probable terms) within the topic, represented by the red bars. The red bars are stacked on blue bars, representing the total distribution of that specific term in the corpus. This lets the user verify how the terms composing the topics relate to the corpus and other topics.


Panel A

Panel B

Selected Topic: 5

λ = 1

Figure 4. 4 LDAvis output. Visual exploration of relationship between terms-topics-corpus.

Figure 4.4 shows how selecting a term (in this case, 'charlie') in Panel B, the system automatically highlights all the topics that include that term in Panel A. In this way it is easier to assess whether the topic identified computationally actually have coherence and value, by exploring the terms in a context understandable to the human eye.

As a further method to explore the composition of topics, Cortext output also presents the list of the ‘most relevant’ documents, i.e., the specific topics are more present. I could also skim the list of documents that the system attached to the topic.

In the study, to identify the *k*, I empirically tested the script's parameters, observing the differences in the coherence scores result. I then explored the visual representation of the topics in the graphs and the more relevant associated documents. I ended up relying on the optimization of topics offered by the script (i.e., 0 for automatic search, min. 10 max. 40). This choice, which uses the system's definition/calculation of coherence, was revealed to have a higher coherence score.

Furthermore, the visual exploration of the graphs shows a clear relationship between how the terms are associated with topics and different documents. The combination of all these other elements provided by LDAvis gave me sufficient confidence in the reliability of the analysis.

Semantic network analysis

Cortext can also create network mapping representing the semantic relationship between topics and terms using the elements identified with terms extraction and topic identification. This can be done by selecting different fields (e.g., terms, topics) as nodes. In this study, I used network analysis to link terms extracted from newspaper articles to topics, then the topics and terms to specific publications.

Temporal distribution analysis

Using the demography tool in Cortext, I was able to map the raw frequency of topic occurrences through time (Cortext documentation, 2021c). Cortext demography script processes each field of the corpus (in this case, topics) and counts the raw number of occurrences of the top items. I used this tool to perform a temporal analysis of the frequency of useful topics to study the changes through time (time granularity: months from January 2015 to April 2018). I used this method with an awareness of two main limitations: firstly, each document may have more than one topic; secondly, each topic for one document is not representative of the content with the same intensity. Some topics may be strongly present in the documents, while others are marginal. For this reason, Cortext demography script does not show the real evolution of the importance of a topic in the content of the documents. However, it can explain the evolution of topic frequencies (i.e., the raw frequency) in the documents since 2015. This type of analysis provides insight into the life of topics as if they were isomorphs, which is in line with the ANT assumptions. Therefore, it is helpful to give an overview of the development of the life of the controversy.

4.8.3 Samples used in analysis

In the next chapters (chapter 5 and 6), I will provide a more detailed description of the construction process of the corpus of web pages and newspaper articles. Here I summarise the main information about the corpus and the samples that I have used for the quantitative and qualitative analysis.

The original corpus of web pages collected with keywords via google.co.uk comprises 756 web pages containing the statements from the actors in the controversy. I have expanded the original list with the action of Hype and reached a total of 2503 URLs, connected via hyperlink through the crawler. I used this larger corpus to code actors into groups and perform network analysis. To identify issues, I used qualitative analysis of texts, and I analysed 354 web pages, approximately 80 per year, except for 2018, which had a slightly higher number (100). I selected the sample from the first results returned in the search engine, relying of the search engine's definition of 'relevance', as explained in §4.5.4.

In the case of newspapers, I have created a newspapers dataset collecting articles using Lexis Nexis repository. From the Lexis Nexis' UK publications dataset, I have downloaded approximately 1000 articles per year from 2015. I have used the following list of keywords: "social networking" OR "online social networking" OR "social sites" OR "social network*" OR "social media*" OR "networking sites") AND ("freedom of expression" OR "freedom of speech" OR "free speech").

The list of keywords list is the same as the one on Google. The choice is supported by the fact that it gives quite a large variety of possible ways in which an issue is described in a specific public arena (press) and by the fact that it was checked on the role of the tool (archive/search engine) used to collect the data. Moreover, it fits with the internal indexing system. In this way, the keywords reflect the main interest of the research at the same time being as open as possible to avoid influencing the results/ to capture the different elements composing the controversy. The corpus comprises articles from UK newspapers, selected as a non-specialised list of publications (for instance, it does not include specialised magazines such as The Economist or Wired). To perform computational analysis, I used the entirety of the corpus (i.e. 3014).

For the qualitative analysis of texts, I read and coded the first 200 documents per year except for 2018. I coded 100 articles (since the articles covered a shorter period, i.e. up to April 2018). In this way, I created a sample of 700 articles. This sampling was necessary as it would have taken too much time to manually code all the documents, which amounted to about 1000 per year. The justification behind the choice to code the first 200 records is as follows:

1. The dataset itself presented articles ordered by relevance (i.e. the first records are the ones most closely related to the topic).

2. Several documents reported in lower positions were the same articles published in different editions of the same newspapers.

3. The sampling respected the proportion of publications (i.e. proportional correspondence between the number of publications in the sample and the entire dataset checked via chi-square test).

To avoid a completely arbitrary choice, I have checked whether the subset I selected respects the original proportion of publications. To check the comparability of my sample with the larger collection of articles, I have run a chi-square test to check if the publications included in my subsets are comparable to the distribution in the larger sample. The tables with the chi-square tests values are available in Appendix 44-A.

Performing this test, I have found that the publication proportion has changed in a few cases: the Guardian/MailOnline are slightly more represented than in the original sample. In order to maintain the same 'proportion', I weighted the subset to reflect the distribution of the publications present in the sample. The final result is a subset that reflects the proportion of the original list of publications.

4.9 Limitations of controversy mapping with web pages

When approaching the more analytical part of controversy mapping with hyperlinks, the phases that have to do more with Social Network Analysis (SNA) and analysing the narratives, a number of limitations emerge. The labour behind research with digital data is gigantic.

Firstly, the data collection of hyperlinks, either via search engine or crawler, favours quantity over quality, meaning that the number of statements that can be collected is quite big (hundreds of different documents, for statements from websites alone). However, even though the initial search can be tailored and the digital tools are accurate, a large part of the results will have to be filtered out, either because they are duplicates, or because they are not really semantically related to the topic (e.g. on a web page online, the keywords appear be in a very small part, in a collection of other articles on different topics). Or, as became evident in the empirical process of reconstructing the networks of hyperlinks connecting the pages, several links are there because of the structure of the web pages which now all include advertisements, or link to the main SM platforms as a way to share their contents. As much as this is indicative of technological dynamics of the media, in order to study the content of the statement I had to clean the dataset of non-related pages. Cleaning the dataset implies an immense effort on the part of researchers to eliminate the large presence of ‘noise’ and pages found only through commercial ad links. Even after this procedure, the corpus of documents is still too large to be the focus of a deep and rich qualitative analysis. For this reason, it can help to use computation tools to identify topics and keywords on large datasets using algorithmic detection of topics or other relevant semantic elements. However, this is another way to ‘black-box’ decisions in the research procedure, since tools for automatic recognition of speech work through algorithms that are not always available to researchers. It is thus necessary to balance the results with a more time-consuming qualitative content/discourse analysis, capable of understanding the meaning of that ‘issue’/semantic element. The outcome is that controversy mapping methodology assigns a very central role to the decisions and interpretations made by the researcher.

4.9.1 Mixing methods: centrality of researcher and quantitative approach

According to Venturini, in the process of controversy mapping the researcher undertakes the very central task of unfolding the complexity of controversies by at the same time actively ordering such complexity (Venturini, 2012: 797). As Rogers et al. (2015) stress, the researcher is responsible for:

> properly attributing relevance to all the points of view, and at the same time attributing the significance of some over others, according to the rule of proportionality (p. 43).

In terms of actual research, this rule of proportionality is performed by associating qualitative and quantitative forms of description and analysis of data (see Andreossi et al., 2013; Rogers et al., 2015). Researchers at Sciences Po MédiaLab call it a ‘quali-quanti’ approach. In the UK it is called a ‘mixed methods’ approach. This choice of using a mixed approach to data is also linked to the wide use of digital methods in the field. I have discussed how digital methods serve ANT’s propensity for digital tools of research. Digitised data, because of their ‘numerical’/’quantitative’ nature, suit quantitative forms of manipulation, whether in the form of network or semantic analysis. Quantitative measurements appear for instance in Rogers et al. (2015) in their study of the issue-network built around ‘ageing’ in Europe, based their identification of actors through a qualitative assessment of texts (namely, on the researchers’ identification of actor mentions by third parties and actor inter-mentions within a list of organisations’ websites, Rogers et al., 2015:55). From that qualitative identification, the researchers created a categorisation and a ‘quantification’ of the actors’ presence. They were thus able to ‘count’ actor frequencies and visualise them as word clouds or other forms of visualisations able to render the idea of presence as ‘quantity’ (see Rogers et al., 2015:49-50). Andreossi et al. (2013) followed a similar process as in Rogers et al., (2015) in their ‘What the Frack’ report – an exercise in the application of Venturini’s controversy mapping methodology performed on the issue of fracking. In the study, according to Venturini’s protocol for data collection the students identified a list of actors from a list of statements found on websites selected using specific keywords. They then considered the website’s ownership of each domain as a proxy for ‘actors’ and manually proceeded to assign categories, such as activist, environmental agency, journalist, etc. (Andreossi et al., 2013:46). The manually assigned categories were then used to describe the actors and quantitatively measure their frequency (i.e. describing the composition of actors by percentages). Clearly, there is no mention of statistical relevance of these quantitative measures; however, they are considered as a proxy of the actors’ role in the construction of the controversy.

Another quite established practice in controversy mapping is to exploit the quantifiability of web pages’ hyperlinks as a proxy to assess the centrality of an actor in a controversy. This process is well explained by Ooghe-Tabanou et al. (2018) in their publication ‘Hyperlink is not dead!’. The authors argue that the direction of hyperlinks is a robust indicator of hierarchy effects when measured with adapted metrics and visualisations. This is because the directionality of links reveals asymmetrical associations between the linked documents: the referrer knows the referee but not necessarily the other way around. The tools that have been developed, i.e. Hyphe, as well as the IssueCrawler in Amsterdam, were created to study and visualise these dynamics.
Quantification is used by researchers in the attempt to unfold and at the same time create ‘order’ in complexity (as seen above, Venturini, 2010, 2012), even though it is recognised that it could be misleading to suggest that these forms of quantifications have some sort of statistical representativeness. However, as stressed by Boullier, in this quantification, “the ‘quali’ part outweighs the ‘quanti’ one” (Boullier, 2018:2).

Discussing controversy mapping conducted at Sciences Po, Boullier (2018) criticises the fact that too often the ‘quantitative’ part of the method has been delegated to ‘topologies of networks’ built using hyperlinks, or social media data. He argues that the analogy with the idea of 'network' is misleading, as it creates an emphasis on positions, clusters and distances. It assumes that the topological space created with hyperlinks is analogous to a social space (Boullier, 2018:7). However, considering hyperlinks as a proxy for the social is reductive, since networks of hyperlinks provide no insight on what is being circulated or the significance of the links. Web topologies can offer interpretation of data only in terms of 'positions' and 'communities'. However, the interpretations of alliances of actors or of semantic proximity of topics in these mapping exercises are often like a black-box and do not allow the hypothesis to be tested in a robust way (Boullier, 2018). Boullier criticises especially the desire to restore 'a seamless fabric' of society by following the actors, ignoring the "digital mediations that make it possible to weave the links between these actors, and above all the digital mediations of the observer" (Boullier, 2018:7). By contrast, Boullier, like Marres, recognises the importance of discussing the instruments used in the research; not only 'following the actors', but accounting for the instruments that help social scientists follow the actors and detect them, recognising the conceptual frameworks encapsulated in each of them (Boullier, 2018:6-9). In this regard, he stresses how Sciences Po tools frame the issues under scrutiny in a structural way, focusing the analysis on positions, relationships, network topologies and community detection. However, using this approach researchers are forced to focus on the 'whole' (i.e. in a structuralist way), rather than tracking down the emergence of new entities with agency, i.e. mediators, which are the key actors in ANT methodology (Boullier, 2018:9). In order to face this limitation, Boullier suggests considering not only the position of actors, but above all the circulation and emergence of categories, terms or 'issues' that show the public dynamics of a discussion or a conversation (Boullier, 2018:5). He suggests that the focus should be on the agency of the entities that circulate, the content and the connections/associations that it generates. For instance, he suggests the study of propagation of memes and their ability to make other users replicate them. In this way, it is possible to account for non-human agency both of objects and devices as well as of the messages that circulate between them and humans.

It is clear from these reflections that the use of mixed methods in controversy mapping is different from other forms of social research (where for instance statistical analysis is combined with interviews). In controversy mapping quantification is not intended to be a generalisation about the world, but rather a summary of the data collected. Similarly, as stressed by Boullier, the structuralist heritage embedded in the tools for SNA might lead observers to believe that the distances and relationships between nodes might have a form of objectivity outside the specific visualisation of the data. The study of the controversy focuses on the public mobilised, and this choice has to be remembered in order to avoid being misled in the interpretation. Quantitative tools are used in controversy mapping to present an overview of a large number of actors and a summary of statements that appear through the data collection. The aim is not to reach a statistical representativeness or objectivity, but rather to show specific aspects that are significant for the understanding of the controversy and the interpretation remains fundamentally qualitative. Controversy mapping encourages the experimental use of mix-methods in the sense of moving in different directions to follow the actors. For instance, scholars encourage integrating the mapping with participatory form of social research that provide the opportunity for voices of different actors to be recorded. From this point of view, interviews, focus groups, or other more experimental forms of research (as the datas print workshops, Munk et al., 2019) might integrate and augment the data collected with digital tools (Edwards et al. 2013).

A very challenging element of controversy mapping is building visualisations that allow users to understand and navigate the complexity of a controversy, or "making complexity simple" (Venturini, 2010, Venturini et al., 2015). For this reason, other studies of controversy mapping make extensive use of data visualisation tools, such as word clouds, alluvial diagrams, bubble matrix charts, geographical or chronological distribution of issues (Rogers, 2017). This entails a limitation, which regards the use of these tools, which is not always accessible without a professional background in data visualisation software. Boullier (2018) warns of another type of risk. Given the beauty of data visualisations that can be created with the use of digital tools for mapping controversies online, there is a danger that in the mapping exercise researchers might lose touch with the theoretical background. It is evident for instance, in other controversy mapping exercises, such as Andreossi et al. (2013), where the superb visual aspect of the mapping aspect does not correspond with a theoretical reflection on the 'meaning' or interpretation of the actors' positions. For this reason, I use controversy mapping as a method to empirically collect and visualise a sample of actors involved in governance of speech online and in newspapers (in chapters 5 and 6). However, in chapter 7 I reconnect the data from the mapping to the theoretical interpretations in terms of ANT roles (i.e. spokespersons, mediators, intermediaries, etc).

4.10 Limitations of newspapers

The use of articles from newspapers did not come without issues, either. However, coming from a repository as Lexis Nexis did reduce the amount of noise and data cleaning, as they all were in the same format. Working with articles highlighted a different type of limitation of controversy mapping's methodology. The pathway described by Venturini and in general the idea of retrieving actors from statements works very well only in the online space, where it is possible to follow the equation: web page = author (because although web pages are 'hosted' on domains, the URLs are assigned to specific entities, which become a proxy for authorship). However, in press articles, the journalist is the author of the statements retrieved, even when it is citing other actors. The study of the other actors thus can only be made 'indirectly' through qualitative analysis of the text, and identification of actors that are mentioned in the statements that are reproduced within an article. This complicates the construction of the corpus and the analysis.

4.11 Ethical considerations

In the previous paragraphs I addressed the methodological limitations and implications of this particular research design. Here I focus on the ethical issues connected with working with the specific type of data for controversy mapping. Data from newspapers as they are publicly available documents do not represent a particular issue from the point of view of ethical consideration. Data from web pages collected through Google search engine on the other hand necessitate more careful consideration.

The protection of human subjects is one of the general guiding principles of research ethics, and in this respect, informed consent and privacy are generally necessary for participant protection (Bryman, 2016). Traditional documents such as newspapers and literature, on the other hand, are not considered human subjects, and permission is not necessary to analyse them (Snee, 2013). As stressed by Snee in an ethical reflection on the use of blogs in research (2013), scholars are far from unanimous on how to apply traditional ethics principles to the online context, since the boundaries dividing public versus private, and subject versus author, are more blurred.

4.11.1 Public vs private space

As seen above, the distinction between public and private space is particularly salient in digital social research (Bryman, 2012).

> Just because content is publicly accessible does not mean that it was meant to be consumed by just anyone (boyd and Crawford, 2012: 672ff).

In the digital environment, it is not easy to understand whether the contents publicly accessible on social media and other web pages have to be considered public and exploitable. Several researchers have explored the perception of the 'imagined public', with different conclusions regarding the perception of publicness from the point of view of users (Marwick and boyd, 2011; Oolo and Siback, 2013). According to the Association of Internet Researchers, the greater the venue's accepted exposure, the less risk there will be for researcher in terms of protection of individual identity, anonymity, and the request of informed consent (Ess et al., 2002). And as a general rule, according to Hewson et al. (2003) cited in Bryman (2016):

> data that have been deliberately and voluntarily made available in the public Internet domain, such as newsgroups, can be used by researchers without the need for informed consent, provided anonymity of individuals is protected (p.139).

The general 'principle' is to guarantee the maximum anonymity possible, and when this is not possible to require permission. However, achieving anonymity for materials published online is virtually impossible. This problem has been underlined by Edwards et al. (2013: 256) in the case of "ethical concerns over the potential uses of digital observatories for the purposes of intrusive surveillance". This state of affairs entails some important ethical considerations, notably in more authoritarian contexts; yet as noted by some authors, this same problem has the positive side of increasing the visibility of 'elites' and thus democratic accountability (Edwards et al., 2013:256). On the discussion on anonymity Bounegro and Gray (2014) have adopted the consideration that higher goals (e.g. a phenomenon that needs to be monitored) might justify breaches in anonymity (e.g. in the case of research on the networks of far-right parties). However, they acknowledge that privacy is perceived differently according to different countries.

4.11.2 Subject vs author

The other ethical aspect raised by Snee (2013) concerns the definition of web pages, such as for instance blogs either as proxy of human subjects (i.e. subjects) or are as the expression of human subjects (that in this case would have to be considered as authors). Some studies consider online texts as extensions of persons (Young, 2013). However, the ESRC Framework for Research Ethics defines materials already in the public domain as not human data. This covers intentionally public information given in forums or spaces on the internet and websites (ESRC, 2021). Based on this consideration owners of web pages and blog can be considered as authors rather than subjects in the research (Snee, 2013), However, as Snee points out, it still raises the issues of ownership of the documents. Considering web pages owners or bloggers as authors requires to accompany quotes with references to the original blogs (Snee, 2013).

However, the general advice from internet scholars emphasises the importance of context in making ethical decisions (Ess and Aoir, 2002). Researchers have to consider the expectations of privacy and the role of authors in relation to their websites, as well as data processing and analysis practices and how the data could be used (Snee, 2013:60). As far as the expectation of privacy is concerned, one criterion used by researchers is to assess whether the communication online has been created with the idea of addressing an ‘unknown reader’, since if they are written with such an ‘unknown audience’ in mind, it is possible to consider it publicly available Whiteman (2007). However, as seen above, open to a public does not necessarily mean open to researchers (Boyd and Crawford, 2012). In the face of such a wide range of circumstances, it is up to the researchers' ethical judgement to ensure that the bloggers are accurately portrayed (Snee, 2013).

For instance, Snee (2013) decided to consider ‘gap year’ blogs as public domain and did not look for informed consent from their authors; however, considering the young age of the bloggers and the presence of personal information she preferred to not recognise their status as full authors (as she would have had to mention their names) and to describe the content of the web pages protecting their identities. Also, other studies did not consider it necessary to obtain informed consent from bloggers, as in the study by Leggatt-Cook and Chamberlain (2012) where the authors decided to cite the blogs they used (blogs about weight loss) based on the consideration that according to cultural norms in the ‘blogosphere’ it is more important to ‘send’ people to the blogs, rather than protect the bloggers’ anonymity.

In this study I adopt an approach similar to Leggatt-Cook and Chamberlain (2012), in that I recognise bloggers as authors of statements, and I recognise that they are part of a specific online community. The study gathered a group of bloggers particularly active on the issue of free speech, and it included several statements where they position themselves on the side of a 'liberal' interpretation of the rules of content on the internet. The blogs all include a publicly available description of the blog for the non-acquainted readers that show that they are writing for an audience, that they are all over 18, mostly male individuals, residing in Western countries. For the scope and purpose of the data used in the analysis, they do not belong to any vulnerable group. When citing blogs, I provided quotations from the source and recognised the authorships, because the type of data collection and the use of the data does not imply harm for the participants.

4.12 Conclusion

In this chapter I have presented in detail the methodology that I have used to orientate my data collection and analysis. As stated above, the choice is based on my aims, research questions and theoretical background. First of all, controversy mapping is an empirical application of ANT. Secondly, it allows answering my questions about the elements present in governance initiatives about freedom of expression and social media, following the recommendation to include less 'famous' actors (not always the big names or big states). As this is an inductive methodology, I approach the data without a specific idea of who is going to be the focus of my analysis, as the actors emerge from the collection of statements. The type of data collected, statements, makes it possible to study the narratives associated with technology, in a critical perspective.

In the next two chapters I present the results of the data collection. I used websites and newspaper articles to study the public gathered around free speech on SM platforms as a 'matter of concern', using the tools developed within the controversy mapping research community. I have outlined the main actors and issues/public shocks that define the controversy over social media as a vehicle for freedom of speech but also of threats and abuse.

The first empirical chapter (chapter 5) focuses on the findings from the analysis of web pages, while the second one (chapter 6) discusses the findings that emerged from the analysis of newspaper articles. I have structured both chapters by following the steps for controversy mapping described above in this chapter, adapting the methods to the different types of data sources: i.e.

web pages and press articles. I used controversy mapping to empirically detect actors, starting from statements. However, the different technological affordances of different media made necessary an adjustment of the methodology. In the empirical chapters I will start with an overview of the actors and main public shocks constituting the controversy about governance of speech online. In the study of the controversy on web pages, I collected actors from URLs and crawlers. In the study of the newspapers, I collected actors from the publications and from the content of the articles. The chapters will also include an analysis of the narrative elements, and I will provide an overview of the contents appearing more frequently, following the recommendation by Boullier (2018) to focus on the content rather than the position of actors, and what actors are connected to others and how.

5. Mapping controversies using web pages

5.1 Introduction

As discussed in the methodology chapter (chapter 4), in this part of the study I am interested in understanding how the controversy around SM platforms and content regulation is staged in the public space offered by Google.co.uk. Following the pathway of controversy mapping described previously, in this chapter I identify actors and issues animating the controversy around social media platforms and freedom of expression using statements on web pages. I also offer a reflexive description of how I have applied the instructions for controversy mapping described in the methodology. Below I explain how I used crawlers to collect hyperlinks and identify actors and their relationships, and how I used qualitative discourse and quantitative content analysis to identify issues emerging from the texts in the web pages collected.

5.2 Construction of corpus

Drawing on Rogers and Marres and other controversy mapping studies, I started my mapping exercise by collecting a corpus of documents containing statements on the issue 'Freedom of speech and social media'. The corpus was to be created from statements of actors on web pages. I collected the statements using the UK version of Google (i.e. Google.co.uk). For all actions in this part of data collection, I used a browser 'repurposed' for research following Rogers' instructions in the DMI dataset.[6] After selecting a list of keywords (using the method discussed in chapter 4[7]), I used the DMI tool Google Scraper to collect a list of URLs. I performed the scraping until exhaustion of the pages (i.e. when the search engine itself communicated that going forward there would be only duplicate elements). This scraping procedure provides confidence that a fair amount of the URLs linked to the keywords, available at that moment on Google.co.uk, have been collected. However, the results are not the totality of existent web pages containing the keywords,

[6] Following the DMI instructions, I created a separate research profile in Firefox and Google, cleaned of all history, cookies and tracking and personalisation devices#.

[7] The output was a list of synonyms and related concepts for 'social media' and 'freedom of expression' to use for the query in the search engines: 'Freedom of speech' OR 'freedom of expression' AND 'social media' OR 'social network*' OR 'social media site*' OR 'social media platform*' OR 'business social media' OR 'new social media' OR 'social media content' OR 'social media governance'.

and are presented in Google according to its relevance algorithm; a methodological aspect and limitation of the research tool that I discussed in the methodology chapter (chapter 4).

In chapter 4 I mentioned an important limitation of using web pages as data source: i.e. the internet does not work well as an archive of events that happened in the past. URLs and web pages can be updated several times and change their content (Rogers et al. talk about “unstable media” and “ephemerality of content”, 2015:31). In order to obtain a coherent set of web pages linked to the controversy, I had to perform different data collections from December 2015. For the final corpus, I only kept the URLs lists collected in December 2015 and 2016, September 2017, and April 2018 (last data collection) as they cover the largest time span in the most ‘regular’ way (see Table 5.1).

Table 5. 1 - Summary of URLs collected after removal of duplicates

Year	URLs	Collection date
2015	151	December 2015
2016	169	December 2016
2017	219	September 2017
2018	226	April 2018
Tot. URLs in corpus	765	

As part of the methodology, and also as a way to further expand my list of actors I passed the list from Google.co.uk into a crawler. I used the crawler developed by social and computer scientists at Sciences Po Paris, called ‘Hyphe’.[8] The first outcome of this second phase of the data collection was an expansion of the web pages from the original list.

Table 5. 2 - Expansion of list of actors after the crawl

Year	URL original list	URL after crawl
2015	151	453
2016	169	543
2017	219	667
2018	226	840
Total	**765**	**2503**

[8] Described in chapter 4

Table 5.2 shows the changes in the number of URLs per year, after the crawl. The final list includes 2503 URLs, connected via hyperlink to the original web pages containing the statements from the actors in the controversy.

5.3 Presentation of findings

Below I describe what this list of URLs can tell us about the actors and their relationships (based on the network of hyperlinks).

5.3.1 Identification of actors

Controversy mapping online conventionally considers URL domain names as a proxy for the 'authors' of statements and thus includes them as actors in the controversy (this is the methodology used in the DMI University of Amsterdam and developed from Sciences Po MédiaLab).

Table 5.3 shows how I extracted the actors from the URL.

Table 5. 3 - Example of identification of actor from URLs

URL	Home page	Actor name
https://www.autism.org.uk/~/media/nas/nasschools/radlett-lodge-school/documents/policies%20(nas)/policies%20(local)/2017-18/rls-online%20safety%20policy-sep17.ashx?la=en-gb	https://autism.org.uk	National Autistic Society

As explained above, the creation of the list of actors took place in two steps. In the first step, I built a list of URLs scraping the results of Google.co.uk. In the second step, I expanded the list through the use of a crawler. The final list includes 2503 URLs (see Table 5.2). The original list contains statements from the actors, and the second list shows how these actors are connected to each other, and to other actors. In order to make sense of such a large number of actors, I had to assign a label to actors, grouping the pages according to similarities. Bearing in mind ANT criticism of the interpretation of data through the adoption of pre-existing categories, I chose to describe groups only as 'group-like' formations whose boundaries have to be constantly redefined (Venturini, 2010; Rogers et al., 2015).

I created two groups from actors that could be ascribed to civil society: the first one including web pages belonging to activists, either part of NGOs/advocacy groups or in their 'individual' capability (i.e. bloggers), the second including web pages from research organisations, such as academia and think tanks. Other groups include news media, social media companies and private companies, and international organisations. I also divided the group 'Public bodies' into 3 sub-groups: public bodies (such as schools and police), government and government bodies and politicians. Table 5.4 presents a description of the groups and sub-groups of actors.

Table 5. 4 - Groups of actors explained

Group	Sub-group	Description
News media	Newspapers, magazines, broadcasting channels,	Online version of newspapers, or websites with clear information purpose
Associations/Activists (Civil Society) Activist	Blogger	Individuals with no public role active on blogs
	Activist	Individuals with official public role
	NGOs/Associations	Non-profit entities (might be foundations, or charities)
Academia (Civil Society)	Academia, research centres and think tanks	Universities, academic research groups and university student associations, think tanks
Social media companies, private companies	Private companies	Social media companies (e.g. Quora, Reddit) but also law firms, private research centres such as think tanks, web designers but also television channels, or platforms with a clear commercial purpose (i.e. e-commerce platforms, Amazon, eBay...)
International organisations	International/ European organisation	International or European organisation (e.g. UN, Unesco, EU...)
Public bodies	Public bodies	Schools, NHS departments, also law enforcement agencies
	Politicians	Individuals with political commitment and public role
	Government and government bodies	Government or Parliament/ Ministries/Agencies

First list

Based on this grouping system I explored the frequencies of each group. Table 5.5 reports the summary of the overview of actors by group and divided by year.

Table 5. 5 - List of actors from list 1 – scrape of Google.co.uk

Actors	2015	2016	2017	2018	Total
News media	43	58	103	78	282
Academia	41	48	44	64	197
Association/ Activists	29	19	29	42	119
Social Media/ Private company	16	25	27	22	90
Public body	20	19	12	2	53
Int org and EU institutions	2		4	18	21
Grand total	151	169	219	226	765

From these initial figures presented in Table 5.5 it can be observed that in this first data collection the majority of web pages correspond to actors from the group of news media (n=282, i.e 37% of total actors), academia, associations and activists, while public bodies and political organisations are less present. The proportion among groups also remains more or less the same considering the different years of data collection (see Table 5.5). This indicates a sort of stability in the proportion of type of actors (we do not know if this might be a result of the tool). Another of the initial findings is that the great majority of web pages captured represent unique examples of publications from a specific source, a sort of 'one shot' expression of interest in the controversy from the corresponding actors rather than a large amount of content produced by the same actor. This results empirically in a long list of URLs to articles or pages from home domains which appear only once in the dataset. This is not surprising in an online environment, where often distribution of content follows power distribution, with a small number of elements producing exponentially more content than the majority.

Second list

As described in chapter 4, using the crawler Hyphe I retrieved other websites connected to the first list of web pages described above. This step significantly increased the number of URLs to be coded as actors. However, after the coding, the overall proportions of the groups of actors remained stable: news media remains the first group by number of web pages, followed by associations and NGOs, academia, private companies, and finally international and European institutions. Table 5.6 presents the results after the crawl. The crawler expanded the number for all groups, and the final result is a list with similar proportions across groups. Comparing the final list with the initial one, the main change is that the crawler expanded the presence of pages from NGOs and associations, slightly above the pages from academia. The final list does not include 1480 web pages retrieved which became 'not assigned'. They mainly represent non-related links of advertisements (bets, football websites) but also private companies offering software, design for website, marketing and social media management, or pages that the crawler included that did not correspond to the year of the data collection.

Table 5. 6 - Expanded list of actors[9]

Actors	2015	2016	2017	2018	Grand Total
News media	141	197	308	347	993
Association/Activists	111	126	118	160	515
Academia	103	80	118	170	471
Social media/Private company	56	95	72	59	282
Public body	36	40	27	65	168
Int org/EU institution	6	5	24	39	74
Grand Total	453	543	667	840	2503

5.3.2 Overview of actors

News media

'News media' represent the largest group of web pages collected. The overview shows a wide spectrum of publications, from different countries and created for different audiences. Examples of the most frequent publications are articles from national newspapers in their online version, for instance the Guardian, Independent, Telegraph and Daily Mail as well as digital publications from the BBC website and the Huffington Post (Table 5.7). Among the news media actors with more than one publication there are also publications specialised in the topic of technology, such as computer science or technology publications like Wired and CNet (some others not in the table but worthy of mention are PCmag or Arstechnica).

[9] News media: in 2015 141 (43), in 2016 197 (58) in 2017 308 (103) in 2018 347 (78) total 993 (282)
Association/NGOs/Bloggers in 2015 111(29), in 2016 126 (19), in 2017 118 (29) in 2018 160 (42) in total 515 (119)
Academia in 2015: 103(41), in 2016 80 (48) in 2017 118 (44) 2018 170 (64) total 471 (197)
Social Media/Private companies in 2015 56 (16), in 2016 95 (25), in 2017 72 (27) in 2018 59 (22) in total 282 (90).
Public body in 2015 36 (20), in 2016 40 (19) in 2017 27 (12) in 2018 65 (2) in total 168 (53)
International organisation/EU institution in 2015 6(2), in 2016 5 (-), in 2017 24 (4), in 2018 39 (18) in total 74 (24).

Table 5. 7 - News media publications that appear most often, across the years (first list)

Actor	2015	2016	2017	2018	Grand Total
theguardian.com	3	3	7		13
bbc.co.uk	4	2	3		9
independent.co.uk	2	1	4	1	8
wired.co.uk	1	3	2	1	7
dailymail.co.uk	2	2	2		6
mirror.co.uk	2	2	2		6
politico.eu	2		2	2	6
telegraph.co.uk	1	2	2		5
cnet.com	1	2	1		4

Table 5.7 does not include the vast majority of web pages collected from news outlets with only a single page in the sample. They include a wide variety of actors, from tabloids to local newspapers (e.g. Walesonline, Belfast Telegraph), as well as groups of small or 'independent' publications outlets. Some of these are commercial websites,[10] and they include news mixed with a large number of advertisements. Others are openly partisan publications, both from left and extreme right politically affiliated groups, such as the news outlet Radix Journal, published by the alt-right activist Richard B. Spencer. The crawler expanded the list by mostly connecting the specific pages of the articles to the main homepage of the publications.

[10] A rapid overview of the texts contained in these web pages revealed very poor quality of content (no references to studies or detailed exploration of the issue), often repeated across different pages. The suspicion is that these pages might belong to companies that publish large amounts of textual content specifically designed to satisfy algorithms for maximal retrieval by automated search engines. Websites such as 'The Odyssey' and 'The Blaze' can be considered of this kind. Their presence raises a problematic aspect in the study of the controversy, as these websites copy content from other websites with no real 'editorial' rationale. In a way they can only be considered as a signal amplifier for the content produced by others.

Associations and activists

This group includes articles and publications online from websites that belong to NGOs, advocacy groups, bloggers and activists. With regard to frequency, it is the second largest group in terms of the number of web pages retrieved. Table 5.8 shows an overview of the websites from NGOs / associations and bloggers that appear the most in the list scraped from Google.co.uk. The group is very large and varied, and also in this case a small minority has produced most of the content. These are NGOs or advocacy groups active in the protection of human rights and freedom of expression in particular (Human Rights Watch, European Digital Right Initiative (EDRi), Pen International and Front Line Defenders; IFEX (International Freedom of Expression), Article 19, and the Electronic Frontiers Foundation, as well as groups focusing on children, minorities, and women).

Table 5. 8 - Summary of NGOs with more than 2 pages collected in the dataset

Actor	2015	2016	2017	2018	Grand Total
article19.org	1	2	1	5	9
EDRi.org		2	2		4
en.wikipedia.org				3	3
privacyinternational.org	1	1	1		3
statewatch.org	1	2			3
apc.org				2	2
cipesa.org			1	1	2
eff.org				2	2
gp-digital.org			1	1	2
hfhrpol.waw.pl				2	2
hrw.org				2	2
ifj.org	1		1		2
rankingdigitalrights.org				2	2

The crawler expanded the list of NGOs and academic institutions, by connecting the original link of a document to the main home page of the organisations. However, among the discovered pages there were also NGOs which were not present in the original list, especially groups active in freedom of speech and freedom of press advocacy: ACLU, Reporters Sans Frontiers, Human Rights Watch, Freedom House, and the Committee to Protect Journalists. In this case the crawler actually expanded the list of actors.

Bloggers

In the same group I also included pages produced by bloggers. As explained above, I decided to include them with NGOs and advocacy groups, since the majority of blog pages collected belong to individuals who are particularly passionate about a specific cause and can be considered as ‘activists’. In the group there are two main types of bloggers: those who are members of advocacy groups and NGOs or political parties, and those who are not associated with any specific group. The latter group show extreme political views (either left- or right-wing). Among the ‘activist’ bloggers many are affiliated with NGOs and are either academics or specialists, active in raising awareness on human rights. An example is Privacy Surgeon, the blog by Simon Davies, founder of Privacy International. Others are individuals active in the technological field, such as http://technollama.co.uk (TechnoLlama is the online persona of Dr Andrés Guadamuz, a technologist). However, the second most typical type of blog collected belongs to politically interested individuals, from extreme sides of the political spectrum from leftists (MoronWatch) to white supremacy bloggers such as Fortress of Faith (a blog with mostly Islamophobic content). The crawler did not increase the list of bloggers; however, as explained below, it did connect them to social media platforms and other websites.

Academia and think tanks

Several web pages collected are from academic and think tank publications and journal articles. For the analysis, I preferred to keep journal articles and long documents separated, as their length exceeds by far the average content of a web page. The categorisation of actors in this case included only web pages, usually including abstracts of publications or the description of research projects. Table 5.9 shows that the majority of the web pages collected belong to universities and research centres in the UK and the US; however, some are from India and other geographical areas.

Among the academic publications, Oxford University Press, the London School of Economics and the University of Warwick have the largest visibility, together with the big publication houses such as Emerald Insight and Springer (see Table 5.9). In the dataset appear several research projects, such as the Free Speech Debate from Columbia University and the Digital Wildfire project managed by researchers from Oxford University (Oxford Internet Institute), and other universities (including Cardiff). Table 5.9 shows the academic or think tank groups with more than 2 pages collected in the dataset.

Table 5. 9 - Most frequent academia and think tank actors

Actor	2015	2016	2017	2018	Grand Total
academic.oup.com	2	2	2		6
emeraldinsight.com	1	1		3	5
wrap.warwick.ac.uk	2	2		1	5
blogs.lse.ac.uk	1	1	1		3
eprints.lse.ac.uk	1	1	1		3

As in the case of NGOs, the crawler did expand the list of academic institutions and think tanks, mostly by including the home pages of the universities, but also in some cases introducing to the list original new entries such as the Citisenlab (a research centre in Toronto investigating states' control over digital communication and breach of human rights).

Public bodies

Public bodies, and governments in particular, are also present in the web pages collected, although in lower numbers compared to the other groups. Among the web pages, the UK government and publications from Parliament emerge (see Table 5.10). Other public bodies collected in the list are police or school publications relating to content management policies applicable within the organisations.

Table 5. 10 - Most frequent public bodies actors

Actor	2015	2016	2017	2018	Grand Total
gov.uk	2	2	1		5
courtsni.gov.uk	1	1			2
derbyshire.police.uk	2				2
publications.parliament.uk		1	1		2

Table 5.10 shows the public bodies with more than 2 pages collected in the dataset. The groups with lower numbers of pages in the first part of the data collection are private companies and international organisations or European institutions. The crawler did expand the list, since it added to the original list important pages, such as the Information Commissioner Officer and the Crown Prosecution Service. It also included several other governments' pages, from the German Ministry of Justice to the US Justice Department, Copyright department, Congress and White House.

Social media and private companies

In the group 'Social media and private companies' I included different types of actors (see Table 5.11). The vast majority are web pages from commercial companies offering counsel about social media policies for private businesses (e.g. blog posts from law firms) or social media management or design for websites. Big social media platforms (Facebook, Twitter, YouTube) surprisingly feature rarely on the list, and if they appear it is with one page. Others such as Reddit or Quora have more pages than the giants. One reason might be related to the timespan of the data collection. Since 2018 SM platforms have been called in front of several national parliaments (see US/UK interrogations) and they had to produce much more original material relatable to the issue at study. A second reason might be related to the collection through keywords, and the privacy protection of users in SM platforms. Pages that can only be seen behind a password are not retrievable through Google.co.uk. Table 5.11 shows the social media or private companies with more than 2 pages collected in the dataset.

Table 5. 11 - Most frequent private companies

Actor	2015	2016	2017	2018	Grand Total
youtube.com				3	3
quora.com				3	3
yaircohen.co.uk		1	1		2
woodfines.co.uk	1			1	2
medium.com			2		2

Yaircohen and Woodfines are law firms with a particular interest in social media.

The most interesting finding for this group relates to the action of the crawler. The Big Tech names appear after the expanded list created after the crawl. The expanded list includes a larger presence of pages from or linked to social media platforms, such as Facebook, Instagram and WhatsApp, Twitter, Google and YouTube and Reddit. From 2016 Snapchat, Tumblr, Microsoft and Amazon and Gab.ai are added. They are a minority (see Table 5.6) in terms of the overall number of URLs and most are home pages (which does not include statements related to the issue of free speech). However, as I explain below, they occupy a pivotal role in the network of hyperlinks holding the web pages together.

International organisations (IO) and European institutions

A minority of statements on web pages were published on web pages belonging to international and European organisations. Here we find pages published on websites that belong to the EU Commission, the Council of Europe, and the Organisation for Security and Cooperation in Europe (OSCE) alongside the UN and Unesco. The group also includes multi-stakeholder initiatives, such as the Internet Governance Forum (see Table 5.12) and other mixed initiatives such as the Global Network Initiative (GNI), already described in the literature as some of the loci where governance of freedom of speech takes place (Gorwa 2019a, 2019b, 2020).

Table 5. 12 – Most frequent International organisations or European institutions

Actor	2015	2017	2018	Grand Total
intgovforum.org			3	3
osce.org		1	2	3
unesco.org		1	2	3
coe.int		1	1	2
ohchr.org			2	2
cordis.europa.eu			1	1
ec.europa.eu			1	1
europa.eu			1	1
globalnetworkinitiative.org/			1	1
igf2017.sched.com		1		1
oecd.org			1	1
statewatch.org	1			1
unesdoc.unesco.org			1	1

The crawler made interesting additions to the original lists: the ICT Coalition for Children Online, and the Manila Principles (also mentioned in the literature as one of the regulatory initiatives between private actors and civil society).

5.4 Types of documents

The web pages collected in the original list present a wide variety of types of content: they range from the online version of newspaper articles to reports by NGOs or public bodies, official statements by political organisations (as in the case of the European Parliament) to academic project descriptions and abstracts to blog or social media posts (as in Quora). The documents vary in terms of length and level of specificity, from short articles published on news sites to extremely specialised publications, such as peer reviewed academic studies. In between, there are 'work' documents, such as NGO policy reports, and opinion pieces from experts in blogs or on news media.

As mentioned above, only after the second crawl did the list include posts or pages from social media platforms (mainly Facebook, Twitter, LinkedIn, GooglePlus). These documents do not necessarily include statements related to the controversy and present a challenge to the traditional methodology of controversy mapping with web pages, since the URL or web page belongs to the social media platform, but they ‘represent’ other actors (for instance, the Guardian page on Twitter, or bloggers’ Facebook accounts, etc.). Often, specific pages and accounts are not visible if protected by privacy settings on social media. Even if these pages do not tell much in terms of the most *substantive* elements of the controversy, they represent interesting findings concerning the structure of the public sphere online.

5.5 Study of hyperlink relationships

The overview of the group of actors highlighted some groups (such as news media, NGOs and civil society) that produced more statements on the controversy than others. It could be argued that a larger presence of statements published online concerning the issue can be indicative of a higher level of influence (i.e. more materials produced, greater chances of mobilisation of other actors). However, controversy mapping and ANT distinguish actors' power on the basis of their capacity to either ‘mobilise’, i.e. the actors capable of challenging or stabilising forms of order, or assign roles to the others. As highlighted in the literature review, influence in online controversy mapping is more often interpreted either as the product of connection of a web page (i.e. hyperlinks as in Google Page Rank and in Jacomy et al., 2018) or as the frequency with which certain specific contents, topics or issues are reproduced (i.e. as in Boullier, 2018).

In this part of the study I focus on the study of connections. I applied the controversy mapping technique, using the idea that hyperlinks reveal the network of relationships connecting the actors composing the ‘public’ of the controversy. As stressed by Venturini, the network visualisation is interesting as it helps reveal what is ‘visible’ to each of the actors in the controversy. Below I focus on the contents, and in particular on those cases and storylines repeated by different actors (as in Boullier, 2018). Actors can be connected in different ways: mentions, references, similar opinions or vocabulary when talking about specific issues. In this section, I present the analysis of connections between actors based on the hyperlinks that connect the different URLs.

5.5.1 Interpreting the networks

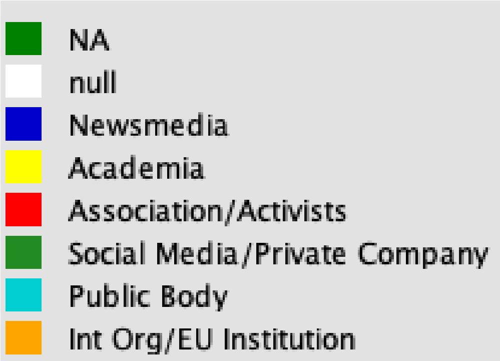
I used Hyphe to study hyperlinks connecting the URLs collected. The images below (Figures 5.3 to 5.10) were produced by visualising the network of hyperlinks connecting the different URLS of the web pages collected in the scrape. The circular shapes represent what in social network analysis are called nodes. The lines represent what in social network analysis are called edges. The words are the labels for nodes, and show the name of the corresponding actor or group. For reasons of readability of the graphs I include only the labels only of the nodes that I am describing (either because they occupy a central position in the network or because of relevance).

In the network, the nodes represent the websites from the original and expanded list. The edges represent hyperlink citations; that is, when one page inserts a link to another page. In Ooghe-Tabanou' et al. methodology (2018) it is assumed that hyperlinks correspond to 'mentions', i.e. it indicates that they are cited as sources by other actors, hence as a measure of influence. The sise of each circular shape represents measures of centrality (or influence) of the node, either in-degree or out-degree (which I describe in greater detail below). In the diagram, the larger the sise, the higher the value.

In graph analytics, measures of centrality are used as measures of importance. In directed graphs, it is possible to distinguish the direction of the relationships, with one node acting as source and another as target of a hyperlink. It is possible to distinguish between in-degree and out-degree. A node with high in-degree stands for web pages that are cited a lot. In social media settings, the nodes with high in-degree centrality will be the nodes that have a huge number of followers or retweets and could be ideal candidates to influence the public (such as the President of the United States) or promote commercial products. In network analysis they are called 'authorities' (Venturini et al., 2017). By contrast, a high out-degree is indicative of an account or web page that cites a lot. In network analysis these are called hubs. Both authorities and hubs are considered influential as they are in contact with more elements than other actors.

The colours of the circles in the graphs below represent the groups of the web pages. The groups are the same as in the coding described above: i.e. News media, civil society in the form of

Academia, NGOs/advocacy groups and think tanks, Social media companies and private companies, Public bodies and International organisations. Figure 5.1 shows the colour coding for each group.



🟩 NA
⬜ null
🟦 Newsmedia
🟨 Academia
🟥 Association/Activists
🟩 Social Media/Private Company
🟦 Public Body
🟧 Int Org/EU Institution

Figure 5. 1 - Colours assigned to each group

The thickness of a line on the graphs represents the weight of the edge (i.e. the frequency of hyperlink connection). The layout (i.e. the placement of the nodes in the space) has been decided by one of the algorithms available in Gephi, OpenOrd, starting from a Random Layout. These layouts are called force-driven placement algorithms because they place the nodes only in function of their links, i.e. the algorithm groups together nodes that are more connected and separates them spatially from the ones with which they have fewer links in common, with the aim of making clusters more visible. Being an algorithm based on the 'force' of the links, it ignores other attributes (such as the colour of the group of the actors). It works iteratively by having all nodes repulse each other and connected nodes attract each other (as in physics simulations). The resulting projection is said to be isotropic: it has no specific axes and could be turned or flipped without losing its features. It is supposed to be interpreted in terms of relative distances. The final position of nodes has been adjusted so that they do not overlap, bringing a minor bias but optimising readability during visualisation.

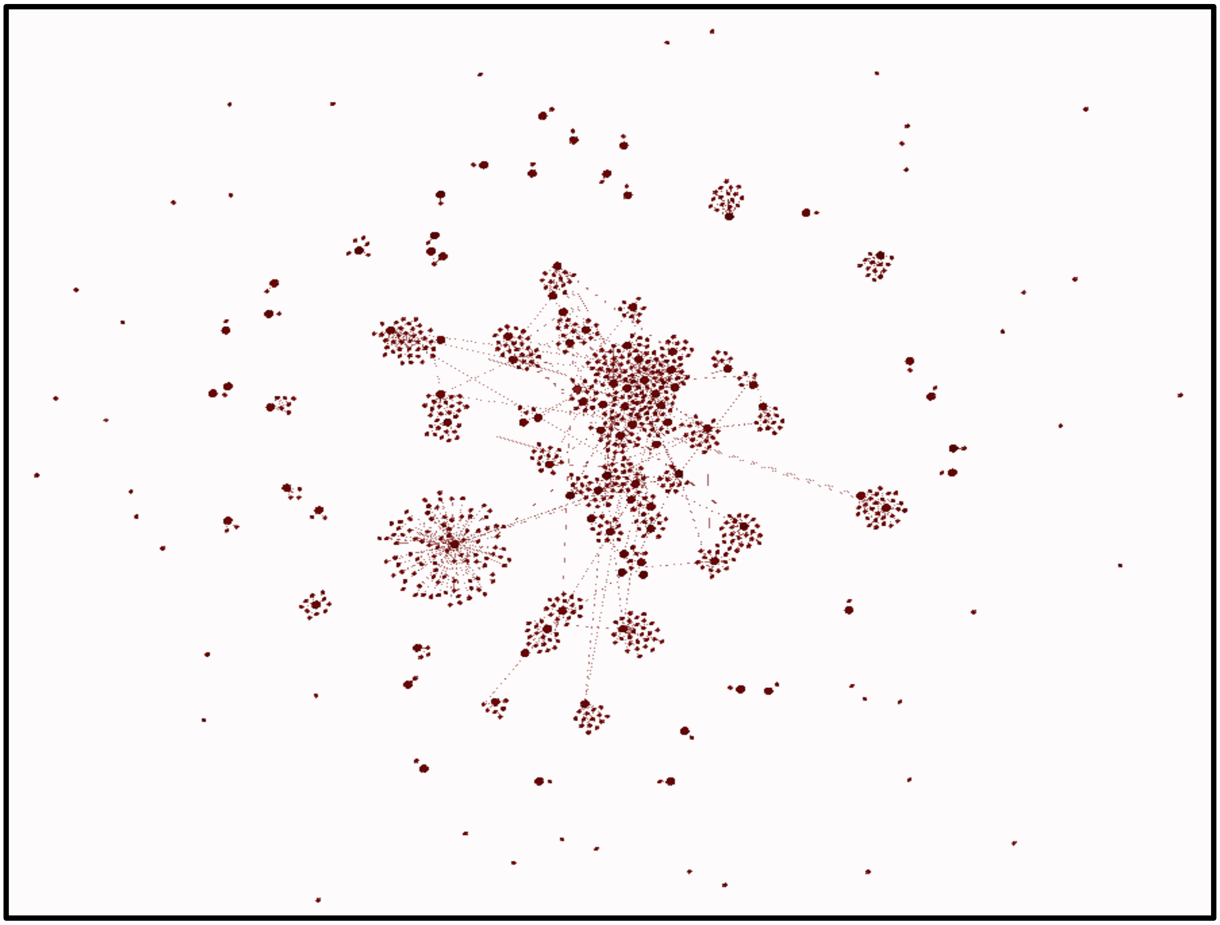


Figure 5. 2 - Example of overview of a network

Network representing the links between 2015 URLs from the initial list and the ones added by the crawler.

Table 5. 13 - From URLs to networks

Year	URL original list	URL after crawl	Nodes	Edges
2015	151	877	877	947
2016	169	1126	1126	1310
2017	219	1420	1420	1945
2018	226	1544	1544	1977

All network visualisation in each year presents a core component deeply connected, with a crown of isolated nodes (see Figure 5.2). All networks in all years have a 'floral' shape: some nodes are the centre of communities, just as the pistil is at the centre of petals. The floral distribution tells us that only a small amount of content is repeated across communities, rather than flowing without constriction across the whole network. This is an important element, as it confirms Boullier's (2018) idea that controversy mapping and ANT is about following the semiotic elements that travel across different actors, and the different associations that are thus created. Newspapers are the only nodes that really connect to each other (the others do not have many links).

5.5.2 Analysis of measures of in-degree/out-degree

The figures below (Figure 5.3- 5.6[11]) represent an overview of the networks of original and discovered web pages in the four different years, with a focus on the measure of in-degree. It is possible to distinguish the part of the network with a higher number of connections across nodes, as the messy part in the centre of the image. However, larger shapes (i.e. nodes with highest in-degree) emerge. These are the URLs of web pages that have been cited the most by other entities. In Ooghe-Tabanou' et al.'s (2018) methodology and in general in controversy mapping methodology, it is assumed that web pages that are the target of many hyperlinks correspond to 'mentions', i.e. that they are cited as sources by other actors, and hence are a measure of influence.

[11] Larger visualisations are available in the appendix

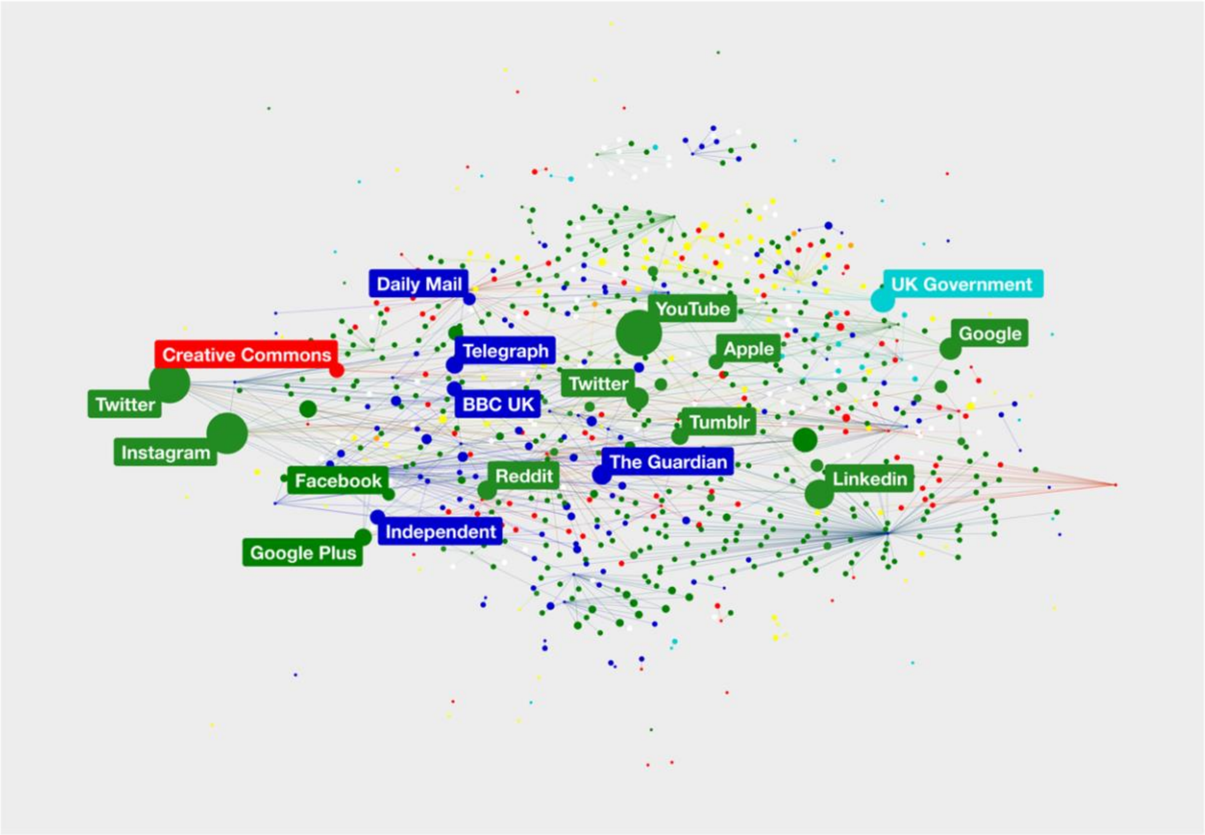


Figure 5. 3 - Nodes with higher in-degree in 2015, coloured by groups

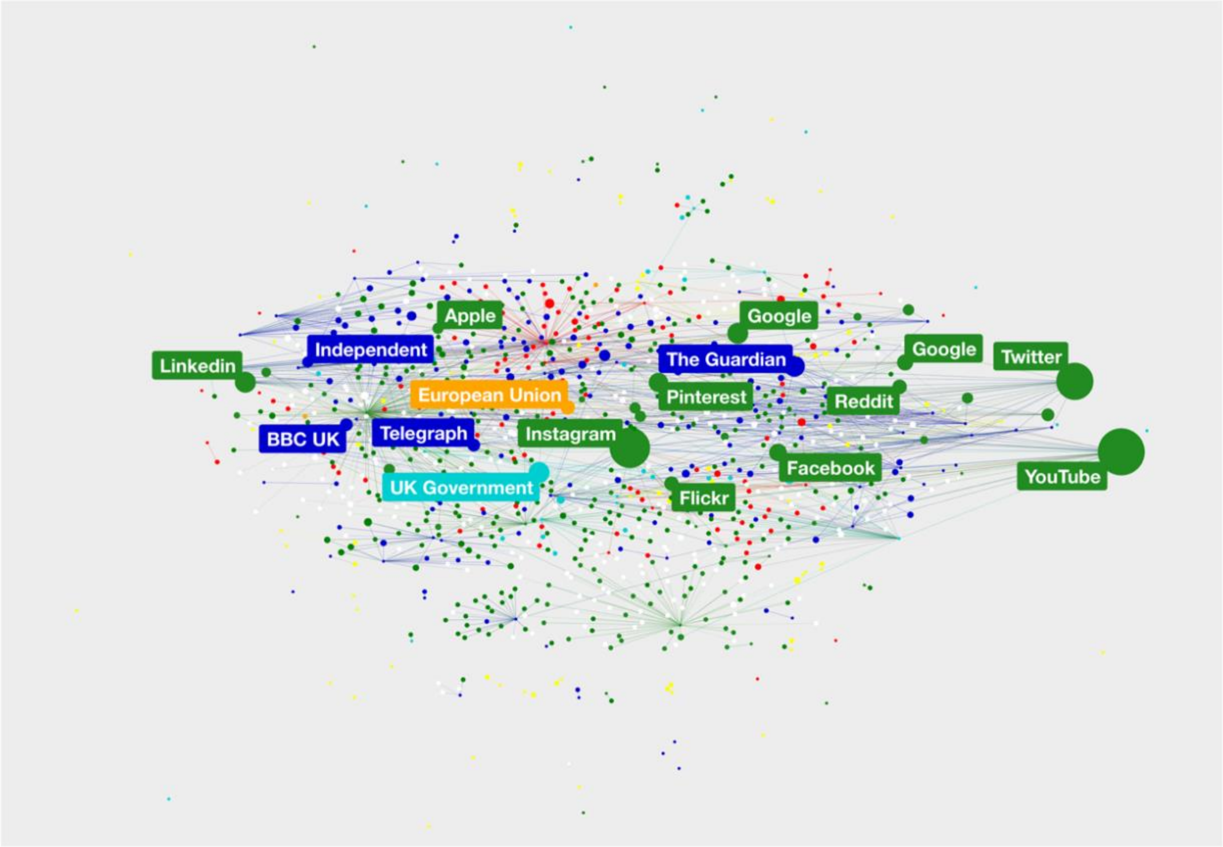


Figure 5. 4 - Nodes with higher in-degree in 2016, coloured by groups

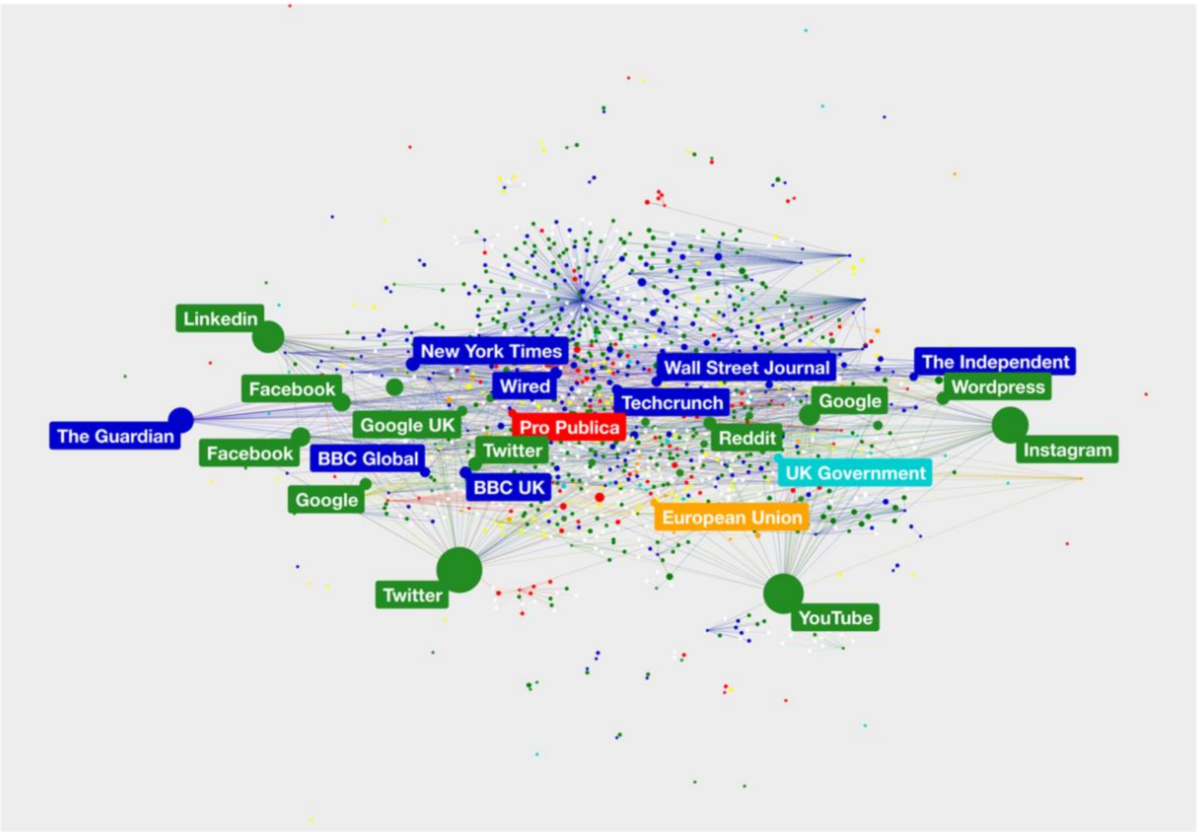


Figure 5. 5 - Nodes with higher in-degree in 2017, coloured by groups

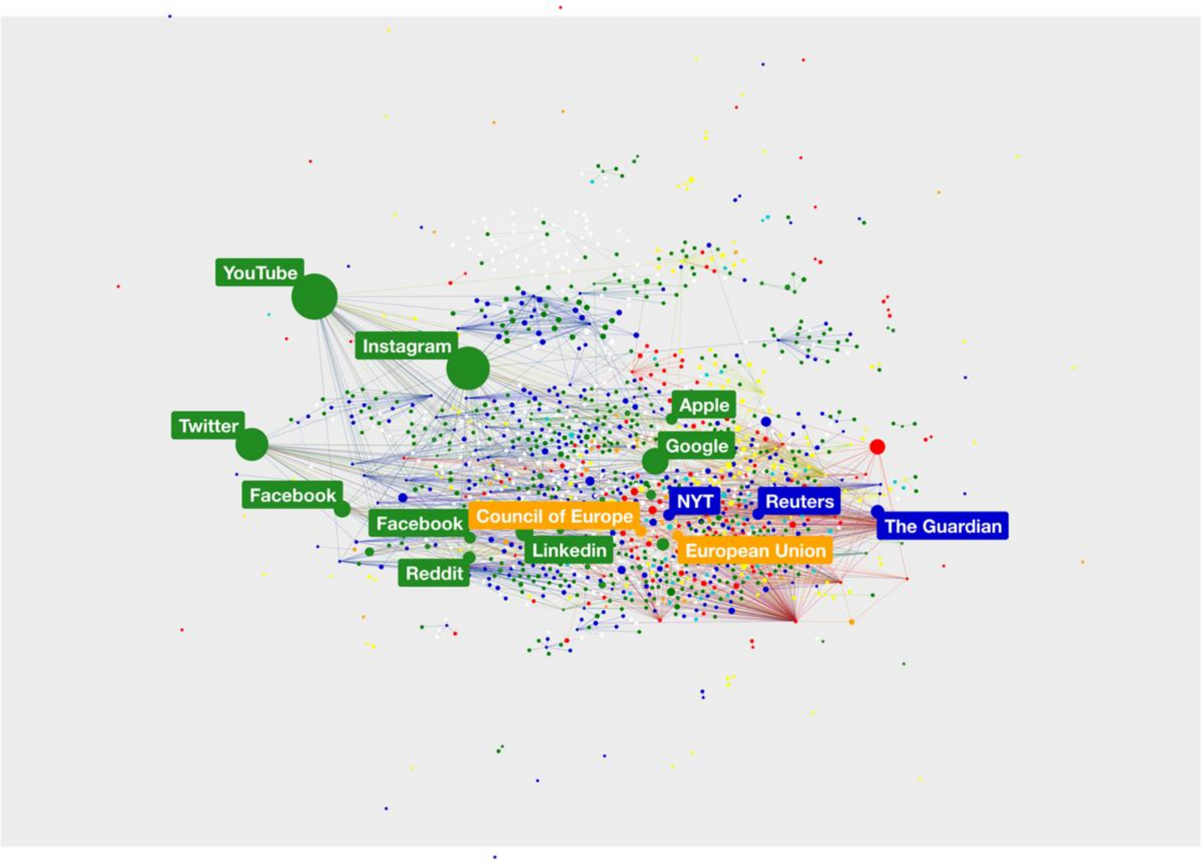


Figure 5. 6 - Nodes with higher in-degree in 2018, coloured by groups

If we compare each image with the analysis of the out-degree (Figures 5.7 to 5.10 below), it is possible to see how the two measures are complementary, as the nodes that acquire sise in the measure of in-degree lose sise in the out-degree and vice versa.

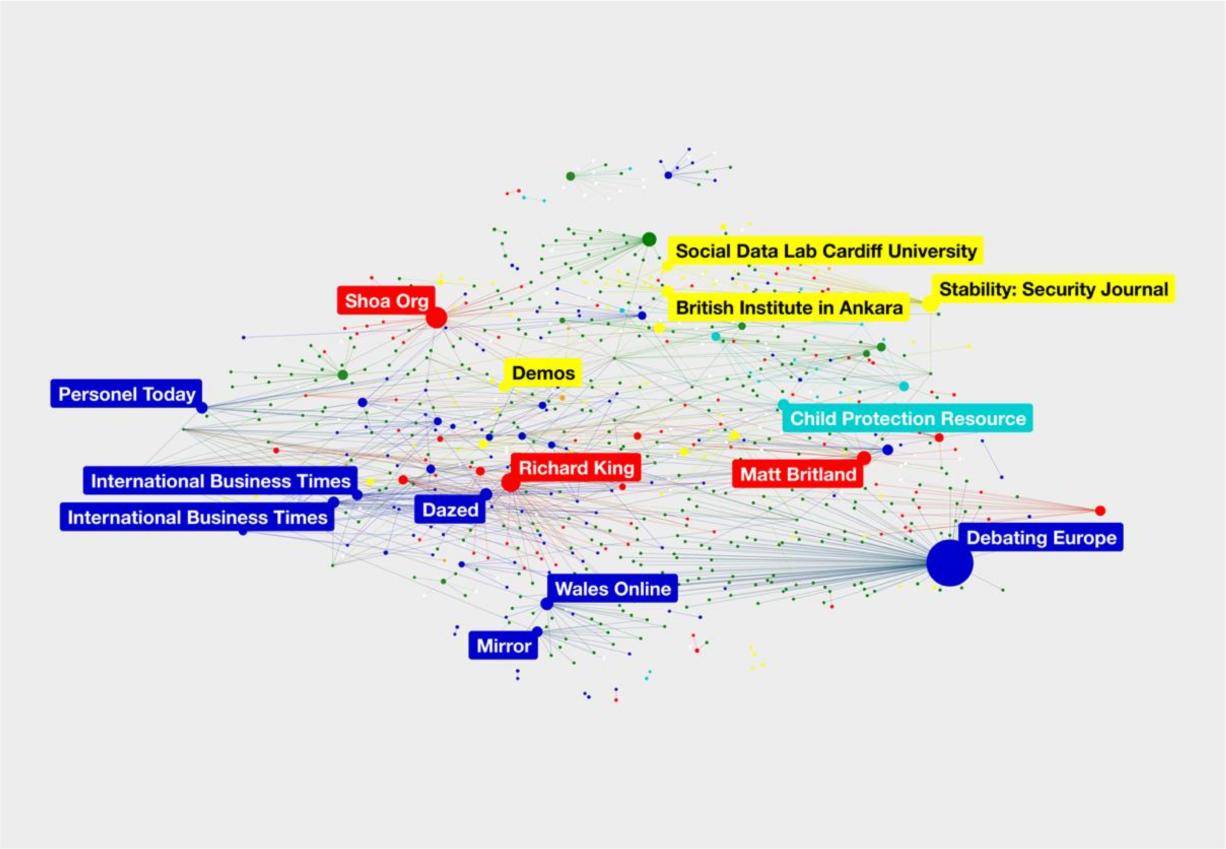

Social Data Lab Cardiff University
Stability: Security Journal
Shoa Org
British Institute in Ankara
Demos
Personel Today
Child Protection Resource
International Business Times
Richard King
Matt Britland
International Business Times
Dazed
Debating Europe
Wales Online
Mirror

Figure 5. 7 - Nodes with higher out-degree in 2015, coloured by groups

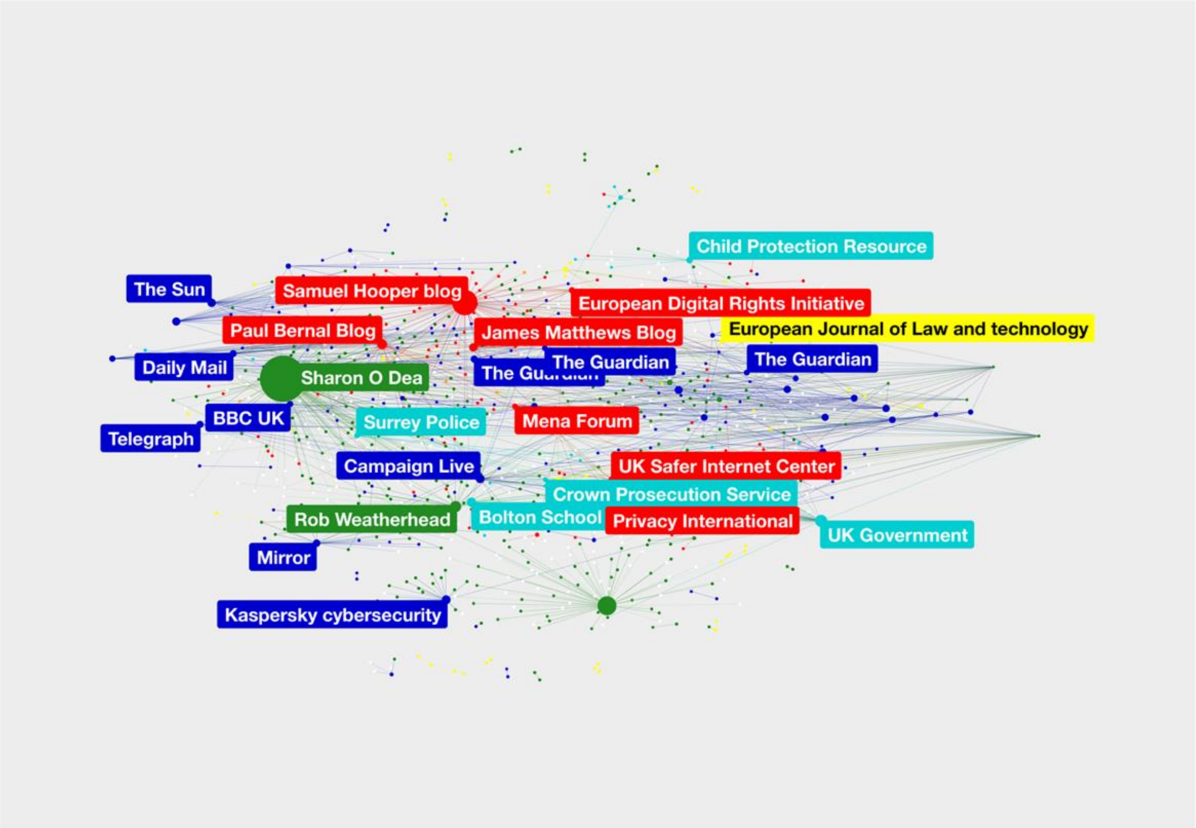


Figure 5. 8 - Nodes with higher out-degree in 2016, coloured by groups

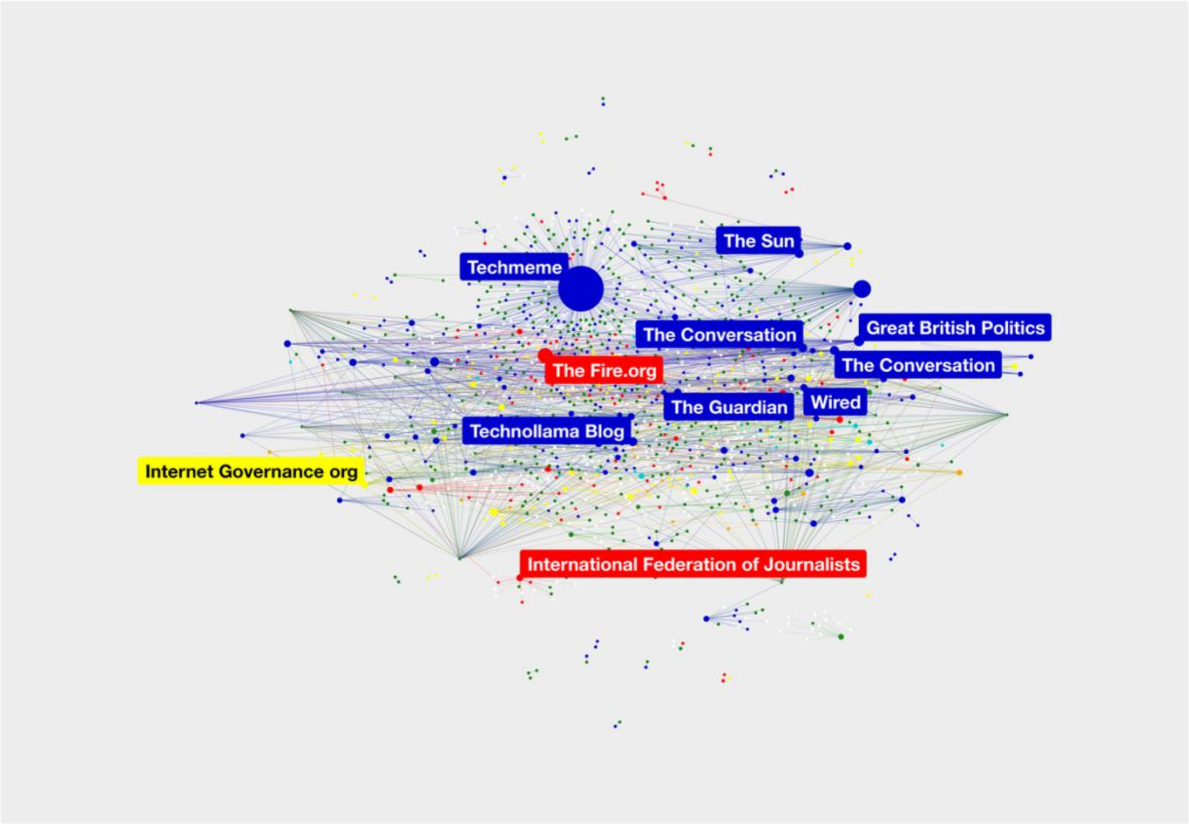


Figure 5. 9 - Nodes with higher out-degree in 2017, coloured by groups

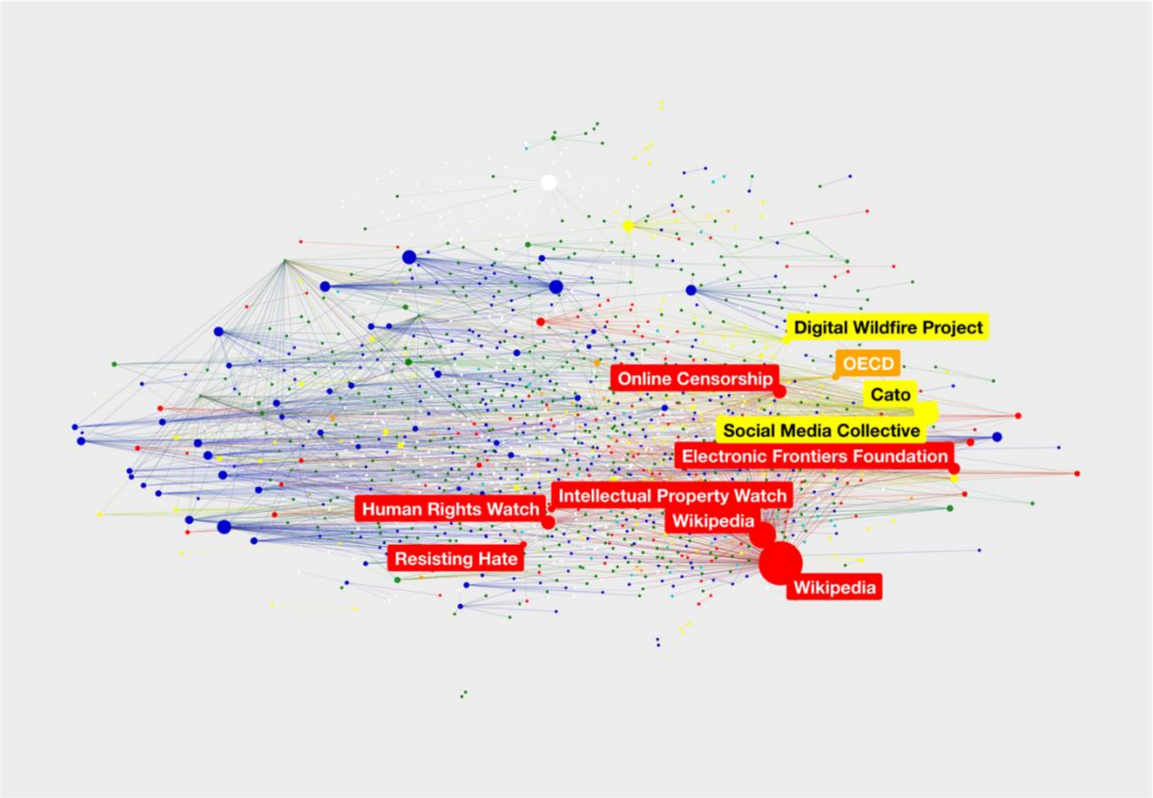


Figure 5. 10 - Nodes with higher out-degree in 2018, coloured by groups

The four visualisations (Figures 5.7, 5.8, 5.9, 5.10) show larger circles for those web pages that cite the most other entities (out-degree). The comparison of in-/out-degree shows a division of the groups, with social media, news media, public bodies and international organisations/ European institutions corresponding to nodes with higher in-degree (e.g. most cited), while on the other side civil society in the form of academia and NGOs, associations and in particular bloggers or individuals stand out as the pages which include the majority of citations. Below I present an overview of the main findings related to the analysis of the structure of hyperlinks, focusing on the different group of actors.

5.5.3 Social media platforms

In the previous part of the study, I highlighted how social media and private companies do not represent a large proportion of the authors of the statements in the controversy. In particular 'big tech' seems to be missing, as it has almost no presence in the lists of URLs. However, the structural analysis of the relationships of the web pages shows a different picture. Across all years, the nodes with higher in-degree values correspond to social media platforms (see green circles in Figures 5.3 to 5.6). This indicates that the most cited web pages are the largest companies: YouTube, Google, Instagram, Facebook, Twitter, Reddit. As seen above, social media platforms were practically absent from the original list of URLs used to build the network, and they are still a smaller number in terms of URLs; however, it is possible to see in Figure 5.11 that even with fewer circle shapes, they have the 'bigger circles' in terms of in-degree.

We know from the previous part of this chapter that these web pages do not include statements about the substance of the controversy. However, the degree of centrality of social media nodes is not only very high, but it is also disproportionate compared with other actors (even newspapers). It is indicative of the monopoly that these companies have on the distribution of content. This type of finding sheds a light on a technological dynamic already highlighted in chapter 3 and 4: these web pages do not say much in terms of the specific issue of freedom of speech and social media, but they show how other actors from other web pages connect and share their contents. It confirms the role of social media as 'publicity devices' (Marres, 2005).

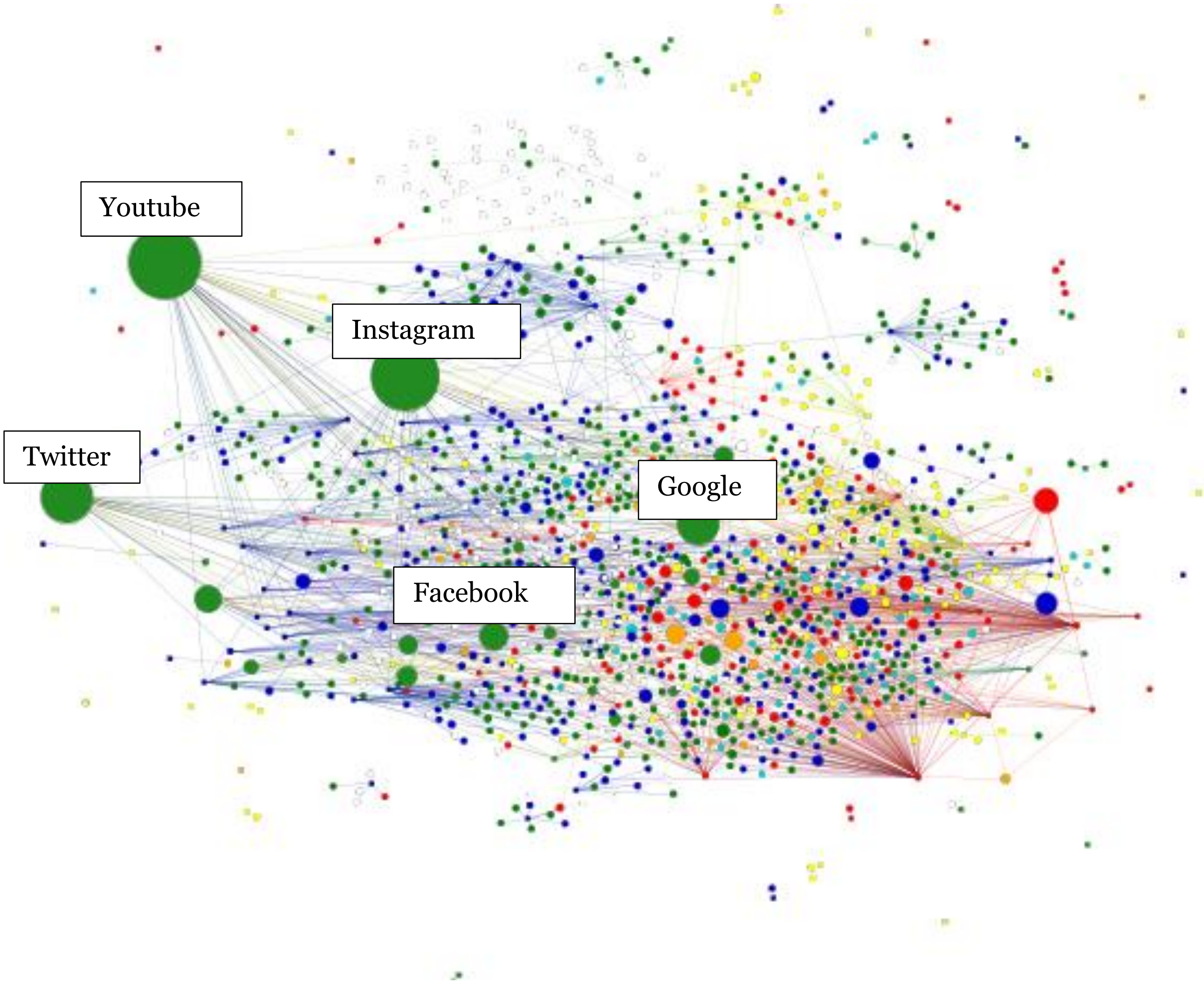


Figure 5. 11 - Detail of the network of URLs in 2018.

Node sise reflects higher in-degree. Social media (green nodes) are out of proportion compared to the other groups.

5.5.4 News media

As noted in the description of actors, news media are the largest category among the actors coded, and not surprisingly they represent the largest number among the more connected nodes (i.e. they tend to have high in-degree). In particular, the Guardian, Telegraph, Independent, Daily Mail, BBC and International Business Times (economy and trade) are the online publications with larger in-degree. Among more technological publications, those that are cited the most are Techdirt, the Register, Ars Technica, PCMag, Slashdot, CNET and the Verge.

The analysis of structure of hyperlinks also shows how reticular is the presence of news media. They are the most connected and distributed group of actors across the years in question. This is visible on larger images of the network, where the blue (news media) nodes and edges appear as the most visible (see Figure 5.12).

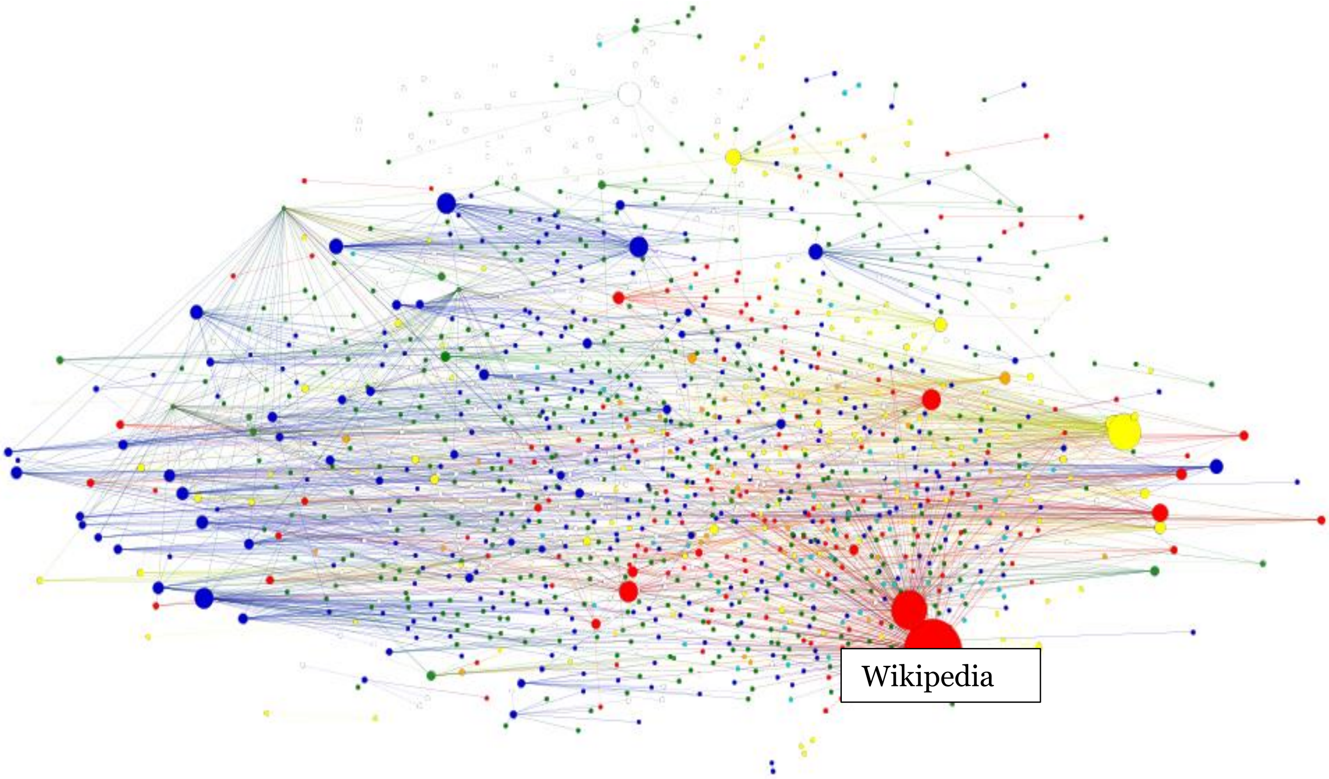


Figure 5. 12 - Overview of hyperlinks connections in 2018.

Larger nodes have higher out-degree. News media are the blue nodes and represent 22.47% of the visible nodes (i.e. the highest percentage of all the groups).

5.5.5 Public bodies and international organisations

Websites of public bodies such as governments and international organisations appear in the group of the most cited web pages. In particular, the UK government stands out as the page of a public body with a high number of citations, together with the Information Commissioner's Office and the Crown Prosecution Service. In other countries, other web pages from public bodies with high in-degree are the US departments such as Justice, Copyright, Congress and the White House.

Among the large sources of content there are also international and European organisations: the European Union and Council of Europe, but also the United Nations, UNHCR, Unesco, OHCHR and OSCE. A very interesting result is also the presence among the most cited 'sources' of actors from the mixed group, such as the ICT Coalition for Children Online, or the Internet Governance Forum, the Global Network Initiative (GNI) and the Manila Principles. Unlike the others, this group of actors has increased its presence since 2015, with progressively higher values on the in-degree scale.

5.5.6 Civil society: NGOs

As noted above, NGOs and associations do not stand out as a group in the analysis of in-degree. Some associations are used as sources, such as: Nesta (which is a charity for funding innovation), the National Society for the Prevention of Cruelty to Children (NSPCC) and Safer Internet Center, a partnership between Childnet International, Internet Watch Foundation and South West Grid for Learning, and coincidentally they are all NGO advocacies or initiatives for child protection.

The list also includes the Independent Press Standards Organisation (IPSO) and other NGOs for the protection of digital rights such as EDRi, ProPublica, Article 19 and the Electronic Frontiers Foundation (EFF), and NGOs largely active on the topic of freedom of speech and protection of journalists, such as ACLU, Reporters Sans Frontiers, Human Rights Watch, Freedom House, Online censorship org, the Committee to Protect Journalists, IFEX, WAN-IFRA and openDemocracy. A special role is occupied by Wikipedia and Wikimedia pages about freedom of expression, which in 2018 become among the most cited websites.

In this group, the node with highest out-degree corresponds to Debating Europe, which is a page of a forum for debate about cyberbullying, collected in 2015. It is an interesting case, as the high number of citations correspond to Facebook posts from participants that took part in the discussion. Other pages of associations citing a high number of others are Shoah.org, the Child Protection Resources, the Birmingham Diocese, African Internet Rights, Get Safe Online, and a large group of free speech actors such as the Index of Censorship, Raif Badawi organisation, Privacy

International, Frontline Defenders, the Great Fire, International Federation of Journalists and European Federation of Journalists.

As noted in the overview of the actors, bloggers have produced many statements, yet they do not get cited very often. The one with the highest citation is a technological blogger: TechnoLlama. The analysis of out-degree shows, however, that bloggers are extremely active in citing other websites (see Figure 5.13). Among the nodes with high out-degree there are also some bloggers, such as Richard King, Matt Britland, Privacy Surgeon, Christopher England and Street Democracy. Bloggers who are activists or ex-activists have connections to pages from NGOs. In terms of distribution of information, it seems that they take the content straight from NGO sources, rather than from newspapers. Examples from 2015 are: Richard Kingdom, who cites Open Rights group, or Privacy Surgeon who cites EDRi and Matt Britland who cites Digital Awareness UK. As stressed by Jacomy (2018), out-degree too can be considered as a measure of centrality, as it is an indicator of which actor 'moves' more contents.

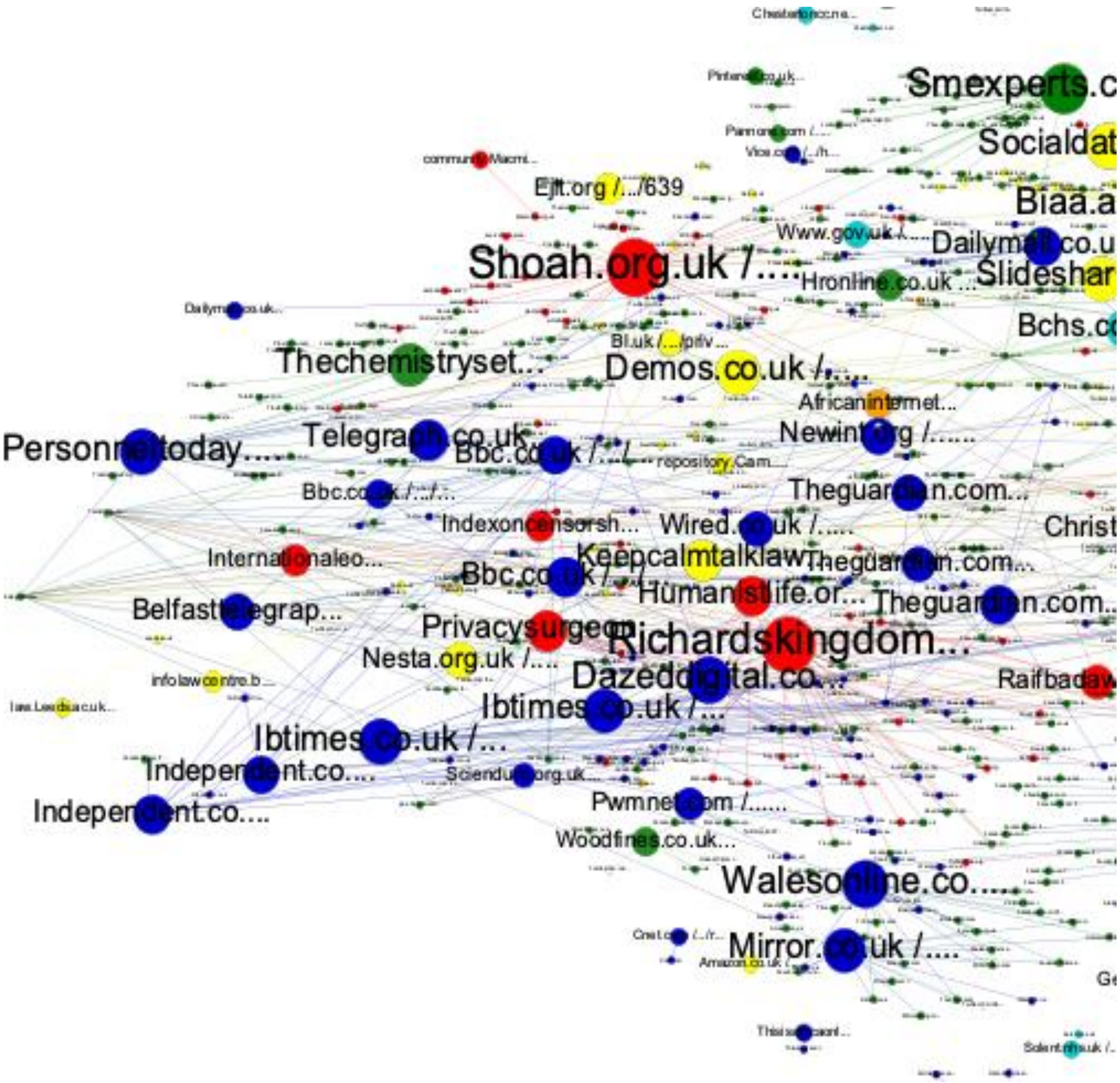


Figure 5. 13 – Detail of nodes with higher indegree in network in 2015

The red circles are NGOs and bloggers, and are among the most active in citing other websites

5.5.7 Academia and think tanks

Academia in the graphs is represented by the yellow circles. The most cited web pages belong to the London School of Economics, Columbia University, and web pages of publishers (Oxford University Press, Routledge, Springers, Typepad) and funding bodies such as the ESRC, University of Kent and University of Oxford. Overall, academia looks a bit disconnected from the other pages; however, this might be because the majority of the original URLs from academic sources are pdfs, which do not necessarily include other hyperlinks. This hypothesis would be confirmed by the high number of hyperlinks connecting to academic publication editors, which probably is the immediate connection that the crawler could find from the pdfs. However, academia appears among the nodes with higher out-degree, especially research centres. For instance in 2015 (Figure 5.14) the Social Data Lab from Cardiff University and the think tank Demos appear among the most active in ‘citing’ other websites.

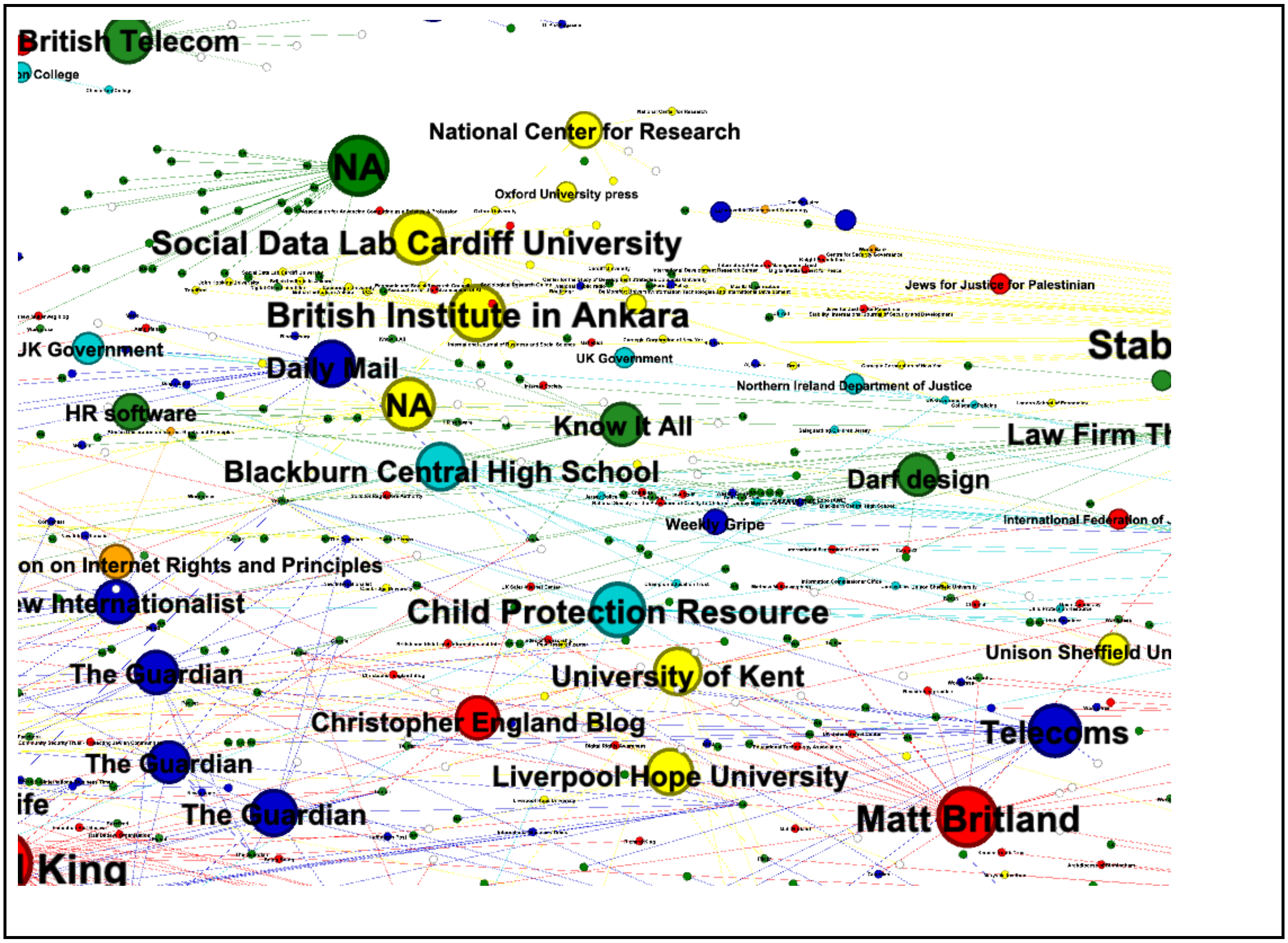

British Telecom
College
National Center for Research
NA
Oxford University press
Social Data Lab Cardiff University
Jews for Justice for Palestinian
British Institute in Ankara
Stab
JK Government
Daily Mail
UK Government
Northern Ireland Department of Justice
HR software
NA
Know It All
Law Firm Tl
Blackburn Central High School
Darl design
Weekly Gripe
International Federation of
on on Internet Rights and Principles
w Internationalist
Child Protection Resource
Unison Sheffield Un
The Guardian
University of Kent
Christopher England Blog
Telecoms
The Guardian
Liverpool Hope University
ife
The Guardian
Matt Britland
King

Figure 5. 14 – Detail of network with higher outdegree in 2015

Academic and research centres with higher outdegreel

Also, from the analysis of the nodes corresponding to web pages in academia, it appears that from 2015 to 2018 there is an increase in the names of authors that appear also in the literature review, such as Zeynep Tufekci, Tarleton Gillespie, Kate Crawford and Rebecca MacKinnon. They appear in the network because they have a personal website where they publish their research.

5.6 Other findings from the analysis of structure of hyperlinks

The analysis of hyperlinks also reveals interesting findings concerning which groups tend to be more connected. For instance, there is a constant relationship between associations and NGOs and international organisations and European institutions.

Figure 5.15 shows the particular focus on the hyperlinks connecting the website of the UN Special Rapporteur for Freedom of Expression, the UN and the OHCHR with a group of NGOs active in the field of free speech and human rights.

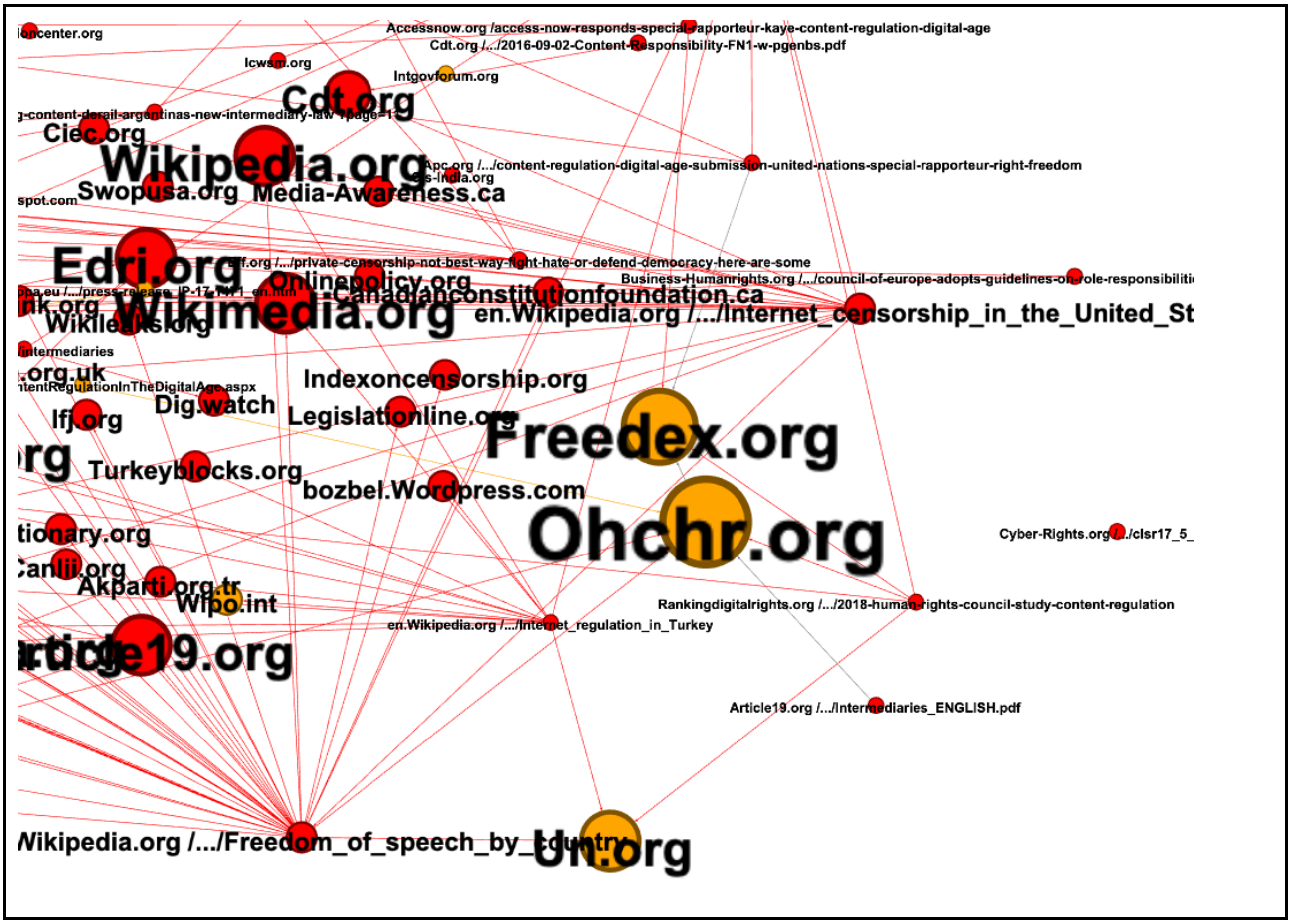

ioncenter.org
Accessnow.org /access-now-responds-special-rapporteur-kaye-content-regulation-digital-age
Cdt.org /.../2016-09-02-Content-Responsibility-FN1-w-pgenbs.pdf
Icwsm.org
Intgovforum.org
Cdt.org
g-content-derail-argentinas-new-intermediary-law-?page=1
Ciec.org
Wikipedia.org
Apc.org /.../content-regulation-digital-age-submission-united-nations-special-rapporteur-right-freedom
Cis-India.org
spot.com
Swopusa.org
Media-Awareness.ca
Edri.org
Eff.org /.../private-censorship-not-best-way-fight-hate-or-defend-democracy-here-are-some
Onlinepolicy.org
Business-Humanrights.org /.../council-of-europe-adopts-guidelines-on-role-responsibiliti
opa.eu /.../press-release_IP-17-1471_en.htm
Canadianconstitutionfoundation.ca
nk.org
Wikimedia.org
Wikileaks.org
en.Wikipedia.org /.../Internet_censorship_in_the_United_St
/intermediaries
.org.uk
ntentRegulationInTheDigitalAge.aspx
Indexoncensorship.org
Dig.watch
Legislationline.org
Ifj.org
Freedex.org
rg
Turkeyblocks.org
bozbel.Wordpress.com
tionary.org
Ohchr.org
Cyber-Rights.org /.../clsr17_5_
Canlii.org
Akparti.org.tr
Wipo.int
Rankingdigitalrights.org /.../2018-human-rights-council-study-content-regulation
en.Wikipedia.org /.../Internet_regulation_in_Turkey
article19.org
Article19.org /.../Intermediaries_ENGLISH.pdf
Wikipedia.org /.../Freedom_of_speech_by_country
Un.org

Figure 5. 15 - Detail of network with higher outdegree in 2018

The network shows high level of connections between associations and international organisations on occasion of the call for proposal issued by the Special Rapporteur on Freedom of Expression

5.6.1 Technological dynamics

The measure of centrality and the direction of the link does not necessarily describe a straightforward relationship linking two or more web pages. Most of the time, meaning emerges only after detailed exploration of all the links and the pages (even switching to reading the hyperlink embedded in the source code). For instance, after an in-depth exploration of the hyperlinks it is possible to say that the large in-degree or out-degree shown by social media platforms is the result of three main possibilities:

1) the web page includes a link to content produced by themselves on a big platform, i.e. web pages mentioning their own social media accounts. This is the case with the page Children Protection Resource, which has embedded the tweets from their account. On the network visualisation this translates as a mention (direct link) towards Twitter/CP Resource account.

2) the web page includes a link to content produced by others on a big platform (direct link towards) and the content hosted on the platform is somehow related to the issue. For instance, Debating Europe is a platform where public debates are hosted and comments can be added also using one's private Facebook account. In the network visualisation, the relationship is described as Debating Europe mentioning content produced on specific Facebook accounts. The content is hosted on Facebook, but it relates to the web page.

3) the web page includes a link to content produced by others on a big platform (direct link towards) and the content hosted on the platform is not related to the issue. For instance, exploring the connections from the Children Protection Resource web page, it is possible to see that a YouTube mention appears at the bottom of the web page, in the section left for the 'extra reading'. In the network visualisation, it does create a link where the Children Protection Resource web page is mentioning YouTube, but the link does not relate to the topic of the web page (issue of interest). Another case that came up is the one linking the Demos research centre to the Telegraph, and the BBC to the Telegraph. In the web page, the BBC was actually citing Demos, but Demos published in the Telegraph, so the direct relation is lost.

The data show that big social media are central in the sharing of the vast majority of content. Every web page existing online now has a link to its sponsors/advertisers or in general to their account on social media or an automatic link to post their content on other people's accounts. In terms of network visualisation, all these possibilities translate to edges pointing at the nodes corresponding to the home pages of these big companies. This confirms a characteristic of the contemporary public sphere, which is embedded in these platforms, and the role of social media as 'publicity devices' used to push content for advertisement purposes (Marres, 2005).

The overview of actors at the beginning of this chapter showed how news media, NGOs and academia are the group of actors producing the most content. The analysis of hyperlink shows, however, how these actors are not used as sources, but rather that they tend to mention other actors. In particular, SM platforms emerge as central elements in the distribution of content. News media also appear as a capillary presence in the network. In this last part of the analysis, following Boullier (2018), I will consider which content has the power of mobilising other actors. Drawing from Pohle (2016) I focus on exemplary cases and storylines recurring across different pages, and the different narratives of freedom of expression, governance and technology that they stimulate.

5.7 Identification of shocks and regulation initiatives

I focused on the texts of the documents collected through the web pages, using a mix of qualitative and quantitative methodology for the analysis of texts.[12] As I described in chapter 4, I adopted the methodology developed in Pohle (2016a,b). I analysed the texts focusing on the emblematic episodes and storylines that are evoked in response to public shocks. I coded the texts, based on the identification of:

1) Emblematic episodes or stories that recur across the texts
2) The type of issues involving freedom of expression and SM platforms they summarise
3) The type of narrative (worldview) of freedom of expression which is intended
4) The type of narrative (worldview) of technology which is intended
5) The type of narrative (worldview) of governance which is intended
6) The type of regulation initiatives they relate to

In this part of the chapter, I analyse the statements produced by the actors in the controversy, excluding the statements from news media. Since news media occupy a particular role in

[12] Note: the texts used in the analysis are extracted from the original list of web pages, since the expanded list included home pages that do not include specific statements. However, the home page is useful to understand how different actors interact (as for instance Demos, and the Telegraph, where Demos published the results of research in the Telegraph, but this type of connection appears only through the hyperlink).

controversy (as explained in chapter 3) and I include newspapers in the empirical analysis of the next chapter, I decided to focus the analysis on the statements of the actors from other groups: academia, NGOs and advocacy groups (including bloggers), public bodies, international and European institutions and private companies.

5.8 Emblematic episodes or recurring topics

I performed qualitative and quantitative analysis of texts, looking for emerging or recurring topics and keywords. For the qualitative analysis of texts, I analysed 354 web pages, approximately 80 for each year, with the exception of 2018 which has a slightly higher number (100). I performed the quantitative analysis with the aid of Cortext Platform. The LDA (Latent Dirichlet Allocation) semantic analysis performs an automatic detection of topics but failed to retrieve a coherent set on websites. I reached this consideration after having used LDAvis tools for assessment of topics coherence, i.e. the visual inspection of graphs and the manual analysis of most representative documents. As I have discussed in the methods chapter (chapter 4), in order to assess coherence of topics I have performed different trials of the tool using different settings to assess differences in the result, I chose the 'optimised' method provided by the machine. In the case of websites, using the affordances of LDAvis to check the visualisation and assess the relevance relevant terms in the topics did not suggest any specific meaningful connection from the point of view of human reading. Also, each of the topics had very low topic-specific frequencies of each term was particularly corpus-wide frequencies of each term. A possible justification for this difficulty is the fact that web pages were saved as pdfs, but lacking a common format, they resulted complicated for the parser.

Instead, I relied on the identification of relevant words (based on chi-square, or TF-IDF) and qualitative analysis of texts. I performed a study of word relevance and compared the_keywords extracted via Cortext with the qualitative analysis of the texts. From the list, I isolated in the texts the exemplary cases and issues with clearer semantic meaning. Retrieving the context of the most relevant words in the documents it was possible to find exemplary cases indicative of 'public shocks' relating to regulation of speech. These episodes correspond to specific 'public shocks', which have shaken public opinion, and that have worked as a catalyst for the actors whose statements appear in the data collection. Table 5.14 shows the list of exemplary cases and relevant stories that emerged from the automatic detection of the most relevant keywords and qualitative analysis of the texts.

Table 5. 14 - Exemplary cases extracted from statements

Year of data collection	**Episodes, storylines identified using Cortext frequent terms and qualitative analysis**
2015	Charlie Hebdo attack, terrorist attacks in Europe
2015	Human rights' activist Raif Badawi's detention in Saudi Arabia
2015/2017	Snowden leaks about the NSA and US Border agents controls of social media
2015	Ashley Madison's website hacked and publication of personal data
2016	Paul Chambers 'Twitter Joke Trial'
2015- 2016 -2017	Harassment against MPs on 2017 general election – Harassment against Caroline Criado-Perez in 2013 – Reclaim the internet campaign
2016	Donald Trump victory on US election, fake news and Russian manipulation of opinion
2017	Charlottesville protest and murder of Heather Heyer. Jo Cox murder in the UK
2017	German government introduction of first law introducing fines for SM platforms failing to take down hate or illegal speech (NetzDG)

1_ Terrorist attacks. Charlie Hebdo

Among the first set of exemplary cases presented by different actors in the controversy appear a group of words related to the wave of terrorist attacks in Europe carried out by members of the Islamist group organisation Daesh, which started in 2015 with the attack on the *Charlie Hebdo* magazine in Paris. The statements collected on web pages show how these attacks have interrupted the 'routine' management of content on SM platforms, and opened questions about freedom of expression and the exploitation of social media tools for the purpose of radicalisation and terrorist recruitment. The specific nature of the first attack in January 2015 aimed at the employees of *Charlie Hebdo* in Paris, opened a discussion concerning the limits to the expression of free speech. The hashtag #Je Suis Charlie that went viral immediately after the attack became the symbol of solidarity and free speech. In the statements of the actors it is also an exemplary case for the larger discussion that started on the limits of free speech.

> I would argue this failure to understand the value of free speech lies at the heart of one of the dilemmas we face in modern democracies where free speech is being gradually eroded – where 'Je Suis Charlie' quickly became 'Je Suis Charlie, but…' (Index of Censorship and Ginsberg, 2015, Document id 26).

Indeed, the terrorist attacks in Europe initiated a wave of anti-terrorism policy initiatives, moving the political agenda from protecting free speech towards an increase in the regulation of speech on and by SM. From the statements, technologies and SM platforms emerge as matters of concern for public authorities, and the target of regulatory initiatives. Terrorist attacks correspond to a fundamental shock, as confirmed in the speech on 'Countering online radicalisation and extremism' delivered by Joanna Shield, UK Minister for Internet Safety and Security, to the George Washington University Centre for Cyber and Homeland Security's extremism programme in April 2017:

> It is no longer a matter of speculation that terrorists and extremists use internet platforms and applications to inspire violence, spread extremist ideology and to plan and execute attacks. Each tragic incident reconfirms it (Shields, 2017, Document id 235).

The attacks are used as background for the necessity of increased regulation of speech, and SM are placed at the centre of these initiatives:

> Terrorists' use of the internet as a sphere of influence will continue to evolve and adapt, and we need new methods to quickly identify and remove terrorist and violent content, and to deliver more effective strategic communications to counter these deadly narratives (Shields, 2017, Document id 235).

However, with the rise of anti-terrorism policy initiatives, there was also an increase in the reaction from the civil society and advocacy groups, in particular from NGOs active in human rights online and the protection of journalists and freedom of expression:

> States are increasingly seeking to seriously limit the capacity of individuals to communicate securely and anonymously. […] Indeed, encryption[13] was the first target of UK Prime Minister Cameron who, in the immediate aftermath of the Charlie Hebdo attack in France, vowed to ban 'a means of communication between people which we cannot read' (Pen International and Clarke, 2015, Document id 49).

Or as described also in the EDRi's Annual report in 2016,

> […] the terrorist attacks in 2015 and 2016 had major implications for our work and impacted the political agenda. […] as a result, we saw a fast-tracking of surveillance and security measures that our analysis found to be in violation of fundamental rights. (EDRi, 2016a, Document id 118)

2_ Exemplary case of Edward Snowden, and Raif Badawi

Other exemplary cases that emerge from the documents are related to the storyline created since Edward Snowden's 'leak' of documents about the US and British mass surveillance intelligence systems, in 2014. NGOs active in the protection of freedom of expression recall Snowden's episode as an example in the description of states' use of surveillance for persecution of freedom of speech and human rights activists.

> Since the Snowden revelations, PEN has worked from the position outlined in the Declaration on Digital Freedom to research the impact of mass surveillance on freedom of expression and writers, and to advocate for surveillance reform (Pen International and Clarke, 2015, Document id 49).

[13] The mathematical process of converting messages, information or data into a form unreadable by anyone except the intended recipient – which protects confidentiality and integrity of content against third-party access or manipulation

Edward Snowden appears when discussing episodes related to surveillance and data protection, as for instance with the introduction of social media checks by the border police in the US in 2017. The blogger Andreas Guadamuz,in his blog TechnoLlama describes the implications:

> […] the US border police may be rolling out a program to harvest contact metadata in an attempt to conduct social network analysis on the subjects, and also to create a social media database. Thanks to Edward Snowden, we already know that the NSA has been involved in surveillance practices that collect data from services, and systems like XKeyscore and Prism are used to gain access to online communications (TechnoLlama Guadamuz, 2017, Document id 332).

This storyline also mentions Raif Badawi, a Saudi blogger who became a human rights case when in 2015 he was sentenced to 10 years in prison and 1000 lashes because of content present on his blog.

> […] In the Middle East and North Africa in particular, PEN has noted the increase in cases of bloggers persecuted under these laws. In June 2015, despite global outcry, the Saudi Supreme Court upheld a sentence of 10 years in prison and 1,000 lashes for the atheist blogger, Raif Badawi, on charges of 'insulting Islam' and 'founding a liberal website'(Pen International and Clarke, 2015, Document id 49).

3_Ashley Madison – data protection privacy

Another exemplary case that emerges from the statements refers to the hacking of the online dating website 'Ashley Madison' in 2016. The episode opened the debate about data protection and exploitation. Richard King, a blogger[14] and an activist involved in the British NGO for the protection of digital rights Open Rights Group (ORG) in his blog 'Richard's Kingdom', uses

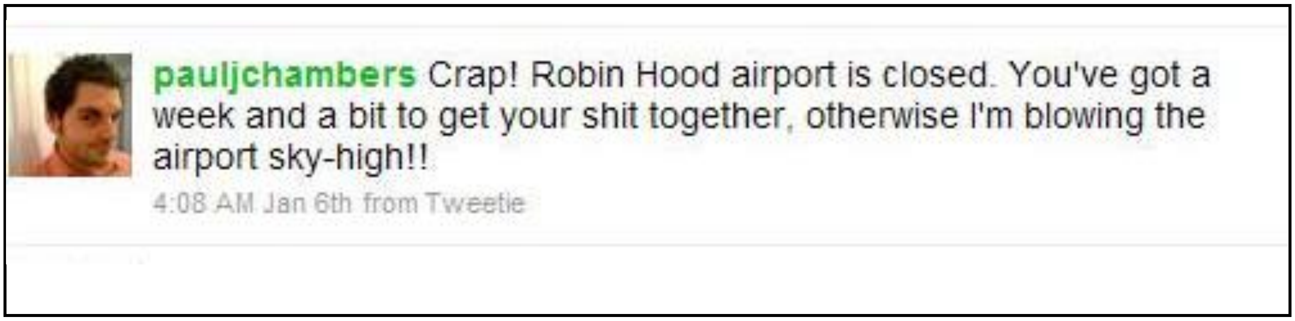
[14] Richard King's site describes him as follows: I am passionate about technology and its impact on society. In 2006 I became involved with the Open Rights Group (ORG) and I started this blog around the same time, partly as a learning exercise, partly to express my thinking on human rights issues as society went digital .I have volunteered variously as an evangelist, copy-writer, editor, newsblogger and wiki maintainer for ORG. I was appointed to their supporters council in 2012 and I started ORG Sheffield – a local chapter of the group that meets regularly to discuss digital-rights issues – in the same year. I'm an active member of the grass-roots technology community in my home town of Sheffield, UK, and part of the nascent makerspace-community in my adopted town of Tromsø, Norway. https://richardskingdom.net/about

Ashley Madison as a way to talk about metadata and data brokers interested in buying metadata, and how the main SM platforms share this type of risk.

> When you sign up for Facebook, Twitter or LinkedIn, you’re nagged to upload your entire contacts list, either by giving them the password to your email account(!), or by letting their app grep your phone’s address book. The benefit you’re offered is automation [...] Stop and think again, but this time instead of considering what these companies could do with the data you’re giving them, reflect on how you’re treating the subjects of that data (King, 2015, Document id 31).

4_Paul Chambers – Censorship

Another exemplary case that emerges from the statement concerns the trial of Paul Chambers as a result of a tweet. Paul Chambers in 2010 tweeted “crap Robin Hood Airport is closed you've got a week and a bit to get your shit together otherwise I'm blowing the airport sky high” when delayed at Robin Hood Airport in South Yorkshire. The tweet led to prolonged legal proceedings, since the anti-terror police charged Chambers for sending a public electronic message that was contrary to the Communications Act 2003 (e.g. either grossly offensive or of an indecent, obscene or menacing character). The development of the legal proceedings became exemplary, involving celebrities and passing into the public discourse. In the end Chambers’ conviction was quashed, but the case has become an example of how the legal identification of internet trolls or offensive material is flawed and how communication on SM can lead to arrests and prosecution. As stated by solicitors Woodfines (who are quoted on the website collected) this type of incident can cause real-life problems for the individuals who go are caught up in this type of experience; on top of having their freedom of expression limited they can lose their job even if cleared of any criminal wrongdoing (Woodfines 2015, Document id 74). Paul Chambers’ case is also used as an exemplary case in academia, in the study of computer-mediated communication, and in terms of the presence of interpretative ambiguities because of the lack of expressional nuances.



pauljchambers Crap! Robin Hood airport is closed. You've got a week and a bit to get your shit together, otherwise I'm blowing the airport sky-high!!

4:08 AM Jan 6th from Tweetie

Figure 5. 16- Paul Chambers' tweet. Source: Daily Mail (2012)

5_ Caroline Criado-Perez

Probably the most exemplary case, which is used the most both in academia and by public bodies is represented by the storyline connected to Caroline Criado-Perez. In 2014, the feminist activist Caroline Criado-Perez and Labour MP Stella Creasy, along with other supporters, were deluged by abuse from Twitter trolls after they successfully campaigned using social media for a female figure to appear on a Bank of England note (i.e. the writer Jane Austen on the £10 note). As a result, the 'trolls' Isabella Sorley, John Nimmo and Peter Nunn were sentenced to jail, according to Section 127 of the Communications Act 2003. Criado-Perez is now a storyline for abuse of female 'public' figures. The case was mentioned again in 2016 and 2017, when other MPs received threats and abuse. Authors in academia and in public bodies mention the Criado-Perez case as the main storyline in a series involving several other episodes of harassment at the expense of female and ethnic minority MPs that took place at the time of the UK general election in 2017.

> There was an English Defence League-affiliated Twitter account – #burnDianeAbbot. I have had rape threats, and been described as a 'pathetic useless fat black piece of sh*t', an 'ugly, fat black b*tch', and a 'n*gger' – over and over again'. Similarly, in 2016 Jess Phillips, a Merseyside MP, spoken of receiving more than 600 threats of rape in one night alone on Twitter and had to have extra security installed in her home, following the abuse she suffered online (Bliss, 2017, Document id 485).

Statements from academia stress how long the problem has been going on for, and connect it with public initiatives, such as the cross-party campaign – 'Reclaim the Internet' (in 2016).

6_US elections

From the documents it is evident that several statements from the actors were created in reaction to the US election in 2016. In particular, the role of SM in promoting fake news, and the impact of fake news on the final result of the election, started a debate involving academia and other actors:

> Fake news has become an important focus for news foundations, democratic interest groups and various journalism academics and researchers, following claims that the US presidential elections may have been influenced by anti-Clinton propaganda created by Russia and shared on social networks (Felle, 2017, Document id 248).

7_Charlottesville

In 2017 another public shock that stimulated a reaction to Big Tech and changes to their policies was the murder of Heather Heyer, a protester at an anti-fascist rally in Charlottesville. She was killed in an attack by James Alex Fields Jr., who deliberately drove his car into the crowd. The terror attack focused attention on neo-Nazi and white supremacist groups active on SM and their use of the platforms to spread propaganda. As a reaction to the murder the big names in social media and digital companies started to withdraw their space for white supremacy supporters:

> The Internet governance implications of Charlottesville are becoming clearer. When a white supremacist protest resulted in the murder of Helen Heyer, the Daily Stormer published repugnant, hate-filled content about her on its website. This provoked numerous Internet service providers (domain name registrars, DNS proxy services, a DDoS mitigation service and a hosting provider) to terminate Daily Stormer's services for a variety of alleged Terms of Service (ToS) violation(s) (Kuerbis, 2017, Internet Governance Project Document id 224).

The expulsion or refusal by large companies to host content from white supremacists opened opportunities for smaller SM platforms, for instance Gab.ai. Gab is a 'smaller' social media company, based in the US, that became famous because a number of representatives of extreme right-wing movements decided to use it after having been expelled from the major platforms. After Charlottesville, Gab gathered the voices of the white supremacists expelled from the major companies, declaring that it had a more absolute interpretation of free speech in its content regulation policies.

> Founded in 2016, Gab.ai aims to be a free speech alternative to the major social networks such as Twitter, Facebook and YouTube. It is a libertarian social network founded in the classical liberal tradition of John Stuart Mill and John Milton, as well as the US First Amendment. Gab supports artistic expression and actively challenges the censorship that takes place on the major social media sites (People's Charter Foundation, 2017, Document id 261).

> When Twitter started banning people for having unwelcome opinions, the founders of Gab saw a gap in the market and started their own version. Both Apple and Google have refused to approve a Gab app until they can ensure nothing which constitutes discriminatory language will be posted, which defeats the whole purpose. Now it appears someone has gone after Gab's domain registration, probably having seen other right-wing sites get their registrations pulled in the aftermath of Charlottesville. So far it's an effective tactic. If the tech giants and domain registrars are the gateway to 99% of communication, denying somebody access is the equivalent of banning them from speaking (White Sun of the Desert – Newman, 2017, Document id 254).

8_NetzDG

The statements from web pages in 2017 highlight another exemplary case which involved the German government approval of the 'Network Enforcement Law', also called NetzDG. This is a law that was introduced in 2017 and came into full effect on 1 January 2018. It introduced fines up to £44m for failure to remove hate speech from SM platforms within 24 hours. This required adaptation from the SM platforms (which included additional features for flagging up controversial content, and hiring more human moderators). In particular, a number of controversial deletions and suspensions bolstered critics from civil society and advocates of free speech:

> Criticism of the new law has intensified over the past six weeks after content from some high-profile users was blocked or their accounts were temporarily suspended, even though some of those actions were submitted due to violations of the company's user rules rather than NetzDG. Users whose speech was censored either by NetzDG or a violation of a

company's user agreement include a leader of the far-right Alternative for Germany party, a satire magazine, and a political street artist (Human Rights Watch, 2018, Document id 387).

Related regulation initiatives

The analysis of public shocks and relevant keywords also highlighted regulation initiatives that are frequently mentioned. One of the most significant for the UK environment is the Crown Prosecution Service Guidelines for speech on social media. The guidelines were first issued in 2013 but were reviewed and updated in 2016. In particular, the revised version introduced a specific section on Violence Against Women and Girls (VAWG) (Crown Prosecution Service, 2016, Document id 154).

Intimidation in public life 2017: another contribution to regulation that emerges from the texts is the Committee on Standards in Public Life's report on 'Intimidation in public life' in 2017. The report was issued as a reaction to the abuse received by MPs during the general elections in 2017 (mentioned above). The report focuses on harassment on social media, and includes a full section on SM responsibilities.

'Future of the Internet' speech 2017: in 2017, at the Internet Governance Forum, the Ministry for Digital Media and Sport gave a speech on the 'Future of the Internet', in which the UK government presented the positions regarding freedom of expression online (Hancock, 2017, Document id 209).

Online harassment and cyber bullying – Parliamentary report: in 2017 the UK Parliament also issued a report on online harassment and cyberbullying. It highlighted how (at the time) Section 103 of the Digital Economy Act 2017 required the Secretary of State to issue guidance to social media providers about action against abuse online (Dent and Strickland, 2017, Document id 293). Even if not strictly regulation initiatives, the parliamentary report and the 'Future of the Internet' speech clarified the Conservative government position on SM and content regulation.

EU Code of Conduct on Illegal Speech: other related regulation initiatives that stand out from the relevant words are the Digital Single Market initiative and the Code of Conduct on terrorism and hate speech, which establish a set of norms for SM platforms in the EU Internet Forum. The code sets out a number of ‘public commitments’, including (among others):

- clear and effective processes to review notifications regarding illegal hate speech on their services, and adaptation of Rules or Community Guidelines clarifying that they prohibit the promotion of incitement to violence and hateful conduct
- review notifications against their rules and community guidelines and where necessary national laws, with dedicated teams reviewing requests
- review the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary
- creation of national contact points designated by the IT companies and the Member States respectively for the notification
  (European Commission, 2016, Document id 486).

UN call for proposal: the study of shocks and relevant words also shows the role of international organisations, in particular the UN and the Special Rapporteur for Freedom of Expression, David Kaye. Their presence is highlighted with the call for submissions issued by the Special Rapporteur, on Content Regulation in the Digital Age (United Nations and Kaye, 2017, Document id 231). The call was opened to states, civil society organisations, private companies and other stakeholders. Several advocacy groups shared positions, since many of the more prominent groups responded: Article 19, Index of Censorship, Access Now, Ranking Digital Rights. In the data they appear as reports and official documents from organised civil society including recommendations for the UN and nation states on what to consider when developing standards for regulation of speech.

5.9 Key issues emerging from qualitative analysis

In order to identify which particular issues involving freedom of expression and SM platforms are summarised in the main topics, I analysed the texts and coded different aspects of the issues.[15] The main codes that emerged from the qualitative analysis in Nvivo are:

[15] I merged 2 lists: Algorithms, B: Fake news, C: Extremism, D: Terrorism/ Radicalisation, E: Cyberbullying F: Harassment, G: Hate speech, H: Children, I: Sexual crimes
In the second round I coded statements concerning:
A: Anonymity, B: Islamophobia, C: Blocking filtering, D: Censorship, E: Far right, F: Net neutrality, G: Remedies (namely, what can be done to obtain remedy in casse of abuse or censorship) H: Responsibility for content (between private companies and states) I: Surveillance, J: Transparency

Quality of content (algorithms and fake news)

Extremism (terrorism, radicalisation)

Harassment and (cyber)bullying

Hate speech

Child pornography and sexual crimes

Censorship (including blocking, filtering)

Islamophobia, far-right ideology

Data protection: anonymity, surveillance

I compared the list with the issues that I coded from relevant words in Cortext ordered on the basis of highest TF-IDF *(*data available in the Appendix, p.31-A).

Censorship

Children protection

Cybersecurity

Data protection and privacy

Extremism

Harassment

Crime

Hate speech

Illegal speech

Accountability

Human Rights

Usage

Net neutrality

Quality of info and fake news

Terrorism

Cyberbullying

I then combined the results of quantitative and qualitative analysis of the texts, and the common themes can be summarised as follows:

- Social media and extremism. This theme includes statements (exemplary cases, storylines) which summarise the main arguments that social media are used by extremist groups or terrorist groups and provide a place for recruitment and radicalisation of people. In this group it is possible to find statements related to the cases of murders during the Charlie Hebdo Islamist terrorist attack and the fascist terrorist attack in Charlottesville described above. An interesting finding, emerging from the statements related to this issue, is the role of schools in enforcing anti-terrorist policies in the UK as assigned by the Counter-Terrorism and Security Act (2015) (Dartington Primary and Nursery School, 2015, Document id 56).

- Social media hate speech and harassment. This theme includes statements (exemplary cases, storylines) which summarise the main arguments that social media are a place for abusive speech addressing groups (hate speech (e.g. racist, religious anti-Muslim speech) or individuals (e.g. misogynist attacks) in a public (cyberbullying) or anonymous way (trolling)). The public shocks most exemplary for these issues are the attacks on Carolina Criado-Perez, Jodie Whittaker, and the female MPs during the 2017 general elections. It also includes the issue of hate speech, and how it has been used to increase regulation, as in the case of the adoption of the NetzDG in Germany.

- Social media, algorithms and fake news (e.g. quality of content). This theme includes statements (exemplary cases, storylines) which summarise the main arguments about the role of technology (i.e. algorithms) in shaping the quality of information available online. The issue has a clear connection with the case of US election results (see above). However, some NGOs did raise the issue before that time (EDRi, 2016b, Document id 110).

- Social media, privacy and protection of personal data. This theme includes statements (exemplary cases, storylines) which summarise the main arguments that states and SM are surveilling, invading privacy and exploiting personal data. The exemplary cases mentioned in the statements of actors are Snowden leaks, Ashley Madison hackers attack and the US border agents investigating social media accounts. This issue involves both states and SM. Considering the debated aspect of anonymity in online communication in 2015 the UN Special Rapporteur on Freedom of Expression David Kaye wrote:

> Journalists and civil society rely on encryption and anonymity to shield themselves (and their sources) from surveillance while artists rely on encryption to safeguard their right to free expression, especially in situations where it is not only the State creating limitations but also society that does not tolerate unconventional opinions (United Nations and Kaye, 2015: 5-6, Document id 487).

Considering SM companies, the theme includes statements stressing the choices taken by companies about the data of users. Above I describe how the theme connects to the Ashley Madison episode, but other statements clearly related this issue to SM platforms.

> When we use social networks like Facebook or video sharing platforms like YouTube, a lot of sensitive data about us is generated and stored. It can be used for different purposes by those companies and by other companies with which the data might be shared. For example, they can decide that, since you accepted their terms of service, they can do ‘research’ based on the information you posted (EDRi, 2016b, Document id 110).

Public bodies however present data retention as a tool to protect from terrorist attacks, or in child protection. The UK government has issued recommendations based on the ICT Coalition for Children Online, a European industry initiative to make its platforms safer for users. In the guidelines for providers of online or mobile social media or interactive services that might attract users under-18 years old, they stress that the government can request data retention through the Regulation of Investigatory Powers Act 2000 (RIPA), or the Data Retention and Investigatory Powers Act 2014/ Investigatory Powers Bill (2016) in order to prevent or detect crime or prevent disorder.

- Social media and censorship, which summarise the main arguments that SM and states are censoring content that should be free. These are issues that emerge especially from the

> interpretation of facts from bloggers. The exemplary cases that emerge are related to censorship in the case of Raif Badawi, or the trial undertaken by Paul Chambers.

The issues that emerge from the public shocks and regulations described above coincide with the ones that emerge from the report of the UN Special Rapporteur for Freedom of expression, in September 2017.

> The spread of 'extremist' content online has triggered legislative and corporate responses that may address serious national security and public order threats but may also limit political discourse and activism. The scourge of online gender-based violence has prompted uneven and excessive regulation that not only fails to address its root causes, but also threatens legitimate content. The perceived urgency to address misinformation through 'fake news' and online propaganda has generated global confusion about what counts as false or misleading – and who decides (United Nations and Kaye, 2017, Document id 231).

5.10 Groups of actors' positions on issues

The qualitative analysis of statements shows that certain issues are particularly dear to specific groups of actors. Figure 5.17 below gives an overview of the distribution of the statements made by each group of actors on each of the issues that emerged from the qualitative analysis. The dimension of the bars represents the number of statements coded under the specific issue in the qualitative analysis. As described in the overview of public shocks and issues, the UK government and public bodies (in yellow in the graph) have been particularly active on the issue of terrorism/extremism and harassment.

The analysis shows that academia (in blue) is among the group of actors more concerned by the issue of surveillance and of quality of content, stressing in particular the role of algorithms and AI, machine learning etc., as well as fake news (see Figure 5.17).

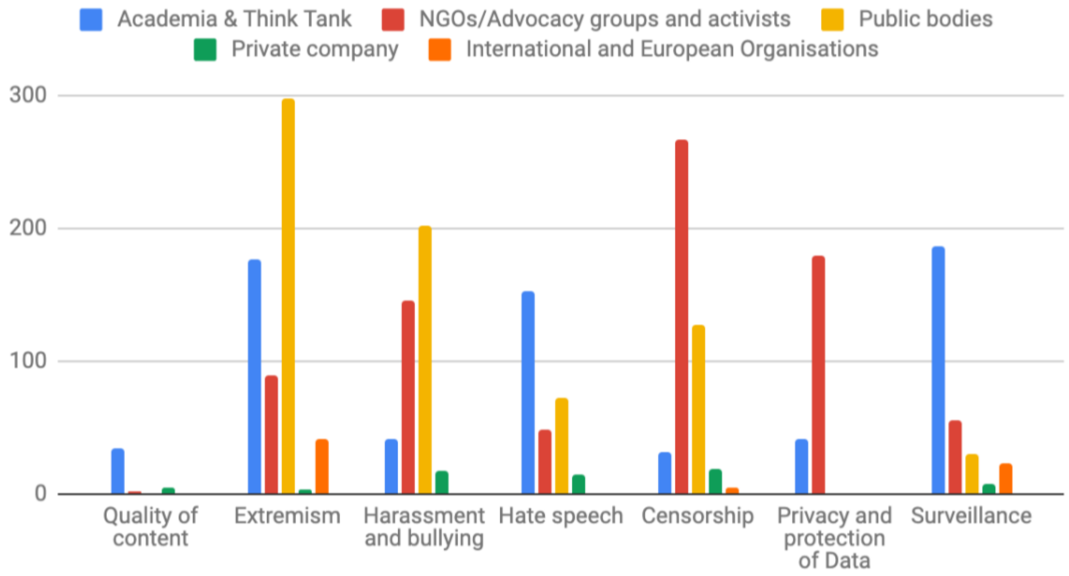


Figure 5. 17 - Actors' statements on issues

NGOs and advocacy groups (in red) are particularly active on the issues of privacy and data protection, as well as censorship. They are the group that have more positions on specific issues, such as anonymity, blocking filtering, censorship, net neutrality, and transparency. In particular, censorship is an issue very much discussed by bloggers, who often take very radical positions in terms of freedom of expression, as it is a topic of interest to far-right/extreme positions. Academia has produced relatively more statements on surveillance and extremism and hate speech, which has seen above are issues in some way related.

Extremism is the most recurrent issue in statements from public bodies, either at the level of national policies (as in the case of the anti-terrorism initiatives) but also at the level of schools and anti-radicalisation programmes (e.g. the application of the Prevent programme in schools). Public bodies and NGOs are also the main groups mobilised by the issue of harassment.

International and European organisations (orange) are active on surveillance, censorship and extremism. The UN with the Special Rapporteur on Freedom of Expression provided statements against state surveillance (in protection of encryption) and censorship. Private companies (in green), especially legal experts, present statements on the issues of censorship, hate speech, harassment and quality of content. They tend to describe the legal implications rather than present a specific interpretation.

5.11 Narratives about free speech

In the analysis of recurring emblematic issues and storylines, I focused on statements that concern the ‘ontology’ of freedom of speech. From the texts I was able to identify four recurring interpretations: 1) Free speech has limits 2) Free speech is absolute, 3) Privacy as a limitation to free speech 4) Privacy as necessary for free speech. This division reflects that some actors tend to see freedom of speech more as an absolute freedom, and any limitations to it as a trade-off, while others do not see a contradiction between freedom of expression and some sort of necessary regulation. The analysis also highlighted an extra layer of complexity, introduced by arguments that see the right to privacy as complementary or against freedom of expression (as in the case of the right to be forgotten).

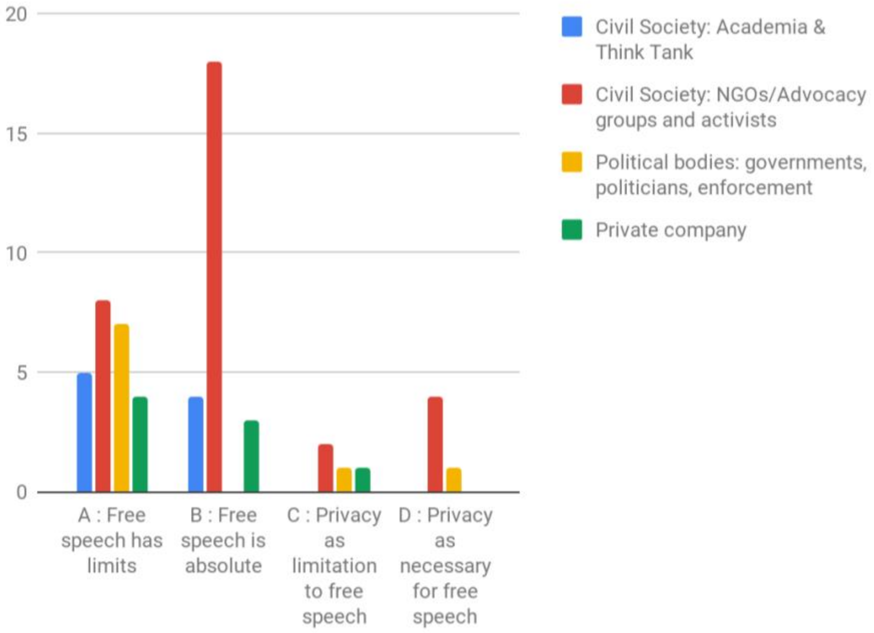


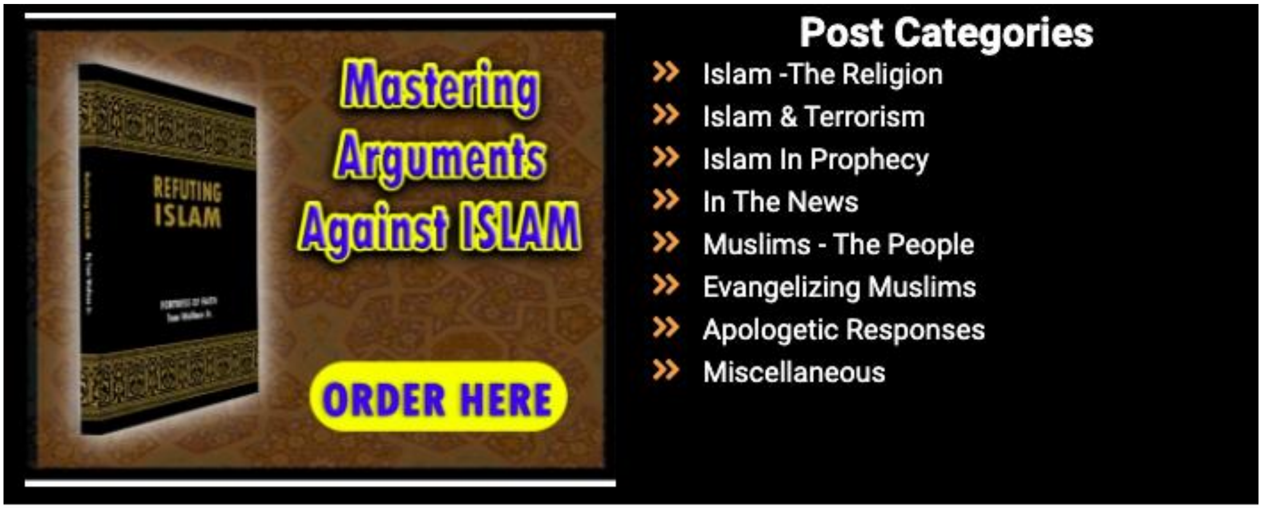
Figure 5. 18 – Narrative about freedom of expression divided by group of actors

Figure 5.18 displays the different interpretation of freedom of expression according to the different groups of actors. The most evident feature in the graph is that the group NGOs/advocacy groups and activists (red) has the most radical stance in terms of absolute interpretation of freedom of expression and of privacy protection. In the previous part of the analysis, the description of actors, I stressed that several web pages in the NGOs/advocacy group belong to bloggers. From the analysis of the text, it emerged that bloggers tend to have a ‘libertarian’ idea of freedom of expression, very much in line with the interpretation of Barlow’s manifesto discussed in the literature review. Bloggers tend to share this position. Even if they are from completely different political backgrounds, most of all they tend to share scepticism towards any form of regulation of speech, including the bloggers addressing hate speech. Christopher England, in his blog “England’s England” defines himself as follows:

> If we have to wear labels, then politically, I am probably right of centre but mixed with this is a (real, not trendy lefty pretend) libertarian or even a touch of anarchist or rebel. I do believe in structure and control of populations (especially the feckin’ stupid ones), but, I like the idea of consensus thinking rather than the historical confrontational divisive rule by fear we have endured and have come to expect. I find the ‘liberal left’ with all its censorship and rules against free speech extremely worrying. I am an atheist. Not so much an evangelising atheist, but I do kick back at those who try to control others via religion, rather than people who have a ‘personal’ religion. I detest the abuse of children by forcing religion onto them (England, 2017).

In another blog: Fortress of Faith, Tom Wallas Jr. comes from a completely different perspective. Extremely religious, the blog presents an extremist Christian and anti-Islam position. The author is concerned with freedom of speech and, like Christopher England, dislikes forms of regulation of hate speech:

> The minute anyone tries to pass a hate speech law your ears should perk up. We cannot allow these kinds of laws to be passed. The text of the law will not say that you can’t say anything specifically against Islam. It will be worded in such a way as to hide what they are really doing (Wallace, 2017, Document id 455).



Post Categories

- Islam -The Religion
- Islam & Terrorism
- Islam In Prophecy
- In The News
- Muslims - The People
- Evangelizing Muslims
- Apologetic Responses
- Miscellaneous

Figure 5. 19 – Example of contents in the blog Fortress of Faith

Another example comes from the opposite end of the political spectrum. Jerry Barnett in the blog MoronWatch consider to be an extreme 'Leftist'. He shares a similar position to Christopher England and Fortress of Faith concerning regulations against hate speech and censorship:

> One problem with censorship is that it is necessarily dumb. Once the ludicrous concept of 'hate speech' had been ruled unacceptable, censors can't tell the difference between genuinely hateful speech, parody, and discussion of hateful speech. Another problem with censorship is that it simply doesn't work. Silencing discussion of a problem doesn't end that problem, it just pushes it into corners where nice, middle-class people can ignore it (or at least, ignore it until it's too late to do anything about it).
>
> [...] Facebook is just one platform, but it is a huge and powerful platform. Increasingly, its methods are leaking into public discourse. Last year, MPs recommended that 'trolls' should be banned from using the Internet. Presumably, this would include people like me, who try to counter far-right extremism online(Barnett, 2016, Document id 190).

Hintz and Milan (2009) did highlight that grassroot movements emphasis on user and technical expert self-regulation has parallels with cyberlibertarian beliefs and private-sector policy preferences. This emphasis tend to show little concern for structural problems such as inequalities and uneven distribution of technical knowledge and concentration of power.

An extreme or 'libertarian' definition of freedom of expression has also been the brand for 'smaller' SM platforms, who have started to attract voices 'expelled' or not tolerated on larger platforms. As noted above, Gab.ai offered alt-right white supremacists a place to gather after they were expelled by the big companies in reaction to the Charlottesville events. Looking at the rhetoric of Gab's spokesperson, it is possible to notice a great reliance on liberal thinkers, such as John Stuart Mill and John Milton, as well as a heavy emphasis on the US First Amendment.

> People of the United Kingdom, this is the state of your country. Listen to your taxpayer subsidised broadcasting agency asking us to censor, vet, control and limit speech on Gab. For shame! This is the great country that produced John Stuart Mill; this is the country that produced John Milton's Areopagitica, this is the country that fought for the Natural Rights of Englishmen in the 1689 Glorious Revolution, this is the country that produced us the best warning manual against dictatorship. (Utsav Sanduja, Chief Communications Officer for Gab.ai, Interviewed by People's Charter in 2017).

The official justification for allowing extreme forms of speech on their platforms is based on the idea that:

> Dangerous extremists are more likely to hang out on the dark web than on social media sites, so such a proposal would inevitably end up penalising innocent people. Furthermore, it is better that dangerous views are out in the open than forced underground where it is much harder to challenge them. (Utsav Sanduja, Chief Communications Officer for Gab.ai, Interviewed in People's Charter Foundation, 2017, Document id 261).

However, as suggested above, the reactions to public shocks and events have been pushing towards a relative interpretation of freedom of expression, and in particular towards the no-platforming of extremist speech by violent groups. The majority of actors in the other groups present an idea of freedom of expression that can be limited.

Less radical views on 'freedom of expression'

The idea that freedom of expression comes with limits is enshrined in most of the legislation and human rights articles. Art. 10 of European Declaration of Human Rights and art. 19 Universal

Declaration of Human rights both include limitations for specific cases. The limitations, however, have to be both necessary and proportionate.

> Freedom of expression and the right to receive and impart information are not absolute rights. They may be restricted but only where a restriction can be shown to be both: a) Necessary; and b) Proportionate. These exceptions, however, must be narrowly interpreted and the necessity for any restrictions convincingly established: Sunday Times v UK (No 2); Goodwin v UK [1996] 22 EHRR 123. See the section below on 'Hate crime' for some examples of exceptions. Accordingly, no prosecution should be brought under section 1 of the Malicious Communications Act 1988 or Section 127 of the Communications Act 2003 unless it can be shown on its own facts and merits to be both necessary and proportionate. (Crown Prosecution Service, 2016, Document id 154).

NGOs active in the protection of human rights and free speech stress that the right of free expression is essential for democracy and applies also to divisive issues/ideas. The analysis of the interpretation of freedom of expression also highlighted how some NGOs such as Privacy International or Index of Censorship (but also ACLU, EFF, Article 19), by 'vocation' more liberal in the understanding of free speech, are extremely wary of limitations to freedom of expression:

> Locke, Milton, Voltaire have all written eloquently on the benefits of free expression, but I think Mill expresses it best when he talks of free expression being fundamental to the 'permanent interests of man as a progressive being.' 'The particular evil of silencing the expression of an opinion,' he argues in On Liberty, 'is that it is robbing the human race… If the opinion is right, they are deprived of the opportunity of exchanging error for truth; if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth produced by its collision with error (Index of Censorship and Ginsberg, 2015, Document id 26).

However, association and NGOs groups active against hate speech (Resisting Hate, for instance) tend to have a clearer stance in favour of regulations:

> The counter argument runs on the lines that there needs to be exceptions to free speech for the safety and good of society. Both myself and Resisting Hate strongly believe that hate

> speech is not free speech. Free speech is not the holy grail of civil liberty. (Carleton-Taylor 2018, for Resistinghate.org, Document id 395).

Of all the actors, political bodies (in yellow in Figure 5.18) present statements more often in line with the idea of free speech as in need of regulation. The majority of public bodies, and the documents produced in recent years, point towards the necessity to impose more limitations. In 2017, the then Minister for the Media and Sport, Matt Hancock, provided a definition of the freedom of expression online according to the Conservative government:

> The Internet is open, not laissez-faire. Liberal, not libertarian. Freedom is a framework. Burke said that liberty 'is not solitary, unconnected, individual, selfish liberty, as if every man was to regulate the whole of his conduct by his own will'. Instead he said liberty is 'social freedom'. 'Secured by the equality of restraint.' In which 'no one man, and no body of men, and no number of men, can find means to trespass on the liberty of any person.'
>
> [...] The fact that we as a society have put these boundaries on acceptable free speech has not undermined our status or credibility as a society that values free speech. No-one can credibly say that because we stop people standing up and spreading racial hatred means that we are on the side of repressive regimes and not free speech. [...]A free and open Internet does not mean an Internet without boundaries or rules. And agreeing as society what those rules should be does not weaken our commitment to freedom (Hancock, 2017, Document id 209).

In the group NGOs and advocacy it is possible to find more statements pointing at the connection between freedom of speech and privacy. In particular, the NGO Index of Censorship stresses the risk of treating freedom of speech and privacy as a trade-off. The stress on protection of privacy in a time of surveillance is legitimate, but in their view it should not come at the price of the right of free expression:

> privacy and free expression are both necessary so that the other can flourish, it would be remiss of me not to caution against any temptation to let privacy rights – which often appear all the more important in both an age of mass surveillance and a bare-all social media culture – trump freedom of expression in such a way that they prevent us, as per the Mill's doctrine, coming closer to the truth. It is for this reason that Index on Censorship opposed the so-

> called ‘Right to be Forgotten’ ruling made in Europe last year. Europe’s highest court ruled in May 2014 that ‘private’ individuals would now be able to ask search engines to remove links to information they considered irrelevant or outmoded. In theory, this sounds appealing (Index of Censorship and Ginsberg, 2015, Document id 26).

Similarly, Harlem Désir, spokesperson for OSCE interviewed by Internet Society said “First, we need to reject the notion that freedom of expression and human rights are detrimental to the security of our societies. I believe the opposite: freedom of expression and human rights positively contribute to security and other interests in our societies” (Internet Society, 2017, Document id 385).

The think tank Demos also expressed some scepticism at the idea of restricting speech as a solution to hate speech. Answering a call for studies published from Facebook, think tank advocates for the use of alternative solutions, such as counter-speech, and self-organised responses from the online communities:

> there is little evidence that censoring or removing content has an effect (or indeed, on what that effect might be). […] A preferable response is a small number of strategic mass take-down efforts, which would make the network harder to reconstruct and allow analysts to study the effect it has on the network. […] There has been a slowly emerging consensus that confronting hate speech with ‘counter-speech’ is a potentially more fruitful approach (Bartlett and Reinolds, 2015, Document id 86).

5.12 Narratives about governance

Analysing the statements and regulations mentioned, it was possible to identify a number of macro positions of actors concerning the ‘ideal’ distribution of roles and responsibilities across actors. I coded the statements according to whether they were leaning more towards the idea of a strong level of responsibility for SM (e.g. editorial responsibility), or the opposite – not wishing to see any specific responsibility of regulation on SM – as well as the middle-ground, of those actors who see SM as having responsibility for regulation, but do not go as far as recognising them as editors. Other actors on the other hand lean towards a model where the state has full responsibility for the regulation of content. These include a specific proportion of statements pointing towards the idea that states should legislate more. On the other hand, there are also those that say that states should have less power (as for instance voices critical of surveillance).

The analysis confirms that this is a very controversial issue, where less agreement than in the other issues can be found within the different groups. Figure 5.19 gives an overview of the distribution of statements (as they appeared in the qualitative analysis of texts).

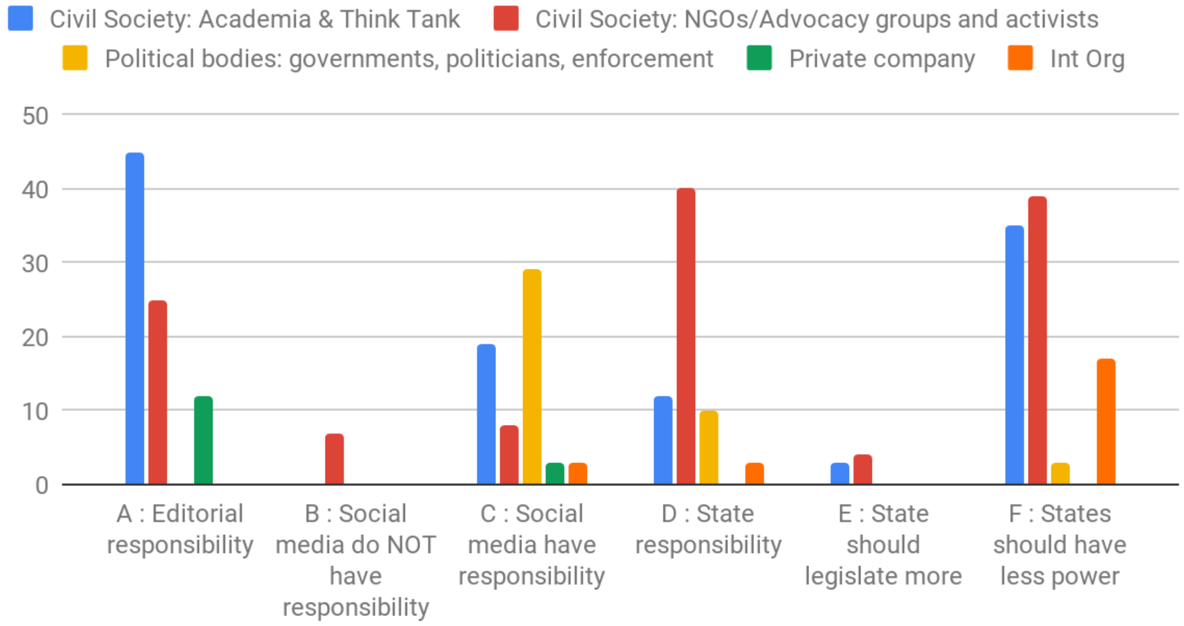


Figure 5. 20 – Narratives about Governance models by groups of actors

From the statements it appears that national governments in Europe have been putting pressure on SM platforms to take regulatory initiatives on policing problematic content (terrorists, extremist, harassment, etc.). Figure 5.19 shows that the majority of statements coded from public bodies (in yellow) refer to an ideal form of governance where SM platforms would have greater responsibility in the regulation of content. At the same time, public bodies do not define the terms of this responsibility, for instance editorial or social responsibility (which, however, academia does, in blue). Civil society and NGOs (red) have statements codified relative to the responsibility of both SM platforms and states.

As far as the position of public actors is concerned, in the UK it is clear that a trend has been developing in recent years to place more pressure on SM platforms. As reported in the 2017 Parliamentary report on online harassment and cyberbullying:

> The Home Affairs Committee published a report on Hate Crime in May 2017 which criticised social media and technology companies for not doing enough. [...] Generally recent governments have tended to favour self-regulation wherever possible, working with the industry to deal with problems that arise. There has been resistance to introducing specific legislation to deal with online harassment and trolling. However, over the past year, arguments for a change in the law seem to have been gaining ground (Dent and Strickland, 2017, Document id 293).

As stated in Joanna Shield's (2017) speech on counter-terrorist regulation, 'united action to tackle this threat is the only way forward. Governments and experts can provide extensive knowledge and a rigorous understanding of the threat but industry is best placed to innovate on technical solutions that address this threat specifically for their own commercial platforms' Shields added: 'It is incumbent upon industry to drive this change' (Shields, 2017, Document id 235). States and companies have been cooperating on the issue of terrorism, at the UK level, with the Counter-Terrorism Internet Referral Unit working with industry and civil society (Shields, 2017, Document id 235).

Other forms of cooperation have also been developed in the field of hate and illegal speech. The EU Code of Conduct on Illegal Speech is the most exemplary initiative of this type which emerges from the data. The initiative is considered a success by members of the EU Commission (EU Commission, 2018, Document id 356).

States have increasingly put pressure on SM, as in the case of the NetzDG described above. The ideal of governance described in the German law sees the state imposing legislation on the private company, and imposing deadlines and fines for failure to comply.

However, this model is not appreciated by representatives of civil society organisations (and bloggers). In particular, EDRi accuses the NetzDG of being in breach of

> Article 14 of the E-Commerce Directive (2000/31/EC) which provides a liability exception for online intermediaries, when they act expeditiously to remove illegal content, according to a notice-take-down procedure (EDRi 2017, Document id 289).

According to EDRi, the German law pushes platforms to over-censor content in order to avoid fines and to regulate using only the big corporations as standard, creating an unreachable standard for smaller platforms (EDRi 2017, Document id 289).

A similar concern is shared, but on a different basis, by bloggers with extreme libertarian views, where the idea of states of pushing responsibility on SM platforms is seen as a risk for extreme speech. For example, in ‘White Sun of the Desert’:

> [T]ech companies will double-up what they’re already doing: pulling down posts and articles willy-nilly if they contain a single word which might upset this year’s designated victim class, yet the stuff calling for shooting cops, punching Nazis, and the destruction of Israel and the west stays up. And if a load of right-wing writers, bloggers, and commentators get caught up in the sweep? Well, that’s a feature, not a bug (Newman, 2017, Document id 254).

As in the case of freedom of expression, NGOs appear divided in their positions on the role and responsibilities of regulation. Regulatory initiatives and cooperation between states and private companies have been criticised by NGOs active in the protection of free speech and digital rights, as they push too much power onto SM. On the other hand, civil society (either NGOs or academia) involved in the protection of victims, as in the case of harassment based on gender and ethnicity, suggest that SM platforms should be asked to do more:

> Social networking companies have been slow in their response to protecting individuals from online abuse. [...] Social networking companies need to be taking more responsibility for what is being posted on their sites, whilst also being more transparent in how they are tackling internet trolls [...] Specific laws, especially aimed at the protection of those being subjected to abuse online, could help better protect individuals (Bliss 2017, Document id 485).

On the side of pushing responsibility towards SM, academia and think tanks are the group with a clear stance for a definition of responsibility for platforms. In particular, some actors from the group of academia see the necessity to treat SM as editors, since they are making choices about what is visible and accessible to users.

> […] we argue that social media companies have a corporate social responsibility to promote a healthy democratic discourse by adopting a code of editorial-like responsibility, including concepts such as the public interest in their content optimization algorithms. Fundamentally this involves applying principles of Responsible Research and Innovation to the design, development and appropriation of technologies (Koene et al. 2017, Document id 243).

However, others prefer to use another definition for the role of SM – that of social editor, since:

> […] social networks like Facebook organise the way in which the public debate around content takes place. It does so by collecting and integrating data from Facebook users into the recommendation process, by calculating popularity and shareability and by offering an entire architecture of tools for users to engage and share (Helberg 2016, Document id 146).

These initiatives do not come without criticisms:

> The result of treating online firms as publishers would be to reduce competition, deter innovation, and threaten the free flow of ideas online. (Adam Smith Think Tank, 2017)

Indeed, some NGOs and activists are more in favour of self-regulation, avoiding state involvement. Article 19 in 2018 stated:

> A model of self-regulation has been the preferred approach to print media. It is considered the least restrictive means available, and the best system for promoting ethical standards in the media. An effective self-regulation mechanism can also reduce pressure on courts and the judiciary. Generally, when a problem is effectively managed through self-regulation, the need for state regulation is eliminated (Article 19, 2018, Document id 447).

And proposed the creation of a self-regulation model for social media, including a dedicated "social media council" – inspired by the effective self-regulation models created to promote journalistic ethics and high standards in print media. We believe that effective self-regulation could offer an appropriate framework to address current problems with content moderation by social media companies, including 'hate speech', providing it also meets certain conditions of independence, openness to civil society participation, accountability and effectiveness. Such a model could also allow for the adoption of tailored remedies, without the threat of heavy legal sanctions (Article 19, 2018, Document id 447).

5.12.1 Private companies

As far as private SM companies are concerned, according to the statements from different actors, they are reluctant to adopt any definition of responsibility. They often cite Mark Zuckerberg's words: 'We are a tech company, not a media company.' At the same time, statements report that SM companies have been adapting to the changes in legislation. In the UK, the Parliamentary report noted the increase in the number of human moderators and the willingness to cooperate with a third-party fact-checking organisation (Dent and Strickland, 2017, Document id 293).

Although SM platforms have been resisting any formal responsibility, or specific role, it is evident that actors consider SM officially involved in policing the content on their platforms.

> In the face of the increased use of social media by extremist and terrorist groups, social media companies themselves– sometimes under pressure from the governments – have made more proactive efforts to remove or reduce the impact of hate speech on their platforms, police content more actively, and remove offending accounts or material more effectively.[xiii] Increased vigilance in both policing and more active social media platform administration has led to higher rates of page, profile and account deletion, stimulating significant changes in the online habits of extremist and terrorist groups (Bartlett and Reinolds, 2016, Document id 86).

5.12.2 The position of other international bodies

International bodies such as the Council of Europe and OSCE, as well as the UN, can only issue recommendations. However, they can set the tone of political discourse. In 2018 the Council of Europe adopted policy guidelines on the roles and responsibilities of internet intermediaries such as search engines and social media. The recommendations include provisions addressing both states and service providers. States are required to respect and protect human rights, and ask for control of content only on the grounds of legislation. Moreover, "Legislation giving powers to public authorities to interfere with Internet content should clearly define the scope of those powers and available discretion, to protect against arbitrary application; When internet intermediaries restrict access to third-party content based on a state order, State authorities should ensure that effective redress mechanisms are made available and adhere to applicable procedural safeguards" (EPRA, 2018, Document id 245; Council of Europe, 2018, Document id 416).

On the other hand, the Council of Europe states that intermediaries should not be considered liable: "When intermediaries remove content based on their own terms and conditions of service, this should not be considered a form of control that makes them liable for the third-party content for which they provide access." (EPRA, 2018, Document id 245; Council of Europe, 2018, Document id 416). From them, the request is to provide:

> A 'plain language' and accessible formats requirement for their terms of service; A call to include outside stakeholders in the process of drafting terms of service; Transparency on how restrictions on content are applied and detailed information on how algorithmic and automated means are used; Any measures taken to removing or blocking content as a result of a state order should be implemented using the least restrictive means. (EPRA, 2018, Document id 245; Council of Europe, 2018, Document id 416).

Harlem Désir is the Operation for Security and Cooperation in Europe (OSCE) Representative on Freedom of the Media / Internet Society 2017 Global Internet Report: Paths to Our Digital Future. He states:

> We must advocate and raise public awareness of the importance of freedom of expression – for democracy, for finding the best answers to society's most pressing challenges, for individuals' and societies' self-realisation. Second, we must hold states to account for imposing illegitimate and unnecessary restrictions on freedom of expression online. Third, we must urge Internet intermediaries to be more transparent about their approaches to taking down content online (Internet Society, 2017, Document id 385).

Some concern has, however, emerged about the lack of plurality in the ecosystem: "The distribution of and access to information depend now for most citizens on very few actors like Facebook and Google [...]" (Internet Society, 2017, Document id 385).

5.13 Narratives about technology

The analysis of documents revealed a special role occupied by technological objects. The most recurring topics include the mentions of technology and technological artefacts, such as trolls (i.e. bullies specifically acting online), abusive tweets, memes, emojis, bots, fake news and algorithms. These are mentioned several times either as issue or solution to the problems concerning SM platforms. Figure 5.20 shows the main technological artefacts mentioned in the texts, and it shows that bots, fake news, abusive tweets against MPs, misinterpreted threats as in the case of Paul Chambers, data spillage, surveillance, and algorithms as a way to influence information, are some of the most relevant terms that appear in the statements.

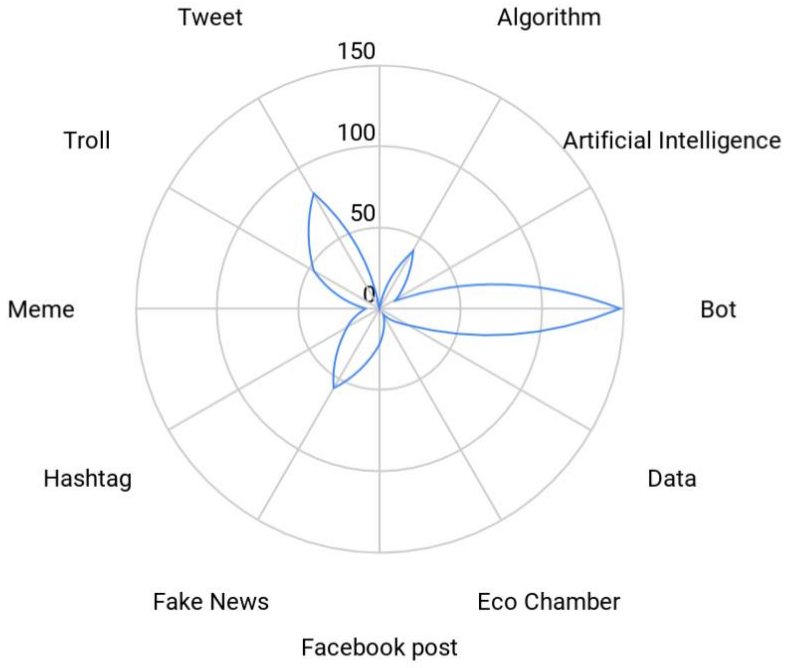


Figure 5. 21 - Most mentioned technological aspects in the texts from web pages

Technology is seen as both part of the issue and a solution. Bots and trolls are regularly mentioned as disruptive force in the public discourse (Murthy et al. 2016, Document id 123). However, the statements show an increased reliance and hope in the solutions offered by technology such as algorithms. For public bodies, automation is the answer in the form of filtering (as in the case of child protection). In particular, they are interested in search algorithms to look for specific words and filter content.

> [I]n terms of technology, we need to improve solutions that classify the language of extremism, automate the identification and removal of dangerous content at scale, and create tools that better tackle automated bots and other techniques that support these propaganda machines.
>
> They [SM] must innovate and automate their response to identifying and removing this vile, hateful material so that together we can ensure that everything possible is done to stop it infiltrating and poisoning a global audience (Shields 2017, Document id 235).

Academia sees algorithms as the fundamental ordering system in content regulation on platforms, and stresses the negative implications of this: “Political actors are using algorithms and automation in efforts to sway public opinion, notably through the use of ‘bot’ accounts on Twitter, Facebook, Reddit, and other social media platforms. Bots are understood to be ‘amalgamations of code that mimic users and produce content” (Woolley & Howard, 2016, para. 1), or ‘automated software agents’ (Geiger, 2016, p. 1)” (Maréchal, 2016, Document id 2015). On this basis, academia calls for public interest to be considered in the design of algorithms.

> [S]ome key concerns regarding forms of algorithmic decision-making and automated processes in the policing of domestic extremism and disorder in the UK, particularly around questions of privacy, freedom of expression and accountability. Moreover, it will question some of the promises of big data for governance that have been prevalent in much debate, particularly around notions of objectivity and efficiency (Koene et al., 2017, Document id 243).

Algorithms and artificial intelligence are seen as an issue principally concerning the lack of transparency behind their functioning.

> Artificial intelligence can pose formidable challenges to freedom of expression, including access to information. It poses questions pertaining to the issue of due process and transparency as we are already observing with reliance on algorithms on social media platforms (Internet Society, 2017, Document id 385).

> The fact that these algorithms are designed by (often private and commercial) actors that lack public accountability and are informed by a set of interests that do not necessarily align with the broader context of law enforcement (and with that, protecting freedom of expression and freedom of assembly) only further highlights this concern.
>
> If the assumed possibility of predictive policing to pre-empt and therefore eliminate an increasing range of criminality means that a risk becomes interpreted as a possible threat, monitoring of, and intervention into, activity based on social media data is likely to expand, with implications for freedom of expression and assembly (Dencik et al., 2016:53)

However, Twitter's spokesperson Nick Pickles stated in 2017 that 'pre-moderation was not possible: Let us be absolutely clear: we are never going to get to a point where internet companies pre-moderate content for the 400 hours of YouTube going up every day and for the 500 million tweets that go up every day. If you want pre-moderation of internet platforms, there may well be no internet platforms' (Dent and Strickland, 2017, Document id 293).

Technology emerges as an ambiguous element, and a possible reason for this concerns the unexpected usage that is made of it. One example was the use of 'political jamming' by pro- and counter-Isis propaganda. Laura Huey from the University of West Ontario (2015) uses Jihadi John in her study of political jamming, i.e. the counterculture practice where mainstream social media culture is disrupted/subverted through the use of satirical parody. Huey shows how SM have been used to rebrand jihadist forms of terrorism into an appealingly 'hip' subculture. Presenting pro-jihadist messages in rhetoric and imagery linked to memes from Western popular culture, Isis propaganda was able to create satirical results, able to make 'jihadi-cool' (2015:2). However, she also points out that, just as SM have facilitated pro-Isis political jams, at the same time they also facilitated counter-speech using the same strategies. She shows how users reacted against Jihadi John's videos photoshopping images of the video to mock Isis, creating a hashtag – #ISISCrappyCollageGrandPrix – and using humour to subvert Isis propaganda (Huey, 2015, Document id 99).

5.14 Findings and literature

The data show that actors that can be divided into six groups (news media, NGOs, academia, public bodies, international and European institutions and private companies). The study of URLs' frequency and the study of hyperlinks confirms previous studies, as it shows the position of power of traditional rules setters (i.e. states) and the increasingly fundamental role of private social media platforms. These confirm that the type of social groups involved in the main governance initiatives about governance of speech are similar to the ones highlighted in the literature (Gorwa 2019a, 2019b, 2020; Gillespie et al. 2020), and that even adopting an empirical approach, the larger players of traditional policy making still emerge either because of the large amount of web pages or because of their position in the network. In particular, the analysis of hyperlinks defines a sort of division of roles between political or policy-making actors (e.g. states and international organisations) and civil society. The UK government and international organisations emerge as key actors for the creation of content (i.e. high in-degree), while news media, NGOs and bloggers are key actors for the citation of content (i.e. high out-degree). The high presence of news media (which represent the majority of web pages collected in the first search) confirm that news media are a fundamental actor for the production and reproduction of the controversy, in line with the definitions from the theoretical framework of media as 'tool of measurement', giving the opportunity to non-experts to engage in the controversy.

The flexibility in the creation of the groups was, however, useful to identify a specific actor typical of the controversy on the web pages, i.e. bloggers. Blogs occupy a small part of the overall ensemble of URLs, but they emerge as some of the most active in terms of citations. As previously mentioned, bloggers often represent narratives of freedom of speech and technology very similar to the original cyberlibertarian ideal.

The analysis of texts shows that more than one public shock has emerged over the years. In this sense, the results of the analysis of texts seems to confirm the claim made in previous literature that governance of freedom of speech happens 'as a reaction' to shocks (Ananny and Gillespie, 2016), and it is performed through discourses and inscriptions.

As far as the narratives about freedom of speech and social media are concerned, the data reveal evidence that the most recurring/exemplary cases can be grouped into larger themes:

- Social media and extremism or terrorist radicalisation.
- Social media and hate speech and harassment.

- Social media and personal data protection (privacy).
- Social media and quality of information (e.g fake news).
- Social media and censorship.

The statements from the actors (excluding news media) highlight how public shocks mobilise or become exemplary cases for different types of narratives about freedom of speech, governance and technology. However, there is a strong ambiguity in some of the exemplary cases and storylines, which have been used in support of more liberal interpretations of free speech or on the opposite, as legitimation for regulatory initiatives. This is the case for Charlie Hebdo, which prompted calls for free speech as well as anti-terrorist regulation of SM, both at the same time.

It is interesting to note how much the idea of absolute free speech on the internet applies to the libertarian interpretation of cyberspace, typically exemplified in the manifesto written by Perry Barlow on the independence of cyberspace. Bloggers and Gab.ai are the only actors that adopt the 'original' libertarian narrative about freedom of speech. However, the statements show the prevalence of visions concerning freedom of speech as limited; visions progressively embraced by the larger SM platforms. The statements from the actors highlight controversial cases, where for every 'solution' to the problem, further levels of complexity are considered. The case of Paul Chambers is an exemplary warning about the limitation of law enforcement bodies.

As far as the findings about the narratives about governance are concerned, it is impossible to identify one dominant narrative in relation to the others, which is indicative of the lack of unity on a single process or decision-making format. However, academia and think tanks are the group with a clearer stance on a definition of editorial responsibility for platforms.

Narratives about technology show a similar division. In the case of technology, it is possible to distinguish two levels of issues associated with technological objects: the first level relates to the *usage* of SM platforms, and concerns tweets, posts, images, as in Paul Chambers or harassment and abuse received by female politicians.

Even though they cover a short time span, the data show changes in the regulatory ecosystem, with the introduction of a number of regulatory initiatives or declarations by regulatory bodies. This is the case of public bodies at national and supranational level, such as the CPS revised guidelines, issued in 2016, or the EU code of conduct also in 2016. However, it also captures changes in the internal policies of SM platforms, such as the decisions taken as result of the harassment of women, US elections, and extremist violence (as in the murders of Jo Cox and Heather Heyer).

One of the effects of regulatory initiatives highlighted in the data is the migration of alt-right members from larger SM platforms to smaller ones such as Gab.ai. A similar effect has also been noticed in other studies, such as Copland 2020, when considering the impact of regulations with regard to hate speech, as in the case of the quarantine introduced by Reddit. On the one hand this has had the result of limiting hate speech on that specific platform, but as a result many users reacted by leaving Reddit for less regulated spaces, with Reddit making this hateful material someone else's problem (Copland, 2020).

The second level concerns more the structure of SM as technological objects, and queries the role of AI and algorithms; in other words, it is more related to the hidden functioning of platforms (i.e. the infrastructure level). NGOs and bloggers (civil society), as well as academia, are more concerned with these latter issues, while the other groups, especially political bodies, see the main issues as being the hashtags, bots, trolls, etc., as technological objects that pertain more to the interface or specific usage of the platforms (see Figure 5.20). From this point of view Huey (2015) as well as the Demos findings show how counter-speech might be more effective than filtering solutions. However, the main public bodies (UK government speech on the Future of the Internet, Joanna Shield's speech, Bew's Report on intimidation in public life, as well as the German NetzDG) all point to the 24-hour deadline to remove content, which by nature favours automatic recognition and filtering of content rather than larger-scale counter-speech projects. This indicates how the narrative of the decision-making bodies tends to be interested in solving problems that lie at the level of the interface or specific usage of the platforms, rather than at the infrastructural level.

Considering the technological dynamics (Marres 2005) revealed by the data, the main finding concerns the consideration of the influence of commercial content in articulating the relationship between web pages. In particular, the results from the crawl and the reconstruction of the network of hyperlinks show how many commercial links are constantly integrated in the majority of web pages (ads, trackers, etc). Similarly, the high number of SM posts or pages that appeared as a result of the crawl (especially after 2016) show that the online presence of actors is no longer exhausted on web pages, but it takes place on several platforms. So it is possible to have a blog or a newspaper page, or an academic page, and find links to the same actor's page on Facebook, Twitter, LinkedIn. This finding does not tell us anything about the substance of the controversy (e.g. freedom of speech, limits, responsibility), but it says a lot in terms of the centrality of SM for the public sphere that we are living and experiencing.

5.15 Conclusions

In this chapter I applied controversy mapping methodology to the study of the SM as controversy on web pages (Venturini 2010, 2012, Venturini et al. 2015; Ooghe-Tabanou et al., 2016). Through the empirical analysis I found that actors on web pages have been mobilised around a limited number of exemplary cases and storylines, all sharing the idea that SM platforms are controversial but focusing on slightly different (even if connected) elements. By comparing the results of quantitative and qualitative analysis of documents I was able to isolate five main ways that SM platforms have been problematised: social media and extremism, social media hate speech and harassment, social media, algorithms and fake news (e.g. quality of content), social media, privacy and protection of personal data, social media and censorship.

The disposition of actors around issues shows a division of actors around the type of concern: government and the news media are mostly concerned with issues that take place at the level of the interface (user abusive usage), while academia and NGOs are more concerned with issues that take place at the level of the infrastructure (i.e. algorithms and data management).
Civil society and especially bloggers are very much divided on the issues concerning the definition of abuse and harassment, or illegal speech online. It appears that the majority of bloggers against regulation adopt an 'absolute' interpretation of freedom of expression. This is true of both extreme right-wing and extreme left-wing bloggers.

In this chapter, I have outlined the main features of the controversy as empirically detected from the observation of websites in different points in time. In order to make the process as transparent as possible, I have documented the different parts of the analysis, based on previous controversy mapping exercises, refraining from adding theoretical interpretations of the dynamics that emerged. As a form of triangulation and integration of my sources, in the next chapter, I am going to integrate this first round of finding with statements from the British press during the same period. I will then analyse the two types of findings through the lenses of the sociology of translation and critical data studies (chapter 7).

6. Mapping controversies using newspaper articles

6.1 Introduction

As explained in the methodological discussions in chapter 4, controversy mapping as a method is often developed to exploit the traceability of digital social data, and I approached the study of controversies using web pages and hyperlinks. In line with this method, in the previous chapter I studied the public shocks and related controversial elements concerning governance of free speech and SM, starting from statements collected from web pages.

Among the findings from chapter 5, I highlighted how news media play a singular role in the controversy. Numerically, they represent the largest group of actors producing statements relative to freedom of expression and social media platforms. The network analysis showed that they are well connected and used as a source for other actors, who cite them via hyperlinks. This centrality of media and press is recognised also at the theoretical level, where media are recognised as a fundamental actor in socio-technical controversies (Barry, 2001, 2013), as they occupy the role of 'tool of measurement' giving voice and informing the public of non-experts (Marres, 2015). Also, they are seen as intermediaries, through which other actors' narratives become more visible and established than others (Barry, 2001, 2013; Latour, 2005b).

In consideration of their specific role, in order to highlight the voices of the other actors, in the previous chapter I excluded news media from the analysis of statements. In this chapter, I re-introduce the voice of the news media, focusing on statements collected from articles specifically belonging to the British press. I will study and analyse the elements of the controversy: i.e. actors, and narratives (e.g. exemplary or recurring cases and storylines) as in the previous chapter. However, focusing on the British press I am interested in observing how news media contribute to the diffusion of specific ideas of freedom of expression and social media.

As described in the methodology, the choice of focusing on the press is also a way to compensate for the limitations created by the use of web pages and digital tools for data collection. In particular, articles from the press work very well as archives, as they are published and collected in datasets without constant updates. It is possible then to study the controversy in an historical perspective.

In the chapter I will present how I have analysed articles, in order to identify actors and public shocks. The findings contribute to isolate the ‘substance’ of the issues from statements that are collected only as ‘dynamics’ of the medium (see Marres, 2015). Below I explain the different stages of data collection and analysis that I have performed in order to identify the public and the issues that emerge from the British general press.

6.2 Data collection

In chapters 4 and 5 I described the different stages of the controversy mapping pathway (Venturini, 2010, 2012). Applying the same ‘steps’ highlighted in the controversy mapping pathway to newspapers is not possible. In particular, the identification of authors of statements in web pages works in a completely different way than in newspaper articles, due to the fact that authorship of statements cannot be inferred in the same way as with websites (e.g. based on the URLs’ domains). In newspapers, the only clear authorship of statements belongs to the journalists or to the publication. In the following paragraphs I will describe more in detail how I adapted the different parts of the method for the analysis of the controversy on newspapers.

6.2.1 Construction of corpus

As in the case of the methodology for online controversy mapping, in order to begin the identification of actors I had to build my initial corpus of ‘statements’. From the LexisNexis UK publications dataset, I downloaded approximately 1000 articles per year from January 2015 until April 2018. They represented the totality of articles available on the repository, selected on the basis of the similar keywords that I used in the previous data collection: ‘social networking’ OR ‘online social networking’ OR ‘social sites’ OR ‘social network*’ OR ‘social media*’ OR ‘networking sites’) AND (‘freedom of expression’ OR ‘freedom of speech’ OR ‘free speech’). Also in this case the list of keywords was based on the contribution of experts and on previous studies, fortified by the double-checking of their presence among the keywords presented in LexisNexis itself. For the collection of statements, I selected the UK national newspapers dataset, rather than issue-specialised magazines such as the *Economist*, because I wanted to collect publications available to the ‘general public’ and ‘non-specialist’ audiences.

LexisNexis only allows the download of approximately 1000 original documents per year; any other result is automatically filtered out by the system on the basis of similarity. This is a limitation that makes it impossible for this study to aim for statistical completeness (i.e. it is not based on the totality of articles produced on the topic in the UK). However, as discussed in the methodology, controversy mapping is a qualitative methodology and the number of articles is still enough to aim at the exhaustion of the main key issues using qualitative analysis.

The theoretical and methodological grounds for the identification of actors are described in the literature review and methods. As in the analysis of data from websites, I used a mix between the methodology of Venturini (2010, 2012) and Marres and Rogers (2005) for the identification of actors in issue-networks, namely exploiting Dewey's definition of 'public'. Differently from controversy mapping online, which uses domain names as a proxy for actors, with newspapers I had to do a qualitative analysis of the texts to find "such an assemblage of actors jointly implicated in an issue" (Marres and Rogers, 2005:8). Using qualitative discourse and quantitative content analysis I was able to extract relevant mentions from other actors, and to interpret their position I have connected them to the groups presented in chapter 5.

Table 6. 1 - Summary of articles collection and coding per year

Year	Articles collected	Articles coded
2015	850	200
2016	875	200
2017	789	200
2018	500	100
Total	3014	700

For the detection of actors and exemplary cases, I relied on the qualitative coding and mentions from the texts of the articles extracted with quantitative and qualitative techniques of analysis of the texts. Using Cortext I extracted a list of the 300 most relevant terms (i.e. based on linguistic techniques for weighting terms from documents, based on frequency per document) and I performed topic analysis using the Cortext LDA tool (the description of the tool is in the methodology chapter, chapter 4).

In previous studies, the association between actors was studied as the association/links connecting different websites (URLs). It is more complicated in the case of actors extracted from press articles. The association in this case can only be ‘semantic’ or based on the relationship that links the publication to the actors extracted. In this part I present the visualisation of networks connecting terms and topics and publication. Using Cortext’s tool for building heterogeneous networks (a description of the tool is given in the methodology chapter), I built networks of semantic relationships linking terms to the topics, then the topics and terms to specific publications. The network figures that appear in this chapter (Figure 6.2 to 6.7, 6.9 and 6.10) show the associations connecting the topics from the previous paragraphs to the different terms that emerge from the analysis of text. The sise of the nodes represents the co-occurrence, i.e. how many times the terms have appeared together in the documents.

6.3 Presentation of findings

6.3.1 Identification of actors

Compared to the identification of actors with web pages, the number of ‘authors’ retrieved from newspaper articles is much lower. Figure 6.1 shows the proportion of articles, according to the publications. In total, the dataset includes articles from ten different newspapers: The Guardian (different editions), The Mail (different editions), The Independent (different editions), Telegraph (different editions), The Times (different editions), Daily Mirror, Express Online, The Sun, The Observer, The Express. Within these ten publications, many articles are produced by four publications: The Guardian (22.9% of articles), Mail (21.1% of articles), Independent (19%) and Telegraph (13.4%).

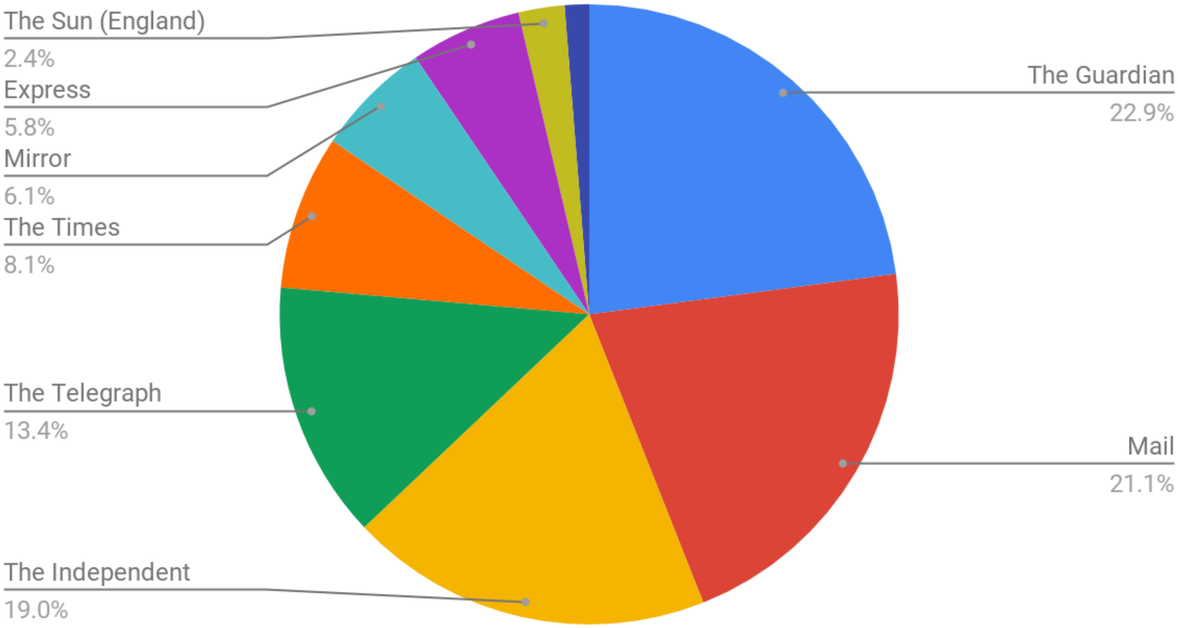


Figure 6. 1 - Percentage of articles in the dataset per publication

In the UK, national newspapers declare political support for specific political parties. Newspapers in the UK are recognised to have a political affiliation (i.e. Norris, 2001). I tried to consider this in the analysis, especially when considering the salience/relevance of topics. In Table 6.2, I have associated the previous percentages of publications' frequencies to their specific political position, based both on their support in the last general elections (2017) and on the perception of the UK public.

Table 6. 2 - Publications and political orientation

Publication	**Political Orientation ***	**Percentage**
The Guardian (different editions)	Liberal/Centre-left	22.91
The Mail (different editions)	Right-wing, conservative	21.07
The Independent (different editions)	Liberal, centrist	18.95

Telegraph (different editions)	Centre-right, conservative	13.44
The Times (different editions)	Centre-right, conservative	8.06
Daily Mirror	Centre-left	6.08
Express Online	Right-wing, Eurosceptic	5.52
The Sun	Right-wing, conservative	2.40
The Observer	Centre-left	1.27
The Express	Right-wing, Eurosceptic	0.28

* based on support expressed in 2017 political elections

The results show that publications are spread almost 50-50 between left-wing and centrist publications and right-wing publications. The articles composing the dataset belong more or less to a varied political spectrum, even if slightly leaning towards right-wing positions. The overview of publications shows that four larger media players have published more statements regarding the controversy, both online and on printed press. As in the previous chapter, where: The Guardian, Independent Telegraph, Daily Mail as well as digital publications from the BBC website, the Huffington Post (see Table 5.7, chapter 5) were among the most present URLs.

As explained above, the other actors that I have identified as part of the ‘public’ emerged from the analysis of the texts. For this reason, I will introduce and discuss them in the next part of the chapter, where I present the result of the analysis of public shocks, issues and narratives.

6.3.2 Identification of shocks and issues they summarise

To triangulate the findings from the web pages, I performed quantitative and qualitative analysis of the texts to identify the main shocks, exemplary cases and storylines. The format of newspaper articles is more standardised than that of web pages and it was possible to use the tool for automatic topic analysis provided in Cortext (based on the LDA algorithm described in the methodology). The automatic detection identified 10 topics on the basis of groups of words statistically recurring together across different documents. I also performed an automatic detection of the most relevant terms, using the measure TF-IDF (as in the previous chapter). I selected a list of exemplary cases comparing the results of the automatic recognition with the result of the qualitative analysis of texts. Table 6.3 presents the summary of the topics identified with the automatic recognition with Cortext (further documentation is available in the Annexes, from page 32 A).

Table 6. 3 - Cortext LDA topics output and exemplary cases from qualitative analysis

Years	**LDA topics**	**Exemplary cases from qualitative analysis**
2015	Topic 2 – Government – terrorism and encryption	
2015	Topic 5 – Charlie Hebdo	*Charlie Hebdo – SM – terrorist attacks – Anjem Choudary – Lee Rigby*
2016	Topic 6 – Facebook – Fake news and privacy	
2016	Topic 4 – Jo Cox, abuse online, hate speech	*Labour MP Yvette Cooper – Labour MP Jo Cox – EU referendum – Remain campaign – Twitter – Trolls (2016)* *The controversy involving EU referendum – Leave campaign – Twitter – Facebook – Fake news and the controversy involving US election – Donald Trump – Russian trolls – Fake news (2016)*
2017	Topic 7 – Twitter abuse/ harassment – trolls and blocks	*Exemplary case of harassment with Independent game-maker Zoe Quinn (also called Gamergate) (2014/2015)*
2017	Topic 1 – Milo Yiannopoulos – Twitter_-Leslie Jones	*Exemplary case of harassment Leslie Jones – Milo Yiannopoulos-Twitter – Ghostbuster movie – trolls (2016)*
2017	10_content removal_hate speech, Germany	*Migrant crisis and Introduction of NetzDg in Germany*
2017	Topic 3 – Hate groups – Reddit – alt-right – Trump – Charlottesville	*The controversy involving EU referendum – Leave campaign – Twitter – Facebook – Fake news and the controversy involving US election – Donald Trump – Russian trolls – Fake news (2016)* *Cambridge Analytica, Alexander Kogan, Christopher Wylie, Steve Bannon*

		Facebook, EU referendum, US elections (2018)
2018	Topic 8 – University Student Union – No Platform – Jordan Peterson – Milo Yiannopoulos_women	*Katie Hopkins-Twitter-LBC radio- terrorist attack in Manchester (2017)*
2018	9_China, Turkey Internet regulation and censorship	*SM agreements with states for censorship*

Topics are numbered by the algorithm but chronologically ordered

As mentioned in the methodology chapter, I based the selection of the number of topics on the assessment of coherence included in the script, combined with visual inspection of the output (see Appendix, p.42 A) as well as manual analysis of texts extracted as sample of the document attached to the topic by the algorithm.

Overall, the results fit with the ones already described in the analysis of web pages in the previous chapter. However, the analysis of newspapers highlights other exemplary cases, completing and reinforcing the structure of episodes and issues found from the statements on the web pages. In the next paragraphs I present an overview of the specific exemplary cases and stories that emerge from the articles as well as the main shocks and issues they exemplify.

The main issues described by the exemplary cases and storyline can be summarised as:

A – public shocks related to terrorism, extremism;

B – public shocks related to hate speech on SM;

C – public shock and cases of harassment, especially misogyny and gender violence on SM;

D – public shock related to SM and fake news and manipulation of users;

E – public shock related to SM and censorship.

A – Public shocks related to terrorism and extremism

As in web pages, in newspapers several articles also reported episodes of public shocks as a reaction to episode of terrorism. The articles confirm that the Charlie Hebdo attack represents a pivotal event in the global public discourse about freedom of speech. On that occasion, SM, and in particular Facebook, took a public stance in defense of free speech:

> Facebook 's CEO, Mark Zuckerberg, in a post published on his personal profile page on Friday morning, called for a rejection of ‘extremists trying to silence the voices and opinions of everyone else around the world’. ‘I won't let that happen on Facebook,’ he wrote. (Parkinson, 2015, document: The Guardian 2015-01-09).

As seen in the statements from web pages in chapter 5, newspapers also confirm government preference for increasing regulation and increasing the responsibilities of SM for radicalised and hate speech. British newspapers provide further information on how a narrative of security started to develop. They show that the storyline goes back to controversial British ‘radicalised individuals’, as in the case of Islamic preacher Anjem Choudary. Choudary has been considered the origin of the radicalisation of the murderers of Fusilier Lee Rigby in 2013. In 2015, the British government accused SM of "consciously failing" or “being in denial” (Parsons, 2015; document: Mirror 2015-01-28) about their role in combating extremism and hate groups using their services.
In these story lines newspapers reinforce the government narratives presented in chapter 5 calling for greater control over speech on SM. Some as in the Times, have a milder approach:

> There is always a delicate balance to be struck between free speech and security. The revelation that social networking sites resisted requests from the police to delete radical and extreme content posted by Anjem Choudary and his followers suggests that the companies are not getting this delicate balance right (Hamilton et al. 2016, document: The Times 2016-08-18).

Others, such as the Mail, present a more explicit criticism:

> Didn't Choudary's kid-glove treatment go far beyond tolerance of free speech? Why were Twitter and YouTube so loath to close down his accounts? [...]Isn't it almost as if the West's liberal elite, in its craven terror of offending minorities, has a death-wish for our values and way of life? (Daily Mail Comment, 2016; document: Mail 2016-08-17).

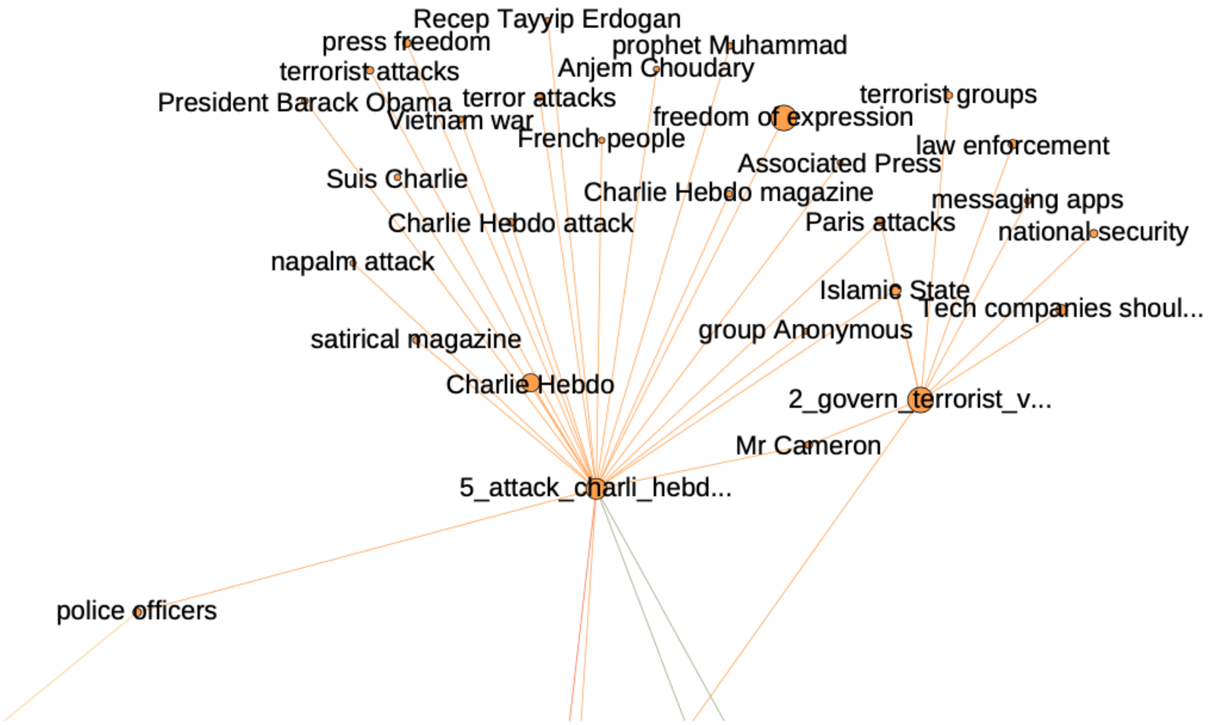


Figure 6. 2 - Semantic connection between Topic 5 and 2 (yellow links) and most relevant terms Mostly present in the Guardian, Express and Times

When looking at the visual representation of topics, we see that also from the automatic recognition of topics in Figure 6.2 Anjem Choudary appears related to the Charlie Hebdo attack. The automatic recognition also caught the expression "SM: Tech companies should…", which shows how SM are connected to a need for further responsibility.

In newspapers it emerges that the migrant crisis and terrorist attacks are mentioned together as trigger events for hate and more specifically Islamophobic speech. Between 2016 and 2017 several regulation initiatives were taken both by states and SM to reduce Islamophobic hate speech, and right-wing leaning publications (such as the Mail) mentioned the migrant and refugee crisis (as a result of the Syrian and Libyan conflicts) in relation to freedom of expression, and in particular presenting criticisms of regulation of speech online adopted as protection of migrants and minorities.

B – Public shocks related to episodes of hate speech

In chapter 5, we saw how episodes of hate speech were mentioned as justification for regulatory initiatives as in the German law NetzDG. From the newspapers, other UK regulatory initiatives on the topic of hate crime emerge (such as the UK Parliament publication on hate crime (2017)) and the Internet Safety Strategy green paper (2017). Newspapers present the government initiatives in opposition to actors critical of regulations of hate speech. As with the web pages, there is a correspondence of criticism of regulation between free speech activists and far-right groups. Far-right groups have been petitioning against hate speech policies, on the basis of their right to free speech:

> [...] After (far right group) Britain First's suspension, Facebook was criticised by the far right group's leaders as 'fascist' and said the move denied freedom of speech to 1.1 million people who have liked the page. The page was restored an hour and a half later – and a Facebook spokesman said it had been taken down due to an 'error'. (Mirror 2015-12-09)

With regard to the position of journalists on newspapers, what emerges is a tendency to be critical of restriction to speech:

> Clearly, we should all be interested in creating a society without hate and xenophobia. Yet silencing people will only send negative sentiments into the underground, where they will fester and rot into an even more disgusting form. [...] At this stage in history, us Europeans need an open forum more than ever. There are millions of migrants moving across Europe, wars igniting on our borders and terrorists carrying out vile atrocities on the streets of cities considered to be centres of the free world. This is not a time to be silenced. (Hamill, 206; document: Mirror 2016-01-19)

Despite increased regulation, hate speech episodes have continued and several attacks from white nationalists brought to the point of rupture between big platforms and far-right groups. Newspapers confirm the finding from the previous chapter where the attack on the protesters in Charlottesville played a pivotal role in the approach to hate speech of the big SM companies, which introduced stricter regulation of speech:

> For Silicon Valley companies that must balance the right to free speech with the risk of empowering and broadcasting abhorrent beliefs, the violence in Charlottesville has been a clarifying moment (White, 2017; document: The Independent 2017-08-20).

This choice created a separation from smaller companies like Reddit and Gab, with the latter, as explained in chapter 5, using the opportunity to diversify their users and gather the voices of the groups expelled from the other platforms. Newspapers reinforce the idea that 'alternative' platforms (including Reddit) are mostly occupied by white supremacists, far-right sympathisers and Trump supporters (McGoogan, 2016, document: The Telegraph 2016-11-14; Parry, 2017, document: Mail 2017-08-17).

> Silicon Valley is cracking down on neo-Nazis and white supremacy in the wake of the deadly Charlottesville rally. Despite years of net and political neutrality, tech giants like Google, Apple and Facebook have announced they want to make it harder for the alt-right to spread its hateful rhetoric using their services (Parry, 2017, document: Mail 2017-08-17)

> Internet companies are generally reluctant to police the political nature of their content in the interests of freedom of speech and expression, which Silicon Valley as a whole champions. But recent events in Charlottesville, which Donald Trump maintains were 'both sides' fault', have shaken the industry out of silence and into definitive action. The stance poses questions over the tech industry's control of free speech and effective censorship of content in daily means of communication. As conventional services are cut off, white supremacists are turning to other methods of communication, namely the 'alternative' social networking site Gab (William, 2017, document: The Independent 2017-08-21).

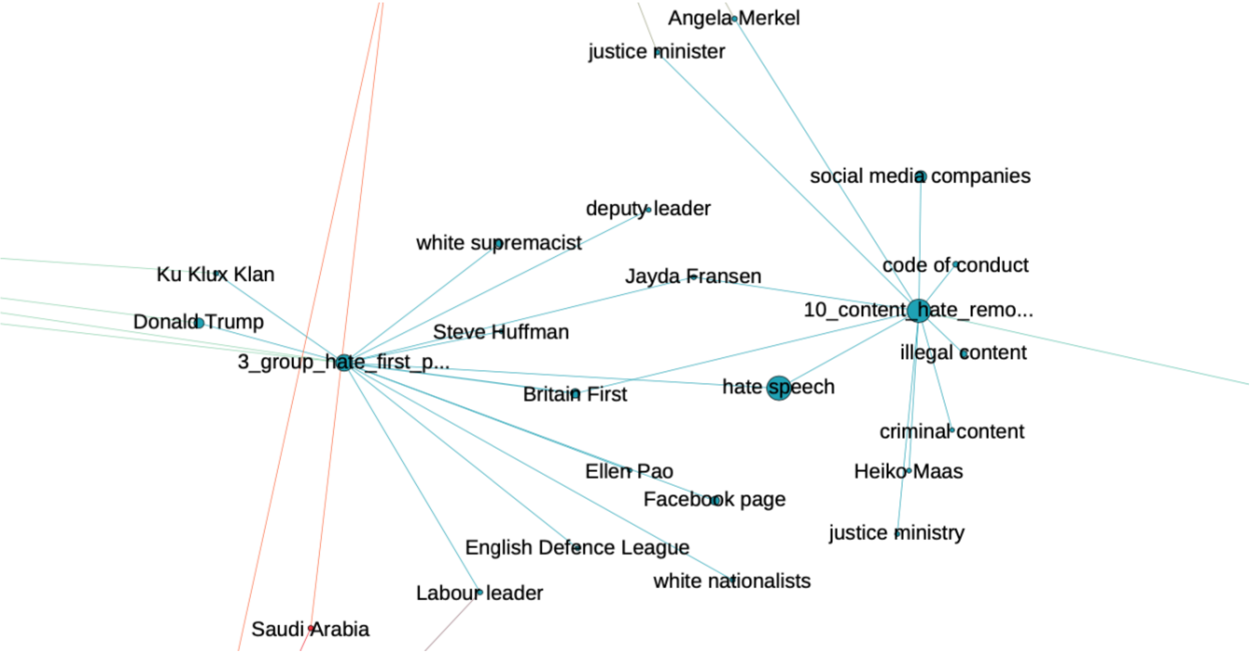


Figure 6. 3 – Semantic connection between Topic 3 and 10 (both in blue links) and most relevant terms

Mostly present in the Times, Express, Observer

In Figure 6.3 it is possible to see the visual representation of the topic and issues connected with hate speech, neo-nazis and white supremacy groups. In the group fall the nationalist and xenophobic groups Britain First and the English Defence league (Jayda Fransen). Also, as introduced above, the (former) President of the US is an exemplary case of hate speech and inflammatory content, where his presence on SM has supported white supremacist groups, and white nationalists. In the network of semantic connection, he acts as a bridge between hate speech and trolls. The topic of hate speech is also connected to the Code of Conduct on Illegal speech, which was discussed in chapter 5. The names of two German politicians, Angela Merkel and German minister of justice Heiko Maas, are among the most relevant words, and show the connection between the adoption of the famous NetzDG and the prevention of hate speech.

From the analysis of relevant words emerges also the names of Steve Huffman and Ellen Pao, respectively CEO and former CEO of Reddit. In 2015, Ellen Pao was fired as a result of a petition online, after she introduced anti-harassment policies, a regulation against revenge porn and banned abusive forums. Her case also became exemplary since she filed sex discrimination claims against the Silicon Valley venture capital firm Kleiner Perkins (The Independent, 2015, document: The Independent 2015-07-12). The episode connects the issue of hate speech to another main group of exemplary cases, created by episodes of bullying and harassment based on gender.

C – Public shock and cases of harassment, especially misogyny and gender violence

In the study of statements on web pages, episodes of misogyny and violence based on gender emerged as public shocks which mobilised the development of forms of regulation of speech. Of all the publications, The Guardian appears as the one producing more content, framing the issue of harassment as the foundation for stronger regulation of speech (Jeffries 2016, document: The Guardian 2015-04-16). Some of the most frequent cases mentioned in newspapers in the discussion of the issue of harassment on SM have already appeared in the analysis of the web sites. These are the case of the harassment received by the feminist activist Caroline Criado-Perez and Labour MP Stella Creasy campaign (the Jane Austen on the £10 note campaign) and the harassment and threats addressed to female MPs in connection with the Remain campaign and UK general elections in 2017. With newspapers, other exemplary cases and storylines of gender-based harassment emerge, such as the story of the death threats received by Yvette Cooper and Jo Cox, and the storyline created with the Gamergate case, and the case of harassment against women celebrities such as Leslie Jones.

As seen in the web pages, in 2016 female Labour MPs received several death threats on Twitter for their support of the Remain campaign in the EU referendum. In the same year, Labour MP Jo Cox was killed by a right-wing extremist and English nationalist. The event shocked public opinion and drew attention to the issue of safety for female MPs. Newspapers reproduced the content of the threats in articles. One of the figures who received abuse was the Remain-supporting black London MP David Lammy, who called police after reportedly receiving a death threat via social media. In one message he was reportedly told "I hope your kids get cancer and die" and "I wish you the same fate as that b*tch got stab" – a reference to the Labour MP Jo Cox who was killed during the referendum campaign (Lusher, 2016, document: The Independent 2016-08-14).

These cases were mentioned on different occasions as examples of the necessity for more regulation of speech. As seen in the previous chapter, these episodes are mentioned in the revised

Guidelines for anti-hate and harassment online published by the Crown Prosecution Service in 2017. Newspapers also link the murder of Jo Cox and the death threats against BAME and female MPs to the creation of the Online Hate Crime Hub: a body of volunteers recruited and trained by the Metropolitan police to identify and appropriately respond to hate speech online.

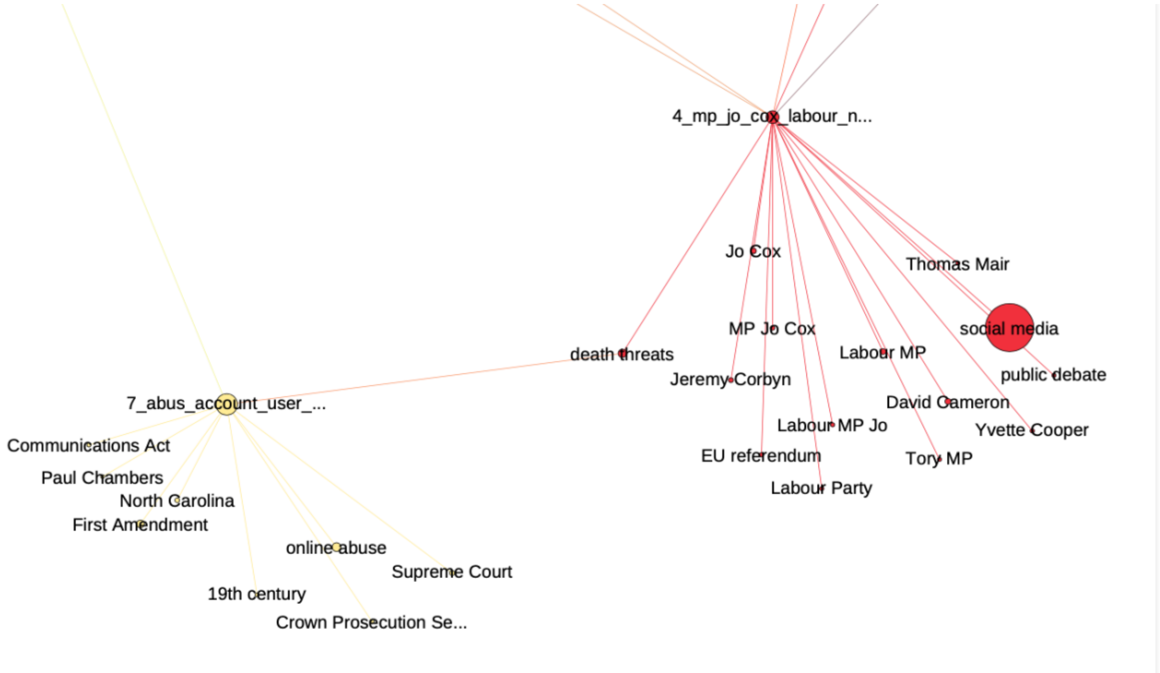


Figure 6. 4 – Semantic connection between Topic 4 and most relevant words (red links)

Mostly present in the Telegraph, Times, Guardian

Figure 6.4 shows how the topic of hate crime and the exemplary case represented by Jo Cox's death and other death threats are connected to the EU referendum. The name of MP Yvette Cooper also stands out, as she was also a target of misogynistic harassment. She is also the organiser of the 'Reclaim the Internet' campaign initiative already emerged in the web pages, and the chairperson of the Home Affair Committee on hate speech (2017). The node 'death threats' connects the topic to another where the Crown Prosecution Service Guidelines also appear. As we saw in the statements from web pages in those years the CPS produced the revised guidelines inclusive of threats against women. Also, from the figure it is possible to see how death threats and CPS are also connected to the episode of Paul Chamber, which is the example of the excess result of state regulation and policing of online environments.

A symbolic episode which is used in the narrative presented by newspapers is the harassment case against Zoe Quinn. This is an episode that took place before the data collection, and as in the case of Anjem Choudary and Caroline Criado-Perez it is used as a storyline in the description of episodes of harassment that happened later. In August 2014, a gender-based harassment campaign targeted several women in the video game industry; notably game developers Zoe Quinn and Brianna Wu, as well as feminist media critic Anita Sarkeesian. Zoe Quinn's former boyfriend wrote a disparaging blog post about her, and users using #gamergate hashtag started harassment campaigns against Quinn and others included doxing (i.e. public sharing of private information or identity about an individual or organisation), threats of rape, and death threats. Gamergate proponents (‘Gamergaters’) organised anonymously or pseudonymously on online platforms such as 4chan, Twitter, and Reddit. The Guardian connects Zoe Quinn's story from 2014 to events taking place in 2017, and in particular it stresses the clash between alt-right groups and women’s and feminist movements.

> The action taken against Google now echoes similar campaigns of harassment and activism by the ‘alt-right’ during the 2016 presidential election and by Gamergaters in 2014. Both movements were dominated by angry men, sometimes using anonymous online identities, who felt disenfranchised and wanted to see a radical overhaul of the status quo. [...] (Wong, 2017, document: The Guardian 2017-08-11).

The Guardian is one of the few publications openly taking a stance against abusive misogynistic speech. The Guardian explicitly called out the Daily Mail for perpetuating harassers' tones when discussing the case of harassment on barrister Charlotte Proudman when a senior lawyer, Alexander Carter-Silk, sent her a sexist message on LinkedIn.

> The Daily Mail went in for the kill, calling her a ‘feminazi' barrister’ The obvious other precursor to Charlotte Proudman is the case of Zoe Quinn (Williams, 2015, document: The Guardian 2015-09-16).

A similar opposition emerges also in the case of Leslie Jones and Milo Yiannopoulous (July 2016). On that occasion almost all the UK publications described the racist and misogynistic harassment received on Twitter by actress Leslie Jones because of her part in the remake of the movie ‘Ghostbusters’. The attack started after the ‘professional troll’ Milo Yiannopoulos (a former editor of the alt-right media Breitbart) wrote a review and posted comments against the actress and in general against feminism. His followers and sympathiser users continued the attack and pushed Jones to withdraw from Twitter. As a result, Twitter took the decision to permanently suspend Yiannopoulos's account, the first case of a 'celebrity' account suspended on the basis of abusive speech. The case created an opposition between publications welcoming greater regulation, and others framing the suspension as an attack on free speech. As in the case of hate speech some newspapers used the case to argue that regulations are counter-productive if the result is to leave to these groups the monopoly of free speech discourse:

> Yiannopoulos himself described the suspension as a 'cowardly' act and suggested that it amounted to an attempt by the 'totalitarian left' to make Twitter a 'no-go zone for conservatives'. Now that may be nonsense, but the narrative in which the conservative right is fighting a 'culture war' – as Yiannopoulos put it – against leftie liberals is increasingly prevalent[...] (Gore, 2016, document: Independent 2016-07-20).

> Vulnerable people should not be left to the mercy of the mob – either on Twitter or on the street. But by enabling ostentatious rabble-rousers like Yiannopoulos to present themselves as martyrs in the cause of liberty, there is a danger that Twitter shifts the focus away from their misdemeanours, instead of holding a mirror to them (Gore, 2016, document: The Independent 2016-07-20).

D – Public shock related to SM and fake news and manipulation of users

In the previous chapter, we found statements from actors on the controversies that emerged after revelations that SM have played a role in both the Brexit referendum and the US election results, by allowing the viral diffusion of fake information and by hiding political use of advertisement and commercial space on users' feeds. Similar statements emerge from newspapers. The magnitude of the implications for the public and SM only started to be discussed in 2018, the final year of the data collection. The articles discuss the involvement of the Russian bots in the Leave campaign in the EU referendum and Donald Trump's campaign in the 2016 US election and the use of SM for the creation of fake news and manipulation of public opinion. In particular, Facebook appears as the most targeted platform on fake news.

> At the centre of the backlash against the dissemination of fake news was Facebook. The publishing and technology company, which was criticised for allowing its automated systems to send fake stories to 1.8 billion users, has updated its algorithms to downgrade fake news and clickbait on users' news feeds (Gupreet, 2017, document: Times 2017-02-11).

> Facebook has a fake news problem, a rampant epidemic of phony and outrageous headlines in which a fraction-of-a-penny-per-click gets traded for lies (Weinstein, 2016, document: Mirror 2016-12-07)

The issue of fake news showed the weakness of the absence of regulation of political advertising on SM. In spring 2018, Cambridge Analytica became another exemplary case and addition to the storyline on misinformation. In spring 2018 it was leaked that Cambridge Analytica – a company close to the British Conservative Party which collects personal data to create psychological profiles of users for use in micro-targeted marketing campaigns – was suddenly suspended by Facebook on charges of having used data collected on the social network that did not belong to it. The company also appears in the investigation concerning the manipulation of opinion that happened during the Leave campaign for the EU referendum and Trump campaign in 2016. The Guardian (2018-03-17) and the New York Times published articles accusing Facebook of having made the collection possible, albeit not actively, and of having then underestimated or hidden it. For the first time, the national press started to discuss the economic return from the users' data and advertisers and their possible role in the manipulation of information.

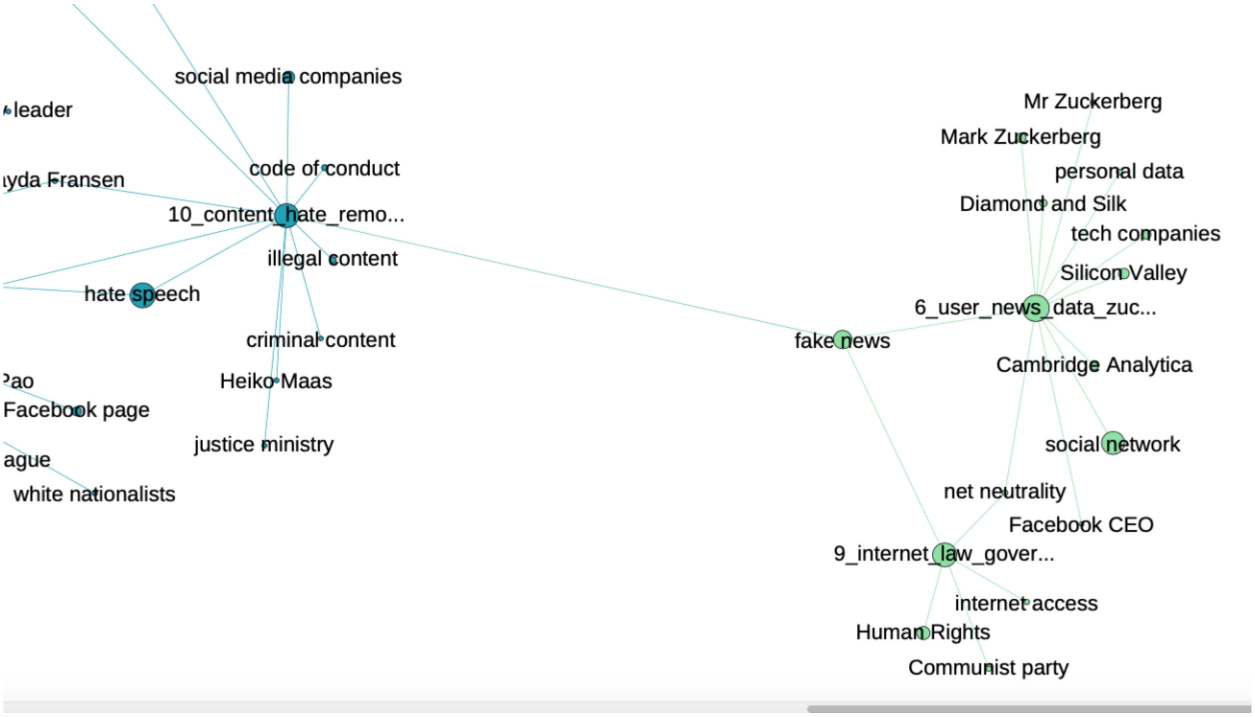


Figure 6. 5 – Semantic connection between Topic 6 and 9 (green and dark green) and most relevant words

Mostly present in Observer, Guardian, Mail

Figure 6.5 shows the connection between the issues relative to fake news and the large topic of users' data protection. Cambridge Analytica appeared in newspapers as a revolutionary moment, a 'wake up call for a generation' (Hastings, 2018, document: Mail 2018-03-21). In the narrative created by newspapers, the recurrent aspect is that 'something has changed' in the relationship between states and platforms, and in general the way in which society looks at platforms. Newspapers had already defined SM as 'trojan horses' in 2017, on the occasion of the first round of investigations taking place after the suspect Russian infiltration during the US election campaign (Borger et al., 2017, The Guardian 2017-10-22). Newspapers strengthened the idea that SM are no longer free speech defenders, and sustained the government pressure for more regulation of SM.

> But there is an inescapable sense that this year, something is different. That the slew of troubles and criticisms that Google, Facebook and others have faced are reflective of a general sea change, a growing feeling that they may not be the good guys (Titcomb, 2017, document: The Telegraph 2017-12-27).

What emerges from the articles is the beginning of a new way of narrating the cases, associating the episode to a general problem related to the business model of the platforms.

> Bot accounts furiously sharing messages happened on such a grand scale that it's hard to believe the platforms didn't notice. Algorithms ‘left to their own devices’ mean that content generated by any random individual – with no journalistic track record, no fact-checking or no significant third-party filtering – can reach as many readers as, say, the BBC. And that's a critical problem. [...] Facebook insists it is not a media company but merely a ‘neutral technology pathway’ facilitating connections between people. It is a misconceived and dangerous position. It is a media company with enormous influence in shaping someone's worldview about whom to trust. And it is profit-driven (Botsman, 2018, document: The Guardian 2018-02-11).

> Social media micro-targeting has become another battleground [...] As with mass data collection, perhaps it may eventually be concluded that that reach is simply incompatible with democratic and human rights. [....] At the very least, we must now seriously question the business models that have emerged from the dominant social media platforms (Joseph, 2028, document: The Independent 2018-04-03).

E – Public shocks related to censorship

As seen above and in the previous chapter, the same exemplary episode can be used by actors to discuss different issues. In the case of death threats, some actors (as for instance MP Yvette Cooper) used the episode as a case to request further regulation. While other actors, used the topic of death threats online to discuss the case of Paul Chamber and his trial with the Crown Prosecution Service in 2014. In chapter 5 I described the episode, in which Paul Chambers became an example of the risk of censorship related to policing SM using legislation and government tools. In the articles it still appears as one of the most relevant issues, as can be seen in Figure 6.7.

.

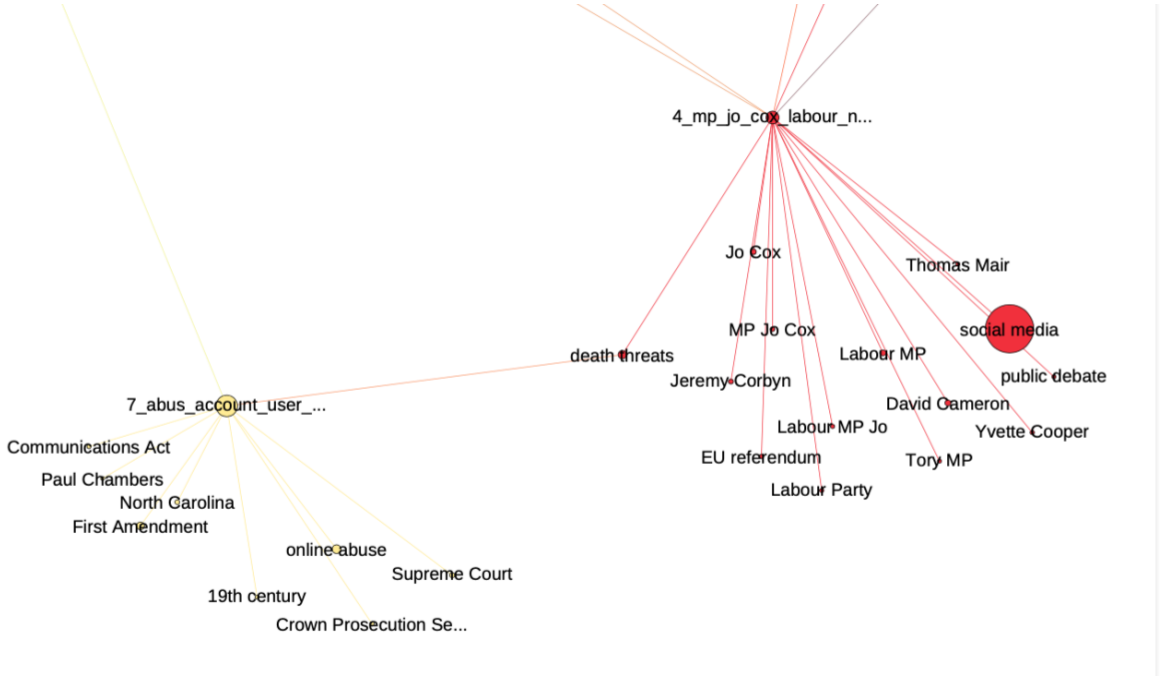


Figure 6. 6 – Semantic connection between Topic 7 (yellow) and 4 (red) and most relevant words Mostly present in Mail, Mirror and Sun

Similarly, the episode of Charlie Hebdo is linked to the issue of terrorism and radicalisation, and on web pages it is also associated to state surveillance (with the discussions over the protection of encryption), while on newspapers it appears linked to the issue of the ambiguous relationship between SM platforms and authoritarian states as Turkey and China (see fig. 6.3). Other episodes are used to stress the role of platforms in editing the contents and criticise the SM role of regulators, and in general point out the incongruences between SM declarations in defence of free speech, are at odds with their role in speech regulation. One of the most relevant appears to be the Norwegian prime minister post of the famous picture of the Vietnamese Girl running away from Napalm (fig. 6.7) censored by Facebook. The case was taken by the editor of a Norwegian publication (the Aftenposten), who challenged the rules established by Facebook (in this case, the automatic censorship of nudity because of children pornography regulation) resulting in this way in unjustified censorship. The Norwegian editor-in-chief later referred to Facebook CEO Mark Zuckerberg as "the world's most powerful editor" (Brown, 2016, document: Express 2016-09-09).

Figure 6. 7 - Nick Ut's photograph of Kim Phúc

The automatic recognition of topics highlights another aspect of the controversy, with articles stressing the incongruence between SM internal policies and declarations in favour of free speech. Particularly, after the attack to Charlie Hebdo, Facebook was criticised for publicly taking a stance in favour of free speech while obeying Turkey’s requests to censor the platform on its territory. The Guardian defined as a “murky relationship” (Hern, 2015, document: The Guardian 2015-01-13), and the Independent as “disingenuous” (Dewey, 2015, document: The Independent 2015-01-28) Facebook decision to collaborate with states’ requests and at the same time attempting to present the company as a protector of freedom of expression. The semantic connection in Figure 6.2 shows that the episodes relate to the relevant issue of Turkey.

The contradiction appears also in other topics automatically detected (Topic 9 in Figure 6.8 below), where semantically Facebook and fake news are related to the issue of China, and internet access. The case refers to another episode where Facebook agreed to collaborate with a state without questioning the issue of freedom of expression in the country concerned. Facebook has been barred from China since allowing separatist movements to post material opposing the Communist Party. However, according to The New York Times, Facebook created algorithms to block censored content from appearing in users' feeds in particular areas of China, as a way to appeal its ban to the Chinese government (Bridge, 2016, document: Times 2016-11-24).

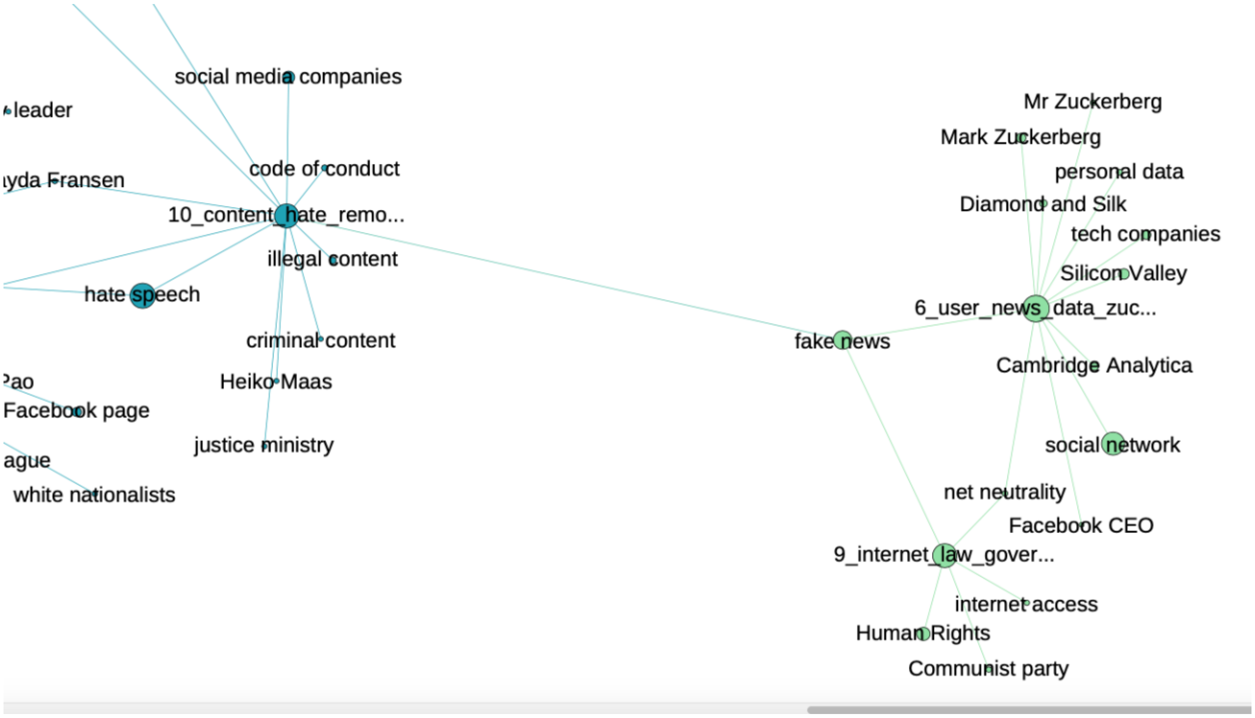


Figure 6. 8 – Semantic connection between Topics 6 and 9 (green and dark green) and topic 10 (blue) and most relevant words

Mostly present in Observer, Guardian, Mail

6.3.3 Narratives about free speech

As highlighted in the previous chapter, across the years of the data collection it is possible to notice a change in the general narrative adopted by SM companies, that from declarations in defence of free speech in occasion of Charlie Hebdo, ended up with declaring that regulation of speech has become inevitable:

> Last week, Sinead McSweeney, a senior [Twitter] executive in Europe, told MPs that 'it is no longer possible to stand up for all speech' (Titcomb, 2017, document: The Telegraph 2017-12-27).

The main finding about the narrative about freedom of expression emerging from the articles is that the debate is framed as between left- versus right-wing supporters, the first more in favour of regulation and the second more sceptical and asking for protection of free speech. The position in favour of more control is called by newspapers 'no-platform', and they include episodes of disputes that arose as a result of 'divisive' characters being invited to speak at universities or in public spaces. Student unions in different universities, and their choice to no-platform a number of speakers such as professor Tim Hunt, Mary Beard, Kate Smurthwaite, Julie Bindel, Alan Perkins, Germaine Greer and Jacob Rees-Mogg are examples used by newspapers to present opposing sides: on the one hand the necessity to protect minorities and vulnerable groups. On the other, the 'state of censorship' and restriction of free speech. Some publications such as the Guardian, focus more on the aspect of the victims and the need of protection and regulation of speech, and others such as the Mail, are more in favour of 'abusive' speech, insofar as it is 'free speech'.

> The National Union of Students (NUS) [...] has banned its institutions from hosting a myriad of speakers, including those from the English Defence League (EDL), British National Party (BNP), some members of Ukip, [...] It's time to defend the right to be offensive: not allowing debate on campus is dangerous. Free speech is a basic human right, and an essential tool for a functioning democracy. [...] Men should be allowed to debate abortion, student media should be uncensored, and Katie Hopkins should be allowed a platform to share her views – however ludicrous they may be. George Orwell was right when he said: 'If liberty means anything at all, it means the right to tell people what they do not want to hear' (Pearson-Jones, 2015, document: The Independent 2015-12-28).

While others present the issue as relative to a specific group of people: the 'useful liberals' (The Guardian 2018):

> Most freedom of speech debates now start on the false premise that denying someone a platform is censorship. [...] The disappeared of Egypt, the jailed and flogged blasphemers of Saudi Arabia, the arbitrarily detained bloggers and journalists of China are being denied freedom of speech. It's an insult to their ordeals that we equate them with shutting down Milo Yiannopoulos's Twitter account. [...] In On Liberty, John Stuart Mill, one of the great defenders of free speech, says a struggle always occurs between the competing demands of authority and liberty. He argues that we cannot have the latter without the former[...] Freedom of speech is no longer a value. It has become a loophole exploited with impunity by trolls, racists and ethnic cleansing advocates. They are aided by the group I call useful

> liberals – the ‘defend to the death your right to say it’ folk (Malik, 2018, document: The Guardian 2018-03-22).

Some characters, such as Milo Yiannopoulous, Jordan Peterson and Katie Hopkins are used as symbols of ‘uncomfortable truth’ or ‘free thinkers’ in right-wing publications (Mail and Sun). They have a vision of free speech more in line with alt-right movements. Katie Hopkins’ article on free speech reported in the Sun her position on insults online:

> ‘I hope you get a tumour.’ That was in response to a tweet I wrote on Prime Minister's Questions. You see, when people are given the chance to share their views, they do. And I'm fine with it. The guy on Twitter doesn't really wish I'd get a tumour. He just wants to tell me he disliked what I said (The Sun, 2015, document: The Sun 2015-09-17).

A similar case is that of Jordan Peterson, a controversial Canadian psychologist who has attracted a large group of followers among the ‘alt-right’ and conservative groups. He was interviewed by Cathy Newman in 2017 and the case resulted in a series of attacks and misogyny and abuse online against the journalist, who also received death threats. The Mail describes the episode as follows:

> He dares to say the unsayable on political correctness, feminazis and whingeing millennials. For this, he's demonised by the British Left and shouted at in TV interviews (Sandbrook, 2018, document: Mail 2018-02-10).

Like Jordan Peterson, Milo Yiannopoulos has for years been a key controversial figure in the British public discourse. His storyline is echoed in other relevant cases involving universities and cases of no-platforming of speakers. The articles have different ways of framing the issue. There is a clash between those which describe the Yiannopoulos events as relating to freedom of expression (such as the Mail or the Telegraph) and others as the Guardian which warn against creating a legitimate framing for the character. In 2017 several universities protested the decision to invite Yiannopoulos to present his new book. Some protests were particularly violent, as in the case of UC Berkeley in February 2017. A division appears between those newspapers presenting the protests as part of the debate over free speech, like the Daily Mail and the Telegraph and others, such as the Guardian, which are sceptical of that type of framing.

> Maajid Nawaz describes the students demanding censorship as members of the ‘regressive left’. Milo Yiannopoulos calls them ‘snowflakes’. Nowadays, the only thing that is stopping a student from accessing a new idea is a censorious gag from a student union or NUS apparatchik (Peters, 2017, document: Telegraph 2017-02-17).

> Casting the controversy over Yiannopoulos as one of freedom of speech has been a public relations coup for the right. ‘It has an almost irresistible propagandistic value,’ said Larry Rosenthal, the chair of the UC Berkeley Center for Right-Wing Studies, by providing the right with an opportunity to ‘talk about the hypocrisy of liberals with respect to free speech’. (Wong and Lewin, 2017, document: The Guardian 2017-04-26).

In the case of militants from alt-right groups, the most cited are members of the English Defence League and Britain First, and in their vision the problem is excessive regulation. The overview of free speech interpretation shows how the separation between those requiring less and more regulation corresponds to more right-wing/alt-right positions. Publications such as the Mail utilise the definition of ‘snowflake’ assigned to people that become offended in a system of freedom of expression. Similarly, ‘politeness’ and ‘common sense’ are terms that are increasingly politicised as they are used in support of content regulation positions.

Figure 6.9 shows the aspect of the controversy that moves beyond the strict environment of SM: it includes the ‘no-platform’ issues raised in academia. It is interesting to find that the most representative character for the no-platform issue is Milo Yiannopoulos and that the episodes are linked to the issue of abuse and hate speech (and the actor that semantically acts as bridge is Donald Trump).

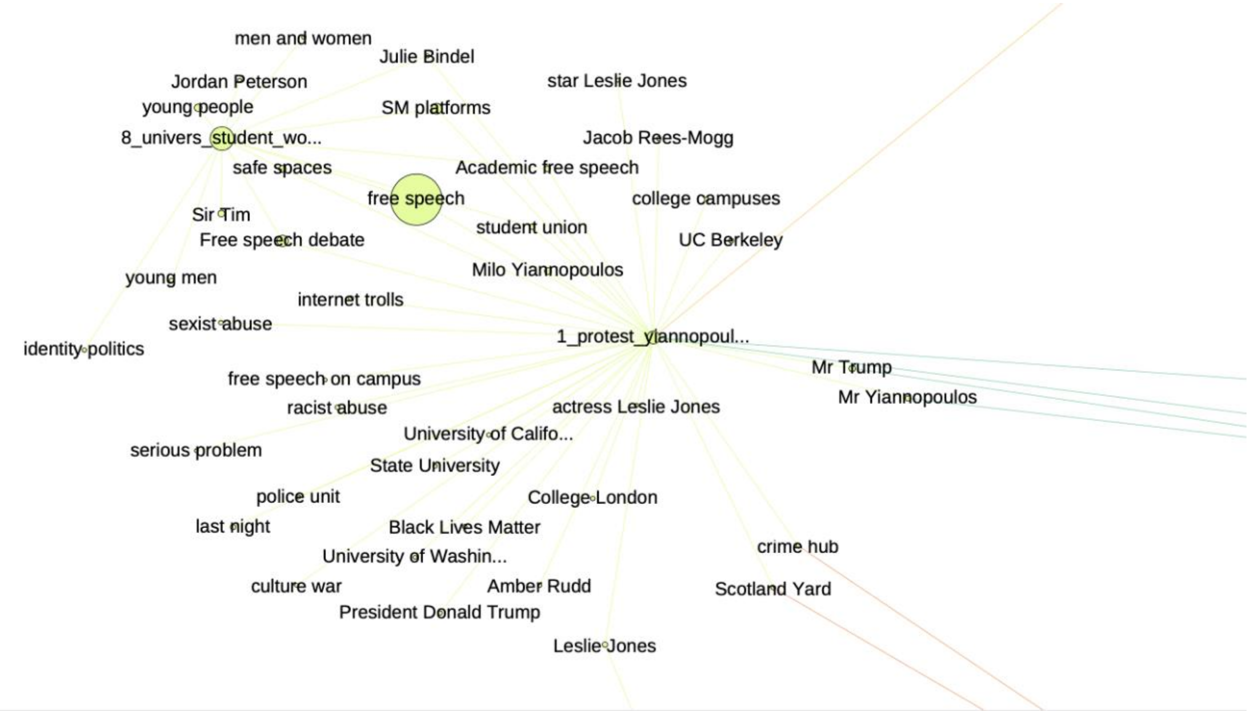


Figure 6. 9 – Semantic connection between Topics 1-8 (yellow) and most relevant words Mostly present in Independent, Mail, Telegraph

As much as newspapers tend to be in favour of less control over speech, the issue of SM censorship calls for more control over SM platforms. The same narrative expands also to the role of SM, and sees SM as an accomplice in the no-platform, or 'cancel culture' against right-wing ideas. Newspapers have reported several conservative and right-wing representatives' complaints against a so-called anti-conservative bias of the main SM platforms (McGoogan and Murgia, The Telegraph 2016-05-09). The question was also raised in the case of the public hearing after Cambridge Analytica; during the Senate hearing of Facebook a conservative Senator asked why Facebook was censoring conservative bloggers. Zuckerberg explained that his team had made 'an enforcement error.' 'And we've already gotten in touch with them to reverse it,' he said, referring to the conservative blogger Diamond and Silk. (Schwabb, 2018, document: Mail 2018-04-11).

> The revelation that Facebook has been censoring news stories they don't agree with is seriously worrying. A seventh of the world's population use Facebook, which means over a billion people are exposed to the sort of Lefty liberalism the social network seems to like. [...] Of course the left will be ecstatic, lauding Facebook for weeding out all the bigots and

> hatemongers, from moderate right-wingers to neo-Nazis. [...] (Leigh-Howarth, 2016, document: The Independent 2016-05-14).

However, The Guardian stresses how the increase in censorship of right-wing posts coincided with the introduction of anti-hate speech standards in SM platforms:

> Facebook, Twitter and Google have recently rejigged their algorithms or tightened their rules in order to address a perceived extremism problem. But this has led to claims that the companies have skewed their platforms to penalise conservative speech. It's also led some on the conservative and radical right to pursue legal action against Twitter, YouTube and others on free speech grounds (Wilson, 2018, document: The Guardian 2018-03-21).

6.3.4 Narratives about governance

The narrative emerging from the articles outlines the introduction of regulatory initiatives. Some have been mentioned in the exemplary cases: the 2018 Honest Advertising Act Initiative, which became two bills at the level of Congress and Senate in the US, the 2016 Reclaim the Internet in the UK, the 2017 NetzDG in Germany. The articles also discuss SM initiatives and highlight an opposition between the request of the governments going in the direction of more control and SM strategies. In particular, the role of SM appears to be at stake, where governments have been introducing legislation pushing for the use of automatic detection of contents, while SM have started introducing tools in the hands of the users. The development of the narrative shows how the discussion has progressively moved towards the adoption of algorithms as the main regulation tool.

Public bodies

From the articles it appears that the general widespread position of public bodies is to increase regulation and get social media platforms to take more responsibility. The representatives of public bodies, and regulatory initiatives, connect episodes of hate speech, abuse, radicalisation and extremism on SM to social unrest, terrorist attacks, and violent acts which hit citizens, after which SM platforms are called on to assume responsibility by targeting SM ‘safe harbour’ provided by art.230 of the CDA.

In the US, a number of initiatives aimed at targeting internet companies emerged from the data: Senator Claire McCaskill's bill Stop Enabling Sex Traffickers Act (SESTA) and Allow States and Victims to Fight Online Sex Trafficking Act (FOSTA) (passed in 2018) introduced liability for internet companies that facilitate sex trafficking, suspending for this specific case the immunity provided by section 230 of the CDA.

From the data the Honest Advertising Act also appears, aimed at enforcing disclosure provisions on internet advertisements and barring foreign nationals from purchasing political advertisements on the internet. By extending existing laws for television and radio stations to internet commercials it aimed at closing the loopholes in campaign finance legislation. The regulation initiatives, together with the summoning of the main companies' CEOs (Mark Zuckerberg, Sundar Pichai, Jack Dorsey) to hearings of the House of Representatives, show a change in the direction of the US approach towards SM companies.

The articles cover several regulatory initiatives taken in the UK: the Investigatory Powers Act of 2016, mentioned in the storyline and exemplary cases of Charlie Hebdo and Anjem Choudary, which expanded the electronic surveillance powers of the British intelligence services and police. Another story is the 'Reclaim The Net' cross-party programme, with its vow to banish misogyny from social media (Glaze, 2016, document: Mirror 2016-07-18).

Alison Saunders, the head of the Crown Prosecution Service "said bullies who incite others by creating derogatory 'hashtags' or by republishing 'grossly offensive messages' will be targeted." (Martin, 2016, document: Mail 2016-10-10).

In March 2017 the Home Affair Committee chaired by Yvette Cooper interrogated Peter Barron, Vice President, Communications and Public Affairs, Google Europe, the Middle East and Africa, Simon Milner, Policy Director for the UK, Middle East and Africa, Facebook, and Nick Pickles, Senior Public Policy Manager for UK and Israel, Twitter on the topic of hate speech.

The establishment of the Global Internet Forum to counter Terrorism in March 2017 (after the Westminster attack) increased the use of automation and machine learning for the purpose of counter-terrorism. On the topic of counter-terrorism, in July 2017, Theresa May demanded that tech firms take down terrorist material within 2 hours (BBC 2017-09-20).

In October 2017, Theresa May tasked the Committee on Standards in Public Life, chaired by Lord Bew, with investigating what could be done to better protect MPs (Telegraph 2017-10-01).
In April 2017, as result of the Home Affairs Committee investigations, the UK Parliament published 'Hate crime: Abuse and extremism online on Social Media', in which MPs strongly criticised social media companies for failing to take down and take sufficiently seriously illegal content. In October 2017 came the publication of the green paper on 'Internet Safety', and to the UK government response to the green paper was published in May 2018.

After the data collection stopped, however, more regulation initiatives have followed: in April 2019, the UK adopted a 'Code of practice for providers of online social media platforms', as part of the Digital Economy Act (2017). In December 2020 the UK government published the Online Harms White Paper, the most updated initiative in its policy on social media, and it introduces two new elements: the duty of care for the companies (as a form of responsibility different from liability) and the establishment of an independent advisor: the Center for Data Ethics and Innovation, which started to present its programme in spring 2019. The centre was created because technology is seen as 'part of the solution', and it recognises that:

> the increased use of data and AI is giving rise to complex, fast-moving and far-reaching ethical and economic issues that cannot be addressed by data protection laws alone. Increasingly sophisticated algorithms can glean powerful insights, which can be deployed in ways that influence the decisions we make and the services we receive. It is essential that we understand, and respond to, barriers to the ethical deployment of AI (Department for Culture Media and Sport and Home Office, 2019).

The press is divided in terms of commentary on these initiatives shifting more responsibility on to SM. The Observer defines section 230 of the 1996 Communications Decency Act as 'vital' and feels that restricting its protection risks putting too much decision-making power on speech in the hands of corporations:

> So now we find ourselves in a strange place where huge corporations are in a position to determine what is published and what is not. In a working democracy, this kind of decision should be the prerogative of the courts. It's as if society has outsourced a critical public

> responsibility to a pair of secretive, privately owned outfits (Naughton, 2017, document: The Observer 2017-09-03).

While the Daily Mail presented a different position in 2018:

> There must be regulation of social media, and every government in the world ought to address itself on how best this can be implemented, without, of course, imposing improper restrictions on free speech. It must be the beginning of wisdom that we understand how wildly excessive and deeply dangerous are the powers of the social media giants, headed by Facebook. They cannot be uninvented, but they must be tamed. Should we fail to do this, these wild beasts will devour our democracies and our individual freedoms (Hastings, 2018a, document: Mail 2018-01-02).

With regard to regulatory tools adopted by the companies, newspapers highlight some problems with SM governance, especially with the use of algorithms and policies for content moderators. In recent years SM platforms have introduced several changes to their internal policies and tools, which are mentioned in the articles. Starting from implementing more opportunities for users to report activity, they ended up employing increasingly more algorithmic detection. Exemplary in this case are the changes adopted in Twitter. They started from tools addressing the users reporting activity, for instance the changes in hateful conduct on Twitter, and then introduced more tools for users to report abusive accounts (from November 2016), and to filter the materials visible, such as the Safe Search, or the option to mute more accounts (Twitter 2016). However, in 2017 the company started to rely more on algorithms for the identification of potentially abusive accounts (Daily mail 2017, document: Mail 2017-03-01).

NGOs and academia

NGOs and academia are associated with sceptical views of regulation of speech and in general on the role assigned to SM platforms. Academia is the one actor that speaks of editorial responsibility from their own pages and on newspapers.

> Jeff Jarvis, journalism professor at the City University of New York, wrote on Friday. [...] Facebook should allow editors of reputable news organizations to make key decisions related to how they use the platform – such as publishing a war photo that may technically violate

> policy.[...]To some in media, a key first step to making Facebook a respectable news publisher is increased transparency. Aside from leaked documents, the public knows little about how Facebook's algorithms work and what role employees play (Levin, 2016, document: The Guardian 2016-09-10).

NGOs and advocacy groups are the groups more concerned with the human rights implication of algorithmic tools of management of content:

> Human rights abuses might be embedded in the business model that has evolved for social media companies in their second decade. Essentially, those models are based on the collection and use for marketing purposes of their users' data. And the data they have is extraordinary in its profiling capacities, and in the consequent unprecedented knowledge base and potential power it grants to these private actors. Indirect political influence is commonly exercised, even in the most credible democracies, by private bodies such as major corporations. [...](Joseph, 2018, document: The Independent 2018-04-03).

Newspapers play a pivot role as they contribute to introduce new elements in the debate. In 2017, The Guardian leaked Facebook's guidelines for moderators (Hopkins, 2017, document: The Guardian 2017-05-21), making public for the first time the training material for human moderators employed in the company. The guidelines appeared to many as unfit, as they leave only 10 seconds to take a decision.

> Thousands of pages of internal documents from Facebook have been leaked, revealing the rules and regulations the social media giant uses to decide what can be shared on its platform. Among the rules detailed in documents obtained by the Guardian are those covering nudity, violence and threats – all things that Facebook has been accused of letting slide in the past. (Shugerman, 2017, document: The Independent 2017-05-21).

> What we've learned from the Guardian 's scoop is that Facebook's baroque, unworkable, ad hoc content-moderation system is unfit for purpose (Noughton 2017b, document: The Observer 2017-05-28).

> A similar leak happened also in June 2017, when the no-profit newsroom ProPublica published the internal documents describing how moderators train the algorithms used by Facebook's censors to distinguish between hate speech and lawful political discourse. The leaks show how algorithms were used to recognise protected categories such as race, sex, gender identity, religious affiliation, national origin, ethnicity, sexual orientation, and disability or disease (White, 2017, document: The Independent 2017-09-15).

The leak also exposed another incongruence in Facebook hate speech policy, where on the one hand the company targets hate speech but on the other it was found that it sold advertisements targeted to antisemitic users (White, 2017, document: The Independent 2017-09-15). NGOs are concerned over Human Rights and social justice implications of technology, while newspapers report the concern over the racial bias:

> ‘Activists in the Movement for Black Lives have routinely reported the takedown of images discussing racism and during protests, with the justification that it violates Facebook's Community Standards,’ reads the letter to Facebook, signed by groups including the American Civil Liberties Union and SumOfUs, a corporate watchdog. ‘At the same time, harassment and threats directed at activists based on their race, religion, and sexual orientation is thriving on Facebook. Many of these activists have reported such harassment and threats by users and pages on Facebook only to be told that they don't violate Facebook's Community Standards’ (Titcomb, 2017, document: The Telegraph 2017-01-19).

Private companies

SM and NGOs have a point of contact in the preference for forms of self-regulation as counter-speech, rather than other forms of regulation of speech. In the general discussions about governance, it is possible to observe a switch from SM companies initial focus on users’ possibilities to counter-speech, or report, to a later focus on the algorithms. Newspapers report a number of initiatives taken with NGOs on the topic of counter-speech, as the report commissioned from Facebook to the think tank Demos (which appeared also in the web pages) and cooperation with NGOs such as Faith Matters (Hinsliff, 2016, document: The Guardian 2016-02-22). However, public bodies (governments, parliaments, law enforcement bodies) requests have gone towards a different direction. As far as SM, web pages did not capture many statements. Newspapers collect more of their narratives, and it is possible to identify a trend based on ‘apologetic and ‘passive’ approach. This is in line with SM refusal for being considered responsible for the content on their platforms or in general to appear as the main regulators. Also, it shows a tendency to create policies as reaction to events.

> [on the censorship of the Vietnamese girl picture] While we recognise that this photo is iconic, it's difficult to create a distinction between allowing a photograph of a nude child in one instance and not others,' a company spokesperson wrote. 'We try to find the right balance between enabling people to express themselves while maintaining a safe and respectful experience for our global community. Our solutions won't always be perfect, but we will continue to try to improve our policies and the ways in which we apply them' (Holmes, 2016, document: Mail 2016-09-09).

> Ed Ho (Twitter) in 2017 We’re learning a lot as we continue our work to make Twitter safer – not just from the changes we ship but also from the mistakes we make, and of course, from feedback you share (Daily Mail 2017, document: Mail 2017-03-01).

> We didn't take a broad enough view of our responsibility, and that was a big mistake, he said. It was my mistake, and I'm sorry. I started Facebook, I run it, and I'm responsible for what happens here.[...] (Daily Mail, 2018, document: Mail, 2018-04-10).

In the general discussions about governance, it is possible to observe a switch from SM companies’ initial focus on users’ possibilities to counter-speech or report (a preference shared with NGOs) to a later focus on the algorithms (a preference shared with public bodies).

Facebook in 2015 stated ‘By working with community groups like Faith Matters, we aim to show people the power of counter speech and, in doing so, strike the right balance between giving people the freedom to express themselves and maintaining a safe and trusted environment’ (Mail 2015-01-03). Newspapers highlight also mixed-actor collaborations, such as the counter-speech report commissioned in 2016 by Facebook to the think tank Demos (Demos 2016) (Hinsliff, 2016, document: The Guardian 2016-02-22).

6.3.5 Narratives about technology

In the analysis of texts, algorithms and filters emerge as an object ‘bending’ the environment, i.e. breaking the routine as in the case of the viral visibility of abusive tweets, the political exploitation of algorithms to distribute fake news, and at the same time the use of algorithm filters as a tool for censorship. In particular, the role of algorithms appears to be at stake, where governments have been introducing legislation pushing for the use of automatic detection of contents, while SM platforms have started introducing tools in the hands of the users. The development of the narrative shows how the discussion has progressively moved towards the adoption of algorithms as the main regulation tool. Private companies present a narrative about technology which is ‘cautious’ and stresses the limits of technology:

> Nick Pickles, Twitter's head of policy in the UK, says creating a technological solution to the problem of online abuse is extremely difficult. ‘No such magical algorithm exists and, if it did, it wouldn't be that simple to implement because of the complexity of understanding sentiment and context.’ [...] Tech companies cannot simply delete misogyny from society,’ he said (Leigh, 2016, document: The Guardian 2016-04-13).
>
> In an appearance in front of two US senate committees on Tuesday, Zuckerberg said removing hateful content from the site was difficult and beyond the capacity of artificial intelligence. He said that by the end of 2018 the company would have 20,000 employees devoted to security and reviewing content (Wong, 2018, document: The Guardian 2018-04-11).

However, newspapers have reproduced the public bodies’ narrative that companies should do more, and in particular, they should adopt technological solutions:

> They are not doing enough, however. All social networks should be prioritising the development of technology to identify and automatically delete this content. That is administratively and technologically fiendish, but the sites' business models rely precisely on the ability to search and analyse huge volumes of data to target advertisements effectively. If they are going to profit from this smart technology, they must use it to protect their users in accordance with the recommendations of the police and security services (The Times, 2016, document: The Times 2016-08-18).

> [A]rtificial intelligence and image recognition software should be used to scan for and track illegal material, reducing the burden on humans who must manually flag and monitor it. (McGoogan, 2017, document: The Telegraph 2017-03-24).

Only a few articles raise the issues of the possible role of algorithms and technology in creating extremist speech.

> Peterson is not an alt-right figure and cannot be held responsible for the 'recommended' content that his viewers come across on YouTube. But YouTube, and its parent company, Google, should be. YouTube algorithms have been criticised for drawing viewers into ever more extreme content, recommending a succession of videos that can quickly take them into dark corners of the internet (Levin, 2017, document: The Guardian 2017-08-13).

On the other hand, the Guardian in 2016 highlighted the narrative also used by public bodies (seen in chapter 5), that technology could be used better:

> [The systems engineer Randi Lee Harper] authored an auto-blocker for Twitter that lets users automatically, pre-emptively block users likely to be associated with harassment groups – people who follow more than one high-profile harasser, or who frequently use hashtags associated with abuse movements.[...] Harper suggests solutions for fixing some of Twitter's most obvious vulnerabilities – fixes she says would require very few engineering hours on the inside but that would offer significant and immediate benefits to users in terms of security, privacy and usability.[...] (Leigh, 2016, document: The Guardian 2016-04-13).

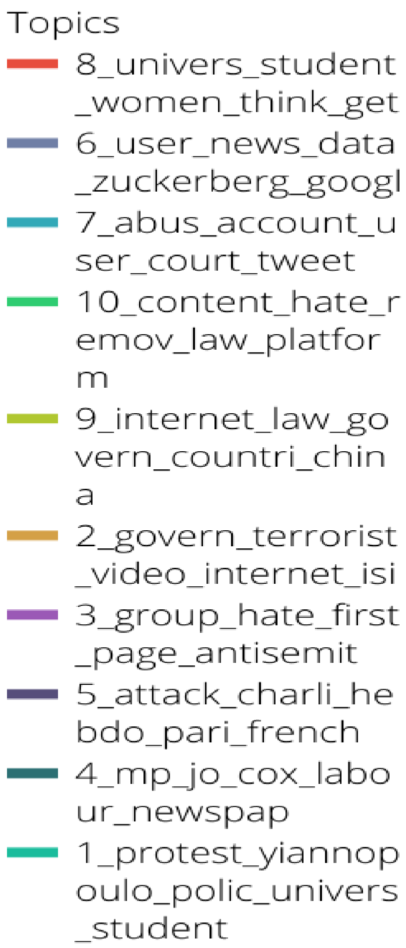
> Jennifer Parry, CEO of the Digital Trust, which acts on behalf of cyber abuse victims, said Twitter was 'really the network of the trolls'. [Parry says] They use their algorithms for their targeted marketing, I would like to see them use some of this technology to identify and deal with abusers' (Leigh, 2016b, document: The Guardian 2016-07-20).

As in the finding from the previous chapter, technology emerges as an ambiguous element. Newspapers do not present a unified vision. Differently from web pages, what emerges are cautious declarations from SM companies.

6.3.6 The temporal dimension

The historical analysis of topics extracted from newspapers shows how the topic identified changes over time and, even with some limitations, can give an idea of the life and death of specific topics. Using the Cortext 'demographic' analysis tool (described in chapter 4), I calculated how the distribution of topics develops through time (time granularity: months. From January 2015 to April 2018). Figure 6.10 represents the temporal evolution of the documents and topics. A colour has been assigned to each of the 10 topics identified with quantitative analysis.

Fig. 6.10a shows the colour associated to each of the topics emerged from the topic analysis:



Topic 8 – University Student Union – No Platform – Jordan Peterson – Milo Yiannopoulos_women
Topic 6 – Facebook – Fake news and privacy
Topic 7 – Twitter abuse/ harassment – trolls and blocks
10_content removal_hate speech, Germany
9_China, Turkey Internet regulation and censorship
Topic 2 – Government – terrorism and encryption
Topic 3 – Hate groups – Reddit – alt-right – Trump – Charlottesville
Topic 5 – Charlie Hebdo
Topic 4 – Jo Cox, abuse online, hate speech
Topic 1 – Milo Yiannopoulos – Twitter_-Leslie Jones

Figure 6. 10a Temporal evolution of the topics: colours associated with topics

The position on the y-axis on the left represents the raw number of documents attached to one of the specific topics (the higher the number, the more documents are attached to that topic, in that specific month). The position on the x-axis represents the time, starting from January 2015 (left) and moving towards April 2018 (right). For example, in 2015, the first line from above (i.e. the one appearing on more documents) corresponds to topic No. 5, labelled as 'Charlie Hebdo', and coloured in purple. [The visualisation of the historical distribution in fig.6.10, and the break down by single topic is available in the Appendix, from p. 43 A to p.54 A].

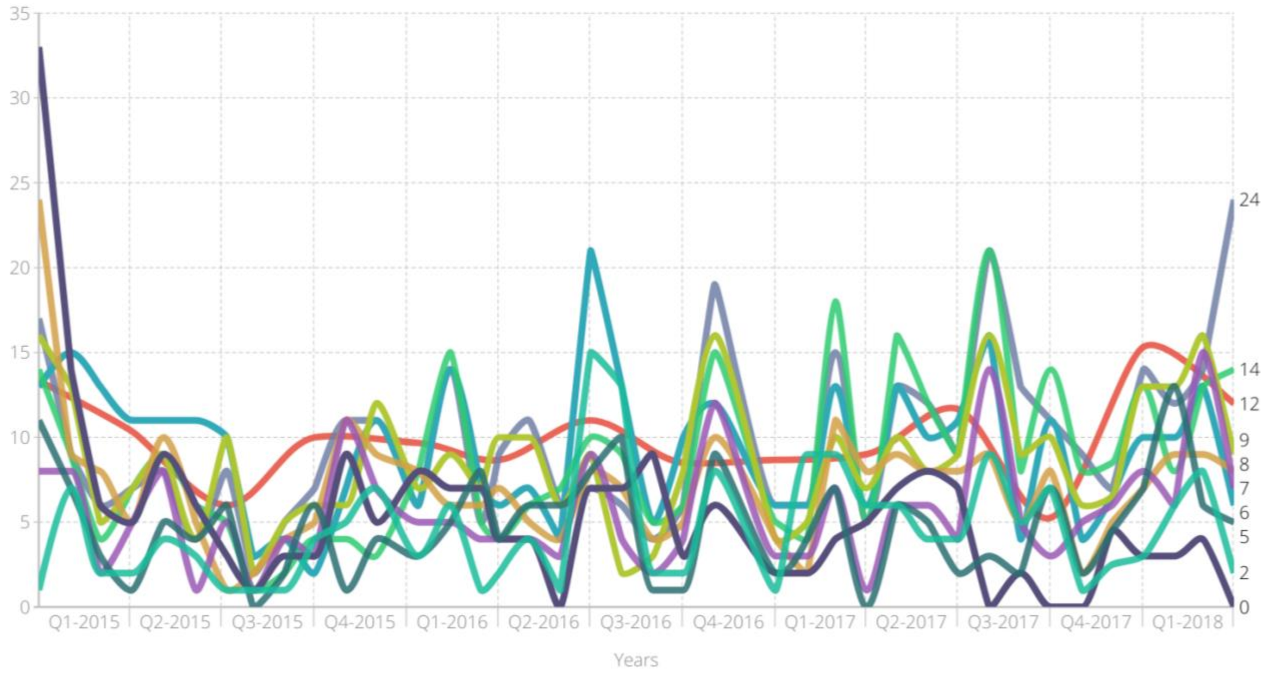


Figure 6. 11b Temporal evolution of the topics[16]*.*

On the x axis is the time, on the y axis is the raw frequency of the documents attached to the topics

A larger version of this image is available in the Appendix from p. 43 A to p.54 A. It is possible to see how in January 2015, the number of documents attached to the topic Charlie Hebdo (topic 5, colour dark blue) is the highest, compared to the other topics (i.e. the highest line on the left of the graph). Similarly, the number of documents attached to topic 2 (i.e. terrorism and national security, in yellow) also peak in January 2015, and then gradually decline to remain a steady

[16] A larger version of the image, together with the visualisation of topics' temporal plots separate and the table with the raw frequency of documents/months are available in the Appendix – 43 A – 54 A

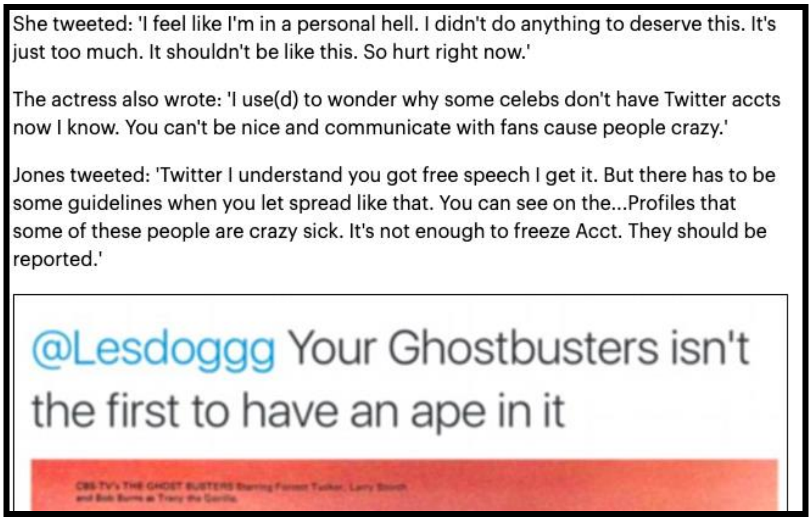
background presence in the subsequent years. These peaks in number of documents show how in January 2015 and the subsequent months, it was easy to find articles (i.e. documents) including references to free speech, encryption, and SM platforms responsibility in the fight of terrorism and national security. This can indicate that terrorism and national security were ‘a matter’ of concern in those months. The temporal decline in the number of documents attached, on the other hand, can indicate how these narratives have been overcame by other matters of concern. In 2016 for instance, the episode in the peaks relates to Yiannopoulos abuse on Twitter and the no-platform debates across Universities in the US and the UK (i.e. topic 1 and 7 in light blue and green). In 2017, the peak is topic 10 (light green) and the associated issue of SM platforms’ policies on hate speech after Charlottesville (US). Finally, in 2018, the topic which appear in more documents is topic 6 (in violet) which represent the issue of data protection and manipulation of users. These last two topics in particular (topic 6 and 10) start with a very low presence across documents at the beginning, but they end up being the most present in the documents at the end of the data collection period.

Other topics are historically examples of more stable matter of concern for the public. For instance, topics relating to misogyny and women abuse (Figure 6.10 topic 8 (red), topic 4 (dark green)), as well as topic 3 (hate speech, in violet) have a consistent distribution across documents in time, in comparison to the others. Overall, the data show a switch in the narratives occupying the dominant positioning the debate, from terrorism and national security to the emergency of hate speech and, in very last position, the issue of data and users’ manipulation. These dynamics indicate that controversies have a life, they can start and peak and then fade away. In general, it shows how the elements that are considered matter of concern are subject to change.

6.3.7 The shape of the controversy

The analysis of newspaper articles’ reveals how ‘substantive’ issues (Marres, 2015) such as the discussion on what constitutes free speech today, free speech in academia, right to offend, etc., are fuelled by or anchored to media-technological dynamics. Analysing the structure of the articles it was possible to notice the constant presence of links to SM platforms’ posts or tweets representative of controversial cases that originated on SM platforms. Newspapers appear to rely on such materials to discuss the larger substantial elements of the controversy around free speech however, they fail to contextualise the cases, by discussing for instance the media-technological environment at the origin of the viral visibility of the people involved. In doing so, they transpose and legitimise their presence in public controversy. In particular, in articles it is possible to note how on several occasions controversial tweets or posts originating from SM are taken and reproduced in newspapers, such as the tweets by Katie Hopkins (where for instance the Independent started to reproduce all the reactions that followed), and even more prominently in the case of Milo Yiannopoulos and Leslie Jones.

This is a specific dynamic of distribution of the controversy, where viral contents produced on SM are reproduced and reinforced in their virality by newspapers choosing to include viral abusive posts and tweets in their articles. In the controversial case created by the opposition between Leslie Jones and Milo Yiannopoulos, the majority of publications chose to reproduce Jones's tweets where she described her decision to leave Twitter as a result of the racist and misogynistic attacks she received, while others, in particular the Mail and the Independent, decided also to publish some of the racist tweets that had been addressed to the actress (Figures 6.12 and 6.13). In both cases, publication included either a direct link to the original tweet (the Independent, Fig. 6.13) or the possibility to share the controversial images via SM platforms (Daily Mail, Fig. 6.12).



She tweeted: 'I feel like I'm in a personal hell. I didn't do anything to deserve this. It's just too much. It shouldn't be like this. So hurt right now.'

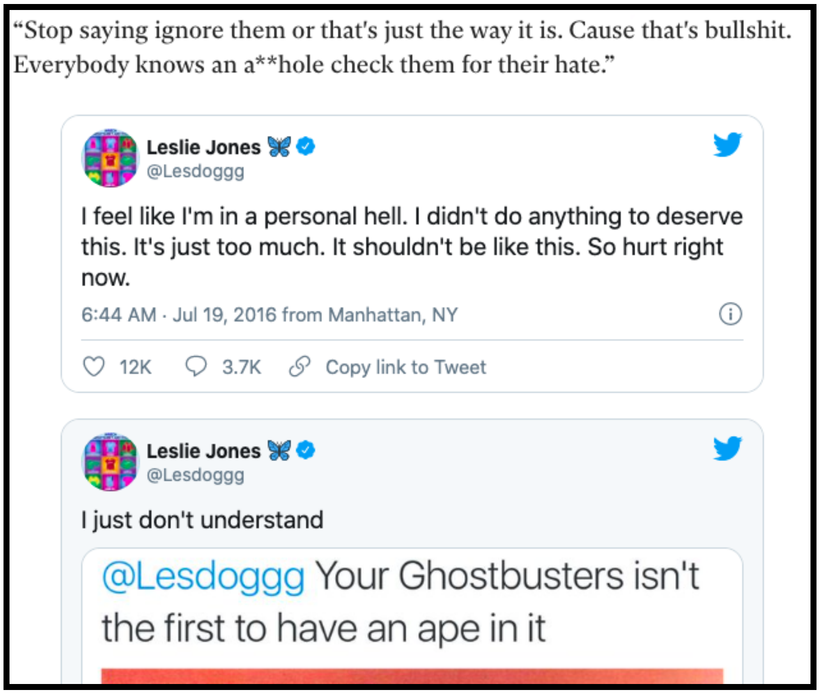
The actress also wrote: 'I use(d) to wonder why some celebs don't have Twitter accts now I know. You can't be nice and communicate with fans cause people crazy.'

Jones tweeted: 'Twitter I understand you got free speech I get it. But there has to be some guidelines when you let spread like that. You can see on the...Profiles that some of these people are crazy sick. It's not enough to freeze Acct. They should be reported.'

@Lesdoggg Your Ghostbusters isn't the first to have an ape in it

CBS-TV's THE GHOST BUSTERS Starring Forrest Tucker, Larry Storch and Bob Burns as Tracy the Gorilla.

Figure 6. 12 Racist tweet (partially represented here) included in the article in the Daily Mail (2016-07-19)



"Stop saying ignore them or that's just the way it is. Cause that's bullshit. Everybody knows an a**hole check them for their hate."

Leslie Jones 🦋
@Lesdoggg

I feel like I'm in a personal hell. I didn't do anything to deserve this. It's just too much. It shouldn't be like this. So hurt right now.

6:44 AM · Jul 19, 2016 from Manhattan, NY

♡ 12K 3.7K Copy link to Tweet

Leslie Jones 🦋
@Lesdoggg

I just don't understand

@Lesdoggg Your Ghostbusters isn't the first to have an ape in it

Figure 6. 13 Racist tweet (partially represented here) included in the article in the Independent (2016-07-19)

This finding shows a particular technological dynamic which characterises the shape of contemporary public controversies. This is a key insight that demonstrates the inter-dependencies of conventional and SM platforms and how they feed off (as well as challenge) one another.

In the previous chapter, the study of the hyperlink structure connecting the different web pages highlighted SM centrality, since all web pages included a link to the bigger SM companies (Youtube, Twitter and Facebook) as a way to share their own content to a larger public. The exemplary case of the Jones-Yiannopoulos feud shows that SM are not only one of the main channels to diffuse statements of the controversy online, but also one of the main sources of the contents performing the controversy in the press. In reproducing SM viral content, newspapers transpose and amplify SM's voice in the public controversy. Moreover, newspapers take for granted SM metrics and internal algorithmic dynamics of visibility and virality. This reflects a specific editorial choice of newspapers, which tend to prefer 'high attention grabbing' episodes even with the potential of publishing something erroneous, rather than working on the quality of content. A similar dynamic has been observed by Zubiaga et al. (2018) in the case of newspapers and diffusion of rumours, where newspapers seemed to be under pressure to publish quickly, even if the information has not been verified in advance.

6.4 Conclusion

The results of the analysis are very similar in certain aspects to the ones that emerged in chapter 5. Given the different nature of the data, it was impossible to identify actors in the same way as in the controversy on web pages; however, the analysis of the publications with the largest number of articles collected in the corpus confirms the results about news media captured in the online sphere. In particular, The Guardian (different editions), The Mail (different editions), The Independent (different editions), and the Telegraph are the most 'active or prolific' actors in the controversy.

This chapter has showed that newspapers provide two levels of findings: the first one concerns more substantive elements (Marres, 2005), such as the content/narratives about SM platforms and the issues which emerge from the main exemplary cases and storylines. As in the web pages, the storylines and exemplary cases reveal similar groups of issues (terrorist, hate speech, harassment, disinformation, fake news, data protection).

As far as the specific discussions on freedom of expression are concerned, differently from web pages, in newspapers there is a more diffuse general scepticism towards regulation of speech, especially – but not exclusively – from conservative or right-wing positions. Right-wingers or conservatives use divisive statements or actions from professional divisive celebrities such as Milo Yiannopoulos, Katie Perkins and Jordan Peterson as a way to own the issue of free speech. However, in newspaper articles which do not support conservative views, regulation of speech is also presented as negative. On the other hand, even though presenting a sceptical view on limitation of freedom of expression, newspapers in general support the regulation of SM platforms. It is a shared position from both sides, both in publications sympathetic with conservatives (Daily Mail) and more progressist views (the Guardian). Conservatives ask for more regulation since they feel that SM platforms are censoring their expression more than the other parties (since the introduction of hate speech and anti-disinformation regulations in their internal rules). In this regard, even if subsequent to the data collection, the stormy relationship between Donald Trump and Twitter of the last few years is exemplary, where Trump's tweets as President were reported and he was threatened with deactivation on several occasions on the basis of hate speech and anti-disinformation policies. The episodes culminated in May 2020 with Trump signing a Presidential executive order against 'censorship' from Twitter, which added a disclaimer on Trump's tweets against mail votes.

However, it is not only conservatives that challenge the SM role. From newspapers it emerges that members of academia are also talking about full editorial responsibility for the social media platforms. NGOs and advocacies active in the field of freedom of expression appear to be the only sceptical groups with regard to regulation of companies in connection with controlling speech, reiterating that self-regulation tools such as reporting or filtering tools, and counter-speech in the hands of users, are the best options to guarantee free speech. NGOs in particular are the ones more wary of the lack of transparency in internal SM procedures for moderators and algorithmic management of content, which when leaked showcases social injustice.

Their position, however, seems to reinforce the progressive introduction of algorithms and automatic detection systems based on technology. The regulation initiatives mentioned highlight a general agreement across the representatives of public bodies, on the necessity of regulation for platforms. The regulatory initiatives also share similar positions highlighted in the previous chapter, on the role of technology as a tool for the realisation of policies governing free speech. Newspapers report the position of private companies, stressing that SM companies should apologise, and downsise their power on technologies. On the other hand, newspapers present competing visions of experts pushing on the feasibility of technological solutions. Even if newspapers challenge the idea of regulating free speech, they do criticise SM more than states' initiatives. In this sense, they strengthen states' programmes of increasing responsibility on to SM platforms and the use of technology to 'solve' the issues. Since articles are presented in an archived form, it is possible to conduct historical analysis of the narratives. The findings show that topics and consequently their associated narratives follow a specific development across time. In particular, in 2015 the topic of terrorism and national security is the initial and more urgent issue at the beginning of the data collection, but it rapidly gives space to two main issues: hate speech and data exploitation and manipulation by users (fake news) in 2018.

The findings from the historical overview show that in a similar way as with statements from websites, discussions about placing increasing responsibility on SM platforms are not accompanied by discussions on the implications of the use of technology. These latter findings started to emerge at the very end of the data collection.

The second level of finding present in newspapers concerns the technological dynamics. As underlined by Marres (2015), in controversy mapping it is important to study newspapers as actors in the overall controversy. In this controversy, newspapers play a fundamental role with potential to mobilise the public discourse around the controversies arising from the competing translation processes. In these articles, newspapers have linked events to the storylines about SM and freedom of expression, and made of some episodes exemplary cases that can summarise the main issues (e.g. Choudary= terrorism and radicalisation, Yvette Cooper and Jo Cox = hate speech). In this way they act as 'tools of measurement' – a role that I discussed in chapter 3, and that indicates how newspapers are fundamental to diffuse the narratives about technology and free speech for the public of non-experts. At the same time, findings highlight a potentially problematic issue, since the process of creating 'exemplary cases' is largely based on episodes that start on social media. Newspapers use controversial cases that are developed on social media and they make them exemplary for the discussion of freedom of expression. However, they do not include in the presentation of narratives a critical appraisal of the media-technological environment where the exemplary cases originate. For instance, because of the high visibility of the people involved (e.g. Milo Yiannopoulos and Leslie Jones) when discussing viral materials they rarely interrogate the role of technology (and platforms) in making certain aspects more visible than others. In doing so, they transpose and legitimise that type of speech.

As seen in the case of Leslie Jones, newspapers tend to report divisive content taken from SM, without discussing the nature of the virality of the post or tweet (as in the tweet from Katie Hopkins, or other cases). The risk is that they are reinforcing the same viral content that they are accusing the platforms of reinforcing. However, this is understandable since the majority of the publications also have an online presence, and are subject to the 'attention economy' and the same advertisement system as the SM companies use. In this way, newspapers might indicate a technological 'takeover' of the process of issue formation. This is a key insight that demonstrates the inter-dependencies of conventional and SM platforms and how they feed off (as well as challenge) one another. The implication is that, even though social media metrics do not define the issue, de facto our public discourse is influenced by technological metrics, in the sense that highly visible cases (viral contents, followers) have the power of stimulating public discourse and public interventions. It is through newspapers that these metrics, indicators of media and technological dynamics enter the 'official' public discussion.

In the next chapter (chapter 7), I discuss these findings and those of the previous chapter (chapter 5) in the light of the concepts developed by ANT and critical data studies.

7. Discussion

7.1 Introduction

In the previous chapters, I presented the empirical findings from mapping the controversy surrounding freedom of speech on social media. As discussed in the literature review (chapter 2), the governance of freedom of speech is theorised as performative and emerging from episodes of shocks or emblematic issues which break the routine of established regimes, creating the opportunity to introduce change (Pohle, 2016a, 2016b; Hofmann, 2016; Hofmann et al., 2016; Ananny and Gillespie, 2016). This perspective also stresses how governance is performed by a plurality of actors that cross-cut public/private, local/global and formal/informal dichotomies (Wagner, 2013; Balkin, 2016; Gorwa, 2019a, 2019b, 2020).

Using the methodological tools of controversy mapping on websites and newspaper articles, I have identified actors, dynamics and roles. In this chapter, I use the ANT theoretical framework to show how different controversies and their public can impact governance.

In particular, I interpret the findings from the previous chapters using ANT concepts and terminology to identify discursive and material strategies used by actors to influence governance, persuading, or ‘forcing’ other actors to work towards the same objectives and priorities and share the same vision of the governance of freedom of speech (§7.5).

In particular, I show how studying the different stages in the controversy can inform our understanding of norms creation in the complex governance ecosystem. The textual analysis (in chapters 5 and 6) shows how actors' normative interpretations of governance are developed to answer the challenges created by the 'shocks'. By presenting different narratives and storylines about the exemplary cases, actors are de facto creating roles and responsibilities within the socio-technical system, similarly to what Radu et al. (2021) describe as 'norm entrepreneurs'. The different positions in the controversy and the power dynamics studied with the lenses of sociology of translation can inform our understanding of how binding and non-binding instruments for speech regulation have come to be in the last years.

The findings of the analysis can be divided into four layers:

1) In the first layer, I discuss findings related to the morphology of the controversy (§7.2-7.5). I start by discussing the range of actors involved. I then describe the narratives and the dynamics connecting actors, shocks and regulatory initiatives. Finally, I assign ANT roles to actors and observe translation processes in their *problematisation, interessement, enrolment* and *mobilisation* phases (Callon 1986a).

2) The second layer connects the discussion on the different translation processes to a macro perspective. Using the critical data studies approach, I interpret the translation processes from the point of view of broader processes of datafication and platform capitalism (Srnicek 2017), stressing how the lack of problematization of technology and the ambiguous regulatory role assigned to algorithms and AI has important implications for governance.

3) The fourth layer presents the discussion relative to content transmission dynamics and controversies activation. Here I discuss the findings on the technological dynamics holding the controversy and the role of media as 'publicity' devices, privileging certain discourses above others. The economic system based on platform system is visible in the structure of websites (ads, links to SM) and the materials reproduced on newspapers virality. Here I connect the discussion on the importance of narratives in the regulation of media and technology, stressing at the same time how media capacity to produce information and public discourse is affected by the ideology of 'corporate libertarianism' (Pickard 2016).

4) The last part of the discussion is a meta-reflection on the role of researchers and the methodology used. It includes consideration of controversy mapping and the role of the researcher. Here I argue how researchers have to contribute to opening the black-box of technology to expose the social implication of governance choices in algorithmic management.

In the following paragraphs, I describe more in detail these findings, starting from the description of actors and their roles in the controversy.

7.2 Morphology of the controversy

As seen in chapter 5 and 6, the actors that published web pages and statements on the issue of freedom of expression can be summarised into group-like formations (i.e. 'news media', 'public-bodies', 'private companies', 'civil society: academia and think tank', 'civil society: NGOs and advocacy groups' and 'technological objects'). In the next paragraphs, I am going to discuss the results in the light of the literature on online governance of speech presented in chapter 2.

7.2.1 Actors' composition

The findings confirm that governance of free speech is the result of the interaction between a plurality of groups of actors. The statements collected describe the action of national governments, private companies and civil society organisations and individuals– actors that also emerged in the literature review as key in the governance of freedom of expression (Wagner, 2016; Gorwa, 2019a, 2019b, 2020). The analysis of hyperlinks in chapter 5 §5.5 shows how public bodies at the national level (i.e. the UK government), European and international organisations (i.e. European Union and Council of Europe) together with SM platform companies (i.e. Facebook, Youtube, Instagram, Twitter, Google) occupy a very central role in the network, with the highest in-degree values (see fig. 5.3-5.4-5.5-5.6). In the analysis, I have contrasted this high centrality value with the very low frequency of URLs captured for each group. The description of the different types of documents in the overview of the URLs (§5.3.2) show, however, that the web pages associated with public bodies, international organisations and SM companies include examples of legislation, guidelines, policy papers etc. (e.g. Crown Prosecution Service Guidelines, the NetzDG): essential documents that have the 'power' to attribute roles and responsibilities to other actors in the socio-technical system.

Social media companies

The study of actors creates more knowledge on the role of SM platforms, both as a fundamental infrastructure of communication and as a regulatory counterpart of states. In the literature, SM companies are one of the leading actors in the governance ecosystem, as they are the owners of the technological infrastructure with enforcement and regulatory powers. At the same time, they are also an essential tool for every actor in the system to communicate their views (every blog, web page from public institutions or newspaper has links to their official SM pages). However, when considered as participants in the controversies, it is striking that these large companies are almost non-existent as 'authors' of statements on web pages (see § 5.5). Nevertheless, the data show a progressive institutionalisation of SM companies in the governance discourse, capturing the statements produced by the companies representatives in interviews or as respondents to enquiries from public bodies. Furthermore, the data show that articles included SM voices as institutionalised counterparts to policymakers in the discussions of exemplary cases, such as hate speech or terrorism, at the same level as representatives of national governments (see The Guardian and the Mail reporting Baroness Shield speech in the aftermath of terrorist attacks, and Mark Zuckerberg replies in the aftermath of Cambridge Analytica in chapter 6, §6.32). The semantical analysis in chapter 6 and in particular, fig.6.2 and 6.5 display this relationship, showing how SM companies (and Facebook and Mark Zuckerberg) stand out among the most important terms in the topics associated with the exemplary episodes.

Civil society

The overview of actors shows a high presence of members of civil society, in the form of academic institutions and research centres or organisations such as NGOs and advocacy groups §5.3.2. Gorwa (2019a, 2019b, 2020) observed that civil society's presence in the governance of speech does not necessarily translate into regulatory powers, concluding that civil society's role is mainly the watchdog rather than a regulator. In this study, I partially disagree, since what emerges is that: firstly, civil society's involvement in the formulation of governance is more prominent than traditionally considered (for instance, it is not just restricted to the IGF institutions). As highlighted by other authors (van Eeten and Mueller, 2012; Radu, 2019), the places and means of online governance span beyond the 'big' institutions and comprises a large part of discursive norm creation. In the literature, I have discussed how online governance can be interpreted as a *normfare* (Radu et al., 2021), where norms are continuously developed from more traditional and binding legislation to non-binding instruments, such as statements and declarations (Radu, 2019). Radu et al. (2021) argue that in a context such as online governance, normative powers cannot be measured against the power of states but rather within the instituted by practices that make up governance online. In this perspective, it is clear that civil society influence on the governance of speech online can appear through different means and that civil society has a role as that of norm entrepreneurs (Hintz, 2016, Radu et al. 2021). From more 'institutionalised' actions, closer to the traditional milieu of policymaking (as in the case of IGF or other multi-stakeholder organisations) to less institutionalised environments, as in the case of from tech-activism and advocacy run on alternative infrastructures (Milan and Hintz, 2013; Hintz, 2016).

The data in my sample confirm a high presence of traditional policy actions (for instance, all the regulatory changes adopted by states in the field of hate speech and terrorism, one example for all, the NetzDg). These initiatives involve mainly SM and states (as in the case of the anti-terrorism Forum). However, the also data show that civil society has been invited to collaborate on more arenas than the ones belonging to traditional IGF bodies. For instance, the data captured a 'bilateral' collaboration with governments (as in the feedback for the revision of the Crown Prosecution Guidelines, see fig.6.4, or the London Hub for Hate speech), with international organisations (as in the contributions from NGOs in response to the call from the UN Special Rapporteur, see network analysis in fig.5.15) and with private companies (as in the Facebook counter-speech report commissioned to Demos, see §5.12) where civil society organisations and advocacy groups have provided inputs on policies concerning free speech. According to Radu (2019), these are modelling actions where organised civil society can influence policy by presenting their recommendation and positions.

The study of the position of actors on the different issues suggests that normative positions from civil society (i.e. concern about the societal impact of technology) might have gained policy space from informal regulatory initiatives. Thus, for example, the UK White Paper on online harms shows that public bodies have increasingly adopted perspectives on the ethical use of A.I., which historically (i.e. during the previous years) were typical of NGOs and academia.

Another aspect of the findings that has emerged using this methodology concerns the role of 'non-organised civil society actors, such as individuals, bloggers, or controversial 'celebrities'. Their statements align and reproduce a specific normative account of 'what should be the good regulation', which previous literature does not address. Their action on governance is visible in combination with the media technology infrastructure (particularly SM platform technology) as they can influence policy discourse by creating exemplary cases and storylines. Milo Yiannopoulos and Katie Hopkins are examples of individuals who, combined with social media's network effect, have 'shocked' the public discourse and initiated a regulatory process (i.e. Milo Yiannopoulos was expelled from Twitter and became an exemplary case in the regulatory narrative about harassment and free speech).

Technological objects

As far as technological objects are concerned, the findings are twofold. On the one hand, the overview of actors and the hyperlinks connecting web pages in the empirical chapter 5 has shown how fundamental the technological objects created by SM platforms' 'posts, tweets, pages' have become for public expression online. The analysis of hyperlinks already discussed for SM shows a high presence of SM posts and pages (see §5.6.1). SM central position in the network depends on the fact that posts and pages are shared and published by all the other actors in the socio-technical system (see §5.6.1, and fig.5.5, fig.5.13). On the other hand, in the analysis of texts, technological objects such as bots (fig.5.12) algorithms and filters emerge as objects 'bending' the environment, i.e. breaking the routine as in the case of the viral visibility of abusive tweets, the political exploitation of algorithms to distribute fake news and at the same time the use of algorithm filters as a tool for censorship. The data also show that the narrative associated with technology is very ambiguous. On the one hand, it is seen as an issue, on the other as the solution. For instance, algorithmic content moderation is the solution for hate and abusive speech in the public discussion about governance (§6.3.5).

The analysis of the statements, and in particular the analysis of exemplary cases and storylines both from web pages and newspapers, confirms the insight presented in the literature that governance of free speech happens as a reaction to public shocks (Hofmann et al., 2016; Ananny and Gillespie, 2016) rather than by planned organisation. In these shocks, technology and individual users play a fundamental role in breaking the norms and contribute to the creation of new inscriptions (as in the case of legislation such as the NetzDG, and changes in SM internal policies, as in the case of Twitter transparency reports, FB declarations etc.).

7.2.2 Public shocks and regulations

From the analysis of issues and topics in the web pages and the articles, it was possible to identify key public shocks that have constituted the backbone of the larger controversy about freedom of expression and social media across the years. The main result from the overview of the data is that the public controversy does not play out in one debate or controversial case, with opposing groups of actors with stable and clearly defined conflicting interests. On the contrary, the public controversy that is described acquires the shape of a mosaic, where more minor cases or 'tiles' constitute the shape of the broader discourse. In this mosaic, the different tiles made of controversial episodes mobilise and assemble different actors, not always bringing forward new regulatory initiatives. In particular, it was possible to identify seven main public shocks that became exemplary cases and storylines mentioned by actors as representative of issues and brought changes in the regulation.

1) Public shocks and exemplary cases linked to episodes of terrorism, where the most famous cases and storylines are developed from the Charlie Hebdo and other terrorist attacks in European cities from 2015 and are mentioned as examples of the issue of terrorist radicalisation on SM. Terrorist use of SM as a storyline was used to support regulatory initiatives to extend states' control over communication online, such as the Investigatory Powers Act in the UK (see §5.9 and table 6.3). Actors' statements connect these episodes to mixed-actors regulatory initiatives. For instance, terrorism on SM is the connection made by Vera Jurova and the EU Commission as the origin of the Code of Conduct on countering illegal hate speech online, issued in 2016, and of the anti-terrorism Forum. At the same time, Charlie Hebdo was an episode mobilised as a storyline by the UK government as justification for a British change in the regulatory system and the introduction of the Investigatory Power Bill (2015) (chapter 6, §6.3.2).

2) Public shocks and exemplary cases linked to hate speech and episodes of violence as a manifestation of alt-far-right extremism in Europe (as in the Jo Cox murder and Katie Hopkins' dismissal) and in the US (as in the attack in Charlottesville) have become the storyline behind other forms of regulation of speech (§5.8 and 6.3.2). This category also includes the creation of single-actor regulation initiatives, such as the controversial NetzDG in Germany, or the self-regulatory revision of internal policies of the platforms (i.e. Google and Apple erasing alt-right and white supremacist apps from their stores, changes in the policies of Twitter and Facebook). In the UK, exemplary cases such as Jo Cox's murder are mentioned as the origin of the Hate Hub institution involves a collaboration between law enforcement and civil society volunteers.

3) Public shocks and exemplary cases linked to episodes of harassment, especially as a form of misogyny and gender violence. The semantic analysis of articles was able to render the connection visually in fig.6.4. These are storylines that connect to episodes of abuse addressed to female MPs or celebrities in the UK or the US. The exemplary cases isolate the main issue (e.g. harassment).

Among the single-actor regulatory outcomes connected to these episodes are the updates on SM internal policies, as in the case of Twitter's revision of its terms of service, which led to the expulsion of Milo Yiannopoulos. Or in the UK, the Crown Prosecution Service Guidelines for anti-hate speech and harassment online (which represent an initiative taken by law enforcement, however, the data show that NGOs and academia were consulted).

4) Public shocks and exemplary cases linked to episodes of fake news and general manipulation of information of platforms and linked to personal data protection. The exemplary cases connected to these issues are the influence of bots and fake news in connection with the EU referendum and the US election in 2016, as well as issues related to episodes of data breach (as in the hacking of Ashley Madison) and ultimately to the scandal over Cambridge Analytica in 2018. In the semantic analysis, all fake news and Cambridge Analytica appear strictly connected, and they also present a connection with content removal policies for hate speech (fig. 6.5).

The outcomes of the exemplary event included parliamentary interrogations of the prominent SM companies, and in the US, requests for new rules on political ads. The regulatory initiatives were just at the beginning at the time of the data collection. Today several legislative proposals are underway of approval in the US, aiming at introducing regulation for political advertisement on platforms. These initiatives signal a change in the US traditional indifference to the regulation of social media and a tendency towards the modification of art.230 of the CDA. Since the data collection, SM companies did also update their internal rules on the fact-checking and political advertisement (changes that have acquired a particular relevance on occasion of the US election in 2020 and Twitter decisions to fact-check Donald Trump tweets which led to Trump's signature of the Presidential executive order against censorship on SM).

5) Public shocks and exemplary cases linked to over-regulation and censorship implemented based on national requests. Cases were presented as a critical stance on excessive regulation of speech by either states or platforms. Such as the articles mentioning Turkey's content removal requests and China's collaboration with SM platforms (Facebook) on surveillance (fig. 6.6), as well as exemplary cases linked to over-regulation and censorship implemented based on SM internal policies: such as the cases of the Vietnamese girl's picture (fig. 6.7) and the trial of Paul Chambers.

7.3 Dynamics connecting actors

As discussed in chapter 3, the concept of translation (Callon, 1986a) refers to actors' efforts to convince others to embrace their view by showing they have the solution for problems that have arisen and broken the routine. If successful, some actors succeed at imposing their vision of the world on the others: the controversies regarding different interpretations of the issues are settled, and the associated particular ways of thinking and acting become 'fact' (Latour, 1986). In time, the 'fact' can be separated from the network that built it, and it becomes a 'black-box', a knowledge claim which is collectively taken for granted. Describing the translation process, Callon and Latour (1981) stress how elements can resist this process, refusing roles or trying to become the spokesperson too. For this reason, translation processes are not always successful in establishing new orders, and different attempts at translation can overlap. This aspect often emerges in the literature on controversies (Marres, 2015), where it is stressed that not all issues reach the same level of development or have the same 'life span'. The list above shows that not all public shocks have the same potential to stimulate regulation initiatives.

The analysis of the temporal development of the shocks in chapter 6 shows (fig.6.10) that the different shocks and discussions are not the same across the years. As highlighted in the historical development of topics from newspapers, some of the issues presented in the controversial cases have been stable in the documents across the years. For example, it is the case for misogyny and abuse by trolls on SM. What is stable is the shock over violence on social media, political censorship, liability for social media and state intervention (NetzDG). Other issues, such as the case about data management, emerged only from 2018. The variation in the prevalence of specific topics and exemplary cases displays a shift in the 'dominant narrative about freedom of expression and social media technology.

The most evident change in the narrative about freedom of expression shows the movement towards progressively more acceptance of freedom of expression as limited. This transition corresponds to a change in the narrative about content, from discussing whether to regulate or not to regulate (counterposing states and social media platforms), which was the prominent position in 2015, to a sort of general agreement on the fact that regulation of speech on platforms is inevitable. The study of the shocks appears that actors can use the same episode to build narratives and normative accounts of governance reflecting different interests.

7.4 Detection of roles

Here I interpret the data in terms of the sociology of translation, and I assign roles to actors. I have based my interpretation on the findings of different types of analysis: the analysis of web pages and newspapers and the relationships and the position occupied by actors within the issue-network created by the URLs and expanded with the crawler (hyperlink network). In particular, to identify spokespersons and prominent actors, I have combined influence measures from the analysis of centrality in SNA with qualitative analysis of texts. I have assumed that the more central an actor in the network of hyperlinks, the higher their chances of becoming a spokesperson or obligatory passage point (Callon 1986a). On the other hand, I balanced considerations on the nature of the influence based on hyperlinks considering that actors have higher chances to impose their interpretation of order/world vision to the others if their vision is 'repeated' in several arenas.

The content analysis of the issues/public shocks also creates an opportunity to study the role of materials, such as technology and inscriptions, as they appear in the narrative. In this way, it is possible to isolate materials, such as pictures, tweets, videos, and inscriptions such as community standards, regulations, etc., that would not appear in the list of URLs or as authors of news articles.

Below I explain how I identified the actors that play the role of spokespersons. Spokespersons are distinguishable because they give voice to and provide general interpretation for the other actors in the controversy (Latour, 2005b). By contrast intermediaries 'transport[s] meaning or force without transformation' (Latour, 2005b:39) and mediators 'transform, translate, distort, and modify the meaning or the elements they are supposed to carry' (Latour, 2005b:39).

7.4.1 Spokespersons and inscriptions

In the literature review, we have seen a part of studies that explore how norms influence online governance (Radu et al., 2021). Actors in the systems can all play the role of norm entrepreneurs. As seen in chapter 3, §3.4.3, spokespersons can be identified by their position in the network as 'obligatory passage points (OPP)' (Callon, 1986; Pohle, 2016a). An OPP (individually or collectively with other actors) becomes indispensable for all others to achieve their goals[17].

[17] To identify the spokespersons, I analysed the materials produced in the web pages, and the relative positioning of the actors involved in the controversy. Initially, as an indication of influence, I identified actors that in the hyperlink network reconstruction with Hyphe appeared to have a larger in-degree. As I stated in the methodology chapter, in controversy mapping with hyperlinks,

Different actors ‘compete’ for this position in translation processes, using different means to influence governance.

As noted previously, in the processes of translation, when a temporary order is established within the network, it is very unstable and precarious (Callon, 1998; Latour, 2005b:63). A strategy to stabilise the order in favour of a particular actor is to inscribe the order achieved in most durable materials: in the ANT tradition, ‘inscription’ refers to the efforts of an actor to fix an alignment of interests, which has been achieved through various processes of translation, in a stable way (Callon, 1998). In order to do so, actors might try to fix the results of translation efforts in a written text or organisational setting, such as creating a particular procedure to be followed or a project to be launched (Pohle, 2016a, 2016b). Inscriptions can be of different binding nature, from legislations and binding treaties to declarations and recommendations (Radu, 2019).

With this in mind, to identify the actors that succeeded as spokespersons, I considered whether actors displayed a ‘claim’ to define the situation or an effort to push other actors into ‘stabilised’ roles with the aid of devices such as ‘inscriptions’. The data showed a presence of binding and non-binding regulatory initiatives. For example, among the inscriptions with binding power, there is the case of the German government with the NetzDG or the Crown Prosecution Service Guidelines. Among inscriptions without binding power are the comments, policy documents, and recommendations issued by NGOs (for instance, EDRi’ 2017, Document id 289).

Through the NetzDG regulation, the German government defines the issue (i.e. hate speech) and the acceptable way to solve it (i.e. quick take-down of content from the companies). By doing so, they assign roles (i.e. they affirm the liability of companies), and they describe acceptable solutions (i.e. the need for quick take-down of content foresees a system based on automatic recognition as the fastest and most effective tool). The German government law imposes a role on SM platforms in the governance of speech; however, it does not entirely control the results, as for the norm to be successful, SM have to adapt and change their internal rules. This means that the responsibility to identify and take down content is still left on the shoulders of the platforms since SM companies

the number of links (i.e. degree) of a page is used as a proxy for influence, as it indicates that a web page has been used as a source by others (or has used several other pages as sources).

However, as discussed in the previous chapters (chapters 4 and 5), network analysis does not tell us much about the type of relationship beyond the links, and especially does not consider whether these links are actually used for the circulation of content related to the key public shocks. As a way to identify spokespersons I focused on what, according to ANT theory, are the sign indicative of efforts to create durable form of associations between the actors: i.e. inscriptions.

maintain strategic control of the infrastructure. In the German government 'inscription', SM companies are the OPP that can materially affect changes to what is visible.

7.4.2 Intermediaries and Mediators

In chapter 4, I have discussed Callon (1998) consideration that every fixed group of association constantly opens up to new actors because of 'overflows' -or externalities- produced by the initial association of actors. In the case of the chemical plant, the pollution dropping in the river opens the system of chemical production to the involvement of the river, the fish, the people living on the river. In Callon's view, mediators are material, tangible objects effects of 'externalities' created by the associations between other actors. This concept helps to identify the material objects that play the role of mobilising agents or mediators, opening up the original framing and involving 'external' agents (i.e. elements which were not previously involved). Technological artefacts and digital actors such as trolls, tweets, memes, emojis, bots, fake news and algorithms play an important role in the controversy. In the texts, algorithms are said to have influence over the flow of information. It is mentioned that they can alter information, either by automatically deleting certain content (e.g. filtering or blocking speech) or by giving more visibility to other content (i.e. as in profiled ads and newsfeeds). At the same time, they are also mentioned as one of the main solutions to the various issues (i.e. automatic filtering).

The analysis of publics in both datasets highlights how materials and technological objects are capable of 'bending' the space, creating division across the other actors: pictures (as in the case of the Vietnamese girl), videos, but also memes, SM posts, tweets (Paul Chambers, Milo Yiannopoulos), misogynist tweets (Caroline Criado-Perez, Leslie Jones). Technology transversally links actors belonging to different 'groups', creating associations between politicians and trolls (i.e. Yvette Cooper, Luciana Berger, Diane Abbott), private citizens/users and companies, users and political movements (English Defence League). The associations created by material objects (either artefacts or digital technologies) contribute to the larger definition of the translation process, and de facto represent an element in the ordering process.

Mediators are often stabilised in exemplary cases and storylines as justification for norms setting and regulation initiatives presented by different actors. In the case of state, for instance, I have already discussed in the case of the EU Commission Code of Conduct, or in the case of the NetzDg, or Investigatory Power Bill). Similarly, it was on account of the misogynistic tweets from Milo Yiannopoulos, an alt-right celebrity, that NGOs started to organise a feminist reaction against SM as place of harassment (as reported in chapter 6 "Many of these activists have reported such harassment and threats by users and pages on Facebook only to be told that they don't violate Facebook's Community Standards' (Titcomb, 2017, document: The Telegraph 2017-01-19). And again, the tragic events at the anti-fascist rally in Charlottesville are mentioned in the actors' discourses as the exemplary case that pushed SM platforms in Silicon Valley to take action against white supremacists and racist ideologists.

The presence of mediators in the study of governance connects to the insights from the literature from internet governance. In particular, it can inform and be informed by Radu et al. (2021) concept of *normfare,* and see how 'mediators' represent the challenges facing internet governance, and as such they 'stimulate'

> the assiduous development of norms of very different character (public and private, formal and informal, technically mediated and directly implemented) by different actors (platforms, standard-setters, states, etc.) (Radu et al. 2012:2).

7.4.3 Tools for measurement

As seen in the literature, media are usually seen as one way to map non-expert voices in the controversy, i.e. acting as spokesperson for non-expert groups (Schouten, 2014; Barry, 2013; Marres, 2005). Reproducing the main contrasting visions, newspapers provide non-expert actors with a way to measure their interests in the controversy. In this sense, as mentioned by Callon (1998), newspapers can play the role of instruments for the public, to 1) realise that they are involved in the controversy and 2) to define their interests in view of 'negotiations' with other actors with alternative visions.

As far as the specific role, the analysis of the texts shows that news media can act both as intermediaries (transmitting the message from other actors without changing the original message) and mediators (playing a fundamental part in breaking the black box, the routine and the regulation solutions taken for granted with the idea of SM as space for 'free space of expression'). By increasingly associating SM platform with topics such as free speech, the right to offend, right to hate (Islamophobia, hate speech, etc.), protection for the press, and censorship, newspapers have contributed to break the routine and creating new spaces for regulation. In that case, they have reinforced the action of mediators, e.g. events and actors with the power of ‘opening’ the borders of a network of associations, and inserting new ones. Scholars treated in the literature have stressed all media are fundamental in mobilizing public discussion about technology, citizens and public opinion (Barry, 2001). Media are the instruments that represent what is relevant for the collective (Couldry, 2012) are a material site for the exercise of, and struggle over power. Tools for measurements are essential in the competition between narratives, and they can result in certain matter of concern becoming more evident than others and certain world views prevailing over others (Mol, 2002). They are also an actor that has to be considered in the normative conception of governance, as through their action norms can either be reinforced or challenged.

7.5 Translation processes

From the analysis of statements in websites and articles emerge four main ‘normative processes’ in the controversy, i.e. phases in which the normative routine was broken and new norms have been created as response to challenges. These normative processes came as the result of a series of *problematisation* attempts, some of which that led to *interessement, enrolment* and *mobilisation* of other actors in the translation processes. These examples show how narratives and discourse created as reaction to shocks, through the isolation of specific elements in the exemplary cases and storylines, are used as justification for competing normative solutions and mentioned by actors when producing inscriptions. Table 7.1 provides an overview of the main roles played by actors in these moments and of the main normative that followed (as captured in the data).
Below I describe in greater detail different normative moments

Table 7. 1 - Overview of actors’ roles in the controversy

Year	Public Shock/mediators	Principal Spokespersons	Competing Spokesperson	Inscription
2015	Charlie Hebdo/ Terrorists attack	National governments	SM platforms	Investigatory Power Bill

2016	Charlie Hebdo Terrorist attack	EU Commission, SM platforms	NGOs (i.e. EDRi)	EU Code of Conduct on illegal speech
2016	US elections 2016, UK referendum 2016, Fake news – Russian bots	National governments		Change in internal policies of SM – US revision of political ads
2016	Vietnamese girl's picture (post) – automatic recognition of nudity	Facebook	Norwegian Prime Minister, Aftenposten Editor	
2017	Twitter harassment cases, death threats (women MPs)	Crown Prosecution Service	Right-wing supporters (Milo Yiannopoulos), NGOs (Index of censorship), Newspapers	Guidelines for anti-hate and harassment online
2017	The Guardian leak of moderators guidelines (Facebook)	Facebook	NGOs (i.e. Black Lives Matters activists)	Change in internal policies of SM
2017	Hate speech and migrant crisis	German Government	Right-wing supporters, NGOs (i.e. German Journalists Association (DJV)/	NetzDG

			Human Rights Watch	
2017	Charlottesville, Trolls/hate speech online	SM platforms	NGOs (i.e. Human Rights Watch), Gab and smaller platforms, Right-wing supporters	Change in internal policies of SM
2018	State and SM censorship	Turkey, China, Facebook	NGOs (i.e. Ranking Digital Rights)	UN call for proposal
2018	Cambridge Analytica, leaks of personal data	US Senate (and other parliaments)		Change in internal policies of SM – US revision of political ads

7.5.1 Problematisation 1 – terrorism and national security

Since their creation, SM companies have overseen the rules of content regulation on their platforms. Owning their technological infrastructure, SM were the only OPP for the control of content on their platforms. The data showed that Silicon Valley companies have a specific world-vision about freedom of speech and that for years SM have presented themselves as symbols of free spaces for communication, protected by the US law, and in particular section 230 of the CDA. From the point of view of the sociology of translation, for years SM platforms have (quite) successfully translated the interest of a network of other actors by imposing their worldview and rules on content moderation. Indeed, with the only exception of general rules against child pornography, until 2015, SM companies have acted as spokespersons for the whole system of governance of speech online, using their internal rules as instruments of policy (Gillespie, 2018).

The data however show how this process of translation got interrupted by elements acting as mediators. In 2015, terrorist attacks opened the frame of governance of speech (as in Callon 1986b) to a variety of hybrid actors: terrorists and cartoonists, cartoons, tweets, posts, European governments. The historical analysis of topics in chapter 6 (fig.6.10) shows that since 2015 a new problematisation phase has been brought forward by public bodies (in fig. 6.10 it is possible to notice the peak of topic 5 and 2 in the documents). From the statements collected in the mapping, it was possible to understand that public bodies in the UK and other European countries identified the issue of terrorism and radicalisation online as a matter of concern for national security issue. Baroness Shields in the UK and Vera Jourova, representing the EU Commission, were among the first to speak on behalf of other elements in the network. Their declarations are the first observable steps in a new problematisation process, creating new matters of concern and roles for the actors involved (e.g. SM companies, users, and civil society at large). Acting as spokespersons, national and European governments have problematised social media roles, calling for greater control over speech on their platforms. Examples from newspapers show how politicians in the UK have used radicalisation episodes (such as the case of Anjem Choudary in §6.3.2) to problematise the role of SM companies in facilitating terrorism. New roles and norms are created in presenting freedom of speech on SM as a security issue (i.e. a new matter of concern). States and enforcement bodies acquired a central position, demanding a level of regulation and oversight of content online never seen from the creation of the internet.

As seen in Callon (1986a), the data confirm that a problematisation phase stimulates reactions from other actors, either in support or in competition. In the negotiation that unfolded (i.e. interessement), SM platforms did not immediately agree to the roles assigned by public bodies. It is exemplary how throughout 2015, SM platforms, as through the voice of Facebook's CEO Mark Zuckerberg, defended Silicon Valley traditional liberal vision of freedom of expression (as highlighted in §5.8). SM spokespersons presented their narrative in defence of free speech, for instance, refusing to obey governments' requests to open encrypted communications. Invoking Section 230 of the US CDA, SM companies defended their 'neutral' role as intermediaries. The data show that also other actors resisted the problematisation presented by public bodies: NGOs active on the protection of speech online (such as EDRi and ACLU, OpenRights group), as well as some activist bloggers, presented a problematisation where the matter of concern was indeed the increased surveillance and erosion of privacy and free speech.

On the other hand, the data highlight how other actors mobilised (mobilisation) in support of national governments and the EU positions: from web pages and articles, it emerged that the policies of SM platforms were several times mentioned as enablers for terrorist attacks. With this regard, newspapers articles are strategic in reproducing public bodies' narrative of SM as accomplices of terrorists (as in the claim made by David Cameron and reported in newspapers in the aftermath of the terrorist attacks in 2015, in chapter 6). In this mobilisation phase, newspapers act as intermediaries of public bodies, reinforcing their definition of security as the main matter of concern. In order to undermine SM world-vision, newspapers challenged SM narratives by reproducing contentious episodes able to undermine SM credibility (as the exemplary cases of Facebook collaborating with authoritarian countries such as Turkey and China).

In the enrolment phase, new regulatory tools were created by public bodies, for example, in the UK, the Prevent Strategy was applied to all levels of education and introduced the duty for institutions to monitor SM content produced by their students. Previous legislation regulating communication (such as the Communication Act 2003) was increasingly adopted in online communication, and new legislation was created to increase the government power over platforms, as in the Investigatory Power Bill (2015). SM started to be enrolled in mixed-actors initiatives, at the EU level, with the EU Commission Code of Conduct on countering illegal speech and the Anti-Terrorism Forum. These regulatory tools are inscriptions that establish new roles and attribute new responsibility to SM companies in policing content. Through the Code of Conduct on terrorism and hate speech, States and European countries have modified the role for SM, from spokesperson to OPP and linked SM internal rules to their legislation. In this translation, European states successfully acted as spokespersons. Public bodies and particularly national governments challenged the definition of roles originally proposed by SM platforms and pushed for the redistribution of responsibilities. The black-box of governance of speech on SM opened, and governments appeared as primary actors challenging the previous actor-network imposing their position as obligatory passage point (OPP). However, in this translation SM platforms occupy a strategic passage point. Without their involvement, it would be impossible for European states to take any action. Governments must delegate the implementation of a solution for security issues to the private companies themselves, and in doing so, they assign a vital role of public decision making to these platforms. However, as seen in the temporal distribution of the topic (fig. 6.10 in chapter 6), this interpretation of matter of concerns did was not the only one that emerged through the years. The data show how different problematisations emerged in the following periods, opposing different world-views of actors.

7.5.2 Problematisation 2 – hate speech

As in the case of terrorism, the combination of particular events and platforms communications acted as a mediator opening up to negotiations other elements of the status quo. The historical analysis of exemplary cases and the visualization of changes in relevance in fig.6.10 in chapter 6 showed how the migrant crisis and hate speech episodes from right-wing groups have acted as mediators. Especially from 2017, public bodies and NGOs identify the primary matter of concern in the unregulated hate speech (problematisation). They issued claims, as in the request to SM to police the content on their platforms (e.g. Angela Merkel direct request to Mark Zuckerberg to take action reported in newspapers).

In the negotiations (interessement), public bodies from Western European states issued requests to take down content from extremist far-right groups. As in the previous translation process, in the interessement phase, different actors resisted the translation presented by public bodies and interpreted the public shocks for their alternative formulation of the issues. NGOs active in protecting free speech published statements against public bodies decision to impose regulatory responsibilities on SM (in particular, EDRi §5.12). From a different perspective, the decision to take down content based on hate speech was criticised by far-right groups (which get censored) and by newspapers.

Public bodies in Europe forced SM companies to act by creating inscriptions such as the NetzDg, which established roles (and fines) for platforms. For example, in the UK, public bodies looked for cooperation with civil society and established the Online Hate Crime Hub, where police corps started to monitor online communications. As a result, SM companies had to accept the role assigned and adapt their internal policies (as seen in the statements, SM platforms started to expel hate groups such as Britain First and English Defense League because of updates to their internal rules).

From the data, the most decisive mediator element that pushed radical changes in the companies' internal policies about hate speech was the public shock created by the attack in Charlottesville in 2017. As noted in the description of the exemplary cases from web pages, the murder of Heather

Heyer provoked a massive reaction in Silicon Valley, and the large majority of SM officially presented statements against alt-right powerful speech.

Hate and white supremacist acted as mediators, which resulted in updates of the big companies' historical approach to regulation. Before this public shock, SM companies have displayed a preference to operate at the level of users' actions, improving the reporting options. However, as a reaction to the public shock and the 'exemplary case' of Charlottesville SM have started to change internal rules and take actions at the level of account (blocking or suspending automatically).

The data show how from that moment onwards, the bigger social media platforms have officially started to abandon the Silicon Valley imaginary of freedom of expression as absolute on cyberspace in order to move towards a more restricted idea of speech **(§5.9, §6.3.2)**. This provoked a reaction from extreme right-wingers who denounced the partiality of the platforms. The data also shows how this decision to ban the accounts of members of the alt-right and white supremacists had the effect of simply moving more 'extremist' views to other platforms, such as the alt-right friendly platform Gab.ai. This platform emerged as a spokesperson for the fringe users both in web pages and newspaper articles, using the space created by the reactions to Charlottesville to attract users with a more radical vision of free speech.

7.5.3 Problematisation 3 – harassment

On a similar level, SM platforms have been associated with the threat of abuse, harassment and other forms of incendiary speech of the far right. The data showed how in 2016, the rising number of episodes of abuse against female politicians or celebrities (from Yvette Cooper to Diane Abbott, from Jodie Whittaker to Leslie Jones) culminated with acts of violence (as the murder of Labour MP Jo Cox on June 16 2017) have gathered a group of actors from public bodies (Yvette Cooper, and Crown Prosecution Service) and civil society (NGOs and academia), demanding more regulation from SM platforms (§5.12 and §6.3.4). As in the case of hate speech, the negotiation phase saw, on the one hand, public bodies requesting SM to take action to police content from abusive accounts. On the other hand, some actors presented alternative problematisation that focused on protecting free speech (primarily right-wing supporters) (§5.11 and §6.3.3). In these problematizations processes, newspapers plaid against the government positions and tended to embrace the alternative problematisation based on the necessary protection of free speech. They associated a political value to the problematisation, comparing the position favouring regulation to a division left-right wing. Some public bodies managed to enrol other actors issuing new inscriptions, as in the case of the Crown Prosecution Service's guidelines.

Also, in this case, SM platforms accepted the role assigned and updated their internal policies (as after the incident with Milo Yiannopoulos, which led Twitter to update its internal rules). As in 2015 for terrorism, also in 2016 and 2017, public bodies claimed the role of spokespersons to create a number of 'inscription devices', aimed at fixing the roles of different actors involved. As mentioned above, an example of normative action in the UK is the Crown Prosecution Service's Guidelines for anti-hate and harassment online. Similarly, the European Commission initiated the German NetzDG and the Code of Conduct on countering illegal hate speech online. The main thing that these inscription devices have in common is that public bodies, such as the UK Prosecution Service, or the German government or the EU Commission, shifted the responsibility of enforcing their standards onto the platforms (6.3.4). In ANT terms, the actors creating the inscriptions acted as spokesperson, forcing the other actors (SM and users) into a system that was ordered according to their criteria, inscribed into legislation or guidelines.

7.5.4 Problematisation 4 – fake news and misinformation

The Leave Campaign for the EU referendum and the US elections (November 8, 2016) with Donald Trump's victory put SM companies at the centre of another type of issue (i.e. they acted as mediators). The 'discovery' of fake news created another challenge to the SM narrative of independent actors in content regulation. As an initial reaction, public bodies created a new problematisation of speech on SM, focusing on foreign influence. However, since 2018, the Cambridge Analytica controversy further changed the framings and introduced a new matter of concern. The historical data visualisation in chapter 6 (fig. 6.10) shows how since 2018, the most relevant terms in the statements started to include the role of users' data and SM platforms' interest in data profiles and advertisement. These last two problematisation were relatively recent at the time of the data collection, and negotiations were still taking place. State governments made the most visible claims (particularly US senate/UK committee to SM), and requests were made for sharing information.

No formal enrolment moves were made, and no inscription was produced captured in the statement of actors in the data collection. However, we know that since the initial interrogations, the US government is in the phase of discussing legislation aimed at regulating political advertisement and fact-checking on SM, modifying art.230 liability for SM.

In this last phase, newspapers and NGOs sustained public bodies (mobilisation) request for more transparency. Academia and NGOs presented an alternative problematisation that put SM business model at the centre of claims for regulation (§6.3.4). This problematisation mobilising the economic interest of platforms is only recent, and problematises SM role as OPP since it highlights how SM influence on contents is subject to the interests of those buying their ads (as in the case of Facebook's ads including anti-Semitic content or the controversial influences of ads on the US elections or Brexit). The last problematisation phase highlights an issue with the normative process which actors in the data collection have not addressed, i.e. that every attempt to regulate content on SM will clash with the reality of the economic interests of specific groups that use SM as 'publicity devices' (Marres, 2005).

7.6 Translation as a tale of contrasting views

The data show how the controversy has led to the creation of four normative processes, and in particular, the historical contextualisation of the data showed how a system of co-regulation is emerging as a result of the pressure of public bodies, with SM companies occupying the strategic role of enforcers (OPP). As I explain in this chapter using ANT concepts, Western national governments have established themselves and SM as an 'obligatory passage point' in the global governance of freedom of expression. From analysing the public shocks (i.e. exemplary cases and storylines), it is possible to recognise different realities and visions about freedom of speech, its governance, and technology's role. The data highlight how the governance model of free speech appears to be moving towards increasingly stricter regulation, even though some interpretations still aligned with the libertarian approach to freedoms survive, especially within technical communities (e.g. bloggers), SM spokespersons, and conservatives far-right sympathisers. However, this solution emerging from the definition of roles and responsibility prescribed by states and SM lacks an essential element: the definition of the role of technology itself. As exemplified by the NetzDG legislation in Germany, European states have shifted responsibility for policing content to platforms without discussing the implications of SM responsibility and the role of their technology in the enforcement of communication policy. The ambiguity is exemplified by the lack of agreement on what is considered as the origin of most controversial aspects, with public actors more worried about the usage of technology (e.g. stopping trolls, removal of hateful tweets) while others like NGOs and academia more concerned by the infrastructure (e.g. the algorithms and AI). In contrast with the finding on the scarce impact and involvement of civil society in other decision-making settings (as stressed in Gorwa, 2019a, 2019b, 2020), civil society does play a fundamental role of mediator in the controversy, especially the 'identifiable individuals' in association with technological artefacts which create the key cases around which the large controversy about social media platforms and freedom of expression develops.

In the following paragraphs I am going to present the breakdown of the changes in the world-views that this co-regulatory framework has entailed from the point of view of the narratives of free speech, governance and technology.

7.6.1 Free speech

As a result of the four problematisation processes, large SM companies' initial rhetoric about free speech has left the space to a world-vision where companies progressively accept the necessity of regulation and their role in enforcing it. Some network elements have also accepted the co-regulation system created between public authorities and private companies (for instance, NGOs active in the protection of victims of abuse). However, it is still challenged by others (e.g. OpenRights group, Article 19).

On the one hand, representatives of extremist views and far-right groups (as in the case of bloggers encountered in the data collection and the positions stressed in newspapers) criticise the co-regulatory order as censorship of their views. On the other, NGOs, international organisations and human rights activists, whose explicit programme of action is the defence of free speech and human rights (e.g. Article 19, EFF, ACLU, or the UN Special Rapporteur on freedom of expression) stress the dangerous potential for authoritarian action when the government intervenes in the regulation of content and when companies do not have any formal bound to the respect of human rights in their internal processes.

These groups started to question SM companies' content moderation policies bias, either from the point of view of right-wing conservative politicians (who accused Facebook of censoring their voices as a result of the implementation of anti-hate filters, chapter 6) or from the point of view of foreign influence (as in the case of ‘fake news, and the role it played in the last US elections and Brexit), or from the point of view of racial discrimination (as seen in the leaks from the internal guidelines discussed in the newspapers). These groups are pushing for new spaces of contestation, primarily based on the discussion of internal guidelines and decisions taken when reviewing content.

From the data, it appears that civil society, especially bloggers, is most concerned by the definition of limitations to freedom of expression imposed by states. On web pages, most bloggers are against regulation and adopt an ‘absolute’ interpretation of freedom of expression. It is interesting to note that the idea of absolute free speech on the internet relies on the libertarian interpretation of cyberspace, typically exemplified in the manifesto written by Perry Barlow on the independence of cyberspace. In this case the findings are twofold: 1) bloggers’ positions are in line with a definition of cyberspace which ignores the development that has taken place since the complete colonisation of cyberspace by giant corporations (i.e. cyberspace as a free space); 2) the libertarian definition of free cyberspace (ignoring all corporate and society bias) ignores or dwarfs the role that technologies play in reproducing discriminations and injustices. These results somehow correspond to what Hintz and Milan (2009) found concerning grassroots movements emphasis on user and technical expert self-regulation parallels cyberlibertarian beliefs and private-sector policy preferences. This emphasis tends to show little concern for structural problems such as inequalities, uneven distribution of technical knowledge, and concentration of power.

In the public discourse in newspapers, sympathisers of conservative positions (Katie Hopkins, Daily Mail) tend to present an absolute interpretation of free speech. Some newspapers in the UK, such as the Daily Mail and the Telegraph, have contributed to assign the monopoly of ‘free speech to the alt-right, by framing issues as free speech rather than abuse or publicity (as in the case of Milo Yiannopoulos, Jordan Peterson and the no-platform controversy).

7.6.2 Governance

From the translation processes, it emerges that public bodies have succeeded in imposing regulatory duties on platforms, either at the EU level, with the Code of Conduct, or the Directive on hate speech and counterterrorism, or at the national level, with the German NetzDG and more recently in the UK with the Online Harms White Paper (2019) that introduced a ‘mandatory duty of care’. This tendency went against the initial opposition of SM companies, NGOs, and activists, which present a preference for self-regulation, often focusing on the possibility for users to take action either through reporting, blocking content, or developing counter-speech (see chapter 5).

The data show how narratives about governance in SM platforms have moved from focusing on regulation based on users’ self-regulation (i.e. reporting) to embracing forms of governance based on top-down decisions (since Charlottesville and Milo Yiannopoulos). Even with this fundamental switch in the internal regulation system, the study of translation processes highlights how SM’s action in regulation initiatives is extremely ‘passive’ or ‘reactive’ to public shocks, rather than proactive in adopting regulation policies. Similarly, the study of narratives presented in newspapers also stresses the rhetoric of apologies and imperfection put forward by SM spokespersons, which minimises their role as regulators. The data highlight a narrative trend of companies adopting apologetic positions (see SM companies declarations in chapters 5 and 6), which reinforce their position as passive and not responsible for what happens on their platforms. What emerged from the narratives about the model of governance presented in the statements is that governments are pushing for a solution based on technology. Public bodies and the press tended to favour a form of governance based on the regulation of speech on and by SM companies using automated content detection (Shields 2017). On several occasions, governments and the EU have stressed the necessity to use technology to regulate content pro-actively. As filters or algorithmic tools seem the most effective and rapid response and the most efficient tools to respond to states requests (as Germany allows 24 hours to take down content which is considered illegal). This displays another division across groups on the model of governance to which aspire to, with some more interested in actions taken at the level of the user interface (e.g. filtering users abusive usage), while others are more concerned with issues that take place at the level of the infrastructure (i.e. algorithms and data management) recognising a more apparent editorial responsibility for SM. An example is that governments stressed the requirement for SM platforms to take down content as fast as possible, using automatic recognition. On the other hand, EDRi and other NGOs have stressed the impact of algorithms and data management on freedom of expression.

Within the first position, more interested in regulation at the level of users, there is a division between public bodies -which presents the preference for automatic filtering- and academia or

NGOs, which favour interventions in the line of counter-speech and the development of good practices. From the data for instance emerge the findings from Huey (2015) and the Demos (2015) study, showing how counter-speech might be more effective than filtering solutions. However, the main public bodies (e.g. UK government speech on the Future of the Internet, Joanna Shield’s speech, Bew’s Report on intimidation in public life, as well as the German NetzDG), all point to the 24-hour deadline to remove content, which by nature favours automatic recognition and filtering of content rather than longer-term counter-speech projects.

7.6.3 Technology

Narratives about technology show a similar division. In the study of the narratives associated with technological objects, it is possible to distinguish two levels of issues: the first level considers as a matter of concern the usage of technological objects from SM platforms, such as tweets, posts, images, etc., with harmful consequences as in the case of harassment and abuse received by female politicians. The second level considers the structure that holds the technological objects and inquiries into the role of AI and algorithms, i.e. it is more related to the hidden functioning of platforms (i.e. the infrastructure level).

In the first level, technology per se is not a matter of concern. It is the solution. Public bodies require SM companies to develop methods to improve their technological objects, similar to the imperative that Douek (2021) describes as ‘nerd better’. In that case, public bodies, such as the EU Commission and the German government, place the matter of concern within SM platforms (i.e. they make it one of their responsibilities). However, they do not include the infrastructural level of platforms in the matter of concern, pushing the enforcement on the development of ‘better’ technology, for instance, improving filtering algorithms to detect unacceptable speech automatically. On the other hand, some NGOs and bloggers (civil society) and academia are more concerned with this second issue, particularly the implication of shifting governance to technology at the infrastructural level (Koene et al., 2017).

What emerges from the data is that in the co-regulatory forms of order, there is a lack of discussion from the point of view of public bodies of the larger implication for the model of governance of asking SM platforms to intervene on users’ activities through automated tools.

7.7 Macro perspective

A new system

As highlighted in the data, governance and technology have been competing across the years' different narratives and imaginaries of free speech. Events, users, and daily practices (i.e. how users use the platforms) have been acting as mediators challenging pre-existing governance solutions. Public bodies in European nation-states have presented alternative enrolment strategies, based on the problematisation of SM communications in terms of national security, hate speech and harassment. The data confirm the trend already described in the literature review (chapter 2), as a result of pressure from public bodies, the bigger SM platforms have progressively moved from a narrative of 'absolute free speech and a preference for a governance system based on self-regulation without formal editorial responsibility (as 'materialised' in US First Amendment and Article 230 tradition) to the adoption of a definition of 'free speech' and an adjudication system where the consideration of other societal interests limits speech, more in line with European tradition (Tambini et al., 2008; Douek, 2021).

The analysis of the translation processes shows that in three out of four problematisation processes, public bodies claim to be the leading spokesperson for the system by issuing statutory tools forcing SM to remove content using technology (either AI or algorithms) as the primary enforcement tool. However, the data show that in this way, public bodies have bestowed on SM companies the fundamental position of OPP. As mentioned in the literature review (chapter 2), this has been interpreted as a paradigmatic shift (Douek, 2021), where SM platforms are called to apply two new principles in their moderation process: proportionality and probability (Douek, 2021). By applying the principle of proportionality, SM have been asked to incorporate a consideration of larger societal interests that the speech on their platforms might harm in their content moderation process. By applying the principle of probability, SM has been asked to officially acknowledge that automated content moderation at the scale of giant platforms will always involve error and costs. Evidence of this shift also appears from the analysis of statements, which documented the change in SM platforms declarations concerning the need for regulation of speech and the limits of technology (chapter 6).

New problems of legitimacy and transparency

What are the implications of SM platforms occupying the role of OPP and working in forms of co-regulations with public bodies? Public bodies' enrolment of SM platforms has strengthened SM power in the governance of speech and 'normalised' their role of governors. However, this normalisation has opened several questions about the legitimacy of such a solution for the democratic system. How can SM companies balance the different speech and societal interests (proportionality) and at what costs (probability) in the current system? What type of remedies can be given for those who will be interested in the inevitable costs or mistakes? How can measures taken by SM be tackled?

Self-regulation by platforms has always raised some doubts about accountability and legitimacy of the decisions taken (Tambini et al., 2008). In this new regime (Douek, 2021) SM companies do not necessarily have the juridical and political competencies to apply the principles of proportionality and probability requested from them (Hustzi-Orban, 2018; Douek, 2021). As assigned by public bodies, SM platforms' OPP role demands more transparency and accountability and clearer instructions from states and independent oversight (Hustzi-Orban, 2018). This is not a new conclusion, and scholars across the years have been stressing the insufficient level of transparency and accountability at the level of decisions, appeal and remedies in the current moderation system (Crawford and Gillespie, 2016; Gillespie, 2018, Hustzi-Orban, 2018; Roberts, 2019; Suzor et al., 2019; Douek, 2021).

However, increasing demand for transparency and accountability is not yet being met by SM platforms. SM companies do not share the details of their policies (Hustzi-Orban, 2018:236) or information concerning their systematic balancing of rights and calculation of errors. Companies do not need to be accountable for their internal procedures and decisions, and there are no established mechanisms for appeal (Belli et al., 2017; UN and Kaye, 2018). Data in this study show that in case of content moderation adjudication mistakes, SM companies' spokespersons presented rhetoric of apologies, reinforcing the idea that mistakes will be made. However, transparent reflections do not accompany measures and estimates of errors that can be made by automatic content moderation. The data also shows how the only information about bias in content regulation mechanisms

Technology as a black box

As seen in the data (chapters 5 and 6) and the discussion of the problematisation phases, public bodies and other actors tended to problematise users' behaviour or the performance of the technology in filtering content rather than the economic system in which SM platforms' technology is embedded. It is evident from the data that the main message from public bodies to SM platforms was similar to what Dueck call the imperative: to 'nerd better' (Douek, 2021): i.e. a request to solve the issues raised by public shocks through better use of their technology. Especially in the case of public bodies, technology is treated as a black-box. In public regulatory initiatives, SM interventions are asked at the level of usage (i.e. removal of content), and they are required to happen fast (24 hours turnout). This indicates that interventions target what is immediately visible on the surface or interface with no reflection on internal functioning and reliance on automated recognition. SM tend to rely more on automatic detection through AI and algorithms (Gorwa, 2019a; 2019b). This creates issues in terms of costs for freedom of expression, human rights in general, and the governance system's material implications.

As far as freedom of expression and human rights are concerned, scholars stress how content moderation through algorithms, especially at a large scale, will always include a margin of error, where content will be over-or under-censored (Hustzi-Orban, 2018; Douek, 2021). From the moment SM platforms face fines for failing to remove content, with no responsibility in the case of over-censorship, they will be interested in over-regulating with algorithmic means rather than the opposite (Hustzi-Orban, 2018).

Moreover, on top of pushing towards a system over-regulating at a large scale, black-boxing technology as a tool for regulation has created a biased tool. The data confirm the concern of scholars discussed in the literature review (Ziewitz, 2016; Gerrard and Thornham, 2020) and show how SM platforms have been criticised for the biases and mistakes embedded in algorithms for content recognition, as in the case of the leaks on Facebook AI training for the recognition of protected categories (chapter 6).

As far as the material implications of the governance system are concerned, the black-box approach to technology ignores the human rights cost as well as the material costs behind the use of AI as a regulatory tool: as Casilli (2017) and Crawford (2021) put it, there is nothing artificial or intelligent in Artificial Intelligence. On the contrary, every algorithm will reproduce the human expertise of those who make it (Dencik et al. 2018a, 2018b). Using this terminology strengthens the black-box, reinforces the idea of technology's efficiency, and hides the human labour behind it

(Casilli, 2017). It also minimises the argument that human reviewers should be highly trained professionals, while studies show that reviewers are systematically underpaid and geographically exploited communities (Casilli 2017; Hustzi-Orban, 2018). Similarly, it overlooks embedded bias and the implications for society (Hintz, 2016; Redden, 2015, Dencik et al., 2016, 2018a, 2018b) and the environmental costs of maintaining it (Crawford, 2021).

Moreover, by letting the larger companies take care of the technological, regulatory solution, public bodies and other actors ignore the existence of smaller companies and legitimise the current situation of oligopoly. In this way, the idea that the internet corresponds to the Big Tech companies (Google, Apple, Facebook, Amazon, Microsoft, and Twitter) is reinforced. Consequently, major regulations are drawn up with these big companies as standard (as in the case of NetzDG). This creates a situation of inequality and reduced plurality of voices online, with smaller companies incapable of paying the same costs to operate. It also reinforces the idea that smaller companies (like Gab) can avoid regulating hate speech or harassment issues on their platforms, de facto not solving the issue of hate speech or harassment/abuse but simply deviating to smaller companies.

This preference fits the larger dynamic of algorithmic management and datafication underlined in the literature review (van Dijk, 2018). Increasingly more public bodies tend to trust private companies' technology to administer public policies. This strengthens the idea that big, digital data are better data, and that technology works better (Boyd and Crawford, 2012). This way reinforces the idea that 'the code is law' (Lessing, 2000). This tends to disregard the fact that data are socio-technical assemblages that reproduce the power system (Kitchin, 2014; Dencik et al., 2018a, 2018b).

Moreover, it overlooks the surveillance potential embedded in this form of digital technologies (Andrejevic, 2011; Srnicek, 2017) and the influence on content that advertisement and profiling can have. As seen in the literature review in chapter 2, the more a system relies on technology (and AI in particular), the less it guarantees transparency and democratic scrutiny (Sinnreich 2020). The current governance of speech does not leave much space for civil society. Without systematic transparency and scrutiny of the moderation processes and decisions on proportionality and probability, governance will rely increasingly on exclusively private companies (in self-regulation) and (public bodies).

The data have shown how events and users play the fundamental role of mediators, capable of initiating controversies and public shocks. Furthermore, this shows how interactions within members of different online communities originate from the exemplary events and storylines that initiate regulatory discourses. From this point of view, studies of online communities, for instance, studies of counter-speech and online communities self-organisation, can explore an important alternative in the current regulatory panorama, where instead of introducing filters or asking SM platforms to take judicial decisions on public interest, the algorithm is taught to imitate the virtuous behaviour of existing online communities (Housley, 2018; Procter et al. 2019). Similarly, vulnerable groups and communities can be helped in developing resilience (Edwards et al. 2021, Procter et al. 2019), and studies of interactions online can help develop protocols for machine learning open to public scrutiny (Housley et al. 2018, Procter et al. 2013a, 2013b).
However, reflections on communities and regulation should also consider that SM technologies are embedded in a system of power and profit production (Srnicek, 2017), which is based on the extraction of data from users to predict and influence behaviours sell products (Zuboff, 2018). As Milan (2015) found studying social movements organizing on platforms, algorithms significantly alter people’s opportunities and, as a result, steers social actions. ‘Platforms matter, and matter more than activists like to believe’ (Milan, 2015:8). Therefore, it is fundamental to consider the materiality of digital technology to understand the infrastructure of power in society. Algorithms have been designed, developed, and implemented by recognisable social actors and have material implications on human lives. This is the general issue that critical data studies are trying to point out: asking for technological solutions to social problems requires a clear understanding of the power relations and the human expertise embedded in technology and data (Kitchin, 2014; Dencik et al. 2018a, 2018b). The data show that the problem is not just that technology is regarded as neutral – it is just ‘left out by the public discourse and statements created by public bodies. The power of technology and its link with the companies' business model where it is developed is the neglected element in the enrolment process.

From the data, it is possible to notice that an increase in the critical appraisal of the bias in technology started to emerge in 2018 (chapter 6). In the last problematization that followed Cambridge Analytica, representatives of public bodies and newspapers introduced in public discussion consideration on the larger dynamics and components behind SM platforms (an issue that appeared to be a concern more for academia and NGOs). Even if not included in the data collection, it can be hypothesised that there has been an increase in attention towards the ethical implications of technology, for example, in the section on ‘AI ethics’ included in the UK White Paper on online harms (UK White Paper 2019). However, scholars are becoming increasingly more sceptical about the use of ethics as a term associated with AI (Crawford, 2021; Hao, 2021), as it has been done on several occasions only in a formal and not substantive way.

Some scholars argue that SM platforms should receive clearer instructions on how to increase accountability and transparency of the current system and independent oversight on how to orient their moderation processes (Hustzi-Orban, 2018). However, the question remains, how will SM companies be able to assess the criteria of necessity and proportionality or the balance between the individual right of freedom of expression and the more significant societal interest, and enforce them in their technological infrastructure while at the same time still maintaining their business model and economic interests in the interactions that take place on their platforms?

7.8 Technological dynamics

Analyzing web pages and newspapers’ articles has highlighted a specific technological dynamic structuring the controversy (Marres, 2005). SM platforms turn out to be the origin of the exemplary cases and the main channel through which these cases and the related statements are diffused. Furthermore, SM platforms occupy the central position in the network of hyperlinks, demonstrating how every web page has at least one link connection to one of the bigger platforms (Youtube, Facebook, Twitter). Moreover, as highlighted in chapter 6, newspapers also reinforce SM’s central position by reproducing divisive content that originates on the platforms and using it to create stories or exemplary cases within their public discourse about freedom of expression.

Social media users’ divisive materials (memes, videos, cartoons, etc.) attract more interactions on the platforms (Zubaia et al., 2016; Procter et al., 2019). SM platforms have been criticised because of the conflict of interest in policing abusive material while profiting from the increased volume of data produced by interactions around divisive content (Naughton, 2020). Platforms are ‘publicity devices’ (Marres, 2005), and as such, they are explicitly designed to facilitate advertising, similarly, markets will always have an impact on the content which is more visible (Gillespie, 2010). In the last part of chapter 6, I pointed out that by reproducing divisive tweets or posts, images or videos, and by giving the possibility to link back to the original platforms, newspapers are reproducing divisive content without a critical reflection on the technological dynamics that have allowed certain accounts to reach ample visibility, or specific tweets or posts to go viral in the first place. These results confirm a dynamic already highlighted in Zubiaga et al. (2018), who found that, when investigating the spreading of rumours on Twitter, news organisations did nevertheless publish materials later found despite newspapers' efforts to publish well-informed claims to be inaccurate. This tendency was, in their view, reflective of a change in the model of journalism from traditional to online journalism. In this way, newspapers might find an easier way to 'sell' their articles within a market system characterised by an implicit failure (Pickard 2013). This market failure points to a crisis in traditional journalism and the risks connected with leaving the news and information market in the hands of the 'click economy' or 'attention economy'- which are models already adopted by SM platforms as a way to stimulate a reaction from users to extract data (Zuboff, 2019).

7.9 Role of researcher

The study of controversies helps shed light on dominant and neglected elements that contribute to the regulation of speech. Researchers are privileged to observe and understand dynamics of domination and power embedded in a system built on data. As Sinnreich (2018, 2020) states, it is fundamental to investigate the social implications of increasingly adopted algorithms to manage a social life. However, scholars are becoming increasingly more sceptical of using the terms AI ethics (Crawford, 2021; Hao, 2021) as it has been used on several occasions only in a formal and not substantive way, or the meaning has been distorted. Hao (2021), in particular, stresses how a large part of the vocabulary developed concerning AI covers a number of incongruences or hypocrisies (for instance, the keywords of diversity and inclusion used by SM platforms, which do not resonate with the firing of employees that challenge the status quo with works on AI discrimination, as in the controversial case of engineer Timnit Gebru fired by Google) (Hao, 2021). Researchers can help develop a new vocabulary, rethinking concepts to produce a more transparent, open description of the technology that can highlight racially just and political and economic solidarity programmes (Gregory, 2017, 2018; Hintz, 2016; Dencik et al. 2016, 2018; Redden, 2018). Opening the black-boxes pointing at what is missing in this distribution of roles will help find a place for civil society and the online communities affected by decisions taken by

SM companies and public bodies. For instance, stressing the need for transparency on regulations and the right to appeal the decision taken by both states and companies can be accompanied by request for technological transparency and social impact assessment of technological and economic models.

Finally, a role for SM technology cannot be found without a normative discussion on the type of business model that the technology is materialising. Normative discussions about the governance of speech should also stress that media and information infrastructure in a democracy should be treated as a public good. Since SM platforms have started to present newsfeeds, they have increasingly more left the justification for technological companies, and even if they are not yet recognised as media, their role in public life has been assessed. Starting with integrating human rights and developing some form of social responsibility, SM has a different part to play if they want to be part of a democratic society.

7.10 Conclusions

In this chapter, I document the efforts of a range of actors to settle the controversy over the governance of speech online. The data show how the controversy has led to the creation of four normative processes, and in particular, the historical contextualisation of the data showed how a system of co-regulation is emerging as a result of the pressure of public bodies, with SM companies occupying the strategic role of enforcers (OPP). The sociology of translation shows how difficult it is to stabilise a controversy constantly re-opened by introducing new elements (i.e. mediators) changing the shape and boundaries of the associations.

The most evident result is that the normative outcomes of controversies about freedom of expression and content regulation are the outcome of the struggle between competing spokespersons. The analysis has highlighted actors’ argumentation strategies as means of finding compromises or persuading other actors, which in some cases have led to a normative change of the status quo. With the study of exemplary cases, it was possible to observe how norms about the governance of speech are created in a complex assemblage composed of coordinated and dispersed regulations, arrangements, infrastructures, and technical procedures (Schouten, 2014). The actors' role in the different translation processes highlighted how normative processes work within the sociotechnical arrangement. The study of actors’ competing positions concerning freedom of expression, governance, and technology was helpful to understand how certain norms became stabilised, while other concerns (for instance, the issue with algorithmic decision-making) are black-boxed. In particular, the findings show that among the norms developed in the different phases of the translation processes, SM platforms have been assigned the central role of OPP by public bodies. Because of this, their way of moderating content has changed, as they are increasingly asked to make decisions balancing societal interests and technological errors.

The findings highlight some critical issues. Firstly, studying the development of the issues as a tale of competing translations processes, it was possible to notice a paradox in public bodies and civil society asking for more control and responsibility from SM companies. In doing so, they increased the public power of such companies and normalised the idea that private corporations manage public policies through 'non-transparent technology. However, there is a fine line, and corporations still have a business model based on capitalisation and profit rather than serving as a public utility service.

Secondly, the findings show the paradox of the public debate over media, since media are the object of the contestation and at the same time one, if not the main, channel through which the public debate takes place. In chapter 5, the SNA of the websites showed that the centrality of the press is overshadowed by the number of materials that are 'shared' on social media platforms. Social media platforms are essential for the distribution of news. At the same time, the study of the main controversial issues shows that the shocks comprise cases that have almost all taken place on SM and are reported in the press without a critical assessment of the editorial role of the algorithms.

8. Conclusion

In this study, I have applied controversy mapping methodology to study the governance of freedom of expression. Through the empirical analysis of web pages and articles, I found groups of actors mobilised in the governance ecosystem. I was able to isolate public shocks, exemplary cases, and storylines that are used to mobilise actors and enrol them in regulatory initiatives. In the discussion chapter (Chapter 7), I used the findings from the historical overview of the issues and the study of actors' positioning to describe different translation processes. In particular, I identified the actors emerging from the statements who occupy the role of spokespersons (states, SM platforms), mediators (newspapers, technology and users) and intermediaries (newspapers). Using critical data studies, I interpreted the type of governance of speech emerging from the different moments in the translation processes, putting the narrative about technology and data in a materialistic perspective. The critical analysis of translation processes highlighted the social implication of the narratives about free speech, governance and technology presented by the spokespersons (public bodies) and SM platforms. In particular, the analysis of narratives about technology and data shows the emergence of public–private (hybrid) forms of ordering of society and freedom of expression in a co-regulatory system stimulated by public shocks related to episodes of violence, such as radicalization, hate speech and harassment online.

The analysis of narratives shows the relationship between changes at the level of narratives of free speech and regulations. The data shows that on web pages, the historical cyberlibertarian idea of free speech survives in groups of bloggers closer to the technical communities of the internet, while in newspapers, the larger public discourse about free speech is co-opted by alt-right supporters.

The data on the 'normative processes' emerged from the study of the translation processes highlights how technology, particularly algorithms, is ambiguously mentioned as a regulatory black box. Concepts of datafication and algorithmic management in critical data studies help to unpack this black box and the material implication of using technology based on algorithms and data for regulation.

In recent years, companies have progressively accepted more official responsibility for content moderation, and national governments have on several occasions expressed a preference for technological or algorithmic management of social life. However, their statements do not show concern for the 'hidden' costs such as the extraction of value from users' data and content, the labour provided by content moderators who perform a highly destabilizing job without any psychological assistance or the social bias that can be embedded in the technology.

In this thesis, I argue that normative discussions about the governance of speech have to include reflections on private companies as adjudicators, the societal costs in terms of freedom of expression, and the material costs in terms of exploitation and the larger process of datafication touching society as a whole.

8.1 Research questions and objectives

I started this project to answer the questions: how can we study the governance of speech online as an emerging phenomenon without focusing on one actor or specific setting? How do governance initiatives 'initiate' and take form? And what does it mean for the wider governance of freedom of expression and democracy? I have oriented my research on the following operational questions:

- What are the 'public shocks' (not necessarily majors) that contributed to break the routine or pre-existing forms of decision-making concerning public expression on social media?
- What type of governance is taking place as a result of public shocks in the last few years?
- What actors and dynamics of power are revealed when using an approach that does not simply focus on a single platform or actor and including technology in the study of regulation initiatives?
- How do 'public shocks' relate to the narratives associated with free speech and technology? And what is the role of media in reproducing narratives and shocks?

First, I investigated the 'public shocks' that contributed to breaking the governance routine to find an answer. Second, I asked what actors and power dynamics emerged from online statements and what narratives were reproduced in the media. My main goal was to achieve an empirical exploration of actors and narratives of technology involved in regulation initiatives, using controversy mapping methodological tools and interpreting the results through the lenses of the sociology of associations and critical data studies.

From 2015 to 2018, I collected statements of actors from web pages and news media in the UK relating to freedom of expression and SM platforms. It was interesting to find that 'public shocks' indeed emerged as the main break in the routine of decision-making concerning free expression on SM. Statements from different actors shared recurring exemplary cases and storylines, mobilised to create a specific narrative of the matter concerned. As described in earlier chapters, terrorism and hate speech, harassment, quality of content (e.g., fake news), and disinformation have been the main issues stimulating changes in the government of speech.

My initial research questions focused on the morphology of the controversy and the normative processes connected. The data show that the process of problematization corresponds to regulatory initiatives taken either by public bodies or SM. In this sense, the data confirm that governance initiatives are ex-post or, in other words, a 'reaction' to public shocks with the power of mobilizing different groups of actors (mediators). In the study, I have not restricted the focus to a specific actor; however, big SM companies and European and US governments emerged as the more prominent players imposing their world vision on others. In the lenses of ANT, their initiatives can be interpreted as inscriptions and studied as an attempt to assign roles and responsibilities to other actors. The findings also confirm that regulatory initiatives take place within different institutional processes, sometimes simultaneously. There is no coherent institutional space or procedure where the governance of speech is discussed; decisions are taken and power dynamics occur in a diffuse, capillary way. This is negative in terms of the transparency of the process. On the other hand, it creates windows of opportunity, where civil society and media can act as mediators and create change (for example, the leaks about content moderation published by the Guardian).

Observing the specific narratives about technology and the use of storylines and exemplary cases in creating new forms of regulation, I was able to identify significant findings concerning the normative model of governance presented by the actors. Many actors, mainly from national governments, have moved towards a more restricted idea of freedom of expression and see SM as the primary enforcers of speech policies. Here, I answered my last question related to the macro perspective and the normative ideal of the type of governance that are emerging and their implications for the wider governance of freedom of expression and democracy.

The data analysis shows that, as a result of the combination of public bodies occupying the role of spokesperson and assigning the role of enforcement to SM platforms (in ANT terms reinforcing their role as OPP), the current governance system is experiencing an increase in the algorithmic management of content. However, this growth in the use of technological content moderation systems corresponds to a very ambiguous and opaque narrative associated with algorithms and AI, indicating how public bodies in their enrolment processes have failed to assign a clear role for this type of technology. In the discussion, I argue that given the lack of role assigned to technology, this type of solution increasingly legitimises platforms and algorithmic management as governance tools. However, there is a deficit of transparency, accountability and remedies, with broader implications for democracy and social justice.

Future regulatory initiatives should take into consideration that SM companies have changed the nature of moderation processes and require more democratic scrutiny in their assessment of proportionality (i.e., how speech is adjudicated based on larger societal considerations) and probability (i.e., which groups are likely to be penalised by the inevitable mistakes that large scale automated content regulation will make). The study also highlights the role of less traditional actors, such as the SM community of users and newspapers, in originating and diffusing the exemplary cases and storylines used to mobilise governance initiatives. The data suggest that future thinking about the model of governance of speech should take into consideration the dynamics of the interaction of online communities since they have been proved to be at the root of some of the public shocks that have initiated recent regulatory initiatives, as well as the interaction between newspapers and SM in creating public shocks by reproducing content 'edited' by algorithms. In this regard, future reflection on the governance of speech should consider the specific issues raised by the role of technology and automated content management as regulatory tools and, more generally, the economic system connected to it.

8.2 Contribution and significance of the findings

In this study, I contribute to the study of governance, highlighting power dynamics adopting a horizontal and hybrid perspective. SM are central not simply because they are part of the issue but because they physically provide the infrastructure on which the other party's position in the controversy can be shared. The normative reflection on governance must go hand in hand with the unboxing of the narrative associated with free speech and the unboxing of technology and the business model which sustains it. The study shows that there is a lack of planning and, in general, a lack of 'vision' about the type of order we want to live in. Governance and regulation lack normative reflection on the implications of the choices made for human rights and society as a whole. Using a black box in human rights governance is risky in terms of biases and possible unintended consequences.

Although there are positive aspects to flexible and plural governance of speech in terms of the variety of actors involved, the way it has been done does not reflect such diversity because it is modelled on algorithmic management, which reproduces the system of bias of those who created it (e.g., Silicon Valley technologists) (Dencik et al., 2018a, 2018b). In recent years, more has been done to create a dialogue between technologists and human rights (Milan and ten Oever, 2017; Frank Jorgensen et al., 2019); however, technology tends to be closed in black boxes with critical social implications.

This study has also highlighted how black boxing technology impacts the narrative of freedom of expression. The data have shown how conservatives and alt-right groups have co-opted the cyberlibertarian ideal of free speech. Smaller companies such as Gab AI have taken over the narrative and proposed their platforms as strongholds of original internet values. Newspapers have contributed to spreading this idea, presenting controversial exploits from alt-right characters (such as Milo Yiannopoulos) as part of the discussion on freedom of expression, without opening the black box of what is hidden behind their ideas (even to assess the benefit in terms of publicity from divisive events). However, it appears from the data that free speech technology is taken for granted in this discussion. Opening the box would start with acknowledging that the internet is quite different from Barlow's idea. Today, we create speech in private spaces and the business model that holds these places also influences what is visible and what is censored. Smaller companies such as Gab AI are not alien to this system of extraction of value from their users’ data. Discussing free speech as the opposition between leftist and right-wing parties without discussing the structure in which speech takes place in contemporary society only distorts the main issue and contributes to sustaining the positions of highly inflammable and uncritical parts of the population. This is a reflection that touches aspects at every level of society, as platformisation and algorithmic management are the most commonly employed model for developments not just in business but also in the public sector, from citizens’ justice to health to education (see, for example, the recent problem with GCSE results assigned by algorithms in the UK in the time of COVID-19 (Ferguson and Savage, 2020)).

8.3 Limitations of the study, opportunity for further research

One of the aspects that I have emphasised most throughout this work is the complexity of the governance ecosystem for freedom of expression. I have justified my theoretical and methodological choices on the basis that they offered me more tools to ‘pin down’ power dynamics across a plurality of actors. ANT and controversy mapping are particularly useful in making sense of the idea of ‘fluid’ power, embedded at different levels, from the micro controversy about censorship of pictures, to the macro controversy about terrorism and radicalization online. Using controversy mapping, I tried to study forms of order in the making, following actors’ trajectories to explain the emergence of regulatory initiatives. The positive aspect of this theoretical and methodological framework is that it makes room for a larger range of actors than other theoretical perspectives or methods could 'accommodate'. The use of statements and discourses to scope the actors involved in governance made it possible to include highly institutionalised actors (such as governments or politicians), together with non-institutionalised actors (such as individuals or bloggers), and heterogeneous types of actors (such as technological objects). The downside of this is an aspect that is treated in Chapter 3, the limits of isomorphy. As I explained in the theoretical framework and methodology (Chapters 3 and 4), ANT and controversy mapping’s agnostic ontological prescription treats actors as if they were all of the same sise. Similarly, it lets actors deploy their positions (Munk and Abrahamsson, 2012) rather than investigating governance starting from a specific institution or regulatory environment.

This creates a number of issues which I have discussed from the point of view of theory and methodology in chapters 3 and 4. The theoretical issue concerns the fact that not all actors have the same power to mobilise others and that some actors have larger enforcement powers. It creates the methodological problem of interpreting statements from very different actors, such as state representatives, bloggers, etc., and assessing their governance role. I treated the problem of isomorphy in chapter 3 §3.7, where I argued that for the purpose of this study, I have applied ANT isomorphy rule in the empirical data collection, as a method to include possibly overlooked actors in the system. Drawing on studies on the influence of norms on policymaking, I was also able to include in the attention from the actual enforcement power to the more prominent discursive elements that end up being integrated into policies, which have a particular role in internet governance (Radu, 2019). Following this approach, I have shown how exemplary cases and storylines contribute to the larger discourse or worldview construction. However, I also stressed that for the purpose of interpretation, I would have integrated the discussion with a position derived from critical data studies, which stresses the material importance of data and connected technologies. Adopting this point of view, I recognise the prominent strategical role of SM companies, and I stress the important element missing in the discussions i.e. technology as instrument of regulation.

The issue of isomorphy of actors connects to another methodological issue concerning the centrality of the researcher and the role of researchers in the orientation and planning of controversy maps. As a researcher, I often felt the difficulty of orienting my interpretation in line with one of the actors. Within the specific timeframe and resources available for this project, in order to ground my interpretation, I triangulated my sources using computational tools.

As discussed in chapter 4, computational methods using AI and NLP have been employed in social science disciplines and scholars do recognise the constant advancement in the technologies (Tufecki, 2014). However, as I have argued in §4.5.2 digital and computational tools come with a number of biases and limitations. As I have explained in chapter 4, following Marres' (2015) empirical approach, as a way to address these limitations I have considered the uses and the specific characteristics of the sources as part of my investigation, including the possible limitations to my findings.

While performing the data collection and analysis, I had to consider the difficulty of defining the 'entry points' for the data collection. I mentioned part of the academic discussions in §4.5.3 and it took me months of reading and hundreds of attempts before being able to develop a justifiable protocol to design the keywords. I had also to considered the influence of the search engine's algorithms on the ranking of the URLs and on the type of documents that I retrieved (§4.5.4) and the other limitations of the sample, which includes the issue of ephemerality of web pages. Websites are not meant to be studied as archives. This limitation is taken into account in the methodology ('Google is not the internet, the internet is not the world' (Venturini, 2012)), and in my study, I have integrated the results from web pages with data from newspapers in order to mitigate this bias.

Some positive aspects can come from the limitations, when I was cleaning the data set I realised how biased the network of URLs is because of hyperlinks for commercial advertisements and links to SM embedded in the structure of web pages. While this was an exciting finding regarding the ‘publicity’ device operating in the controversy, it also represents a challenge in terms of actual time employed to retrieve relevant information from the pages.

Another type of limitation concerns the quick obsolescence of tools for data collection and analysis. Technology fast growth also impacts researchers using digital tools for data collection and analysis. For example, at the time of the data collection, the issue crawler developed within the Amsterdam University DMI environment could work exclusively on Google. Now Amsterdam's lab has improved, and it is possible to crawl pages from different search engines, creating more opportunities to study the role of the medium in the staging of the controversy.

All these limitations point to another aspect of computational methods which is often not discussed in methodological accounts of research: i.e. the implication of using digital and computational tools for the work of the researcher. An aspect that is often avoided in publications is the amount of time and effort to clean that these actions require. In this study, a large part of the data processing went into the formatting and cleaning the URLs and newspaper dataset of unrelated items.

A few authors have pointed out the massive amount of work that comes with data in digital format (Boyd and Crawford, 2012; Tufecki, 2014). Collecting a large sample of data with a click often means that researchers have to deal with a considerable amount of noise or irrelevant elements to make the dataset ‘readable’ for the tools. At the level of analysis, computational tools still involve considerable ‘work’ on the researcher's side. As discussed in chapter 4, the reliability of the computational analysis of texts is still based on the ‘golden standard’ of the human eye (as in the choice of the topics with the LDA model) (Zhao et al., 2015).

This type of limitation highlights another implication for researchers with a social science background, i.e., the need for specialised training. For example, building a crawler to scrape pages from a search engine can be done quickly with a coding background. However, without a similar background, it becomes necessary for a researcher to rely on tools developed by others. A similar issue can be considered in the use of tools for quantitative analysis of texts. The complexity of algorithms performing text detection using NLP is (usually) beyond the scope of a researcher trained in traditional forms of social research. From this point of view, the dialogue and engagement with academic communities across disciplines (for instance, with researchers involved in the Cortext platform, in MédiaLab in Paris, as well as within the DMI in Amsterdam) were fundamental to discovering the existence of specific tools developed for multidisciplinary work and learning how to use them.

Karen Gregory (2017) argues that this type of labour in research tends to be minimised, creating the illusion of performance and easy results. This is also due to academic journal publication formats, in which the methodological paragraphs are usually quite condensed, and the authors tend to describe what went well rather than what went wrong with their research and use of tools. In this way, researchers contribute to the idea that algorithm is a black box solution to data collection and analysis problems. Pointing out the limitations and the prominent role that researchers have in the analysis performed by automated tools contributes to deconstructing/problematizing this rhetoric. Thus, it is possible to see how computational tools can augment rather than replace traditional forms of social research (Housley et al., 2014).

Another important aspect related to computational tools concern the political implications of using large datasets. Scholarly discussions about the crucial importance of datasets are starting to grow only recently and studies that exposed serious issues and biases in large language models (such as Bender et al. (2021) discussing limitations of BERT, used in Google) have proved to be highly divisive in the AI research community, particularly within corporate environments. Notable actions such as the firing of AI researcher Timnit Gebru from Google's Ethical AI team and the turmoil that followed show how datasets in machine learning are an increasingly hot issue in research and also come with high political and social stakes. As a researcher, it is essential to engage with the black box of computational methods, promoting interdisciplinary exchanges and appreciating tools such as Cortext or Hype, developed in an interdisciplinary context, which provide transparency and opportunities for the researcher to engage with the data in more visual, interactive ways.

The data collected provided valuable insight into power relations in diffuse environments and highlighted the power dynamics embedded in the narratives of free speech and technology, especially those of contemporary society presented in the media. These findings can provide a helpful background for public policies and decision-makers on the normative aspects of the future governance model of speech. As a future direction, and as a way to further balance the centrality of the researcher, it would be interesting to expand the mapping and merge the results of this study with input received from concerned parties (Marres and Rogers, 2005; Venturini, 2010; Yaneva, 2012). Munk et al. (2019) remind us that controversy mapping is becoming increasingly developed as a tool to foster public engagement. Moving towards the participative side of the methodology, this could be done by integrating the data from written statements in web pages and newspapers with data from qualitative interviewing, focus groups and other more conventional research methods able to 'augment' digital social research (Housley et al., 2014).

The governance of speech online, divided between public bodies and SM companies with more and more public responsibilities, has the potential to become the model of governance in a world that is increasingly organised around platforms and digital data. Further research will be fundamental to investigate how SM platforms can take on the responsibility of balancing free speech with other societal interests and how to develop a protocol to assess the costs of algorithmic management within the scope of social justice. The research will be fundamental to opening the black box of governance, formulating a space for online communities and civil society in this system, and helping to reduce the material costs in terms of human labour and the exploitation of people and resources.

References

- Alcouffe, S. et al. (2008). Actor-Networks and the Diffusion of Management Accounting Innovations: A Comparative Study. *Management Accounting Research*. 19. 1-17. 10.1016/j.mar.2007.04.001.
- Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine* [June 23]. Available at: https://www.wired.com/2008/06/pb-theory/
- Andrejevic, M. (2017). To pre-empt a thief. *International Journal of Communication*, 11, pp. 879-896.
- Andrejevic, M.B. (2011). Surveillance and alienation in the online economy. *Surveillance & Society*, 8(3), pp.278-287.
- Andreossi et al. (2013). What the frack. Report. Empirical study of controversy (http://www.whatthefrack.eu/).
- Annany, M. and Gillespie, T. (2016). Public Platforms: Beyond the Cycle of Shocks and Exceptions. [blog] *The internet, policy & politics conferences*. Oxford Internet Institute. http://blogs.oii.ox.ac.uk/ipp-conference/2016/programme-2016/track-b-governance/platform-studies/tarleton-gillespie-mike-ananny.html
- Article 19 (2017). Germany: The Act to Improve Enforcement of the Law in Social Networks [online]. Available at: https://www.article19.org/wp-content/uploads/2017/12/170901-Legal-Analysis-German-NetzDG-Act.pdf [Accessed 14 April 2021]
- Article 19 (2019). Social Media Councils: Consultation. [online]. Available at: https://www.article19.org/resources/social-media-councils-consultation/ [Accessed 14 April 2021].
- Awan, I. (2014). Islamophobia and Twitter: A Typology of Online Hate Against Muslims on Social Media. *Policy & Internet*, 6: 133-150. https://doi.org/10.1002/1944-2866.POI364 [Accessed 14 April 2021].
- Awan, I. (2016). Islamophobia on Social Media: a qualitative analysis of the facebook's walls of hate. *International Journal of Cyber Criminology*, Vol 10, Issue 1 January – June.

Available at: https://www.cybercrimejournal.com/ImranAwanvol10issue1IJCC2016.pdf [Accessed 14 April 2021].

- Balkin, J. M. (2009). The future of free expression in a digital age. *Pepperdine Law Review*, 36(2), 427–444 [Accessed 14 April 2021].
- Balkin, J. M. (2017). Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation. *UCDL Rev. 51*, pp. 1149. [Accessed 14 April 2021].
- Barlow, J. P. (1996). Declaration of the Independence of Cyberspace.
- Barry, A. (2001). Political machines: governing a technological society. London: *Continuum*.
- Barry, A. (2013). The translation zone: Between actor-network theory and international relations. *Millennium*, 41(3), pp.413-429.
- BBC (2013). Google: Alternatives to the search giant. [online] Available at: http://www.bbc.com/news/technology-23318889 [Accessed 14 April 2021].
- Beer, D. (2016). *Metric power*. London: Palgrave Macmillan. [Accessed 14 April 2021].
- Beer, D. (2017). The social power of algorithms, *Information, Communication & Society*, 20:1, 1-13, DOI: 10.1080/1369118X.2016.1216147
- Beer, D. and Burrows, R. (2007). Sociology and, of and in Web 2.0: Some Initial Considerations. *Sociological Research Online* 12(5). pp. 67-79 Available at: http://www.socresonline.org.uk/12/5/17.html [Accessed 14 April 2021].
- Beetz, J. (2016). *Materiality and subject in Marxism, (Post-) structuralism, and material semiotics*. London: Palgrave.
- Belli, L. et al. (2017). Platform regulations: how platforms are regulated and how they regulate us. *Official Outcome of the UN IGF Dynamic Coalition on Platform Responsibility*. Leeds.
- Bender, E. et al. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜 In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, 610–623. DOI:https://doi.org/10.1145/3442188.3445922
- Bennett, L. W. and Segerberg, A. (2012) The logic of connective action, *Information, Communication & Society*, 15:5, 739-768, DOI: 10.1080/1369118X.2012.670661

- Blei, D., and Lafferty, J. 2009. Topic models. In Srivastava, A., andSahami, M., eds., *Text Mining: Theory and Applications.* Taylorand Francis
- Bliss, L. (2017). Can we protect our female politicians from social media abuse? *Political Studies Association*. Available at: https://www.psa.ac.uk/insight-plus/blog/can-we-protect-our-female-politicians-social-media-abuse [Accessed 14 April 2021].
- Bloomfield, B. P. and Vurdubakis, T. (1999). The outer limits: monsters, actor networks and the writing of displacement. *Organization*, 6(4), pp. 625–647. doi: 10.1177/135050849964004.
- Bohman, J. (2004). Expanding dialogue: The Internet, the public sphere and prospects for transnational democracy. *The sociological review 52*(1_suppl), pp. 131-155. [Accessed 14 April 2021].
- Bostdorff, D.M. (2004). The Internet rhetoric of the Ku Klux Klan: a case study of community building run amok. *Communication Studies*, 55(2): 340–361.
- Boullier, D. (2018). Medialab stories: How to align actor network theory and digital methods. *Big Data & Society, 5*(2), pp. 1-13. [Accessed 14 April 2021].
- Bounegro, L. and Gray, J. (2014). Mapping Issues with the Web: An Introduction to Digital Methods [video online]. *Columbia Journalism School*. Available at https://youtube.com/watch?v=PNbmvYk2sks [Accessed 14 April 2021].
- boyd, d. and Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), pp. 662-679.
- boyd, d. et al. (2010). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. *HICSS*-43. IEEE
- boyd, d., et al (2007). Social network sites as networked publics: Affordances, dynamics, and implications. In Papacharissi, Z. (ed). *A networked self: Identity, community, and culture on social network sites.* Routledge, New York, pp. 39–58.
- Bradley, C. and Wingfield, R. (2018). A Rights-Respecting Model of Online Content Regulation by Platforms. White Paper. [online] *Global Partners Digital.* Available at: https://www.gp-digital.org/wp-content/uploads/2018/05/A-rights-respecting-model-of-online-content-regulation-by-platforms.pdf [Accessed 14 April 2021].
- Bruns, A. and Stieglitz, S. (2014). Twitter data: what do they represent? *It-Information Technology*, 56(5), pp.240-245.

- Bryman, A. (2012). *Social research methods*. Oxford, Oxford Univ. Press.
- Bucher, T. (2018). *If... Then: Algorithmic Power and Politics*. Oxford: Oxford University Press.
- Bunz, M., and Meikle, G. (2018). *The Internet of things*. Cambridge, UK, Polity.
- Burgess, M. (2016) UN warns UK's IP Bill 'undermines' the right to privacy. *Wired*. [online] Available at http://www.wired.co.uk/article/un-privacy-ip-bill-not-compliant-international-law [accessed 06/2016].
- Burnap, P., et al. (2014). Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack. *Social Network Analysis and Mining 4*(1), pp. 206. Available at: https://doi.org/10.1007/s13278-014-0206-4 [Accessed 14 April 2021].
- Butcher, M. (2020). The Real Facebook Oversight Board' launches to counter Facebook's 'Oversight Board'. [online] *Techcrunch*. Available at: https://techcrunch.com/2020/09/30/the-real-facebook-oversight-board-launches-to-counter-facebooks-oversight-board/ [Accessed 14 April 2021].
- Callon, M. (1986a). Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay in J. Law (ed.), *Power, action and belief: a new sociology of knowledge?* London, Routledge pp.196-223.
- Callon, M. (1986b). The Sociology of an Actor-Network: The Case of the Electric Vehicle. In: Callon M., Law J., Rip A. (eds) *Mapping the Dynamics of Science and Technology*. Palgrave Macmillan, London. https://doi.org/10.1007/978-1-349-07408-2_2
- Callon, M. 1998. An essay on framing and overflowing: economic externalities revisited by sociology. *The Sociological Review 46*(1_suppl), pp. 244-269. [Accessed 14 April 2021].
- Callon, M. and Latour, B. (1981). Unscrewing the big Leviathan. In K.D. Knorr Cetina and M. Mulkay (eds), *Advances in Social Theory and Methodology* (pp. 275-303). London: Routledge
- Cammaerts, B. (2009). Radical pluralism and free speech in online public spaces: The case of North Belgian extreme right discourses. *International Journal of Cultural Studies*, 12(6), pp. 555–575. doi: 10.1177/1367877909342479.

- Caplan, R. (2019). Content or context moderation? Artisanal, community-reliant, and industrial approaches [Report]. *Data & Society*. Available at: https://datasociety.net/library/content-or-context-moderation/ [Accessed 14 April 2021].
- Carey, S. (2016). Snooper's Charter: What you need to know about the Investigatory Powers Bill [online] *Computer World*. Available at http://www.computerworlduk.com/security/draft-investigatory-powers-bill-what-you-need-know-3629116/ [Accessed 14 April 2021].
- Carmi, E. (2019). The Hidden listeners: Regulating the line from telephone operators to content moderators. *International Journal of Communication,* 13, 440–458. Available at: https://ijoc.org/index.php/ijoc/article/view/8588/0 [Accessed 14 April 2021].
- Casilli A (2017b). Venture labor: how venture labor sheds light on the digital platform economy. *Int J Commun* 11:4
- Casilli, A. (2017b). Lavoro e capitalismo delle piattaforme. Talk given at the *Centro per la Riforma dello Stato,* Roma, 7 November 2017.
- Castells, M. (1996). *The rise of the network society.* Oxford, UK: Blackwell
- Castells, M. (2008). The New Public Sphere: Global Civil Society, Communication Networks, and Global Governance, *The Annals of the American Academy of Political and Social Science,* 616 (1): 78
- Castells, M. (2013). *Networks of outrage and hope: Social movements in the Internet age*. John Wiley and Sons.
- Cha, M. et al. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. *AAAI Conference on Weblogs and Social Media*. 14.
- Collins, H. M. (1975). The Seven Sexes: A Study in the Sociology of a Phenomenon, or the Replication of Experiments in Physics. *Sociology*. 9:2 205-224
- Collins, H. M. (1985). *Changing Order*. London: Sage.
- Collins, H. M., ed. (1981). Knowledge and controversy: studies of modern natural science. *Social Studies of Science* 11:3-158.
- Copland, S. (2020). Reddit quarantined: can changing platform affordances reduce hateful material online?. *Internet Policy Review,* 9(4), pp 1-26. Available at: https://policyreview.info/articles/analysis/reddit-quarantined-can-changing-platform-affordances-reduce-hateful-material [Accessed: 15 Apr. 2021].

- Cortext documentation (2021a). Demography. Available at: https://docs.Cortext.net/demography/ [Accessed: 15 Apr. 2021].
- Cortext documentation (2021b). Terms extraction. Available at: https://docs.Cortext.net/lexical-extraction/ [Accessed: 15 Apr. 2021].
- Cortext documentation (2021c). Topic Modelling. Available at: https://docs.Cortext.net/analyzing-data/topic-modeling/ [Accessed: 15 Apr. 2021].
- Couldry, N. (2012). *Media, society, world: social theory and digital media practice*. Cambridge, Polity.
- Couldry, N. and Hepp, A. (2018). *The mediated construction of reality*. John Wiley & Sons.
- Couldry, N. and Mejias, U.A. (2020). *The Costs of Connection: How Data Are Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press
- Council of Europe (2016). Council of Europe Secretary General concerned about Internet censorship: Rules for blocking and removal of illegal content must be transparent and proportionate [online]. *Council of Europe.* Available at: https://wcd.coe.int/ViewDoc.jsp?p=&id=2432899&Site=DC&BackColorInternet=F5CA75&BackColorIntranet=F5CA75&BackColorLogged=A9BACE&direct=true [Accessed: 15 Apr. 2021].
- Crawford, K. (2021). *The Atlas of AI.* Yale University Press.
- Crawford, K., and Gillespie, T. (2016). What is a flag for? social media reporting tools and the vocabulary of complaint. *New Media Society* 18(3), 410–428
- Cusumano, M. A. et al. (2021) Can Self-Regulation Save Digital Platforms? *Industrial and Corporate Change*, Available at SSRN: https://ssrn.com/abstract=3900137 or http://dx.doi.org/10.2139/ssrn.3900137
- Dahlberg, L. (2001). The Internet and Democratic Discourse. Exploring the prospects of online deliberative forums extending the public sphere. *Information, Communication & Society* 4:4 2001 615–633
- Daniels, J. (2008). Race, Civil Rights, and Hate Speech in the Digital Era. Learning Race and Ethnicity: Youth and Digital Media in Everett, A. (ed.). *The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning*. Cambridge, MA: The MIT Press. 129–154. doi: 10.1162/dmal.9780262550673.129

- Daniels, J. (2013). Race and racism in Internet Studies: A review and critique. *New Media & Society,* 15(5), pp. 695–719. doi: 10.1177/1461444812462849.
- De Filippi, P. et al. (2020). The Declarations of Cyberspace Outlining three essential narratives in the political history of the Internet. [Blog] *Berkman Klein center,* Available at: <https://medium.com/berkman-klein-center/the-declarations-of-cyberspace-ee0c4499de64> [Accessed 14 April 2021].
- de la Bellacasa, M. P. (2011). Matters of care in technoscience: Assembling neglected things. *Social Studies of Science* 41(1), pp. 85–106. doi: 10.1177/0306312710380301.
- De Nardis L., Hackl A.M. (2015). Internet governance by social media platforms. *Telecommunications Policy* 39 (2015) 761–770
- De Nardis, L. (2012). Hidden Levers of Internet Control. An infrastructure-based theory of Internet governance Information. *Communication & Society*, 15:5, 720-738, DOI: 10.1080/1369118X.2012.659199
- De Nardis, L., (2014). *The global war for internet governance*. Yale University Press.
- Deibert, R.J (2003). Millennium. *Journal of International Studies* 32:501 DOI: 10.1177/03058298030320030801
- Dencik, L. et al. (2016). Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data & Society*. 3. 10.1177/2053951716679678.
- Dencik, L. et al. (2018a). *Data Scores as Governance: Investigating uses of citisen scoring in public services.* [Project report]. Cardiff: Open Society Foundations. Available at: http://orca.cf.ac.uk/117517/ [Accessed: 15 Apr. 2021].
- Dencik, L., et al. (2018b). Prediction, pre-emption and limits to dissent: Social media and big data uses for policing protests in the United Kingdom. *New Media & Society*, 20(4), pp. 1433–1450. doi: 10.1177/1461444817697722.
- Department for Digital, Culture, Media and Sport, and the Home Office (2019). *Online Harms White Paper*. Available at: https://www.gov.uk/government/consultations/online-harms-white-paper [Accessed: 15 Apr. 2021].
- Dewey, J. (1927). *The public and its problems*. Athens, OH: Ohio University Press.
- Diamond, L. J. (2010). *Liberation Technology*. Journal of Democracy. 21:3, pp. 69-83
- Diamond, L. J., and Plattner, M. F. (2012). *Liberation technology: social media and the struggle for democracy.* Baltimore, Md, Johns Hopkins University Press.

- Dingwerth, K. and Pattberg, P. (2006). Global governance as a perspective on world politics. *Global governance: a review of multilateralism and international organizations*, 12(2), pp.185-204.
- Douek, E. (2021). Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability. *Columbia Law Review* Vol. 121, No. 1, 2021 Available at SSRN: https://ssrn.com/abstract=3679607 [Accessed 14 April 2021].
- Dutton, W. (2009). The Fifth Estate Emerging through the Network of Networks. *Prometheus*, 27:1, 1-15, DOI: 10.1080/08109020802657453
- Edri (2016a). EDRi and Access Now withdraw from the EU Commission IT Forum discussions [online]. *Edri*. Available at: https://edri.org/edri-access-now-withdraw-eu-commission-forum-discussions/ [accessed 06/07/2016]
- Edri (2016b). EDRi and Access Now withdraw from the EU Commission IT Forum discussions [online]. *Edri*. Available at: https://edri.org/edri-access-now-withdraw-eu-commission-forum-discussions/ [accessed 06/07/2016]
- Edri (2017). EU action needed: German NetzDG draft threatens freedom of expression. *Edri*. Available at: https://edri.org/our-work/eu-action-needed-german-netzdg-draft-threatens-freedomofexpression/
- Edwards et al. (2013). Digital Social Research, Social Media and the Sociological Imagination: surrogacy, re-orientation and augmentation. *International Journal of Social Research Methodology*. 16(3) 245-260
- Edwards, A. (2016). Multi-centred governance and circuits of power in liberal modes of security, *Global Crime*, 17:3-4, 240-263, DOI: 10.1080/17440572.2016.1179629
- Edwards, A. et al. (2021). Forecasting the governance of harmful social media communications: findings from the digital wildfire policy Delphi. *Policing and Society* 31(1), pp. 1-19. (10.1080/10439463.2020.1839073)
- Einspänner-Pflock, J. et al. (2014). Computer-assisted content analysis of Twitter data. In Weller, K. et al (eds.) *Twitter and Society* (pp. 97-108). New York: P. Lang. Available at: https:// nbn-resolving.org/urn:nbn:de:0168-ssoar-54492-0 [Accessed: 14 Apr. 2021].
- Epstein, D. (2013). The making of institutions of information governance: the case of the Internet Governance Forum. *Journal of Information Technology* 28(2), pp.137-149.

- Epstein, D. et al. (2016). Doing internet governance: practices, controversies, infrastructures, and institutions. *Internet Policy Review* [online] 5(3). Available at: https://policyreview.info/articles/analysis/doing-internet-governance-practices-controversies-infrastructures-and-institutions [Accessed: 14 Apr. 2021].
- ESRC (2021). Internet-mediated research. [online] *ESRC.UKRI.org*. Available at: https://esrc.ukri.org/funding/guidance-for-applicants/research-ethics/frequently-raised-topics/internet-mediated-research/ [Accessed: 14 Apr. 2021].
- Ess, C. and the AoIR ethics working committee (2002). Ethical decision-making and Internet Research: Recommendations from the AoIR Ethics Working Committee. [online]. *AOIR.org.* Available at: http://aoir.org/reports/ethics.pdf [Accessed: 14 Apr. 2021].
- European Commission (2016). European Commission and IT Companies announce Code of Conduct on illegal online hate speech. [online] *European Commission*. Available at: http://europa.eu/rapid/press-release_IP-16-1937_en.htm [Accessed: 14 Apr. 2021].
- Facebook (2020). Welcoming the Oversight Board (06 may 2020) [blog]. *Facebook*. Available at: https://about.fb.com/news/2020/05/welcoming-the-oversight-board/ [Accessed: 14 Apr. 2021].
- Facebook (2021). The Online Civil Courage Initiative (OCCI) [online]. *Facebook*. Available at: https://counterspeech.fb.com/en/initiatives/online-civil-courage-initiative-occi/
- Ferguson, D. and Savage, M. (2020). Controversial Exams Algorithm to Set 97% Of GCSE Results. *The Guardian* [online]. Available at: https://www.theguardian.com/education/2020/aug/15/controversial-exams-algorithm-to-set-97-of-gcse-results. Accessed 27 Apr 2021.
- Flyverbom, M. (2011). *Power of networks: organizing the global politics of the internet.* Cheltenham, Edward Elgar Pub.
- Flyverbom, M. (2016). Disclosing and concealing: internet governance, information control and the management of visibility. *Internet Policy Review* [online] 5(3). Available at: https://policyreview.info/articles/analysis/disclosing-and-concealing-internet-governance-information-control-and-management [Accessed: 14 Apr. 2021].

- Frank Jørgensen, R., et al. (2019). Information and communication technologies (ICT): Exploring the human rights impact of the ICT sector in Götzmann, N.(ed.) *Handbook on Human Rights Impact Assessments*. Elgar Publishing
- Freiburger, T., and Crane, J. (2008). A systematic examination of terrorist use of the internet*. International Journal of Cyber Criminology* 2(1), 309–319.
- Froomkin, M. (2004). Technologies for Democracy in Shane, P. M. Ed. *Democracy online: the prospects for political renewal through the Internet*. New York, Routledge.
- Fuchs, C. (2008). *Internet and Society, Social theory in the information age*. Routledge
- Fuchs, C. (2012). The political economy of privacy on Facebook. *Television & New Media*, 13(2), pp.139-159.
- Fuchs, C. (2014). *Social Media, A Critical Introduction*. London Sage
- Fuchs, C. (2017). From digital positivism and administrative big data analytics towards critical digital and social media research!. *European Journal of Communication*, 32(1), pp.37-49.
- Gerbaudo, P. (2015). Protest avatars as memetic signifiers: political profile pictures and the construction of collective identity on social media in the 2011 protest wave, *Information, Communication & Society*, 18:8, 916-929, DOI: 10.1080/1369118X.2015.1043316
- Gerrard, Y., and Thornham, H. (2020). Content moderation: Social media's sexist assemblages. *New media & society*, 22(7), 1266-1286.
- Gillespie T (2017). Governance of and by platforms in Burgess J, et al. (eds) *Sage handbook of social media*. Sage, London
- Gillespie, T. (2010). The Politics of 'Platforms'. *New Media & Society*, 12(3), 347–364. doi:10.1177/1461444809342738
- Gillespie, T. (2015). 'Platforms Intervene', Social Media + Society. 1(1). doi: 10.1177/2056305115580479.
- Gillespie, T. (2018). *Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media.* New Haven: Yale University Press
- Gillespie, T. et al. (2020). Expanding the debate about content moderation: scholarly research agendas for the coming policy debates. *Internet Policy Review* [online] 9(4).

Available at: https://policyreview.info/articles/analysis/expanding-debate-about-content-moderation-scholarly-research-agendas-coming-policy [Accessed: 13 Apr. 2021].

- Gladwell, M. (2010). Small change: Why the revolution will not be tweeted. *The New Yorker*. Available at: https://www.newyorker.com/magazine/2010/10/04/small-change-malcolm-gladwell
- Gore, S. (2003). A Rose by Any Other Name: Judicial Use of Metaphors for New Technologies. *University of Illinois Journal of Law, Technology, and Policy*, Vol. 403, pp. 425-431, at p. 415.
- Gorwa, R. (2019a). The platform governance triangle: Conceptualising the informal regulation of online content. *Internet Policy Review* 8(2), 1–22. https://doi.org/10.14763/2019.2.1407
- Gorwa, R. (2019b). What is platform governance? *Information, Communication & Society* 22(6), 854–871. doi:10.1080/1369118X.2019.1573914
- Gorwa, R., et al. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7(1). https://doi.org/10.1177/2053951719897945
- Gray, J. (2017). Digital Methods and Public Policy: Tracing Networks, Assemblages and Devices. *Paper presented at the 3rd International Conference on Public Policy (ICPP3)* 28-30th June 2017, Singapore
- Gregory, K. (2017). "Digital Labor or Exploitation?" *Digital Media Seminar Series*. University of Stirling. Stirling, UK.
- Gregory, K. (2018). The future of work. *Paper presented at public services international congress*. Geneva, Switzerland. Available at: http://congress.world-psi. org/karen-gregory-talks-about-the-negatives-and-positives-of-computer-platform-capitalism/ [Accessed: 13 Apr. 2021].
- Gregory, K. 2017. "The Labor of Digital Scholarship." Talk given at the University of Edinburgh. Digital Education Seminar. University of Edinburgh. Edinburgh, UK. Audio and Slides available: https://ed.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=41552549-5650-4cdf-bf62-05999534c270
- Habermas, J. (1989). *The structural transformation of the public sphere*. Cambridge, UK: Polity.

- Hao, K. (2021). Big Tech's guide to talking about AI ethics 50-ish words you can use to show that you care without incriminating yourself. *MIT Technology Review* [online]. Available at: https://www.technologyreview.com/2021/04/13/1022568/big-tech-ai-ethics-guide/ [Accessed: 27 Apr. 2021].
- Helberger, N., et al. (2018). Governing online platforms: From contested to cooperative responsibility. *The information society*, 34(1), pp.1-14.
- Heldt, A. (2018). Von der Schwierigkeit, „fake news" zu regulieren [The Difficulty of Regulating "Fake News"] in JuWiss [Blog] *Junge Wissenschaft im Öffentlichen Recht* No. 71/2018. Hamburg: Bucerius Law School.
- Hintz, A. (2015). Social media censorship, privatised regulation, and new restrictions to protest and dissent. In L. Dencik & O. Leistert (Eds.), *Critical perspectives on social media and protest: Between control and emancipation* (pp. 35–52). Lanham, MD: Rowman & Littlefield.
- Hintz, A. (2016) 'Policy Hacking: Citisen-based Policy-Making and Media Reform', in D. Freedman, J. Obar, C. Martens and R. McChesney (eds) *Strategies for Media Reform*. New York: Fordham University Press.
- Hintz, A. (2016). Restricting digital sites of dissent: commercial social media and free expression. *Critical Discourse Studies*, 13:3, 325-340, DOI: 10.1080/17405904.2016.1141695
- Hintz, A. and Milan, S. (2009). At the margins of Internet governance: grassroots tech groups and communication policy. *International Journal of Media and Cultural Politics*, Vol.5 n1&2
- Hintz, A. and Milan, S. (2013) 'Networked Collective Action and the Institutionalised Policy Debate: Bringing Cyberactivism to the Policy Arena?', *Policy & Internet*, vol. 5, no. 1, pp. 7–26.
- Hoffmann, S. et al. (2019). *The Market of Disinformation*. [Report]. Oxford Information Labs; Oxford Technology & Elections Commission, University of Oxford. Retrieved from https://oxtec.oii.ox.ac.uk/wpcontent/uploads/sites/115/2019/10/OxTEC-The-Market-of-Disinformation.pdf [Accessed: 13 Apr. 2021].
- Hofmann, J. (2016). Multi-Stakeholderism in Internet Governance. Putting a Fiction into Practice. *Journal of Cyber Policy* Vol. 1, No. 1, S. 29-49.

- Hofmann, J., et al. (2016). Between coordination and regulation: Finding the governance in Internet Governance. *New Media & Society*, 19(9), pp.1406-1423.
- Housley, W. et al. (2014). Big and broad social data and the sociological imagination: a collaborative response. *Big Data and Society* 1(2)
- Housley, W. et al. (2018). Interaction and Transformation on Social Media: The Case of Twitter Campaigns. *Social Media + Society*, 4(1) doi: 10.1177/2056305117750721.
- Huey, L., (2015). This is Not Your Mother's Terrorism: Social Media, Online Radicalization and the Practice of Political Jamming. *Journal of Terrorism Research*, 6(2). DOI: http://doi.org/10.15664/jtr.1159
- Huszti-Orban, K. (2018). Internet intermediaries and counter-terrorism: Between self-regulation and outsourcing law enforcement. *IEEE*, pp. 227-244. 10.23919/CYCON.2018.8405019.
- Iliadis, A. and Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*. 3(2) doi: 10.1177/2053951716674238.
- Ippolita (2015). The Facebook Aquarium: The Resistible Rise of Anarcho-Capitalism (PDF) in Rasch, M. (ed.) *Theory on Demand, #15*, Amsterdam: Institute of Network Cultures
- Isin, E. F., & Ruppert, E. S. (2015). *Being digital citizens*. London, Rowman & Littlefield
- Jacomy, M. et al. (2016). Hyphe, a Curation-Oriented Approach to Web Crawling for the Social Sciences. *Proceedings of the International AAAI Conference on Web and Social Media,* 10(1). Available at: https://ojs.aaai.org/index.php/ICWSM/article/view/14777
- Jørgensen, R.F. and Zuleta, L., (2020). Private Governance of Freedom of Expression on Social Media Platforms. Nordicom Review, 41(1), pp.51-67.
- Katzenbach, C. (2012). Technologies as institutions: Rethinking the role of technology in media governance constellations. *Trends in communication policy research: new theories, methods and subjects, Intellect*.
- Katzenbach, C. (2013). Media governance and technology. *Routledge handbook of media law*, p.399.
- Kaye, D. (2018). *A Human Rights Approach to Platform Content Regulation* [Report]. New York: United Nations. Available at: https://freedex.org/a-human-rights-approach-to-platform-content-regulation/ [Accessed 14 April 2021].

- Kitchin, R, and Lauriault, T. (2014). Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work. The Programmable City Working Paper 2; pre-print version of chapter to be published in Eckert, J., et al. (eds) *Geoweb and Big Data*. University of Nebraska Press. Available at SSRN: https://ssrn.com/abstract=2474112 [Accessed 14 April 2021].
- Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences.* SAGE Publications.
- Klauser, F. (2009), Interacting forms of expertise in security governance: the example of CCTV surveillance at Geneva International Airport1. *The British Journal of Sociology*, 60: 279-297. https://doi.org/10.1111/j.1468-4446.2009.01231.x
- Klonick, K. (2017). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review* 131(6), 1598–1670. Available at: https://harvardlawreview.org/2018/04/the-new-governors-the-people-rules-and-processes-governing-online-speech/ [Accessed 14 April 2021].
- Kwak, H. et al. (2010). What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web (WWW '10)*. Association for Computing Machinery, New York, NY, USA, 591–600. https://doi.org/10.1145/1772690.1772751
- Latimer, J. and Munro, R. (2006). Driving the Social. *The Sociological Review*, 54(1_suppl), pp. 32–53. doi: 10.1111/j.1467-954X.2006.00636.x.
- Latour B. (1986). The powers of association in Law J (ed.) *Power, Action and Belief: A New Sociology of Knowledge?* London: Routledge.
- Latour, B. (2005a). *What is the style of matters of concern? Two Lectures in Empirical Philosophy. Koninklijke Van Gorcum.*
- Latour, B. (2005b). *Reassembling the social an introduction to actor-network-theory*. Oxford, Oxford University Press. http://site.ebrary.com/id/10233636.
- Law, J. (2008). On Sociology and STS. *The Sociological Review* 56(4), pp. 623–649. doi: 10.1111/j.1467-954X.2008.00808.x.
- Law, J. and Ruppert, E. (2013). The Social Life of Methods: Devices, *Journal of Cultural Economy,* 6:3, 229-240, DOI: 10.1080/17530350.2013.812042
- Law, J. et al. (2011). The Double Social Life of Methods. *CRESC Working paper n.95*

- Levi-Faur, D. (2012). From "Big Government" to "Big Governance"? *The Oxford Handbook of Governance.* DOI: 10.1093/oxfordhb/9780199560530.013.0001.
- Levy, S. (1994). *Hackers: Heroes of the Computer Revolution.* NY: Dell Publishing
- Lotan, G. et al. (2011). The Revolutions Were Tweeted: Information Flows during the 2011 Tunisian and Egyptian Revolutions. *International Journal of Communications* 5, Feature 1375:1405.
- Luokkanen, M., et al. (2014). Geoengineering, news media and metaphors: Framing the controversial. *Public Understanding of Science*, 23(8), pp. 966–981. doi: 10.1177/0963662513475966.
- Lupton, D. (2015). *Digital sociology*. London: Routledge
- Lupton, D. (2016) The diverse domains of quantified selves: self-tracking modes and dataveillance, *Economy and Society*, 45:1, 101-122, DOI: 10.1080/03085147.2016.1143726
- Lupton, D. (2018). How do data come to matter? Living and becoming with personal data. *Big Data & Society*. doi: 10.1177/2053951718786314.
- Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society*. doi: 10.1177/2053951714541861.
- MacKinnon, R. (2012). *Consent of the Networked: The Worldwide Struggle for Internet Freedom.* Basic Books, Inc. New York USA
- MacKinnon, R., et al. (2014). Fostering freedom online: the role of internet intermediaries. [online] UNESCO Publishing. Available at: http://unesdoc.unesco.org/images/0023/002311/231162e.pdf [Accessed 14 April 2021].
- Madsen, A.R. (2012). Web-Visions as Controversy-Lenses. *Interdisciplinary Science Reviews* 37(1), 51-68.
- Maireder, A. and Ausserhofer, J. (2014). Political discourses on Twitter: networking topics, objects and people in Weller, K. et al. eds. *Twitter and Society*. New York, NY: Peter Lang
- Malcomson, S. L. (2016). *Splinternet: how geopolitics and commerce are fragmenting the World Wide Web.* New York and London: OR Books, 2016. 202 pp.
- Margetts, H., et al. (2015). *Political Turbulence: How Social Media Shape Collective Action.* Princeton University Press.

- Marres, N. (2005). Issues spark a public into being. A key but often forgotten point of the Lippmann-Dewey debate in Latour, B. and Weibel, P. (eds.), *Making things public*, Cambridge, MA: MIT Press, 208–217.
- Marres, N. (2015). Why Map Issues? On Controversy Analysis as a Digital Method. *Science, Technology, & Human Values* Vol. 40(5) 655-68
- Marres, N. (2017). *Digital sociology: the reinvention of social research*. Malden, MA : Polity
- Marres, N. and Gerlitz, C. (2016). Interface Methods: Renegotiating Relations between Digital Social Research, STS and Sociology. *The Sociological Review*, 64(1), pp. 21–46. doi: 10.1111/1467-954X.12314.
- Marres, N. and Rogers, R. (2005). Recipe for tracing the fate of issues and their publics on the web in Latour, B. and Weibel, P. (eds.) *Making things public*, Cambridge, MA: MIT Press, 922–33.
- Marwick, A. and boyd, D. (2011a). To see and be seen: Celebrity practice on Twitter. *Convergence*, 17(2), pp.139-158.
- Marwick, A.E. and boyd, D., (2011b). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society*, 13(1), pp.114-133.
- Mayer-Schonberger, V. and Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think.* Boston, MA: Houghton Mifflin Harcourt. 242 pp. ISBN 978-0544002692
- Mayntz, R. (1998). New challenges to governance theory, Florence, European University Institute, *Jean Monnet Chair Papers, 50*. Available at:: http://hdl.handle.net/1814/23653 [Accessed 14 April 2021].
- Mayntz, R., (2003). From government to governance: Political steering in modern societies. *Summer Academy on IPP*, pp.7-11.
- McGoogan, C. (2015). French 'state of emergency' includes website bans, social media blocks [online]. *Wired*. Available at: http://www.wired.co.uk/article/french-emergency-powers-social-media [accessed: 7/06/2016]
- McLaughlin, L. and Pickard, V. (2005) ‘What is bottom-up about global internet governance?’, *Global Media and Communication*, 1(3), pp. 357–373. doi: 10.1177/1742766505058129.

- McLean, C. and Hassard, J. (2004). Symmetrical Absence/Symmetrical Absurdity: Critical Notes on the Production of Actor-Network Accounts. *Journal of Management Studies*, 41: 493-519. https://doi.org/10.1111/j.1467-6486.2004.00442.x
- Meddaugh, P.M. and Kay, J. (2009). Hate Speech or "Reasonable Racism?" The Other in Stormfront. *Journal of Mass Media Ethics*, 24:4, 251-268, DOI: 10.1080/08900520903320936
- Metcalf, J. and Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society* 3(1), p.2053951716650211.
- Mezei, P. and Verteș-Olteanu, A. (2020). From trust in the system to trust in the content. *Internet Policy Review* [online] 9(4). Available at: https://policyreview.info/trust-system [Accessed: 14 Apr. 2021].
- Michael, M. (2017). *Actor-Network Theory: Trials, Trails and Translations*. London: SAGE Publications Ltd. Available at: http://www.doi.org/10.4135/9781473983045 [Accessed 15 Apr 2021].
- Milan, S. (2015). When Algorithms Shape Collective Action: Social Media and the Dynamics of Cloud Protesting. *Social Media + Society* 1(2) doi: 10.1177/2056305115622481.
- Milan, S. and Oever, N. (2017) Coding and encoding rights in internet infrastructure. *Internet Policy Review* 6(1): 1–17.
- Milan, S. and ten Oever, N. (2017). Coding and encoding rights in internet infrastructure. *Internet Policy Review* [online] 6(1). Available at: https://policyreview.info/articles/analysis/coding-and-encoding-rights-internet-infrastructure [Accessed: 27 Apr. 2021].
- Mol, A. (1999). Ontological politics: a word and some questions. In Law, J. and Hassard, J., eds. *Actor network theory and after*, Oxford: Blackwell, 74–89.
- Mol, A. (2008). The logic of care: Health and the problem of patient choice. Routledge.
- Montenegro Meyer, L. and Bulgakov S. (2014). Reflection on ActorNetwork Theory, Governance Networks, and Strategic Outcomes. *Brazilian Administration Review*, V. 11 (1), art 6, 107-124Available at: http: //www.anpad.org.br/bar.

Morozov, E. (2009). The brave new world of slacktivism. *Foreign Policy* 19(05). Available at: http://neteffect.foreignpolicy.com/posts/2009/05/19/the_brave_new_world_of_slacktivism [Accessed 14 April 2021].

- Morozov, E. (2010). *The net delusion: the dark side of Internet freedom*. New York, NY, Public Affairs.
- Mueller M. L. (2002). *Ruling the Root: Internet Governance and the Taming of Cyberspace* Cambridge, MA: The MIT Press.
- Mueller, M. (2010). *Networks and states the global politics of Internet governance.* Cambridge, Mass, MIT Press.
- Mueller, M. (2015). Hyper-transparency and social control: Social media as magnets for regulation. *Telecommunications Policy*. 39(9), pp.804-810.
- Müeller, M. (2016). Assemblages and Actor-networks: Rethinking Socio-material Power, Politics and Space. *Geography Compass* 9/1 (2015): 27–41, 10.1111/gec3.12192
- Muniesa. F, (2015). Actor-Network Theory. *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition), Elsevier
- Munk, A.K., et al. (2019). Data sprints: A collaborative format in digital controversy mapping. *Digital-STS: A Field Guide for Science & Technology Studies*, p.472.
- Munk, A.L. and Abrahamsson, S. (2012). Empiricist Interventions: Strategy and Tactics on the Ontopolitical Battlefield. *Science Studies*. 25. 52-70. 10.23987/sts.55281.
- Musiani, F. (2015). Practice, Plurality, Performativity, and Plumbing: Internet Governance Research Meets Science and Technology *Studies. Science, Technology & Human Values*, 40(2), 272–288.
- Napoli, P. and Caplan, R. (2017). Why Media Companies Insist They're not Media Companies, Why They're Wrong, and Why it Matters. *First Monday*. 22. 10.5210/fm.v22i5.7051.
- O' Neill, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* New York: Crown Publishers
- O'Riordan, K. (2016). *Feminist data visualisation*. Paper presented at the AoIR conference (October 2016), Berlin.

- Ooghe-Tabanou, B. (2018). Hyperlink is not dead! In Proceedings of the 2nd International Conference on Web Studies (WS.2 2018). *Association for Computing Machinery,* New York, NY, USA, 12–18.
- Oolo, E., and Siibak, A. (2013). Performing for one's imagined audience: Social steganography and other privacy strategies of Estonian teens on networked publics. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 7(1), Article 7. https://doi.org/10.5817/CP2013-1-7
- Owen, T. (2019). The Case for Platform Governance. *CIGI Papers No. 231* — November 2019. Available at: https://www.cigionline.org/sites/default/files/documents/Paper%20no.231web.pdf [Accessed 14 April 2021].
- Özarslan, Z. (2014). "Introducing Two New Terms into the Literature of Hate Speech: 'Hate Discourse' and 'Hate Speech Act' Application of 'speech act theory' into hate speech studies in the era of Web 2.0." Galatasaray Üniversitesi İleti-ş-Im Dergisi. (20), pp.53-75.
- Padovani, C. and Santaniello, M. (2018) Digital constitutionalism: Fundamental rights and power limitation in the Internet eco-system. *the International Communication Gazette* 2018, Vol. 80(4) 295–301
- Palladino, N. 2021 The role of epistemic communities in the "constitutionalization" of internet governance: The example of the European Commission High-Level Expert Group on Artificial Intelligence *Telecommunications Policy* 45 (2021) 102149
- Papacharissi, Z. (2002). The virtual sphere: The Internet as a public sphere. *New Media Society* vol 4 (1): 9-27, Sage Publications, London, Thousand Oaks, CA
- Pasquale, F. (2016a) Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power. *Theoretical Inquiries in Law*, 17(2), pp.487-513 Available at SSRN: https://ssrn.com/abstract=2779270
- Pasquale, F. (2016b). *The black box society: the secret algorithms that control money and information.* Harvard University Press
- Peng Hwa, A. (2008). International Regulation of Internet Content: Possibilities and Limits. In Drake, W. J. and Wilson, E. J. (eds). *Governing global electronic networks international perspectives on policy and power.* Cambridge, Mass, MIT Press.

- Perry, B. and Olsson, P. (2009). Cyberhate: the globalization of hate. *Information & Communications Technology Law,* 18:2, 185-199, DOI: 10.1080/13600830902814984
- Pickard, V. (2013). Social democracy or corporate libertarianism? Conflicting media policy narratives in the wake of market failure. Communication Theory, 23(4), pp.336-355.
- Pickard, V. (2015). *America's battle for media democracy: The triumph of corporate libertarianism and the future of media reform*. New York: Cambridge University Press.
- Pickard, V. (2016). Media Failures in the Age of Trump. The Political Economy of Communication, 4 (2), 118-122. Available at: https://repository.upenn.edu/asc_papers/753 [Accessed 14 April 2021].
- Poell, T., et al. (2017). The platformization of cultural production. In 18th *Annual conference of the Association of Internet Research,* Tartu, Estonia (pp. 18-21).
- Pohle, J. (2016a). *Information For All? The emergence of UNESCO's policy discourse on the information society (1990-2003).* PhD thesis in Communication Studies, Vrije Universiteit Brussel
- Pohle, J. (2016b). Multistakeholder governance processes as production sites: enhanced cooperation "in the making". *Internet Policy Review* 5(3)
- Pohle, J., et al. (2017). Analysing internet policy as a field of struggle. *Internet Policy Review,* 5
- Poletti, C. and Gray, D. (2016). Good data is critical data: an appeal for critical digital studies. Good data, 64, p.260.
- Poletti, C. and Michieli, M. (2018). Smart cities, social media platforms and security: online content regulation as a site of controversy and conflict. *City, Territory and Architecture,* 5(1), pp.1-14.
- Procter R, et al. (2013a). Reading the riots: what were the police doing on Twitter? *Polic Soc* 23(4):1–24. doi:10.1080/10439463.2013.780223
- Procter R, et al. (2013b). Reading the riots on Twitter: methodological innovation for the analysis of big data. *Int J Soc Res Methodol* 16(3):197–214. doi:10.1080/13645579.2013.774172

- Procter, R. et al. (2019). A Study of Cyber Hate on Twitter with Implications for Social Media Governance Strategies *arXiv preprint arXiv:1908.11732*. Available at: https://arxiv.org/abs/1908.11732 (Accessed: 19 April 2021).
- Puschmann, C. and Burgess, J. (2013). The Politics of Twitter Data. *SSRN Electronic Journal.* 10.2139/ssrn.2206225.
- Radu, R. (2019). *Negotiating internet governance*. Oxford University Press http://www.oapen.org/download?type=document&docid=1004863.
- Radu, R. and Chenou, J. (2015). Data control and digital regulatory space(s): towards a new European approach. *Internet Policy Review*, 4(2). DOI: 10.14763/2015.2.370
- Radu, R. et al. (2021) Normfare: Norm entrepreneurship in internet governance *Telecommunications Policy* 45 (2021) 102149
- Rahman, K. S. (2018). The new utilities: Private power, social infrastructure, and the revival of the public utility concept. *Cardozo Law Review*, 39(5), 1621–1689. http://cardozolawreview.com/wp-content/uploads/2018/07/RAHMAN.39.5.2.pdf
- Redden, J. (2018). Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society*, 5(2), p.2053951718809145.
- Rhodes, R. A. W. (1996). The New Governance: Governing without Government. *Political Studies*, 44(4), pp. 652–667. doi: 10.1111/j.1467-9248.1996.tb01747.x.
- Richterich, A. (2018). *The big data agenda: Data ethics and critical data studies* (p. 154). University of Westminster Press.
- Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media.* Yale University Press.
- Rogers, R. (2009*). The end of the virtual: digital methods*. Amsterdam, Vossiuspers UvA.
- Rogers, R. (2013a). *Digital methods.* Cambridge, Massachusetts, The MIT Press. http://site.ebrary.com/id/10722729.
- Rogers, R. (2013b). Mapping public Web space with the Issuecrawler. Digital cognitive technologies: *Epistemology and the knowledge economy*, 89–99.
- Rogers, R. (2017). *Digital methods for cross-platform analysis.* The SAGE handbook of social media (2017), 91–110.

- Rogers, R. and Zelman, A. (2002). Surfing for knowledge in the information society. In: Elmer G (ed) *Critical perspectives on the internet.* Rowman & Littlefield, Lanham, pp 63–86
- Rogers, R., et al. (2015). *Issue mapping for an ageing Europe.* Amsterdam, Amsterdam University Press.
- Ruppert, E., et al. (2013). Reassembling social science methods: The challenge of digital devices. *Theory, culture & society*, 30(4), pp.22-46.
- Savage, M. and Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology* 41(5): 885–899.
- Schmidt, H. (2014). Twitter and the rise of personal publics. In Weller K. et al. (eds.) *Twitter and Society*, New York, Peter Lang
- Schouten, P. (2014). Security as controversy: Reassembling security at Amsterdam Airport. *Security Dialogue* Vol. 45(1) 23–42
- Seemab Latif, Z. et al. (2021) Analyzing LDA and NMF Topic Models for Urdu Tweets via Automatic Labeling. *IEEE Access*
- Shields, J. (2017). *Countering online radicalisation and extremism: Baroness Shields' speech.* [online] GOV.UK. Available at: https://www.gov.uk/government/speeches/countering-online-radicalisation-and-extremism-baroness-shields-speech [Accessed 16 April 2021].
- Shirky, C. (2008). *Here comes everybody: the power of organizing without organizations.* New York, Penguin Books. http://rbdigital.oneclickdigital.com.
- Sievert and Shirley, 2014, *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pages 63–70, Baltimore, Maryland, USA, June 27, 2014.
- Sinnreich, A. (2018). Four crises in algorithmic governance. In Joerden, J.C. et al. (eds.), *Annual Review of Law and Ethics* (Vol. 26, pp. 190–199).
- Sinnreich, A. (2020). Moderation, community, and democracy. Democracy cannot survive algorithmic content moderation in Gillespie, T. et al. (2020). Expanding the debate about content moderation: scholarly research agendas for the coming policy debates. *Internet Policy Review* 9(4). Available at: https://policyreview.info/articles/analysis/expanding-

debate-about-content-moderation-scholarly-research-agendas-coming-policy [Accessed: 14 Apr. 2021].

- Snee, H. (2013). Making Ethical Decisions in an Online Context: Reflections on Using Blogs to Explore Narratives of Experience. *Methodological Innovations Online*, 8(2), pp. 52–67. doi: 10.4256/mio.2013.013.
- Srnicek, N. (2017). *Platform capitalism*. John Wiley & Sons.
- Stengers, I. (2005). The cosmopolitical proposal. In Latour, B. and Weibel, P. (eds.) *Making things public*. Cambridge, MA:MIT press
- Sun, B. and Ng, V.T., (2012). Identifying influential users by their postings in social networks. In Ubiquitous social media analysis (pp. 128-151). Springer, Berlin, Heidelberg.
- Suzor, N., et al. (2019). What Do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation. *International Journal Of Communication,* 13, 18.
- Tambini, D. et al. (2008). The privatisation of censorship: self regulation and freedom of expression. In: Tambini, D. et al. *Codifying cyberspace: communications self-regulation in the age of internet convergence.* Routledge / UCL Press, Abingdon, UK., pp 269-289. ISBN 9781844721443
- Thelwall, M. (2014). Sentiment analysis and time series with Twitter. In Weller, K. et al.(eds.) *Twitter and society*, pp.83-95.
- Tucker, J. et al. (2017). From Liberation to Turmoil: Social Media And Democracy. *Journal of Democracy,* vol. 28 no. 4, p. 46-59
- Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *Proceedings of the 8th International Conference on Weblogs and Social Media,* ICWSM 2014.
- Tufekci, Z. (2014). Engineering the public: Big data, surveillance and computational politics, Arun R. et al. (2010) On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In: Zaki M.J., Yu J.X., Ravindran B., Pudi V. (eds) *Advances in Knowledge Discovery and Data Mining*. PAKDD 2010. Lecture Notes in Computer Science, vol 6118. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13657-3_43
- Twitter (2016). *Announcing the Twitter Trust & Safety Council.* [blog]. Twitter. Available at: https://blog.twitter.com/en_us/a/2016/announcing-the-twitter-trust-safety-council.html

- Twitter (2019). *Trust and Safety Council.* [blog] Twitter. Available at: https://about.twitter.com/en/our-priorities/healthy-conversations/trust-and-safety-council
- Ullmann, S., and Tomalin, M. (2020). Quarantining online hate speech: technical and ethical perspectives. *Ethics Inf Technol* 22, 69–80. https://doi.org/10.1007/s10676-019-09516-z
- United Nations and Kaye, D. (2016). Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. [report] Geneva: Human Rights Council. Available at: http://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/Privatesectorinthedigitalage.aspx?platform=hootsuite [Accessed 14 April 2021].
- United Nations and Kaye, D. (2017). Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression : note / by the Secretary-General. [report] New York:, United Nations. Available at: https://digitallibrary.un.org/record/1304394?ln=en [Accessed 14 April 2021].
- United Nations and Kaye, D. (2018a). Report on Artificial Intelligence technologies and implications for freedom of expression and the information environment. [report] Geneva: Human Rights Council. Available at: <https://www.undocs.org/A/HRC/38/35> [Accessed 13 April 2021].
- United Nations and Kaye, D. (2018b). Report on content regulation. [report] Geneva: Human Rights Council. Available at: <https://www.undocs.org/A/HRC/38/35> [Accessed 13 April 2021].
- van Dijck, J. (2013). Facebook and the engineering of connectivity: A multi-layered approach to social media platforms. *Convergence: The International Journal of Research into New Media Technologies* 19(2) 141-155
- van Dijck, J. (2014). Datafication, Dataism and Dataveillance: Big Data Between Scientific Paradigm and Ideology. *Surveillance & Society,* 12(2), pp. 197-208.
- van Dijck, J. and Rieder, B. (2019). The recursivity of internet governance research. *Internet Policy Review*, 8(2), pp.1-10.
- van Dijck, J. et al. (2019). Reframing platform power. *Internet Policy Review* 8(2). Available at: https://policyreview.info/articles/analysis/reframing-platform-power [Accessed: 14 Apr. 2021].

- van Dijck, J., et al. (2018). *The Platform Society. Public values in a connective world*. New York: Oxford University Press.
- van Doorn, N. (2017). Platform labor: on the gendered and racialised exploitation of low-income service work in the 'on-demand' economy, *Information, Communication & Society,* 20:6, 898-914, DOI: 10.1080/1369118X.2017.1294194
- van Eeten
- Venturini, T. (2010). Diving in magma: how to explore controversies with actor-network theory. *Public Understanding of Science*, 19(3), 258–273. doi:10.1177/0963662509102694
- Venturini, T. (2012). Building on faults: how to represent controversies with digital methods. *Public Understanding of Science,* 21(7), 796 – 812. doi:10.1177/0963662510387558
- Venturini, T. and Guido, D. (2012). Once Upon a Text : an ANT Tale in Text Analysis. *Sociologica*. 3. 10.2383/72700.
- Venturini, T. et al. (2015). Designing Controversies and Their Publics. *Design Issues* (Volume:31 , Issue: 3)
- Venturini, T. et al. (2017). Visual Network Exploration for Data Journalists. In The *Routledge Handbook of Developments in Digital Journalism Studies* Rochester, NY.
- Vincent, M. (2015). *To the Cloud: Big Data in a Turbulent World*. Boulder, London: Paradigm
- Wagner, B. (2013). Governing Internet Expression: How public and private regulation shape expression governance. *Journal of Information Technology & Politics*, 10(4), 389–403. doi:10.1080/19331681.2013.799051
- Wagner, B. (2016). *Global Free Expression - Governing the Boundaries of Internet Content.* Cham: Springer.
- Webb et al. (2015). 'Digital Wildfires': a challenge to the governance of social media? *Proceedings of the ACM web science conference*, pp. 1-2.
- Weller, K., et al. (2011). Citation analysis in Twitter: Approaches for defining and measuring information flows within tweets during scientific conferences. In M. Rowe, M. Stankovic, et al. (eds.). *Making sense of microposts* (#MSM2011), workshop at extended semantic web conference (pp. 1–12). Greece: Crete.

- Woods, L. (2019). The duty of care in the Online Harms White Paper, *Journal of Media Law*, 11:1, 6-17, DOI: 10.1080/17577632.2019.1668605
- Wu, T. (2010) *The Master Switch: The Rise and Fall of Information Empires*, Alfred A. Knopf: New York
- Young, K. (2013). Researching Young People's Online Spaces, in Riele, K. and Brooks, R. (eds) *Negotiating Ethical Challenges in Youth Research*. New York: Routledge.
- Zhao, W., Chen, J. J., Perkins, R., Liu, Z., Ge, W., Ding, Y., & Zou, W. (2015). A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC bioinformatics, 16 Suppl 13*(Suppl 13), S8. https://doi.org/10.1186/1471-2105-16-S13-S8
- Ziewitz, M. (2016). Governing Algorithms: Myth, Mess, and Methods. *Science, Technology, & Human Values*, 41(1), pp. 3–16. doi: 10.1177/0162243915608948.
- Ziewitz, M. and Pentzold, C. (2014). In search of internet governance: Performing order in digitally networked environments. *New Media & Society* 16(2):306-322.
- Zubiaga, A. et al. (2016). Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads. *PLoS ONE* 11(3): e0150989. https://doi.org/10.1371/journal.pone.0150989
- Zuboff, S. (2019). *The Age of Surveillance Capitalism. The fight for a human future and the new frontier of power*. New York, Public Affairs

References to web documents (chapter 5)

- Article 19 (2018). *Self-regulation and 'hate speech' on social media platforms*. [online]. Article19.org. Available at: https://www.article19.org/resources/self-regulation-hate-speech-social-media-platforms/ [Accessed 16 April 2021].
- Barnett, J. (2016). *Why Am I Blocked On Facebook?* [Blog] Moronwatch.net. Available at: https://web.archive.org/web/20161108142648/http://moronwatch.net/2016/01/blocked-on-facebook.html
- Bartlett, J. and Reinolds, L. (2015*) State of the Art report 2015*. [online] Demos.co.uk . Available at: https://demos.co.uk/project/state-of-the-art-2015/
- Bliss, L., (2017). *Can we protect our female politicians from social media abuse?* [online] The Political Studies Association (PSA). Available at: <https://www.psa.ac.uk/psa/news/can-we-protect-our-female-politicians-social-media-abuse> [Accessed 16 April 2021].
- Carleton-Taylor, R. (2018). *Hate Speech Is Not Free Speech.* [online] Resistinghate.org. Available at: <https://resistinghate.org/hate-speech-is-not-free-speech/>
- Council of Europe (2018). *Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries.* [online] Coe.int. Available at: https://rm.coe.int/1680790e14 [Accessed 16 April 2021].
- Crown Prosecution Service (2016). *Consultation Interim Revised CPS Guidelines on Prosecuting Social Media Cases* [online] cps.gov.uk. Available at: https://web.archive.org/web/20171118140730/http://www.cps.gov.uk/consultations/social_media_consultation_2016.html (Accessed: 16 April 2021).
- Dartington Primary and Nursery School (2015). *Preventing Radicalisation Policy*. [online] dartington.devon.sch.uk. Originally available at: http://www.dartington.devon.sch.uk/policies/Preventing%20Radicalisation%20Policy.pdf [archived]
- Dencik, L., et al. (2018). Prediction, pre-emption and limits to dissent: Social media and big data uses for policing protests in the United Kingdom. *New Media & Society*, 20(4), pp. 1433–1450. doi: 10.1177/1461444817697722.

- Dent, J. and Strickland, P. (2017). *Online harassment and cyber bullying*, House of Commons Library. Available at: https://commonslibrary.parliament.uk/research-briefings/cbp-7967/ (Accessed: 16 April 2021).
- Doughty, S. (2012). Accountant wins appeal against conviction for airport bomb Tweet after judge realises it was a JOKE. *Daily Mail*, [online] Available at: <https://www.dailymail.co.uk/news/article-2179782/Twitter-joke-trial-Paul-Chambers-wins-appeal-conviction-airport-bomb-Tweet.html> [Accessed 16 April 2021].
- EDRi, (2016a). *Annual Report 2016.* [online] EDRi.org. Available at: <https://edri.org/files/edri_annual_report_2016.pdf> [Accessed 16 April 2021].
- EDRi, (2016b). *Freedom not to be manipulated*. [online] EDRi.org. Available at: <https://edri.org/our-work/freedom-not-to-be-manipulated/> [Accessed 16 April 2021].
- EDRi, (2017). *Recommendations on the German bill "Improving Law Enforcement on Social Networks"* (NetzDG). [online] EDRi.org. Available at: <https://edri.org/files/consultations/tris_netzdg_edricontribution_20170620.pdf> [Accessed 16 April 2021].
- England, C. (2017). Who he? [Blog] *England's England*. Available at: <https://christopherengland.com/who/> [Accessed 16 April 2021].
- European Commission (2016). *European Commission and IT Companies announce Code of Conduct on illegal online hate speech* [online] europa.eu. Available at: https://ec.europa.eu/commission/presscorner/detail/en/IP_16_1937 (Accessed: 16 April 2021).
- European Commission (2018). *Countering online hate speech – Commission initiative with social media platforms and civil society shows progress* [online] europa.eu. Available at: <https://ec.europa.eu/commission/presscorner/detail/en/IP_17_1471> (Accessed: 16 April 2021).
- European Platform of Regulatory Authorities (2018). *Coe recommendation on the roles and responsibilities of internet intermediaries.* [online] EPRA.org. Available at: https://www.epra.org/news_items/coe-recommendation-on-the-roles-and-responsibilities-of-internet-intermediaries
- Felle, T. (2017). Facebook's 'fake news' plan is doomed to failure – social media must do more to counter disinformation. [Blog] *Digital Information Initiative* - City, University of

London, Available at: <https://blogs.city.ac.uk/digitalnews/2017/04/11/facebooks-fake-news-plan-is-doomed-to-failure-social-media-must-do-more-to-counter-disinformation/> [Accessed 16 April 2021].

- Guadamuz, A. (2017). US Border agents to start collecting social media data. [Blog] *TechnoLlama*. Available at: <https://www.technollama.co.uk/us-border-agents-to-start-collecting-social-media-data> [Accessed 16 April 2021].
- Hancock, M. (2017*). The Future Of The Internet: Freedom In A Framework. Minister for Digital Matt Hancock addresses the Internet Governance Forum* [speech] GOV.UK. Available at: https://www.gov.uk/government/speeches/the-future-of-the-internet-freedom-in-a-framework. [Accessed 16 April 2021].
- Helberg, N. (2016). *Facebook is a new breed of editor: a social editor.* [online] LSE Media Policy Project blog. Available at: https://web.archive.org/web/20190924185131/https://blogs.lse.ac.uk/mediapolicyproject/2016/09/15/facebook-is-a-new-breed-of-editor-a-social-editor/
- Huey, L. (2015). This is Not Your Mother's Terrorism: Social Media, Online Radicalization and the Practice of Political Jamming. *Journal of Terrorism Research*, 6(2). DOI: http://doi.org/10.15664/jtr.1159
- Human Right Watch. (2018). *Germany: Flawed Social Media Law NetzDG is Wrong Response to Online Abuse.* [online] hrw.org. Available at: https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law
- Index of Censorship and Ginsberg, J. (2015). *Fighting to speak freely: balancing privacy and free expression in the information age*. [online] Index on Censorship. Available at: <https://www.indexoncensorship.org/2015/10/fighting-to-speak-freely-balancing-privacy-and-free-expression-in-the-information-age/> [Accessed 16 April 2021].
- Internet Society (2018). *Future Thinking: Harlem Désir on Freedom of Expression Online*. [online] Internetsociety.org. Available at:<https://www.internetsociety.org/blog/2018/01/future-thinking-harlem-desir/>
- King, R. (2015). Friends don't share data about friends. [Blog] *Richard Kingsdom*, Available at: <https://richardskingdom.net/friendsdontsharedataaboutfriends> [Accessed 16 April 2021].

- Koene, A., et al. (2017). Editorial responsibilities arising from personalization algorithms. *ORBIT Journal*, 1(1).[online] Available at: https://doi.org/10.29297/orbit.v1i1.26
- Kuerbis, B. (2017). *After Charlottesville: Registrars, content regulation and domain name policy*. [online] Internet Governance Project - Georgia Tech School of Public Policy, Available at: <https://www.internetgovernance.org/2017/08/30/after-charlottesville-registrars-content-regulation-and-domain-name-policy/> [Accessed 16 April 2021].
- Maréchal, N. (2016). Automation, algorithms, and politics| when bots tweet: Toward a normative framework for bots on social networking sites (feature). *International Journal of Communication*, 10, p.10.
- Murthy, D., et al. (2016). Automation, algorithms, and politics| Bots and political influence: A sociotechnical investigation of social network capital. *International journal of communication,* 10, p.20.
- Newman, T. (2017). Germany's Suppression of Free Speech Online. [Blog] *White Sun of the Desert.* Available at: <http://www.desertsun.co.uk/blog/5573/> [Accessed 16 April 2021].
- Pen International and Clarke, S. (2015). *Mass surveillance and online censorship—the PEN International perspective.* [online] Peninternational.org. Available at: <http://www.peninternational.org/07/2015/masssurveillanceandonlinecensorshipthepeninternationalprespective/> [Accessed 16 April 2021].
- People's Charter Foundation (2017). *Gab.ai, the free speech alternative to Twitter*. [online] People's Charter foundation. Originally available at: <https://www.google.com/url?q=http://peoplescharter.org/gab-ai-the-free-speech-alternative-to-twitter/&sa=D&source=editors&ust=1618574476614000&usg=AFQjCNECKlSL7v_GCUu0X06wYeDihGxjqw> [archived].
- Privacy International (2017). *Social Media Intelligence*. [online] Privacyinternational.org. Available at:<https://privacyinternational.org/explainer/55/social-media-intelligence>
- Shields, J., 2017. *Countering online radicalisation and extremism: Baroness Shields' speech.* [online] GOV.UK. Available at: <https://www.gov.uk/government/speeches/countering-online-radicalisation-and-extremism-baroness-shields-speech> [Accessed 16 April 2021].

- United Nations and Kaye, D. (2015). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.* [report] Geneva: United Nations. Human Rights Council. Available at: http://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/Privatesectorinthedigitalage.aspx?platform=hootsuite
- United Nations and Kaye, D. (2017). *Call for Submissions: New Study of Content Regulation in the Digital Age* [online] Freedex.org. Available at: https://freedex.org/2017/09/15/call-for-submissions-new-study-of-content-regulation-in-the-digital-age/ (Accessed: 16 April 2021).
- Wallace, T. (2017). Islam's Attack On Free Speech Through The UN. [Blog]. *Fortress of Faith.* Available at: <https://fortressoffaith.com/index.php/islams-attack-on-free-speech-through-the-un/> [Accessed 16 April 2021].
- Woodfines Solicitors (2015). *Social Media Freedom of Expression or Crime?* [online] Originally available at: <http://www.woodfines.co.uk/blog/socialmediafreedomexpressionorcrime> [archived].

References to newspapers' articles (chapter 6)

- Borger, J. et al. (2017). Tech giants face Congress as showdown over Russia election meddling looms. *The Guardian* [22 October 2017]. Available at: https://www.theguardian.com/technology/2017/oct/22/facebook-google-twitter-congress-hearing-trump-russia-election
- Botsman, R. (2018). Dawn of the techlash. *The Guardian* [18 February 2018]. Available at: https://www.theguardian.com/commentisfree/2018/feb/11/dawn-of-the-techlash
- Bridge, M. (2016). Facebook 'ready to help Beijing censor content'. *The Times* [24 November 2016]. Available: https://web.archive.org/web/20201126135957/https://www.thetimes.co.uk/article/facebook-ready-to-help-beijing-censor-content-cxfnhwhsx
- Brown, A. (2016). Norway declares war on Facebook: PM and newspapers unite to attack Zuckerberg 'censorship'. *Express* [09 September 2016]. Available at:

https://www.express.co.uk/life-style/science-technology/708942/Facebook-CEO-Mark-Zuckerberg-Delete-Iconic-Napalm-Girl-Image

- Daily Mail (2017a). Elite California school forced to cancel speech by right-wing firebrand Milo Yiannopoulos as students set off fireworks at cops, tear down barricades and set the campus ablaze in protest. *Daily Mail* [02 February 2017].
- Daily Mail (2017b). Twitter tweaks its algorithms to try and automatically hunt down abusive accounts. *Daily Mail* [01 March 2017]. Available at: https://www.dailymail.co.uk/sciencetech/article-4271922/Twitter-algorithms-clamp-abusive-content.html
- Daily Mail (2018). Silicon Valley Wunderkind Zuckerberg In Eye Of The Storm. *Mail Online* [10 April 2018] Available at: https://www.dailymail.co.uk/wires/afp/article-5600239/Silicon-Valley-wunderkind-Zuckerberg-eye-storm.html. Accessed 18 Apr 2021.
- Daily Mail Comment (2016). A hate preacher and the craven liberal elite. *Daily Mail* [17 August 2016] available at: https://www.dailymail.co.uk/debate/article-3744355/DAILY-MAIL-COMMENT-hate-preacher-craven-liberal-elite.html
- Dewey, C. (2015). Mark Zuckerberg's Facebook censors images of the Prophet Mohamed in Turkey – two weeks after he declared 'Je Suis Charlie'. *The Independent* [28 January 2015]. Available: https://www.independent.co.uk/news/world/asia/mark-zuckerberg-s-facebook-censors-images-prophet-mohamed-turkey-two-weeks-after-he-declared-je-suis-charlie-10007929.html
- Glaze, B. (2016). Women being "drowned out" by online trolls, Yvette Cooper tells #ReclaimtheInternet seminar. *Mirror* [18 July 2016]. Available at: https://www.mirror.co.uk/news/uk-news/women-being-drowned-out-online-8441703
- Gore, W. (2016). In suspending Milo Yiannopoulos' account, Twitter played the man rather than the ball. *The Independent* [21 July 2016]. Availabe at: https://www.independent.co.uk/voices/suspending-milo-yiannopoulos-account-twitter-has-played-man-rather-ball-a7146901.html
- Gurpreet, N. (2017). Step up the battle against fake news, urges Apple boss. *The Times* [11 February 2017]. Available at: https://www.thetimes.co.uk/article/step-up-the-battle-against-fake-news-urges-apple-boss-0pzk3vf07

- Gutteridge, N. (2016). EUROPE'S elite have announced a sweeping crackdown on freedom of speech online which has been branded "lamentable and Orwellian" by pro-democracy campaigners. *Express* [01 June 2016]. Available at: https://www.express.co.uk/news/world/675535/EU-referendum-Brexit-Brussels-blasted-Orwellian-crackdown-online-criticism-UKIP
- Hamill, J. (2016). Facebook's censorship plans threaten to destroy free speech in Europe - and we should all 'dislike' them. *Mirror* [19 January 2016] available at: https://www.mirror.co.uk/news/technology-science/technology/facebooks-censorship-plans-threaten-destroy-7203995
- Hamilton, F. et al. (2016). Choudary hate videos are still available on YouTube. *The Times* [18 August 2016] available at: https://www.thetimes.co.uk/article/choudary-videos-still-on-youtube-99x28v7jh
- Hastings, M. (2018). MAX HASTINGS: Social media giants are wild beasts devouring freedom and democracy. We MUST tame them *Daily Mail* [02 January 2018]. Available at: https://www.dailymail.co.uk/news/article-5227455/Social-media-giants-wild-beasts-devouring-freedom.html
- Hastings, M. (2018). MAX HASTINGS: The best way to fight back against greedy predators like Facebook? Stop laying bare our lives online. *Daily Mail* [20 March 2018]. Available at: https://www.dailymail.co.uk/debate/article-5525041/MAX-HASTINGS-warns-against-laying-lives-bare-Facebook.html
- Hern, A. (2015). Mark Zuckerberg says he believes in freedom of speech. Does Facebook? *The Guardian* [12 January 2015]. Available: https://www.theguardian.com/world/2015/jan/12/mark-zuckerberg-freedom-speech-facebook
- Hinsliff, G. (2016). Play Nice! How The Internet Is Trying To Design Out Toxic Behaviour. *The Guardian* [22 February 2016] Available at: https://www.theguardian.com/technology/2016/feb/22/play-nice-how-the-internet-is-trying-to-design-out-toxic-behaviour.
- Holmes, S. (2016). "Facebook Criticised For Deleting Photo Of Vitetnam Girl Fleeing Bomb". *Mail Online* [09 September 2016]. Available at: https://www.dailymail.co.uk/news/article-3781566/Facebook-condemned-censoring-

iconic-Vietnam-War-photograph-naked-girl-fleeing-napalm-attack.html. Accessed 18 Apr 2021.

- Hopkins, N. (2017). Revealed: Facebook's internal rulebook on sex, terrorism and violence. *The Guardian* [21 May 2017]. Available at: https://www.theguardian.com/news/2017/may/21/revealed-facebook-internal-rulebook-sex-terrorism-violence (Accessed: 18 April 2021).
- Jeffries, S. (2016). Sue Perkins, Zayn Malik, Tony Hall: how did death threats become so casual? *The Guardian* [16 April 2015]. Available at: https://www.theguardian.com/society/2015/apr/16/sue-perkins-zayn-malik-tony-hall-how-did-death-threats-become-so-casual
- Joseph, S. (2018). Why Facebook's business model is incompatible with human rights. *The Independent* [03 April 2018]. Available at: https://www.independent.co.uk/life-style/gadgets-and-tech/facebook-business-human-rights-data-breach-social-media-personal-information-a8286021.html
- Leigh-Howarth, J. (2016). Facebook is censoring our views – and this is feeding extremism. *The Independent* [14 may 2016]. Available at: https://www.independent.co.uk/voices/facebook-censoring-our-views-and-feeding-extremism-a7029251.html
- Leigh, A. (2016a). "Online Abuse: How Women Are Fighting Back". *The Guardian* [13 April 2016]. Available at: https://www.theguardian.com/technology/2016/apr/13/online-abuse-how-women-are-fighting-back. Accessed 18 Apr 2021.
- Leigh, A. (2016b). Milo Yiannopoulos: Twitter banning one man won't undo his poisonous legacy. *The Guardian* [20 July 2016]. Available at: https://www.theguardian.com/technology/2016/apr/13/online-abuse-how-women-are-fighting-back. Accessed 18 Apr 2021.
- Levin, S. (2016). 'Facebook needs an editor': media experts urge change following photo dispute. *The Guardian* [10 September 2016]. Available at: https://www.theguardian.com/technology/2016/sep/10/facebook-news-media-editor-vietnam-photo-censorship
- Levin, S. (2017). James Damore, Google, And The Youtube Radicalization Of Angry White Men". *The Guardian* [13 August 2017]. Available at:

https://www.theguardian.com/technology/2017/aug/13/james-damore-google-memo-youtube-white-men-radicalization. Accessed 18 Apr 2021.

- Lusher, A. (2016). Scotland Yard to use civilian volunteer 'thought police' to help combat social media hate crime. *The Independent* [14 August 2016].
- Malik, N. (2018). Hate speech leads to violence. Why would liberals defend it? *The Guardian* [22 March 2018]. Available: https://www.theguardian.com/commentisfree/2018/mar/22/hate-speech-violence-liberals-rightwing-extremists
- Martin, A. (2016). Crackdown on trolls who egg on others: New guidelines mean bullies who encourage harassment could be prosecuted. *Mail* [10 October 2016]. Available at: https://www.dailymail.co.uk/news/article-3830019/Crackdown-trolls-egg-New-guidelines-mean-bullies-encourage-harassment-prosecuted.html
- McGoogan, C. (2016). Reddit's plan to tackle trolls and beat Facebook. *The Telegraph* [14 November 2016]. Available at: https://www.telegraph.co.uk/technology/2016/11/13/reddits-plan-to-tackle-trolls-and-beat-facebook/
- McGoogan, C. (2017). Three ways internet companies like Google and Facebook can prevent the spread of extremism. *The Telegraph* [24 March 2017]. Available at: https://www.telegraph.co.uk/technology/2017/03/24/three-ways-internet-companies-like-google-facebook-can-prevent/
- McGoogan, C. and Murgia, M. (2016). Facebook hides conservative news from its homepage. *The Telegraph* [09 May 2016]. Available: https://www.telegraph.co.uk/technology/2016/05/09/facebook-hides-conservative-news-from-its-homepage-former-employ/
- Murray, D. (2015). Steering clear of mocking religion is not ‘good taste’; it’s cowardice. *The Sunday Times* [01 March 2015]. Available: https://www.thetimes.co.uk/article/steering-clear-of-mocking-religion-is-not-good-taste-its-cowardice-pgt78xg0zq3
- Naughton, J. (2017a). Why this woman strikes fear into the net’s big boys. *Observer* [03 September 2017].

- Noughton, J. (2017b). Move Fast, Zuckerberg, Or Hate Will Kill Facebook. *The Guardian* [28 May 2017] Available at: https://www.theguardian.com/commentisfree/2017/may/28/hate-speech-facebook-zuckerberg-content-moderators.
- Parkinson, H. (2015). Mark Zuckerberg, Apple and Google respond to Charlie Hebdo attack. *The Guardian* [09 January 2015]. Available at: <https://www.theguardian.com/technology/2015/jan/09/mark-zuckerberg-apple-google-respond-charlie-hebdo-attack> [Accessed 18 April 2021].
- Parry, H. (2017). Silicon Valley wages war on neo-Nazis: Tech giants including Google, Apple and Facebook crackdown on racists spreading hateful rhetoric online in the wake of the Charlottesville rally. *Daily Mail* [17 August 2017]. Available at: https://www.dailymail.co.uk/news/article-4800088/Silicon-Valley-wages-war-neo-Nazis.html
- Parsons, J. (2015). Facebook and Twitter are terrorist "accomplices" if they fail to remove extremist content, says French president. *Mirror* [28 January 2015] available at: https://www.mirror.co.uk/news/technology-science/technology/facebook-twitter-terrorist-accomplices-fail-5056377
- Pearson-Jones, B. (2015). Not allowing free speech on-campus is dangerous - universities need to defend their right to be offensive. *The Independent* [28 March 2015]. Available: https://www.independent.co.uk/student/istudents/not-allowing-free-speech-campus-dangerous-universities-need-defend-their-right-be-offensive-a6788426.html
- Peters, C. (2017). The student Left's culture of intolerance is creating a new generation of conservatives. *The Telegraph* [17 February 2017]. Available: https://www.telegraph.co.uk/education/2017/02/17/student-lefts-culture-intolereance-creating-new-generation-ofconservatives/
- Sandbrook, D. (2018). Pilloried for speaking sense? He says the unsayable on political correctness - and for this, Canadian professor and avowed culture warrior Jordan Peterson is demonised by the Left and shouted at on TV. *The Daily Mail* [09 February 2018]. Available at: https://www.dailymail.co.uk/news/article-5374295/DOMINIC-SANDBROOK-Pilloried-speaking-sense.html

- Schwabb, N. (2018). Sorry we censored you, Zuckerberg tells Trump-loving video stars Diamond and Silk as Republicans hammer Facebook boss for anti-conservative bias. The *Daily Mail* [11 April 2018]. Available at: https://www.dailymail.co.uk/news/article-5604161/Lawmakers-grumble-Zuckerberg-Facebooks-treatment-Diamond-Silk.html
- Shugerman, E. (2017). Facebook leak: Internal documents show how social media giant handles violence, threats and nudity. *The Independent* [21 May 2017]. Available at: https://www.independent.co.uk/news/world/americas/facebook-rules-violence-threats-nudity-censorship-privacy-leaked-guardian-a7748296.html
- Stolworthy, J. (2017). Katie Hopkins leaves LBC 'immediately' - Twitter reacts; The radio station announced the former Apprentice star is to leave her role 'immediately'. *The Independent* [26 May 2017] Available at: https://www.independent.co.uk/arts-entertainment/tv/news/katie-hopkins-sacked-lbc-final-solution-manchester-atatck-explosion-daily-mail-a7756941.html
- The Independent (2015). The fall of 'Chairman Pao'; The businesswoman who ran Reddit and tried to rein in online hate has become its latest victim. *The Independent* [12 July 2015].
- The Sun (2015). Will Facebook fans like dislike button? The Sun [16 September 2016]. Available at: https://www.thesun.co.uk/archives/news/115846/will-facebook-fans-like-dislike-button/
- The Times (2016). "Net Hate". *The Times* [18 August 2016]. Available at: https://www.thetimes.co.uk/article/net-hate-wtz9k26ts. Accessed 18 Apr 2021.
- Titcomb, J. (2017). 2017 was a bad year for Silicon Valley. Next year could be even worse. Telegraph [27 December 2017]. Available at: https://www.telegraph.co.uk/technology/2017/12/27/2017-bad-year-silicon-valley-next-year-could-even-worse/
- Weinstein, M. (2016). Mark Weinstein: Did Facebook elect Trump President? The role it played with 'fake news'. *Mirror* [07 December 2016]. Available at: https://www.mirror.co.uk/tech/mark-weinstein-facebook-elect-trump-9401898
- White, J. (2017). Taking a stand on Charlottesville, technology companies seek balance with free speech. *The Independent* [20 August 2017]. Available at: https://www.independent.co.uk/news/world/americas/tech-charlottesville-daily-stormer-cloudflare-google-paypal-apple-godaddy-a7901491.html

- White, J.B. (2017). Facebook Reportedly Sold Adverts Targeted At Users Interested In 'How To Burn Jews'. *The Independent* [15 September 2017]. Available at: https://www.independent.co.uk/news/facebook-ads-jewish-antisemitic-haters-keywords-sells-target-users-a7947656.html.
- William, R. (2017). Computer says no: neo-Nazis losing their online voice; INTERNET Tech leaders worry about their power to police content online. *The Independent* [21 August 2017].
- Williams, Z. (2015). Feminazi: the go-to term for trolls out to silence women. *The Guardian* [15 September 2015]. Available at: https://www.theguardian.com/world/2015/sep/15/feminazi-go-to-term-for-trolls-out-to-silence-women-charlotte-proudman
- Wilson, J. (2018). Cambridge Analytica sets quandary for right: hate Facebook, love Trump. *The Guardian* [21 March 2018]. Available at: https://www.theguardian.com/us-news/2018/mar/21/cambridge-analytica-burst-your-bubble-media-facebook
- Wong, J. C. (2018). "Mark Zuckerberg Faces Tough Questions In Two-Day Congressional Testimony – As It Happened". *The Guardian* [11 April 2018] Available at: https://www.theguardian.com/technology/live/2018/apr/11/mark-zuckerberg-testimony-live-updates-house-congress-cambridge-analytica. Accessed 18 Apr 2021.
- Wong, J.C. and Lewin, S. (2017). Ann Coulter cancels speech (again) – but battle for Berkeley's political soul rages on. *The Guardian* [26 April 2017]. Available: https://www.theguardian.com/us-news/2017/apr/26/uc-berkeley-far-right-speakers-free-speech-protests
- Wong, J.C. et al (2017). Google cancels staff meeting after Gamergate-style attack on employees. *The Guardian* [11 August 2017]. Available at: https://www.theguardian.com/technology/2017/aug/10/google-cancels-meeting-james-damore-memo-alt-right-gamergate

Appendices

The following appendices contain the preview and link to the dataset of coded documents, as well as the tables associated with the figures presented in chapters 5 and 6. All data used in the chapters are available in the repository at this link.

Appendix: List of tables

Table A. 1 - Table Data for Figure 5.17 – Actors' statements on issues 4
Table A. 2 - Data from Figure 5.18 Actors' narrative on free speech 5
Table A. 3 - Data from Fig. 5.19 Actors' governance model 6
Table A. 4 - Data from Fig.5.20 most mentioned technological objects 7
Table A. 5 – Cortext demography output 33

Appendix: List of figures

Figure A. 1 - Web pages dataset, part 1 2
Figure A. 2 - Web pages dataset, part 2 3
Figure A. 3 - Graph for year 2015 – Indegree nodes are bigger 10
Figure A. 4 - Graph for year 2015 – Outdegree nodes are bigger 11
Figure A. 5 - Graph for year 2016– Indegree nodes are bigger 12
Figure A. 6 - Graph for year 2016– Outdegree nodes are bigger 13
Figure A. 7 - Graph for year 2017– Indegree nodes are bigger 14
Figure A. 8 - Graph for year 2017– Outdegree nodes are bigger 15
Figure A. 9 - Graph for year 2018– Indegree nodes are bigger 16
Figure A. 10 - Graph for year 2018– Outdegree nodes are bigger 17
Figure A. 11 – Newspapers' dataset 19
Figure A. 12 - Output of Cortext keywords extractions 20
Figure A. 13 - Coretext demography visualisation – time plotting of topics 34
Figure A. 14 - Cortext demography visualisation - time plotting divided by topic...................... 35

Appendix. 1

1) Web pages dataset

Below is reproduced an extract from the dataset of web pages, with the associated texts, and coding. The dataset with the coded texts can be accessed via the following link.

The first column ‘Document ID’ represents the document id that I have assigned to be able to find them more easily from the references in the texts. The column ‘Actor’ contains the main page of the URLs collected with the scraper.

The Column ‘Title’ represents the initial lines or titles of the web pages.

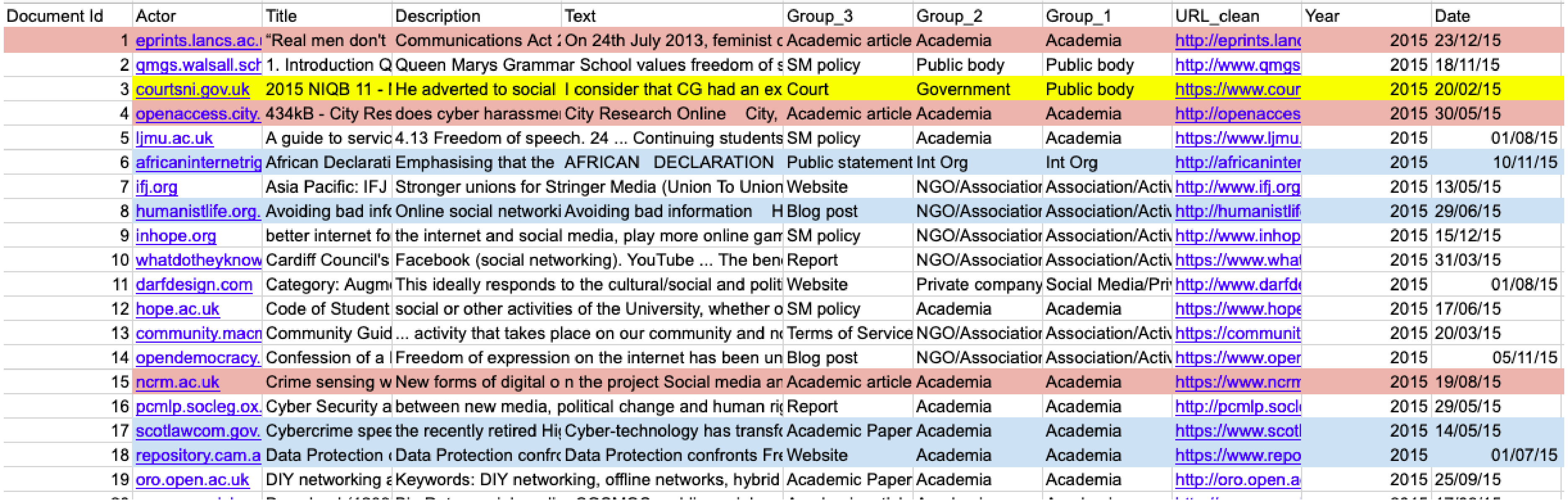
Columns ‘Description and Text’ are two columns reporting some extracts from the web pages texts containing the associated keywords. This column is automatically created by the scraper.

Group, Group 2 and Group 3, are columns that I have used to identify actors and associated them to groups.

The column ‘URL clean’ is the URL that links to the web page collected by the scraper.

Year and Date contain information on the year and specific date of the original publication of the web page.

Figure A. 1 - Web pages dataset, part 1

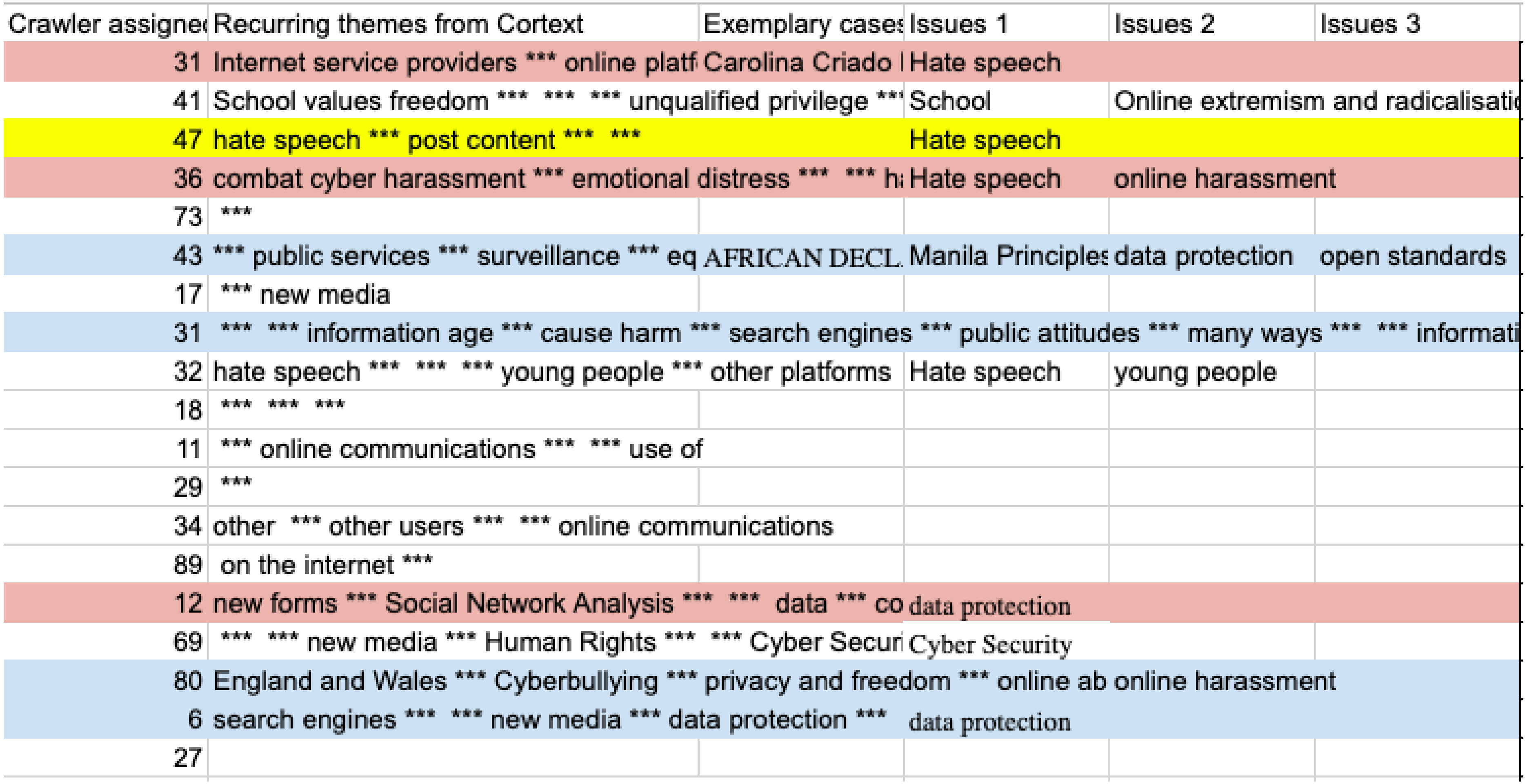


Document Id	Actor	Title	Description	Text	Group_3	Group_2	Group_1	URL_clean	Year	Date
1	eprints.lancs.ac.	“Real men don't	Communications Act :	On 24th July 2013, feminist c	Academic article	Academia	Academia	http://eprints.lanc	2015	23/12/15
2	qmgs.walsall.sch	1. Introduction Q	Queen Marys Grammar School values freedom of s		SM policy	Public body	Public body	http://www.qmgs	2015	18/11/15
3	courtsni.gov.uk	2015 NIQB 11 - I	He adverted to social	I consider that CG had an ex	Court	Government	Public body	https://www.cour	2015	20/02/15
4	openaccess.city.	434kB - City Res	does cyber harassme	City Research Online City,	Academic article	Academia	Academia	http://openacces	2015	30/05/15
5	ljmu.ac.uk	A guide to servic	4.13 Freedom of speech. 24 ... Continuing students		SM policy	Academia	Academia	https://www.ljmu.	2015	01/08/15
6	africaninternetrig	African Declarati	Emphasising that the	AFRICAN DECLARATION	Public statement	Int Org	Int Org	http://africaninter	2015	10/11/15
7	ifj.org	Asia Pacific: IFJ	Stronger unions for Stringer Media (Union To Union		Website	NGO/Associatior	Association/Activ	http://www.ifj.org	2015	13/05/15
8	humanistlife.org.	Avoiding bad infc	Online social networki	Avoiding bad information H	Blog post	NGO/Associatior	Association/Activ	http://humanistlif	2015	29/06/15
9	inhope.org	better internet fo	the internet and social media, play more online gan		SM policy	NGO/Associatior	Association/Activ	http://www.inhop	2015	15/12/15
10	whatdotheyknow	Cardiff Council's	Facebook (social networking). YouTube ... The ben		Report	NGO/Associatior	Association/Activ	https://www.what	2015	31/03/15
11	darfdesign.com	Category: Augm	This ideally responds to the cultural/social and polit		Website	Private company	Social Media/Pri	http://www.darfd	2015	01/08/15
12	hope.ac.uk	Code of Student	social or other activities of the University, whether o		SM policy	Academia	Academia	https://www.hope	2015	17/06/15
13	community.macn	Community Guid	... activity that takes place on our community and n		Terms of Service	NGO/Associatior	Association/Activ	https://communit	2015	20/03/15
14	opendemocracy.	Confession of a	Freedom of expression on the internet has been un		Blog post	NGO/Associatior	Association/Activ	https://www.oper	2015	05/11/15
15	ncrm.ac.uk	Crime sensing w	New forms of digital o	n the project Social media an	Academic article	Academia	Academia	https://www.ncrm	2015	19/08/15
16	pcmlp.socleg.ox.	Cyber Security a	between new media, political change and human ri		Report	Academia	Academia	http://pcmlp.socl	2015	29/05/15
17	scotlawcom.gov.	Cybercrime spee	the recently retired Hi	Cyber-technology has transfc	Academic Paper	Academia	Academia	https://www.scotl	2015	14/05/15
18	repository.cam.a	Data Protection	Data Protection confr	Data Protection confronts Fre	Website	Academia	Academia	https://www.repo	2015	01/07/15
19	oro.open.ac.uk	DIY networking a	Keywords: DIY networking, offline networks, hybrid		Academic Paper	Academia	Academia	http://oro.open.a	2015	25/09/15

In part 2 the dataset continues, and it includes a column with ‘Crawler assigned ID’. These are the ID that indicate the order of collection from Google. Columns ‘Recurring Themes from Cortext’ include the keywords identified via quantitative analysis of texts they are the most relevant terms associated with the specific document.

The columns Exemplary cases, Issues 1, Issues 2 and Issues contain the summary of the qualitative coding (that took place on Nvivo, the file is available in the link provided above).

Figure A. 2 - Web pages dataset, part 2



Crawler assigne	Recurring themes from Cortext	Exemplary case	Issues 1	Issues 2	Issues 3
31	Internet service providers *** online platf	Carolina Criado	Hate speech		
41	School values freedom *** *** *** unqualified privilege **		School	Online extremism and radicalisati	
47	hate speech *** post content *** ***		Hate speech		
36	combat cyber harassment *** emotional distress *** *** h		Hate speech	online harassment	
73	***				
43	*** public services *** surveillance *** eq	AFRICAN DECL.	Manila Principles	data protection	open standards
17	*** new media				
31	*** *** information age *** cause harm *** search engines *** public attitudes *** many ways *** *** informati				
32	hate speech *** *** *** young people *** other platforms		Hate speech	young people	
18	*** *** ***				
11	*** online communications *** *** use of				
29	***				
34	other *** other users *** *** online communications				
89	on the internet ***				
12	new forms *** Social Network Analysis *** *** data *** co		data protection		
69	*** *** new media *** Human Rights *** *** Cyber Secur		Cyber Security		
80	England and Wales *** Cyberbullying *** privacy and freedom *** online ab			online harassment	
6	search engines *** *** new media *** data protection ***		data protection		
27					

2) Data from Qualitative analysis of texts

The following table summarise the coding exercise conducted with Nvivo (data available in Nvivo format in the repository following the link).

Table A. 1 - Table Data for Figure 5.17 – Actors' statements on issues

Groups	**Quality of content**	**Extremism**	**Harassment and bullying**	**Hate speech**	**Censorship**	**Privacy and protection of Data**	**Surveillance**
Academia & Think Tank	35	177	41	152	32	42	187
NGOs/Advocacy groups and activists	2	89	146	48	266	179	56
Public bodies	1	297	202	73	127	0	30
Private company	5	3	17	15	19	0	7
International and European Organisations	0	41	0	1	5	1	23

Table A. 2 - Data from Figure 5.18 Actors' narrative on free speech

Groups	A : Free speech has limits	B : Free speech is absolute	C : Privacy as limitation to free speech	D : Privacy as necessary for free speech
Civil Society: Academia & Think Tank	5	4	0	0
Civil Society: NGOs/Advocacy groups and activists	8	18	2	4
Political bodies: governments, politicians, enforcement	7	0	1	1
Private company	4	3	1	0
Int Org	1	0	1	1

Table A. 3 - Data from Fig. 5.19 Actors' governance model

Groups	**A : Editorial responsibility**	**B : Social media do NOT have responsibility**	**C : Social media have responsibility**	**D : State responsibility**	**E : State should legislate more**	**F : States should have less power**
Civil Society: Academia & Think Tank	45	0	19	12	3	35
Civil Society: NGOs/Advocacy groups and activists	25	7	8	40	4	39
Political bodies: governments, politicians, enforcement	0	0	29	10	0	3
Private company	12	0	3	0	0	0
Int Org	0	0	3	3	0	17

Table A. 4 - Data from Fig.5.20 most mentioned technological objects

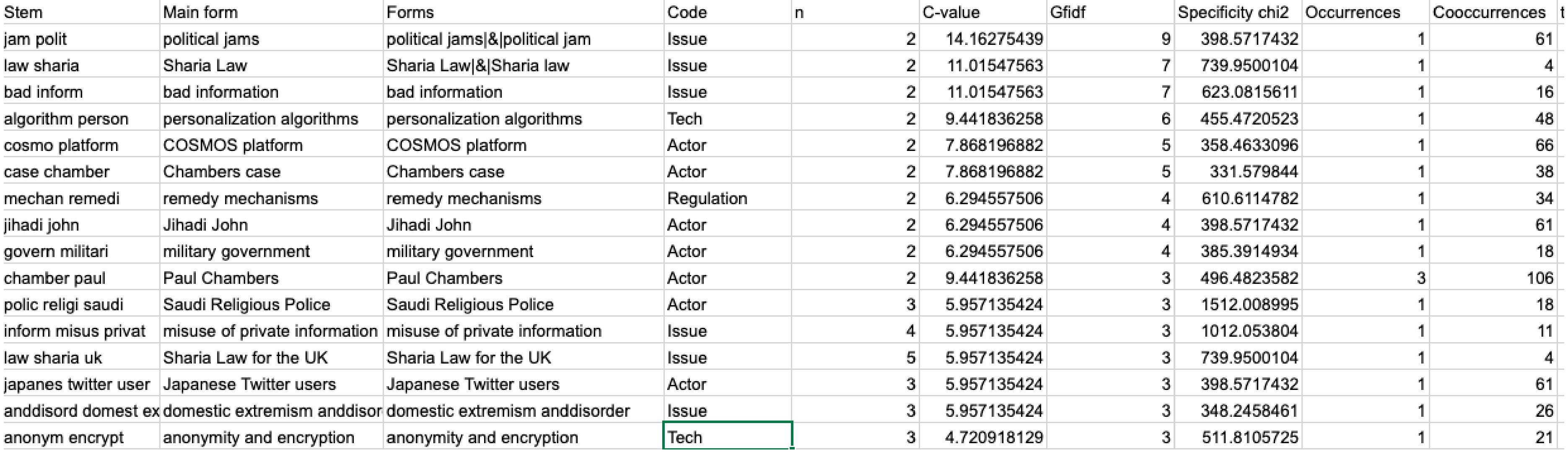
Technological artefacts	**Frequency**
Algorithm	41
Artificial Intelligence	11
Bot	148
Data	19
Eco Chamber	4
Facebook post	22
Fake News	57
Hashtag	23
Meme	8
Troll	46
Tweet	82
Youtube video	2
Grand Total	463

3) Data from quantitative analysis of texts

Terms extraction from web pages - TF-IDF output from Cortext

Below is reported the output of the terms extraction algorithm from Cortext. For the purpose of the analysis I considered the GF-IDF, which is one of the most employed measure of relevance of terms in documents. The full list is available at this link

Figure A. 3 - Output of Cortext keywords extractions



Stem	Main form	Forms	Code	n	C-value	Gfidf	Specificity chi2	Occurrences	Cooccurrences	t
jam polit	political jams	political jams|&|political jam	Issue	2	14.16275439	9	398.5717432	1	61	
law sharia	Sharia Law	Sharia Law|&|Sharia law	Issue	2	11.01547563	7	739.9500104	1	4	
bad inform	bad information	bad information	Issue	2	11.01547563	7	623.0815611	1	16	
algorithm person	personalization algorithms	personalization algorithms	Tech	2	9.441836258	6	455.4720523	1	48	
cosmo platform	COSMOS platform	COSMOS platform	Actor	2	7.868196882	5	358.4633096	1	66	
case chamber	Chambers case	Chambers case	Actor	2	7.868196882	5	331.579844	1	38	
mechan remedi	remedy mechanisms	remedy mechanisms	Regulation	2	6.294557506	4	610.6114782	1	34	
jihadi john	Jihadi John	Jihadi John	Actor	2	6.294557506	4	398.5717432	1	61	
govern militari	military government	military government	Actor	2	6.294557506	4	385.3914934	1	18	
chamber paul	Paul Chambers	Paul Chambers	Actor	2	9.441836258	3	496.4823582	3	106	
polic religi saudi	Saudi Religious Police	Saudi Religious Police	Actor	3	5.957135424	3	1512.008995	1	18	
inform misus privat	misuse of private information	misuse of private information	Issue	4	5.957135424	3	1012.053804	1	11	
law sharia uk	Sharia Law for the UK	Sharia Law for the UK	Issue	5	5.957135424	3	739.9500104	1	4	
japanes twitter user	Japanese Twitter users	Japanese Twitter users	Actor	3	5.957135424	3	398.5717432	1	61	
anddisord domest ex	domestic extremism anddisor	domestic extremism anddisorder	Issue	3	5.957135424	3	348.2458461	1	26	
anonym encrypt	anonymity and encryption	anonymity and encryption	Tech	3	4.720918129	3	511.8105725	1	21	

4) Data from SNA - Graphs visualisations

Below I report the graph visualisation from the SNA presented in the chapter, in a larger dimension.

The original graph files can be access via this link

An interactive visualisation of the networks can be found here:

Network of actors in 2015

Network of actors in 2016

Network of actors in 2017

Network of actors in 2018

Figure A. 4 - Graph for year 2015 – Indegree nodes are bigger

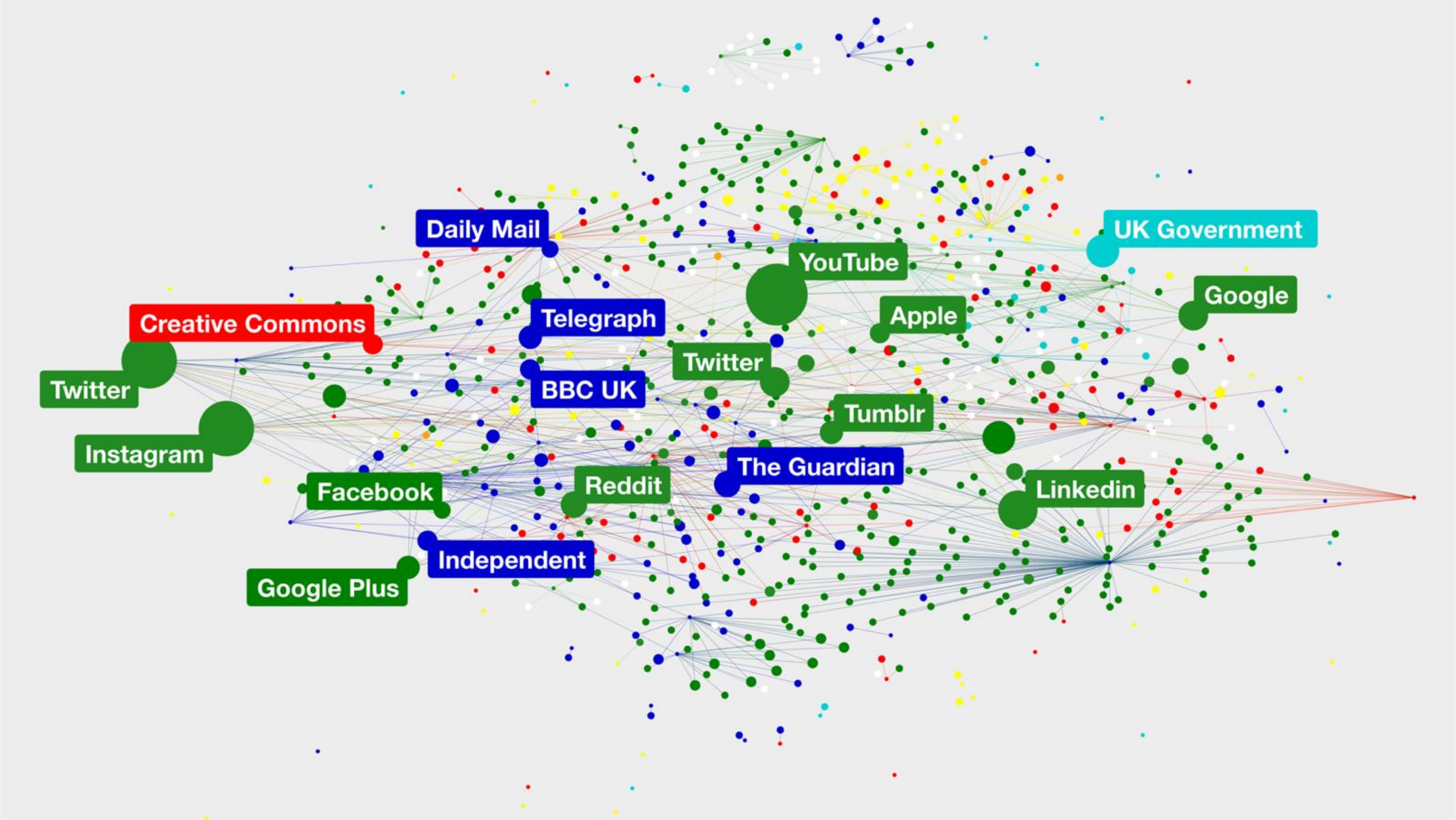


Figure A. 5 - Graph for year 2015 – Outdegree nodes are bigger

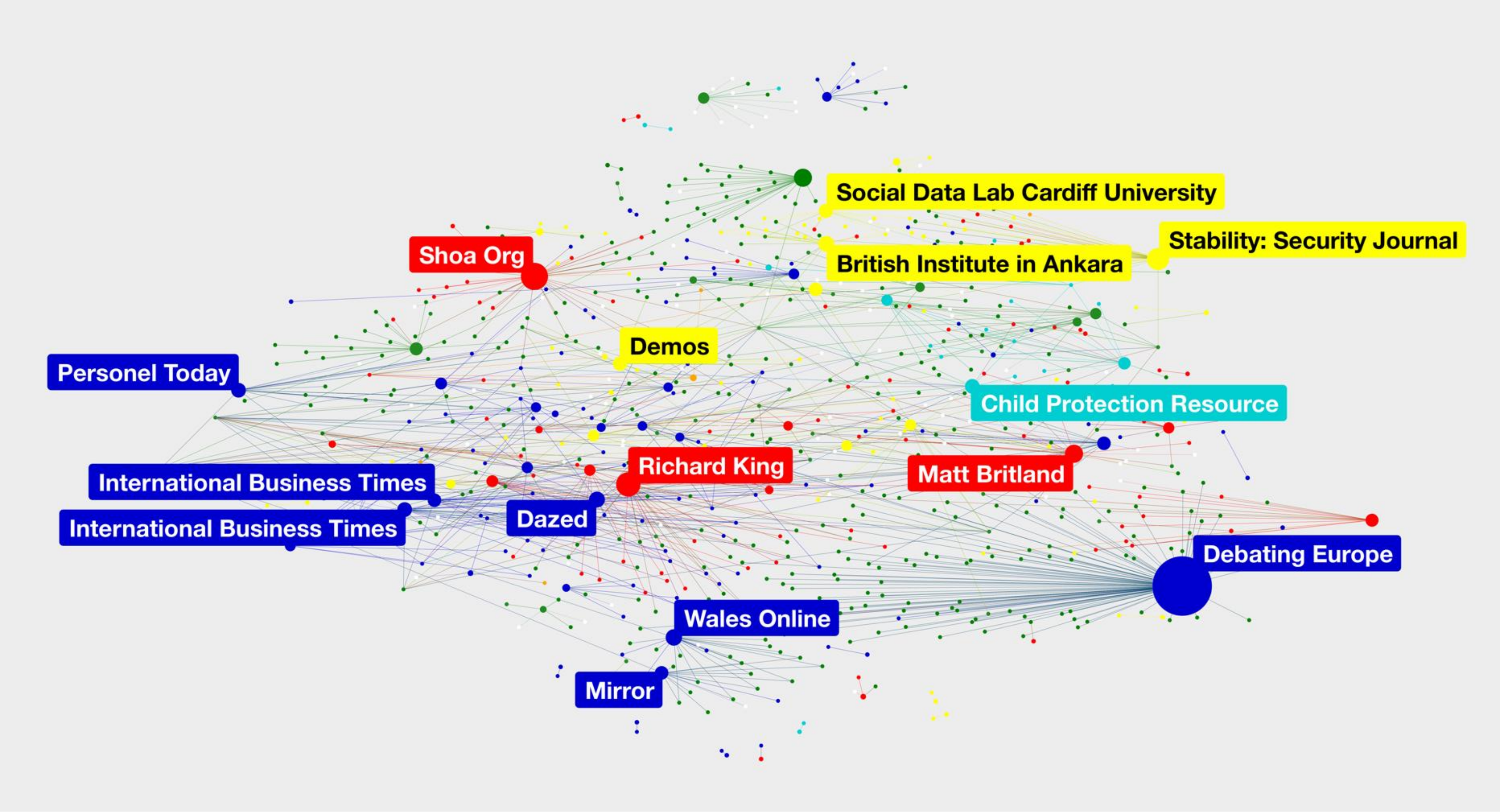

Social Data Lab Cardiff University
Stability: Security Journal
Shoa Org
British Institute in Ankara
Demos
Personel Today
Child Protection Resource
Richard King
Matt Britland
International Business Times
Dazed
International Business Times
Debating Europe
Wales Online
Mirror

Figure A. 6 - Graph for year 2016– Indegree nodes are bigger

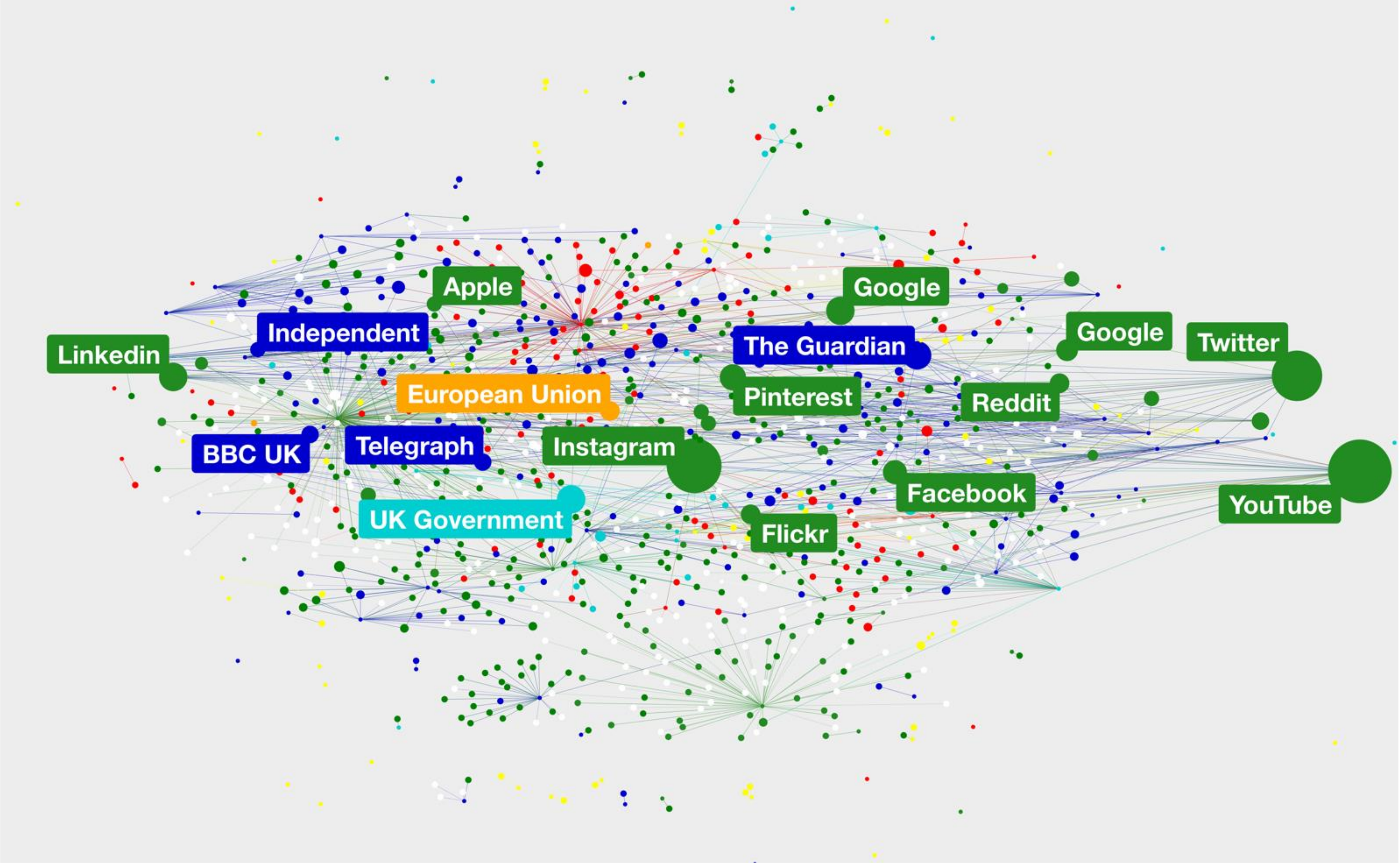


Figure A. 7 - Graph for year 2016– Outdegree nodes are bigger

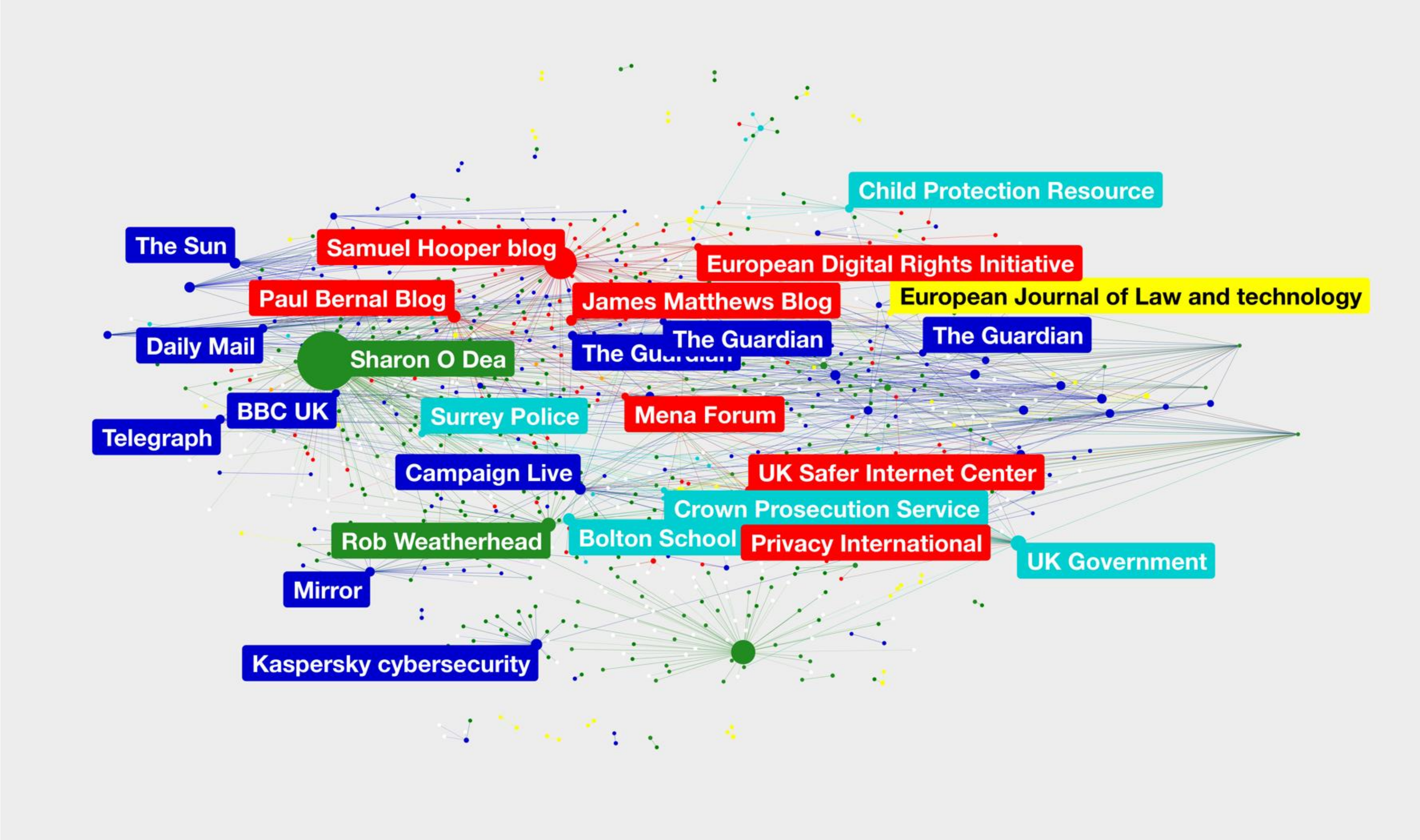


Figure A. 8 - Graph for year 2017– Indegree nodes are bigger

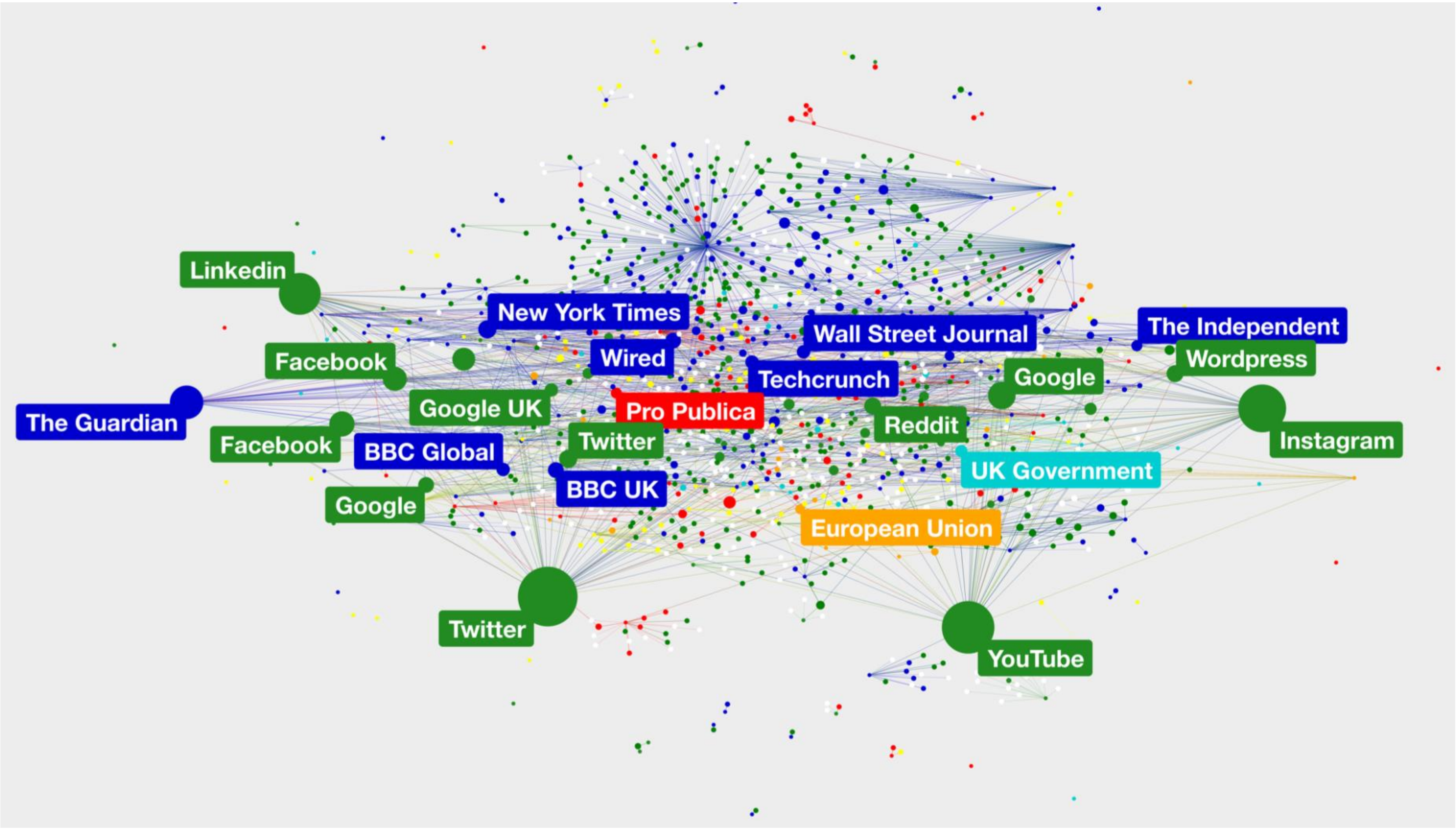


Figure A. 9 - Graph for year 2017– Outdegree nodes are bigger

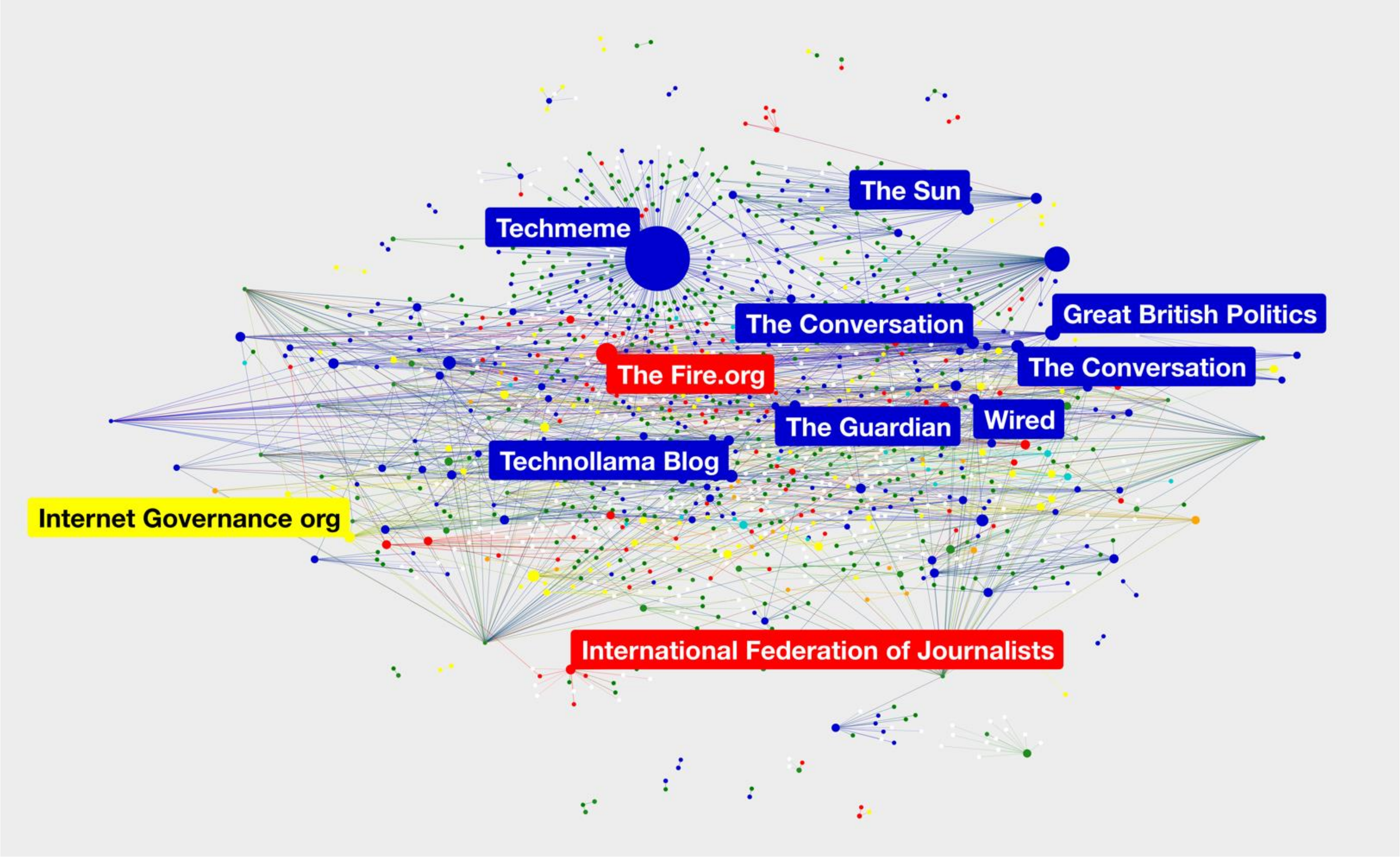


Figure A. 10 - Graph for year 2018– Indegree nodes are bigger

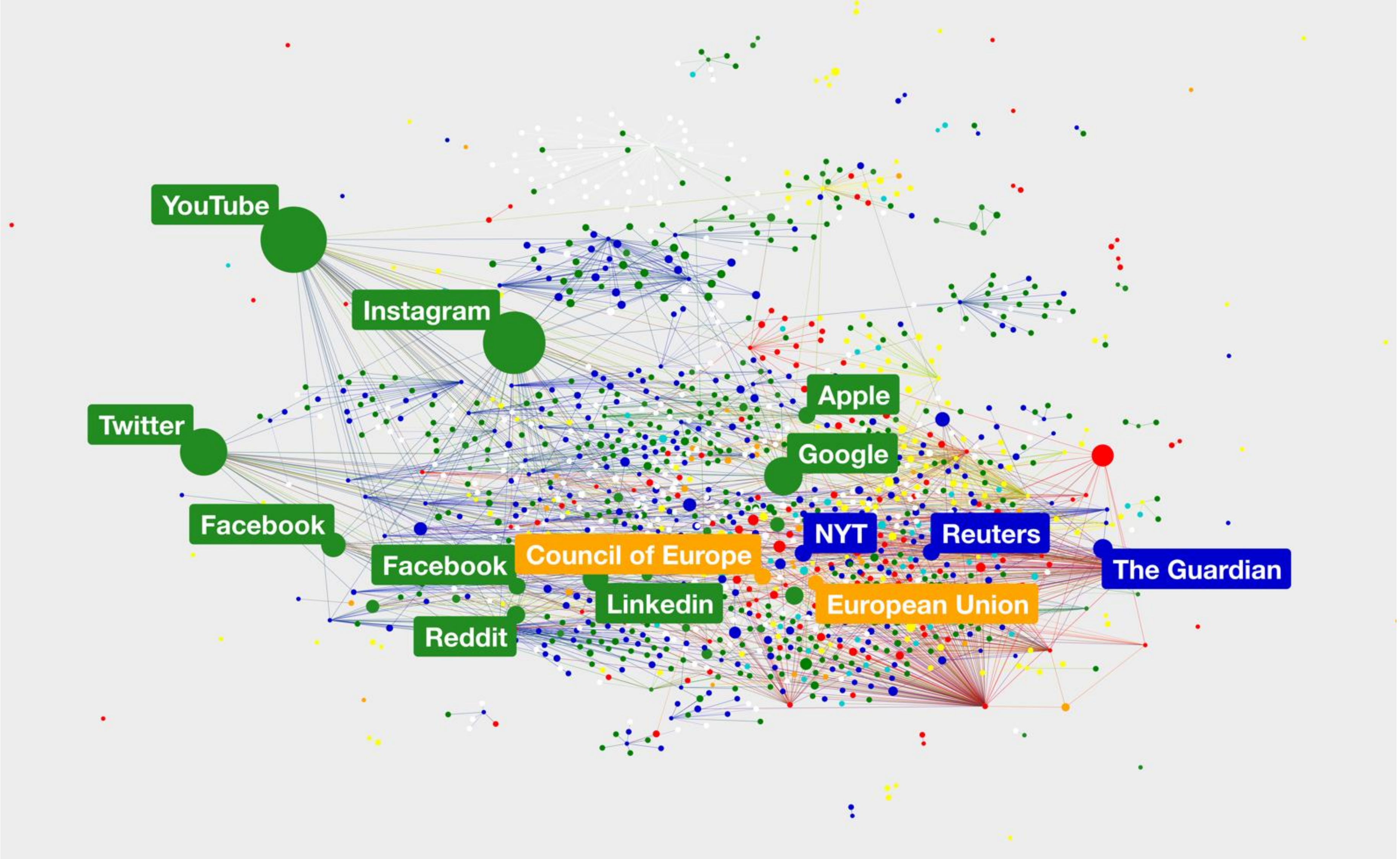
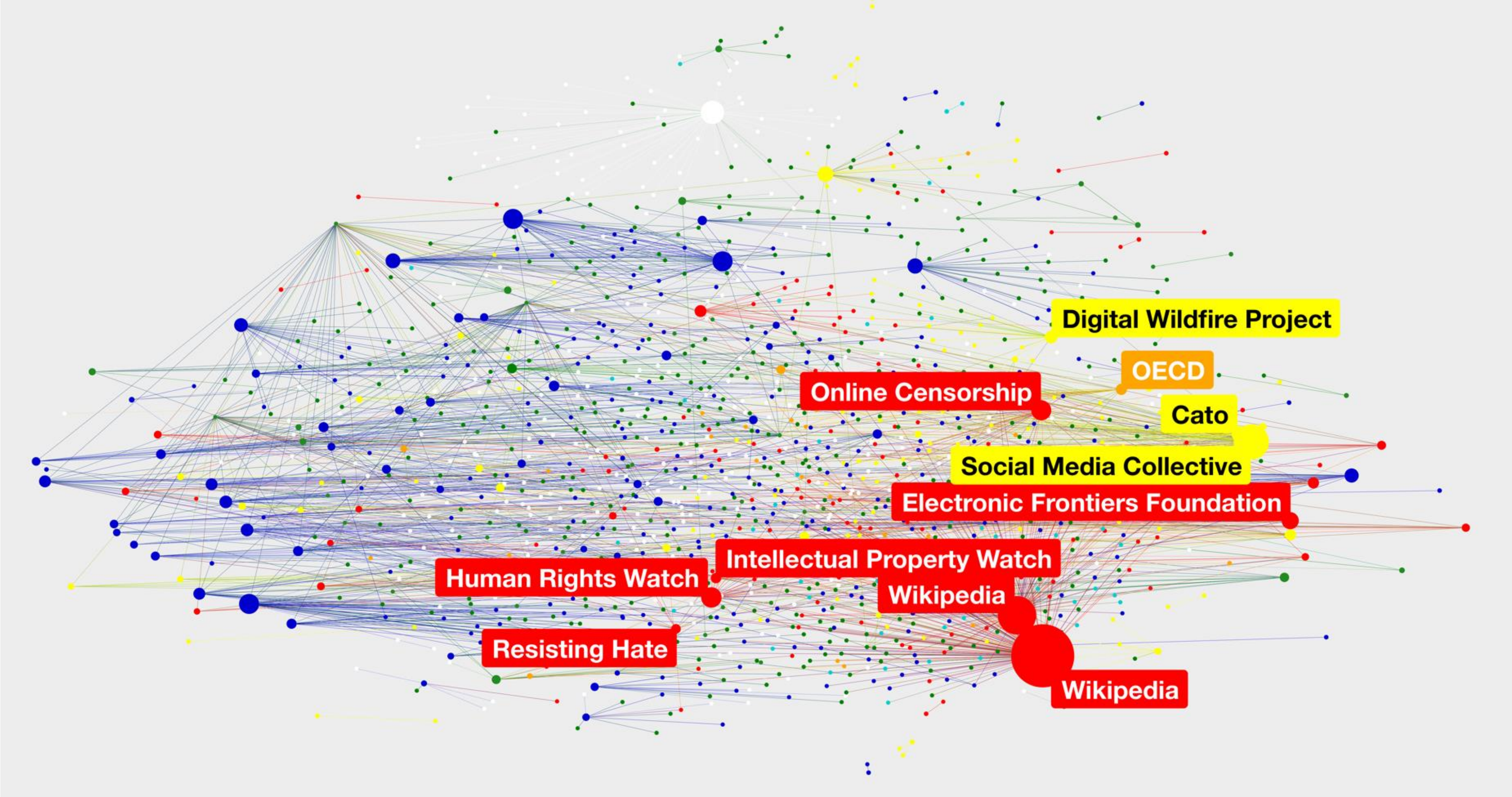


Figure A. 11 - Graph for year 2018– Outdegree nodes are bigger



Appendix 2

1) Newspaper's dataset

Below I report the link to access the dataset and the tables that I have used for mapping the controversy on newspapers. The dataset with newspapers articles metadata, texts and coding can be accessed at this link

In column File is reported the unique identifier assigned from the repository Lexis Nexis.

Column publication contains the name of the publication house.

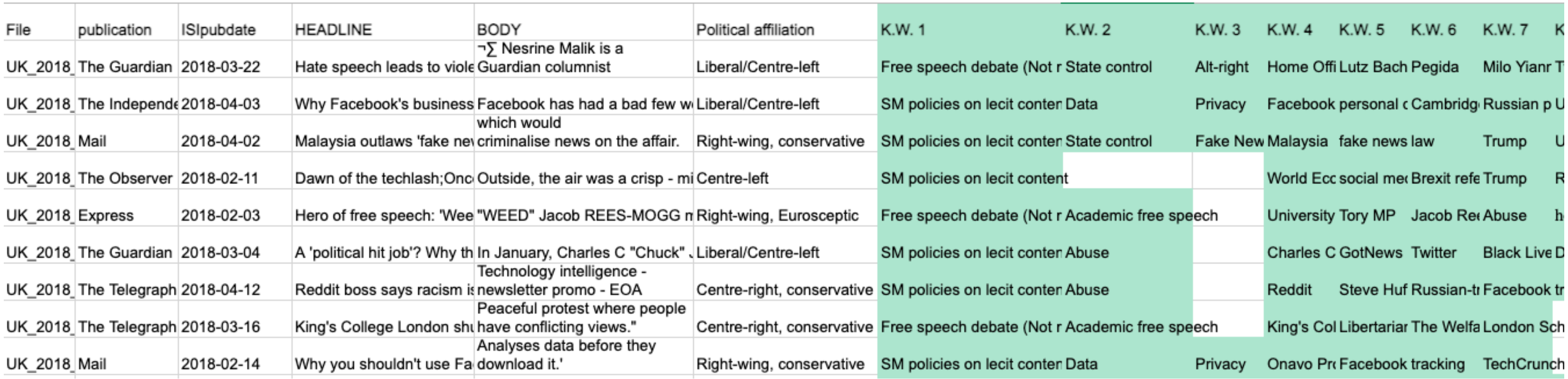
ISIpub date reports the publication date.

Headline is the title and Body contains the text of the articles. All these are columns created by the repository at the moment of the download.

Column political affiliaiton was added by me, and it connects the publications to their political affiliation as declared on occasion of the 2017 General Elections.

The following columns, from K.W.1 to K.W. 45 are columns that I have used to code qualitatively the texts looking for exemplary cases, and storylines.

Figure A. 12 – Newspapers’ dataset



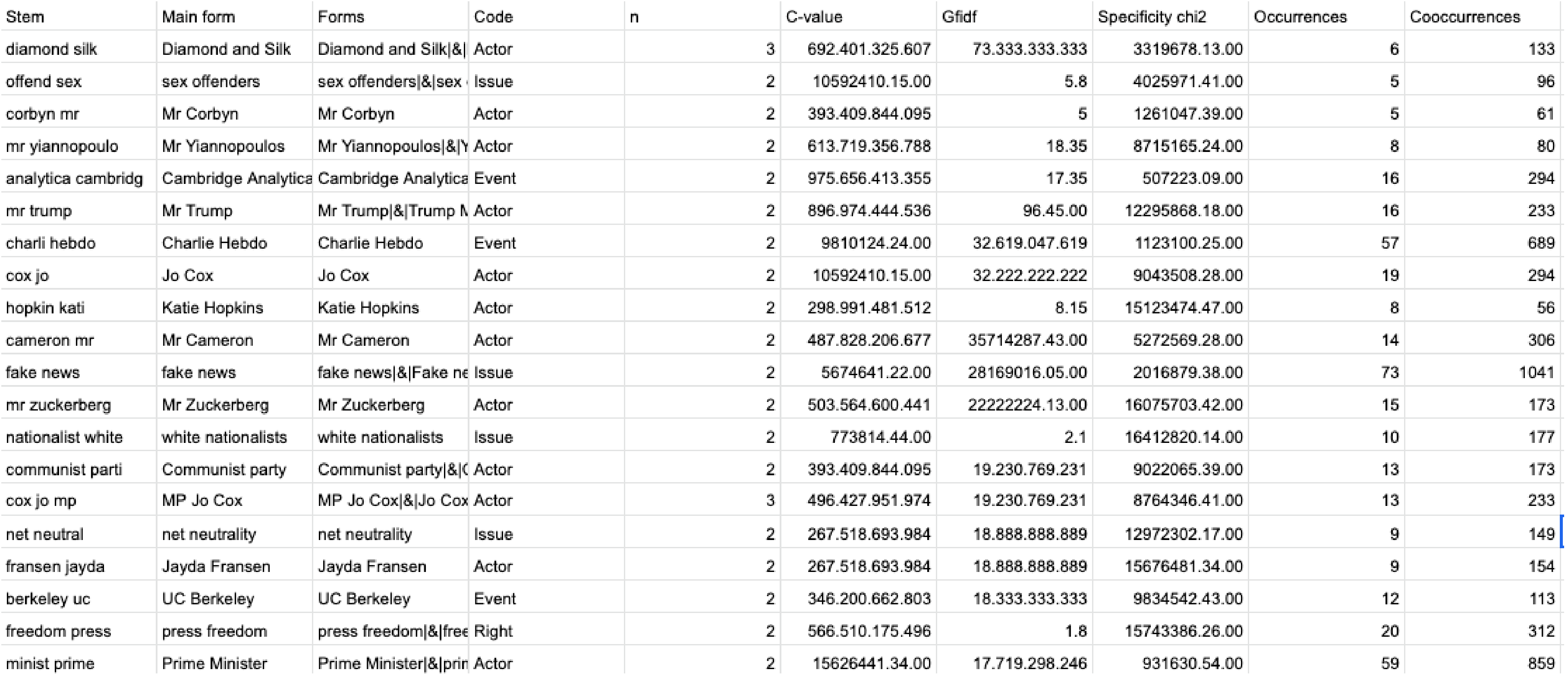
File	publication	ISIpubdate	HEADLINE	BODY	Political affiliation	K.W. 1	K.W. 2	K.W. 3	K.W. 4	K.W. 5	K.W. 6	K.W. 7	K
UK_2018_	The Guardian	2018-03-22	Hate speech leads to viole	¬∑ Nesrine Malik is a Guardian columnist	Liberal/Centre-left	Free speech debate (Not r	State control	Alt-right	Home Offi	Lutz Bach	Pegida	Milo Yianr	T
UK_2018_	The Independe	2018-04-03	Why Facebook's business	Facebook has had a bad few w	Liberal/Centre-left	SM policies on lecit conten	Data	Privacy	Facebook	personal c	Cambridg	Russian p	U
UK_2018_	Mail	2018-04-02	Malaysia outlaws 'fake nev	which would criminalise news on the affair.	Right-wing, conservative	SM policies on lecit conten	State control	Fake New	Malaysia	fake news	law	Trump	U
UK_2018_	The Observer	2018-02-11	Dawn of the techlash;Onc	Outside, the air was a crisp - mi	Centre-left	SM policies on lecit content			World Ecc	social me	Brexit refe	Trump	R
UK_2018_	Express	2018-02-03	Hero of free speech: 'Wee	"WEED" Jacob REES-MOGG n	Right-wing, Eurosceptic	Free speech debate (Not r	Academic free speech		University	Tory MP	Jacob Re	Abuse	h
UK_2018_	The Guardian	2018-03-04	A 'political hit job'? Why th	In January, Charles C "Chuck" .	Liberal/Centre-left	SM policies on lecit conten	Abuse		Charles C	GotNews	Twitter	Black Live	D
UK_2018_	The Telegraph	2018-04-12	Reddit boss says racism is	Technology intelligence - newsletter promo - EOA	Centre-right, conservative	SM policies on lecit conten	Abuse		Reddit	Steve Huf	Russian-tı	Facebook	tr
UK_2018_	The Telegraph	2018-03-16	King's College London shu	Peaceful protest where people have conflicting views."	Centre-right, conservative	Free speech debate (Not r	Academic free speech		King's Col	Libertariar	The Welfa	London Sch	
UK_2018_	Mail	2018-02-14	Why you shouldn't use Fa	Analyses data before they download it.'	Right-wing, conservative	SM policies on lecit conten	Data	Privacy	Onavo Pr	Facebook tracking		TechCrunch	

2) Data from quantitative analysis of texts

Newspaper TF-IDF output from Cortext

Below is reported the output of the terms extraction algorithm from Cortext. For the purpose of the analysis I considered the GF-IDF, which is one of the most employed measure of relevance of terms in documents. The full list is available at this link

Figure A. 13 - Output of Cortext keywords extractions



Stem	Main form	Forms	Code	n	C-value	Gfidf	Specificity chi2	Occurrences	Cooccurrences
diamond silk	Diamond and Silk	Diamond and Silk|&|	Actor	3	692.401.325.607	73.333.333.333	3319678.13.00	6	133
offend sex	sex offenders	sex offenders|&|sex	Issue	2	10592410.15.00	5.8	4025971.41.00	5	96
corbyn mr	Mr Corbyn	Mr Corbyn	Actor	2	393.409.844.095	5	1261047.39.00	5	61
mr yiannopoulo	Mr Yiannopoulos	Mr Yiannopoulos|&|Y	Actor	2	613.719.356.788	18.35	8715165.24.00	8	80
analytica cambridg	Cambridge Analytica	Cambridge Analytica	Event	2	975.656.413.355	17.35	507223.09.00	16	294
mr trump	Mr Trump	Mr Trump|&|Trump M	Actor	2	896.974.444.536	96.45.00	12295868.18.00	16	233
charli hebdo	Charlie Hebdo	Charlie Hebdo	Event	2	9810124.24.00	32.619.047.619	1123100.25.00	57	689
cox jo	Jo Cox	Jo Cox	Actor	2	10592410.15.00	32.222.222.222	9043508.28.00	19	294
hopkin kati	Katie Hopkins	Katie Hopkins	Actor	2	298.991.481.512	8.15	15123474.47.00	8	56
cameron mr	Mr Cameron	Mr Cameron	Actor	2	487.828.206.677	35714287.43.00	5272569.28.00	14	306
fake news	fake news	fake news|&|Fake ne	Issue	2	5674641.22.00	28169016.05.00	2016879.38.00	73	1041
mr zuckerberg	Mr Zuckerberg	Mr Zuckerberg	Actor	2	503.564.600.441	22222224.13.00	16075703.42.00	15	173
nationalist white	white nationalists	white nationalists	Issue	2	773814.44.00	2.1	16412820.14.00	10	177
communist parti	Communist party	Communist party|&|C	Actor	2	393.409.844.095	19.230.769.231	9022065.39.00	13	173
cox jo mp	MP Jo Cox	MP Jo Cox|&|Jo Cox	Actor	3	496.427.951.974	19.230.769.231	8764346.41.00	13	233
net neutral	net neutrality	net neutrality	Issue	2	267.518.693.984	18.888.888.889	12972302.17.00	9	149
fransen jayda	Jayda Fransen	Jayda Fransen	Actor	2	267.518.693.984	18.888.888.889	15676481.34.00	9	154
berkeley uc	UC Berkeley	UC Berkeley	Event	2	346.200.662.803	18.333.333.333	9834542.43.00	12	113
freedom press	press freedom	press freedom|&|free	Right	2	566.510.175.496	1.8	15743386.26.00	20	312
minist prime	Prime Minister	Prime Minister|&|prin	Actor	2	15626441.34.00	17.719.298.246	931630.54.00	59	859

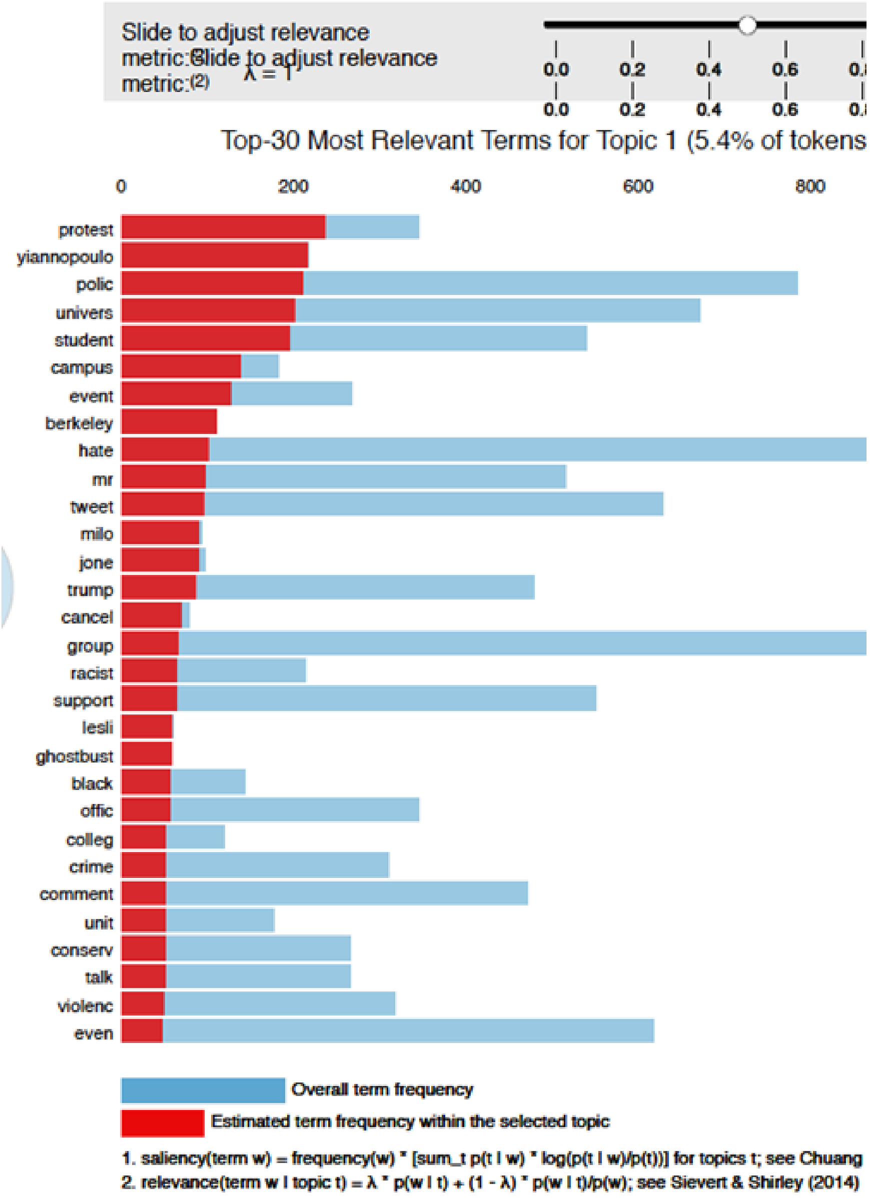
5) Data from quantitative analysis of texts

Parameters of LDA topic modelling tool in Cortext:

```
2019-02-18 23:22:34 INFO :
Data Description:
Fields:
- BODY
Number of Topics - (0 for automatic search): '0'
Minimum number of topics: '10'
Maximum number of topics: '40'
Steps: '10'
Custom name for storing topics: Automated LDA
Text Cleaning Parameters:
Lower Case: true
language: English
Stop-words Removal: true
Remove punctuation: true
Stemming: true
Minimum frequency of words: '5'
Maximum frequency of words (in percentage of the total corpus): '50'
LDA algorithm parameters:
Alpha: symmetric
Number of iterations for learning the model: '20'
```

Newspapers LDA topic analysis from Cortext

Topic 1 – Milo Yiannopoulos - University Student Union -



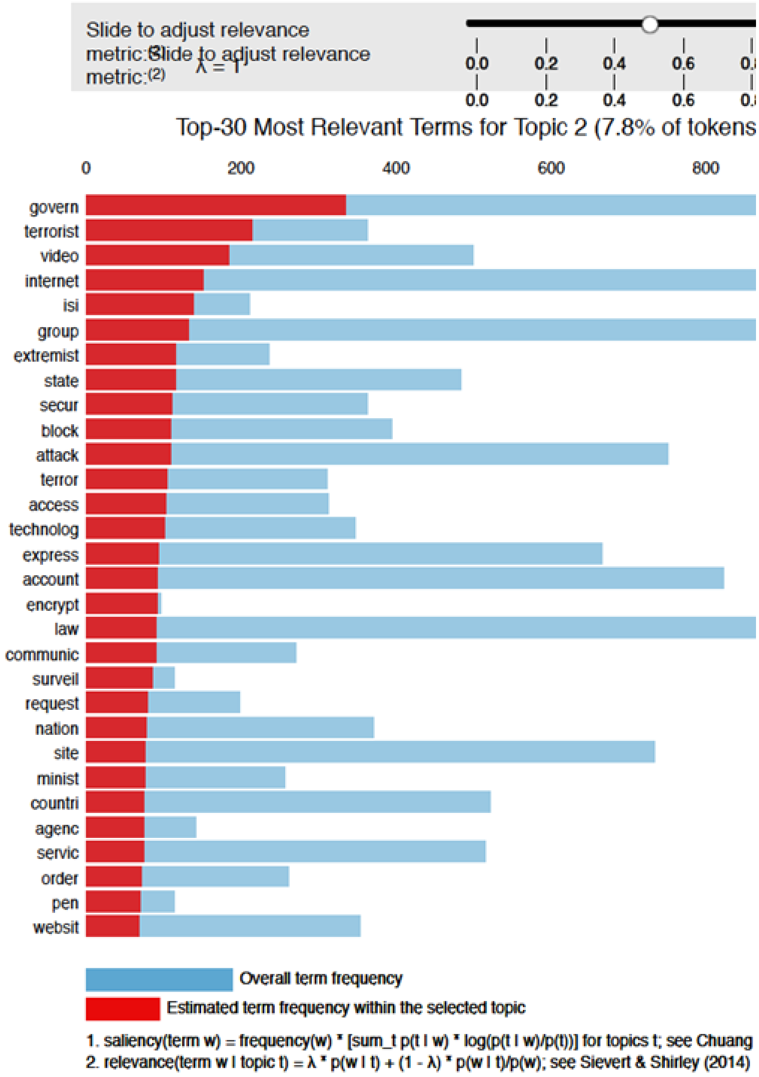
Slide to adjust relevance metric:Slide to adjust relevance metric:(2) λ = 1

Top-30 Most Relevant Terms for Topic 1 (5.4% of tokens

1. saliency(term w) = frequency(w) * [sum_t p(t I w) * log(p(t I w)/p(t))] for topics t; see Chuang

2. relevance(term w I topic t) = λ * p(w I t) + (1 - λ) * p(w I t)/p(w); see Sievert & Shirley (2014)

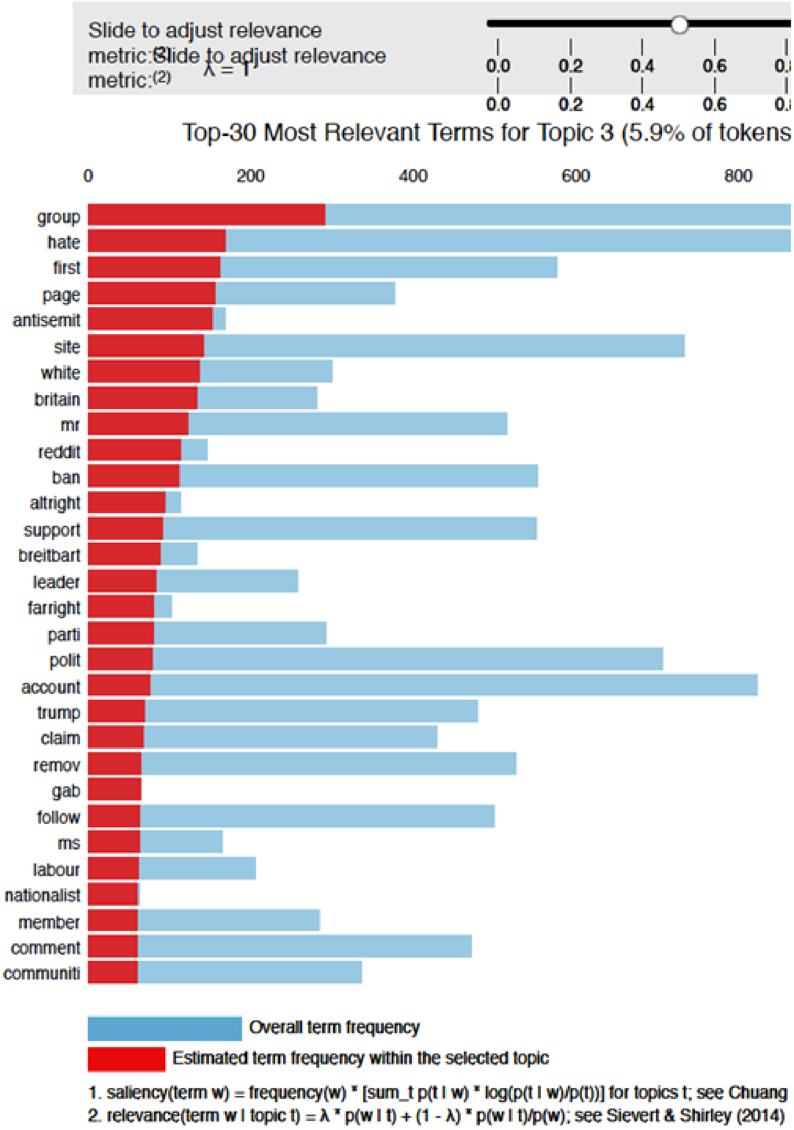
Topic 2 - Government - terrorism and encryption



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang

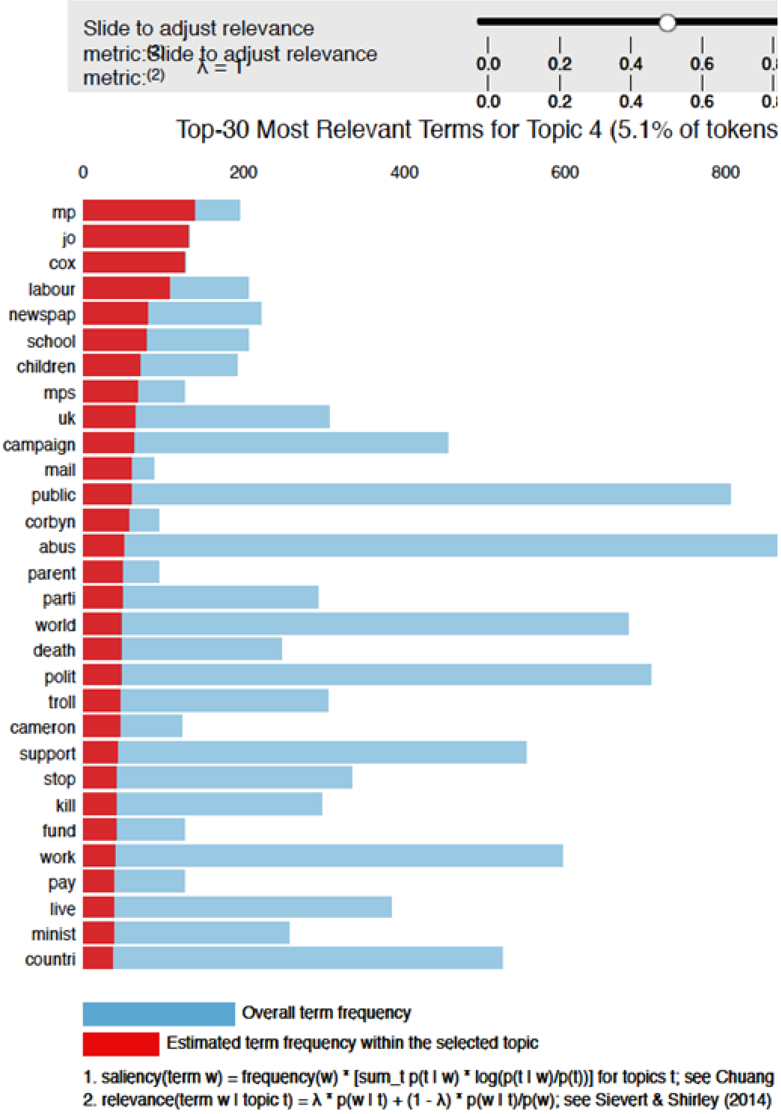
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Topic 3 - Hate groups -reddit - alt-right - trump



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Topic 4 - Jo Cox, abuse online



Slide to adjust relevance metric:(2) Slide to adjust relevance metric:(2) λ = 1

Top-30 Most Relevant Terms for Topic 4 (5.1% of tokens

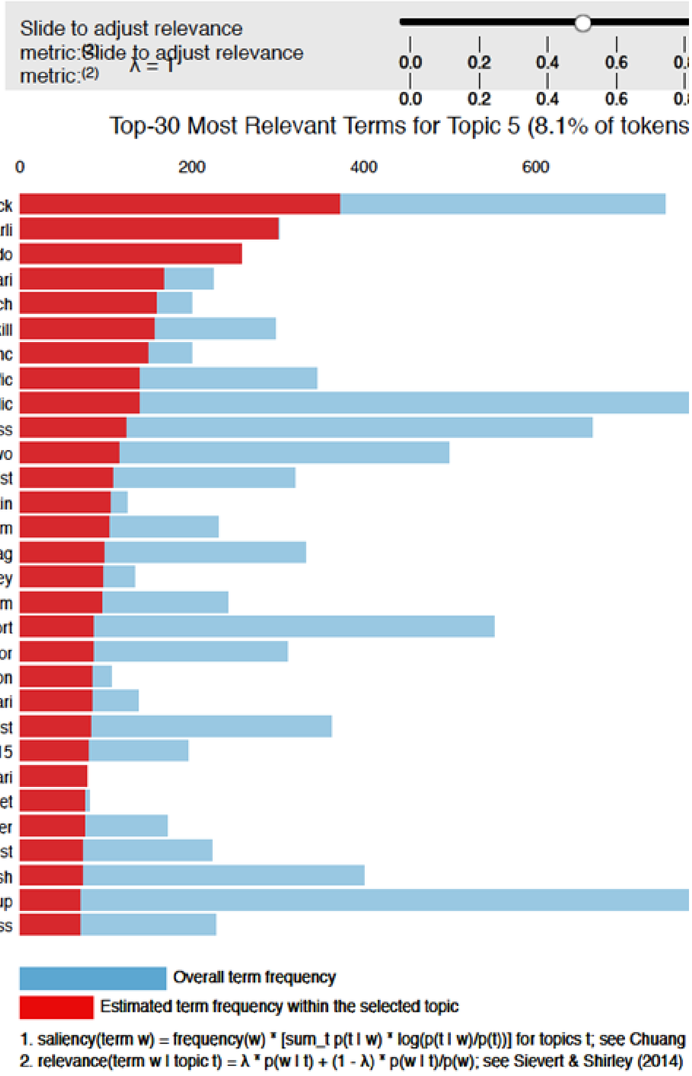
🟦 Overall term frequency

🟥 Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t I w) * log(p(t I w)/p(t))] for topics t; see Chuang

2. relevance(term w I topic t) = λ * p(w I t) + (1 - λ) * p(w I t)/p(w); see Sievert & Shirley (2014)

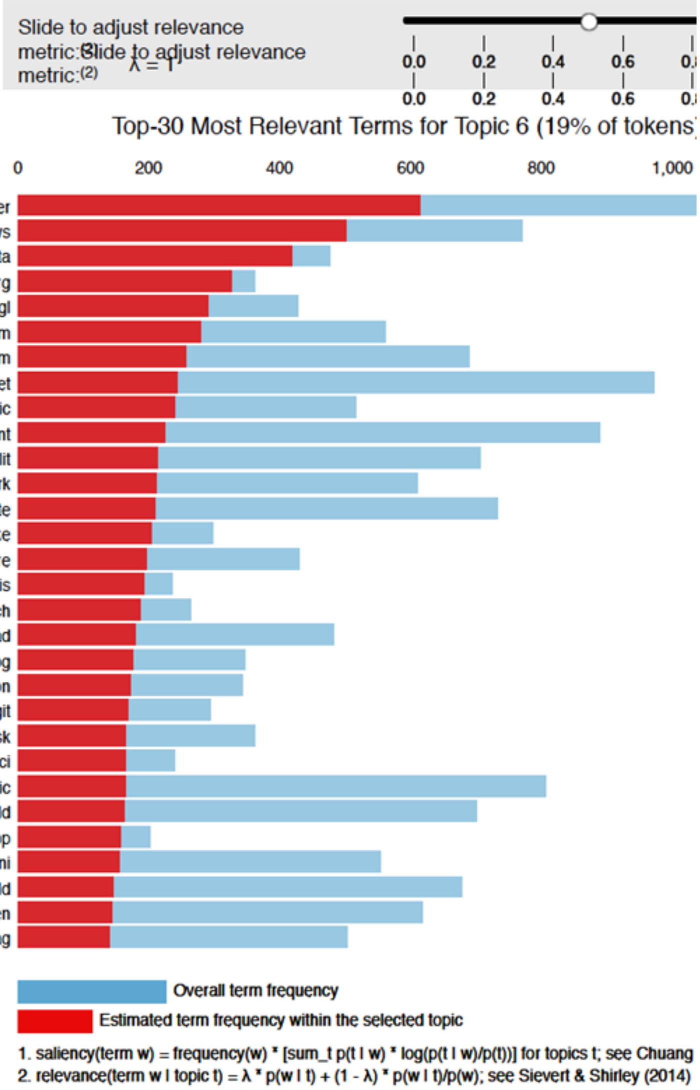
Topic 5 - Charlie Hebdo



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang

2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Topic 6 - Facebook - Fake news and privacy



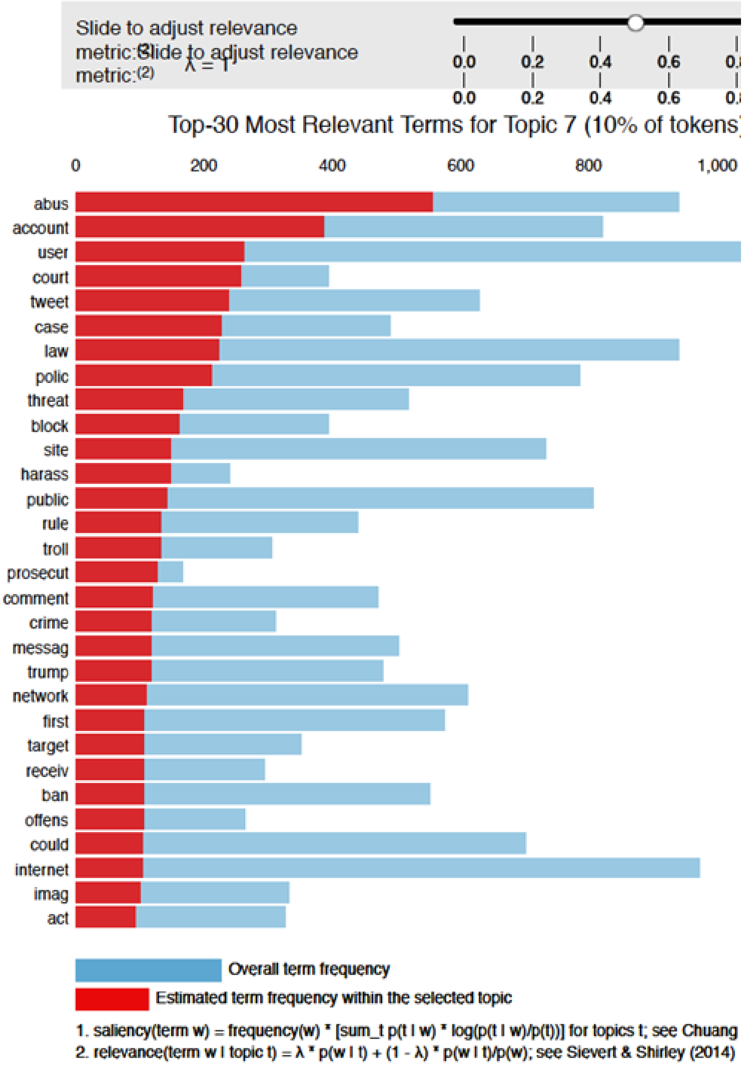
Slide to adjust relevance metric:(2) λ = 1

Slide to adjust relevance metric:(2)

Top-30 Most Relevant Terms for Topic 6 (19% of tokens

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Topic 7 - Twitter abuse/harassment - trolls and blocks



Slide to adjust relevance metric:Slide to adjust relevance metric:(2) λ = 1

Top-30 Most Relevant Terms for Topic 7 (10% of tokens

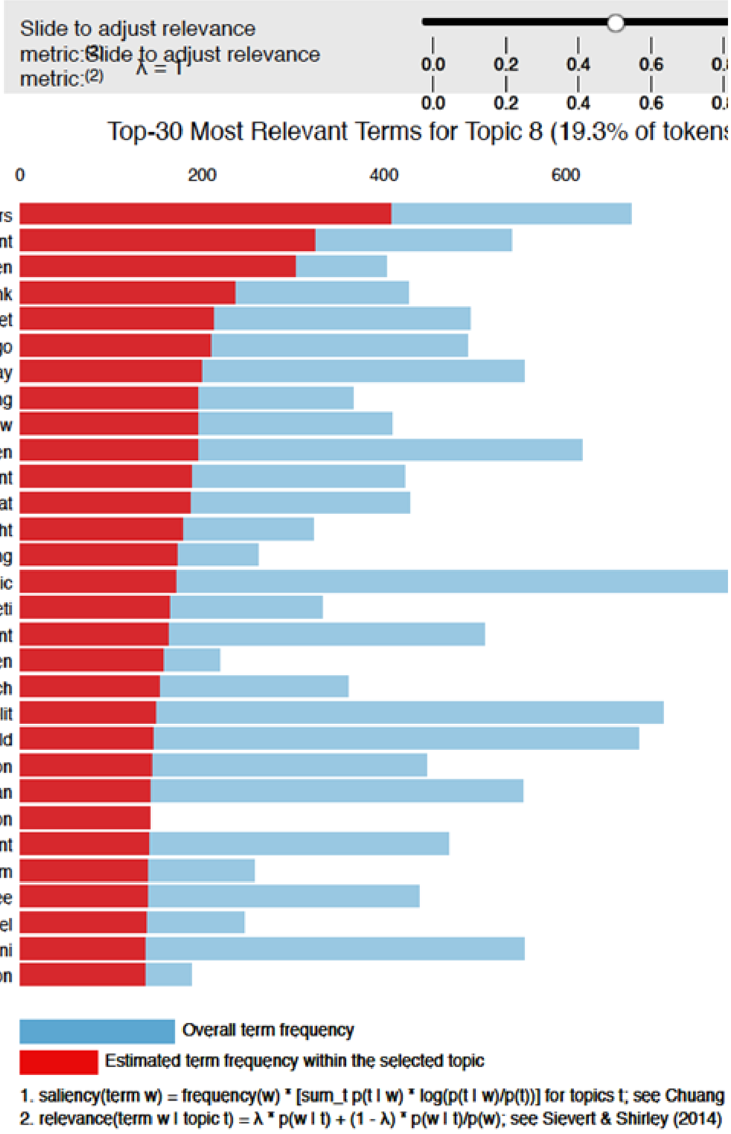
🟦 Overall term frequency

🟥 Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t I w) * log(p(t I w)/p(t))] for topics t; see Chuang

2. relevance(term w I topic t) = λ * p(w I t) + (1 - λ) * p(w I t)/p(w); see Sievert & Shirley (2014)

Topic 8 - University Student Union - No Platform - Jordan Peterson - women

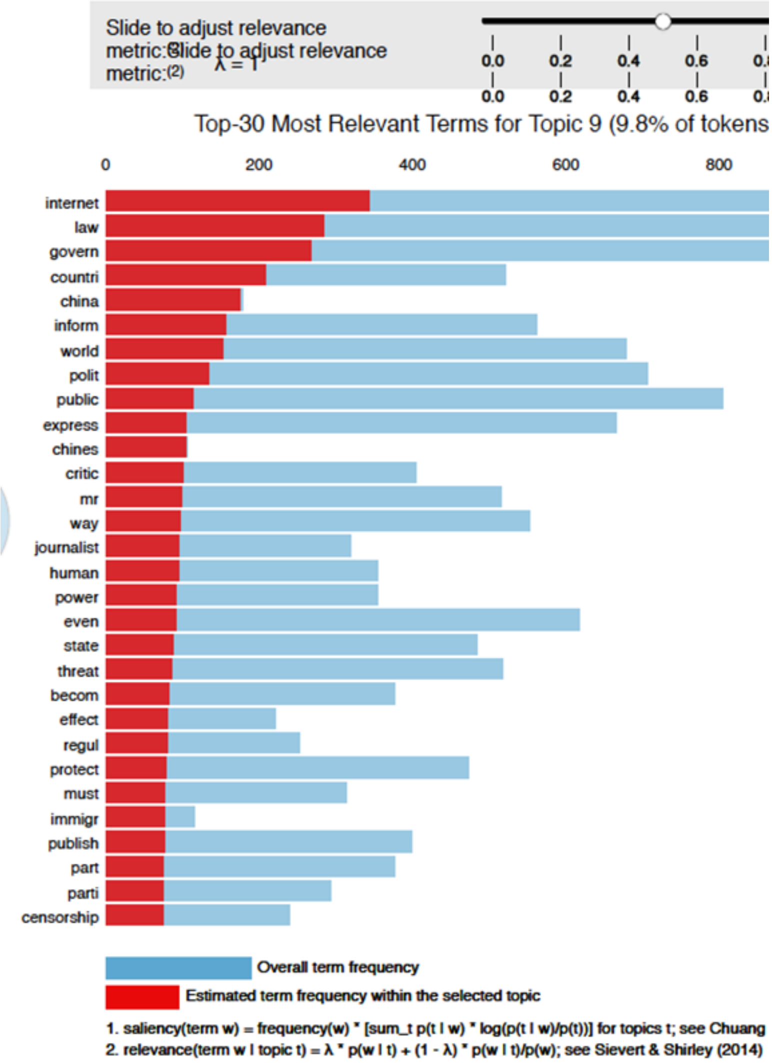


🟦 Overall term frequency

🟥 Estimated term frequency within the selected topic

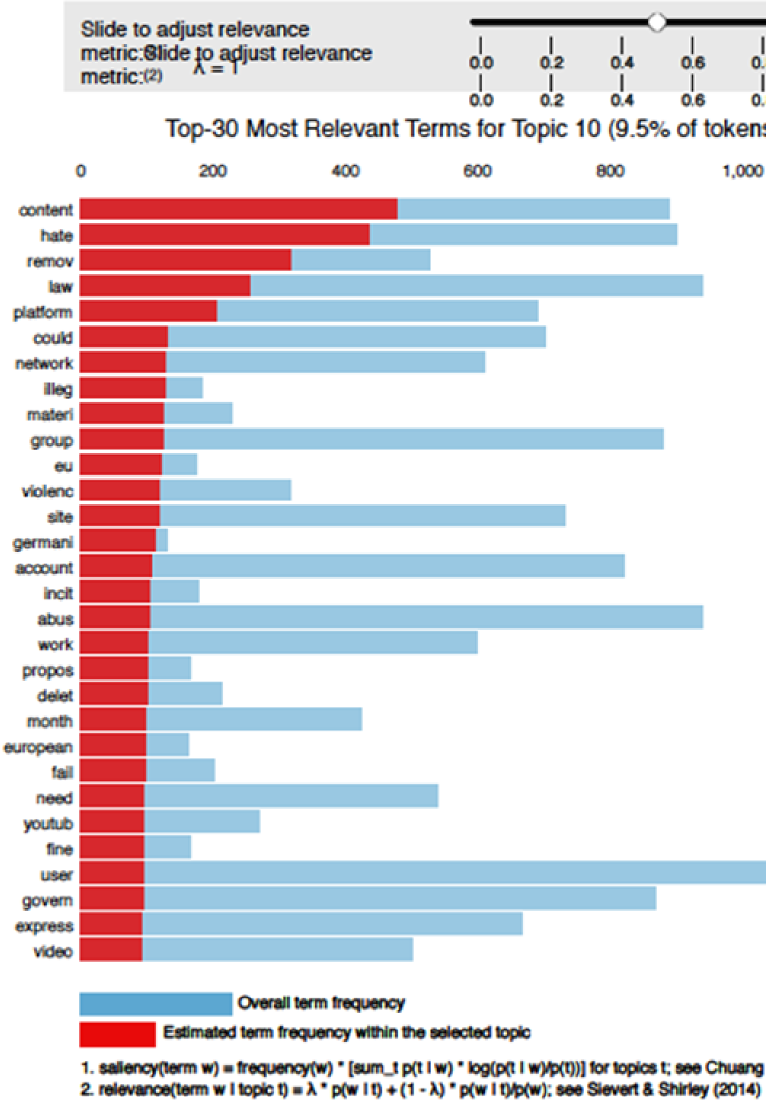
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Topic 9 -China, Internet regulation and censorship



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Topic 10 - content removal_hate speech, Germany



Slide to adjust relevance metric:Slide to adjust relevance metric:(2) λ = 1

Top-30 Most Relevant Terms for Topic 10 (9.5% of token

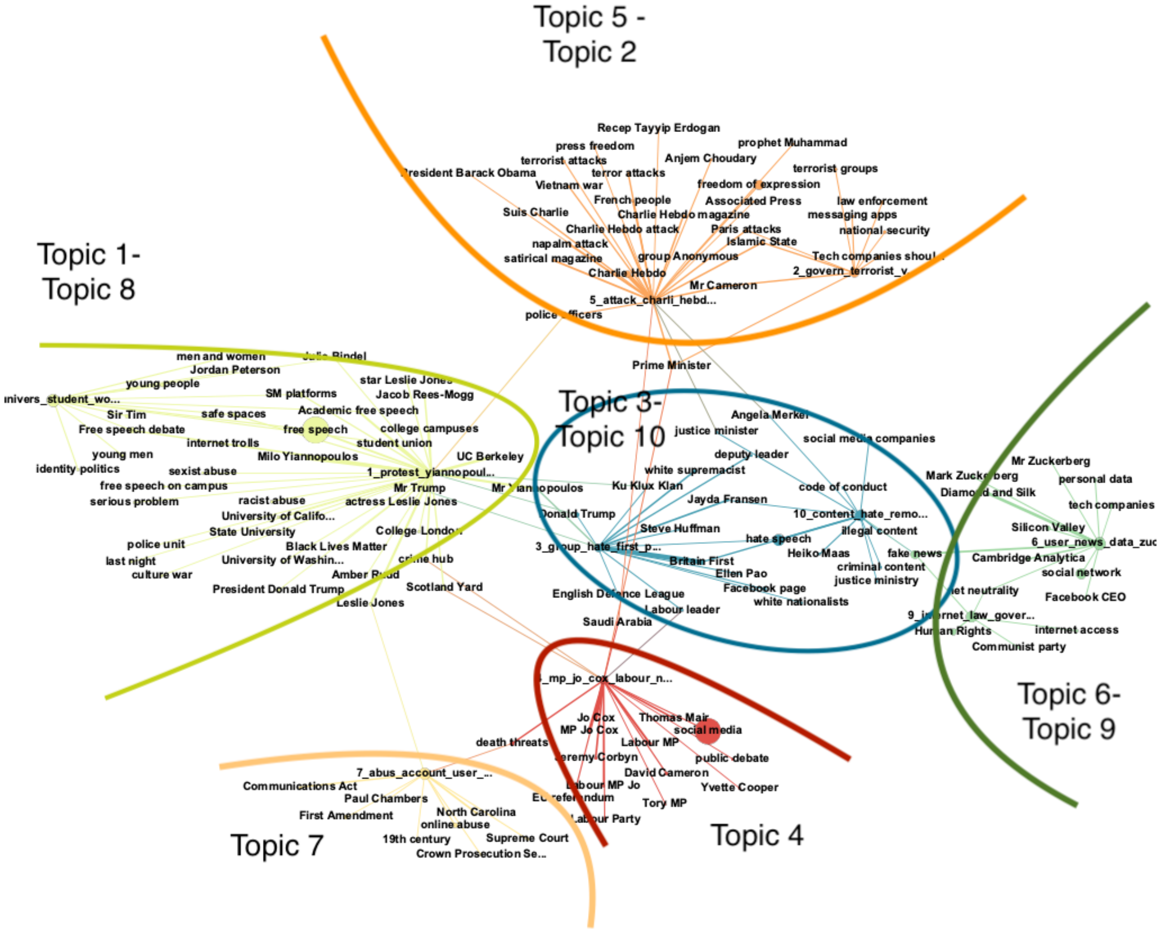
Overall term frequency

Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang

2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

4) Newspapers: Semantic relations topics-keywords



5) Time plot of the topics

Below is the table with the data from the demographic analysis performed in Cortext used in the temporal plotting of the topics.

Table A. 5 – Cortext demography output – raw frequency of documents associated with the topics divided by months (in initials)

	2015												2016												2017												2018				Tot
	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	
Topic1	1	7	2	2	4	3	1	1	1	4	5	7	3	6	1	2	4	1	15	13	2	2	8	4	1	9	9	6	6	4	4	9	5	7	1	1	3	6	8	2	180
Topic2	24	9	8	5	10	5	1	2	4	5	11	9	8	6	6	7	5	4	8	7	4	5	10	5	4	2	11	8	9	8	8	9	5	8	2	5	7	9	9	8	280
Topic3	8	8	2	5	8	1	5	1	4	3	11	7	5	5	4	4	4	3	9	4	2	4	12	6	3	3	7	1	6	6	4	14	5	3	5	6	8	6	15	7	224
Topic4	11	7	3	1	5	4	6	0	2	6	1	4	3	5	8	4	6	6	8	10	1	1	9	3	2	4	7	0	6	5	2	3	2	7	2	6	7	13	6	5	191
Topic5	33	14	6	5	9	6	3	1	3	3	9	5	8	7	7	4	4	0	7	7	9	3	6	4	2	2	4	5	7	8	7	0	2	0	0	5	3	3	4	0	215
Topic6	17	9	6	7	8	4	8	2	5	7	11	11	6	14	5	9	11	6	7	6	4	6	19	6	6	6	15	5	13	12	9	21	13	11	9	8	14	12	14	24	386
Topic7	13	15	13	11	11	11	10	3	4	2	7	11	6	14	7	6	7	4	21	13	5	10	12	7	6	6	13	6	13	10	11	16	4	11	4	6	10	10	13	6	368
Topic8	16	16	8	8	8	15	8	3	7	11	7	12	6	15	8	8	9	9	21	10	2	4	13	4	5	5	16	6	10	11	8	20	7	5	5	7	10	16	20	12	391
Topic9	16	13	5	7	9	4	10	2	5	6	6	12	7	9	7	10	10	6	9	2	3	8	16	4	4	5	10	7	10	8	9	16	9	10	6	9	13	13	16	9	340
Topic10	14	9	4	7	9	6	5	1	2	4	4	3	8	15	5	4	6	7	10	9	5	6	15	5	5	4	18	5	16	12	9	21	8	14	8	12	13	8	13	14	343

Figure A. 14 - Cortext demography visualisation – time plotting of topics

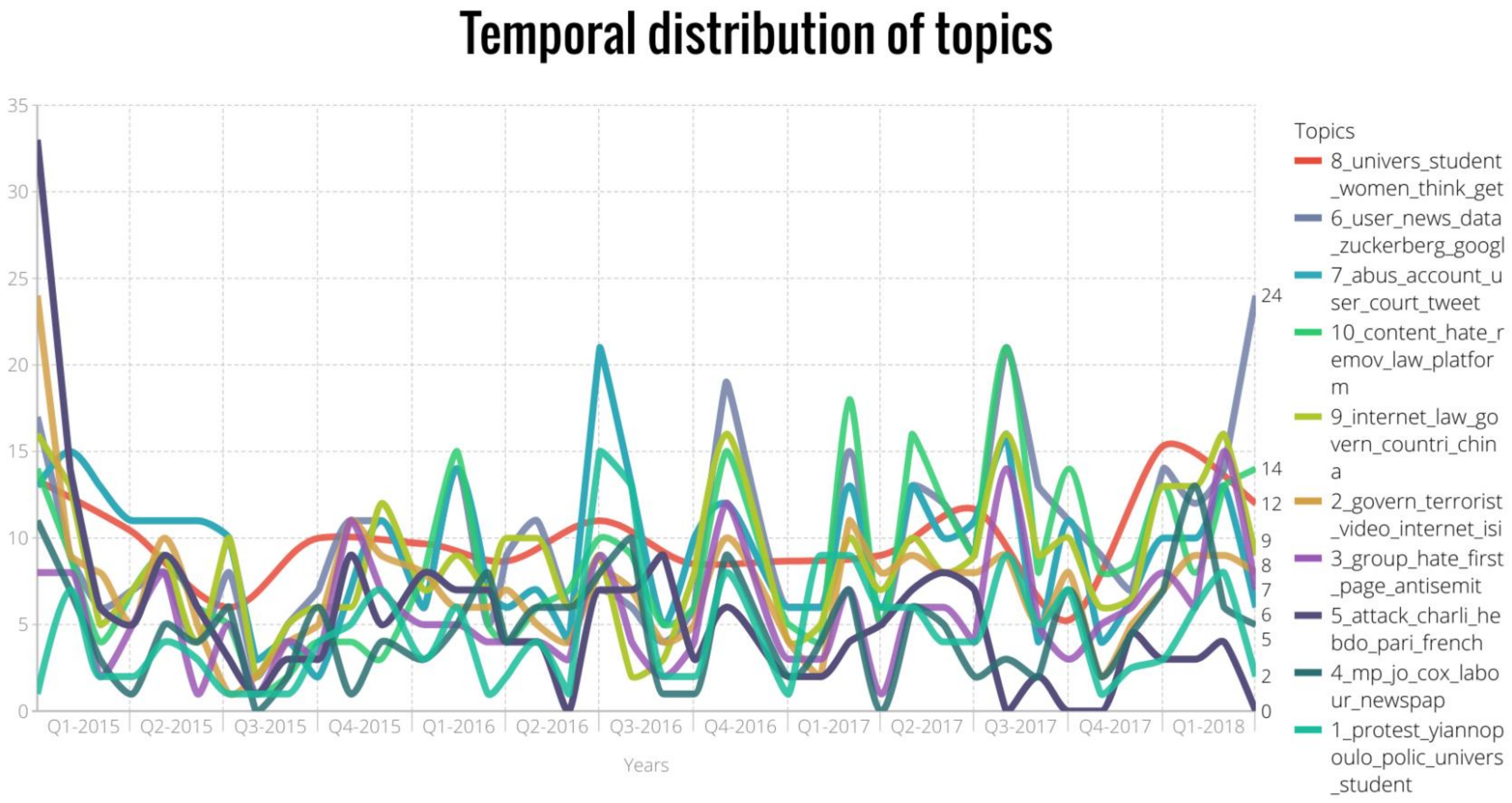
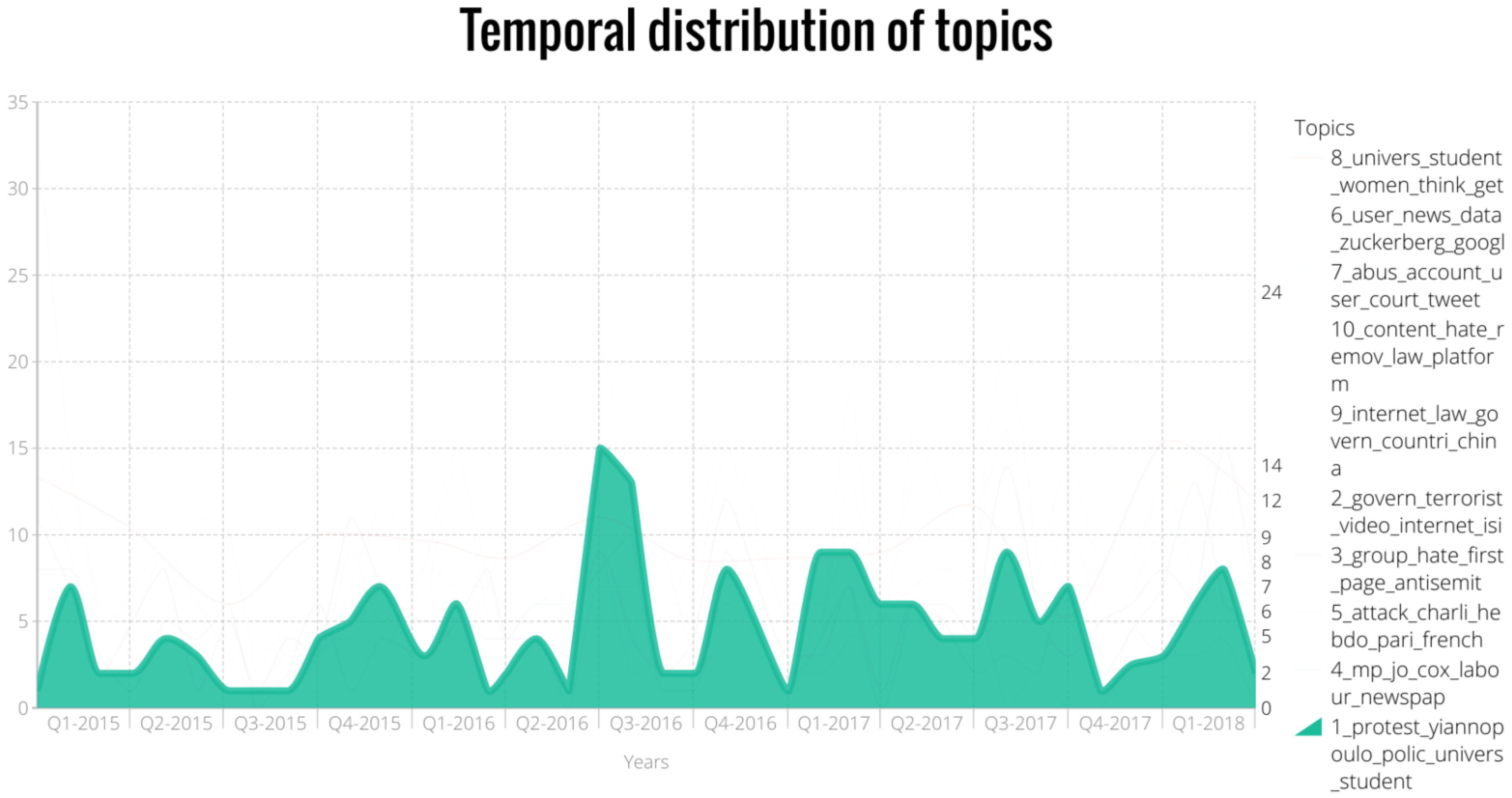
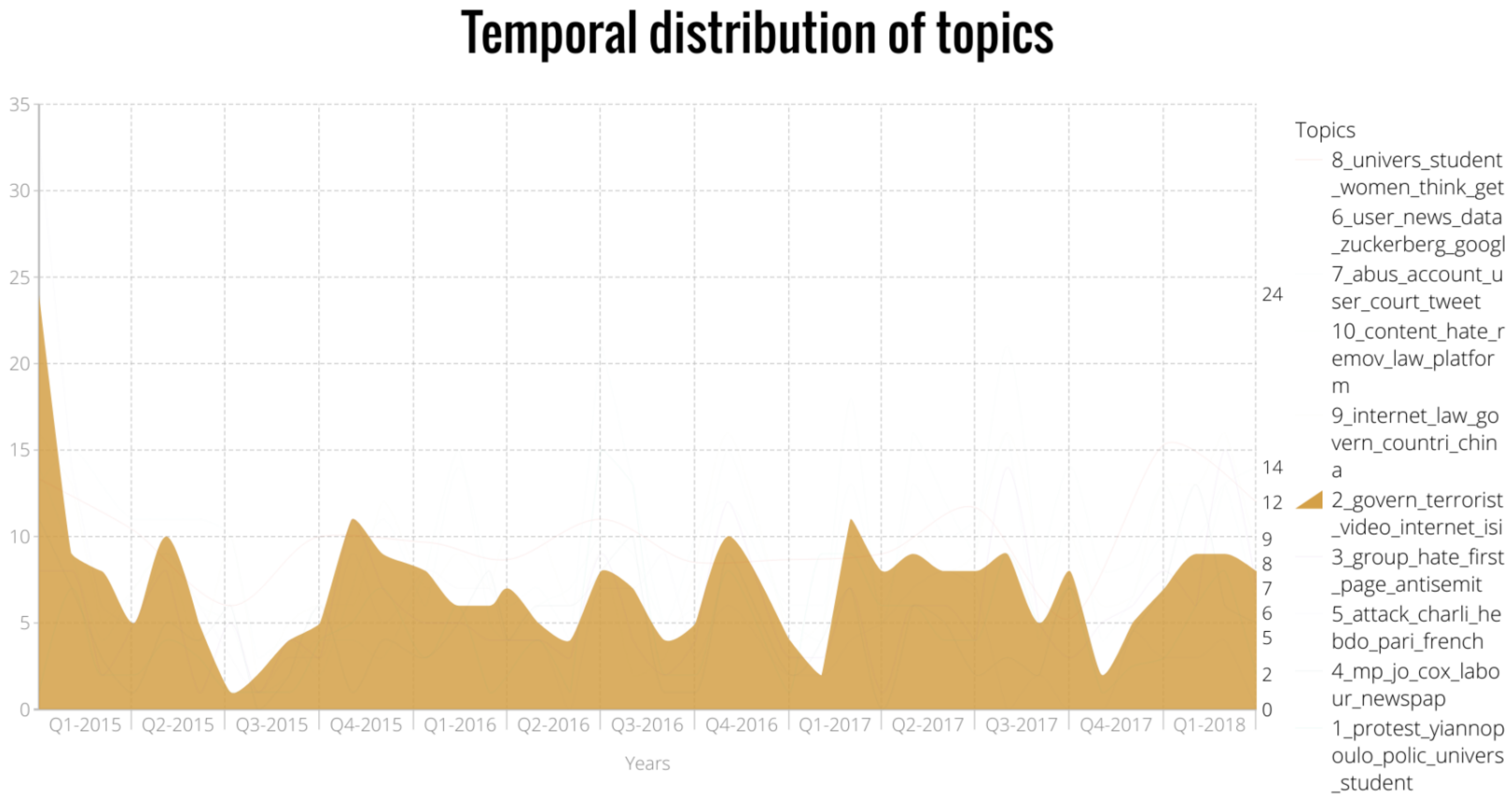


Figure A. 15 - Cortext demography visualisation - time plotting divided by topic

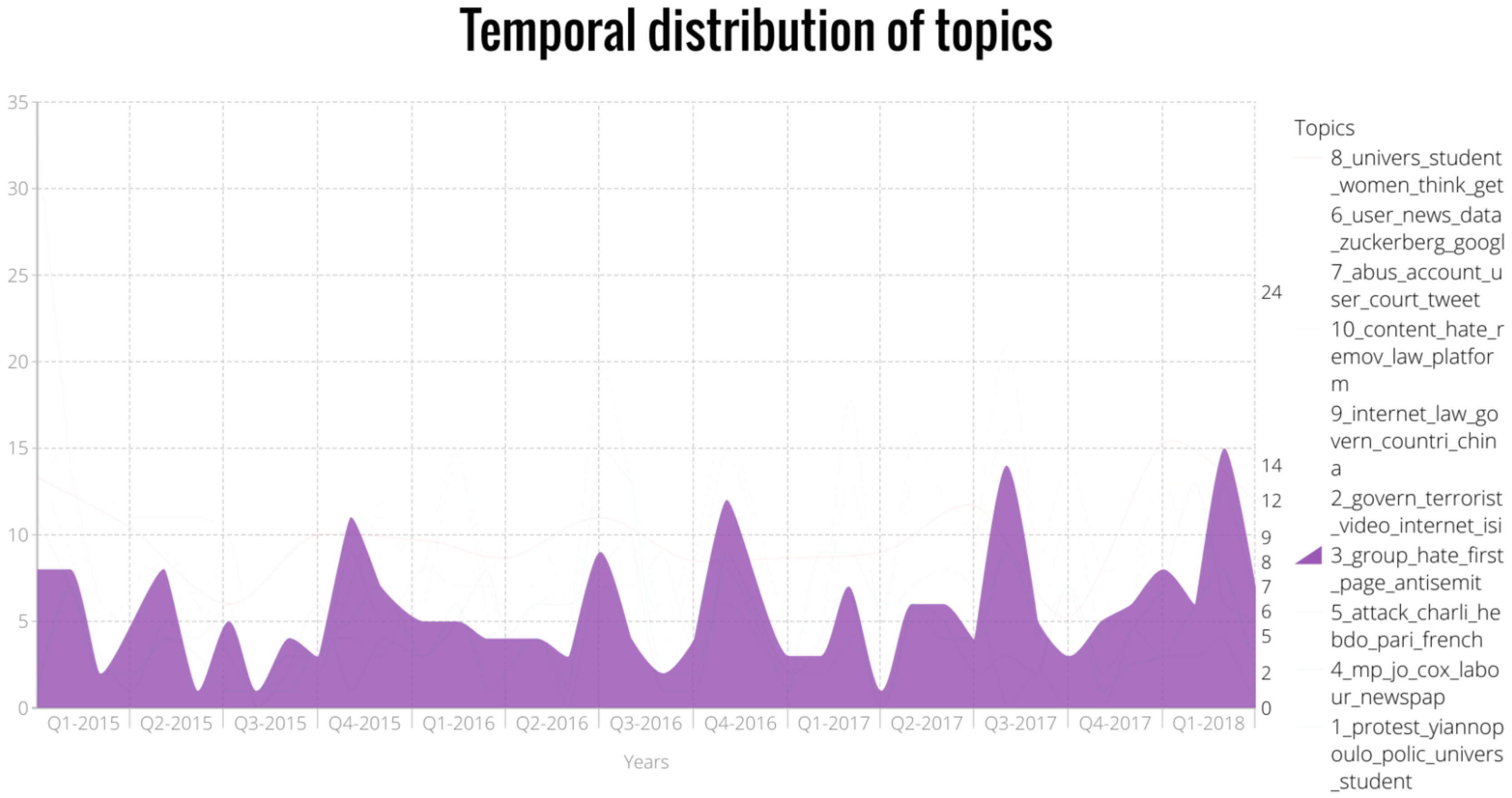
Temporal distribution - Topic 1



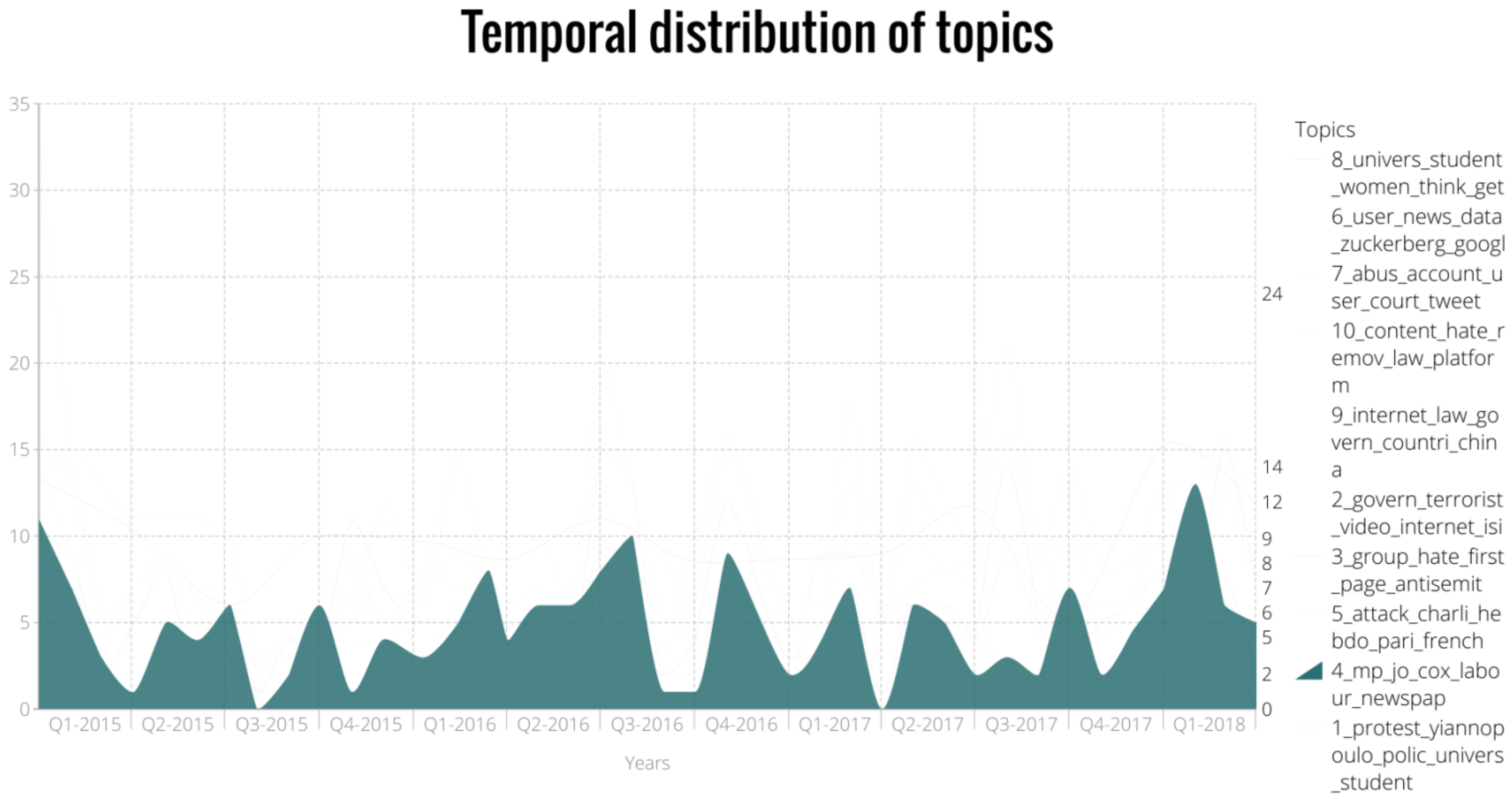
Temporal distribution - Topic 2



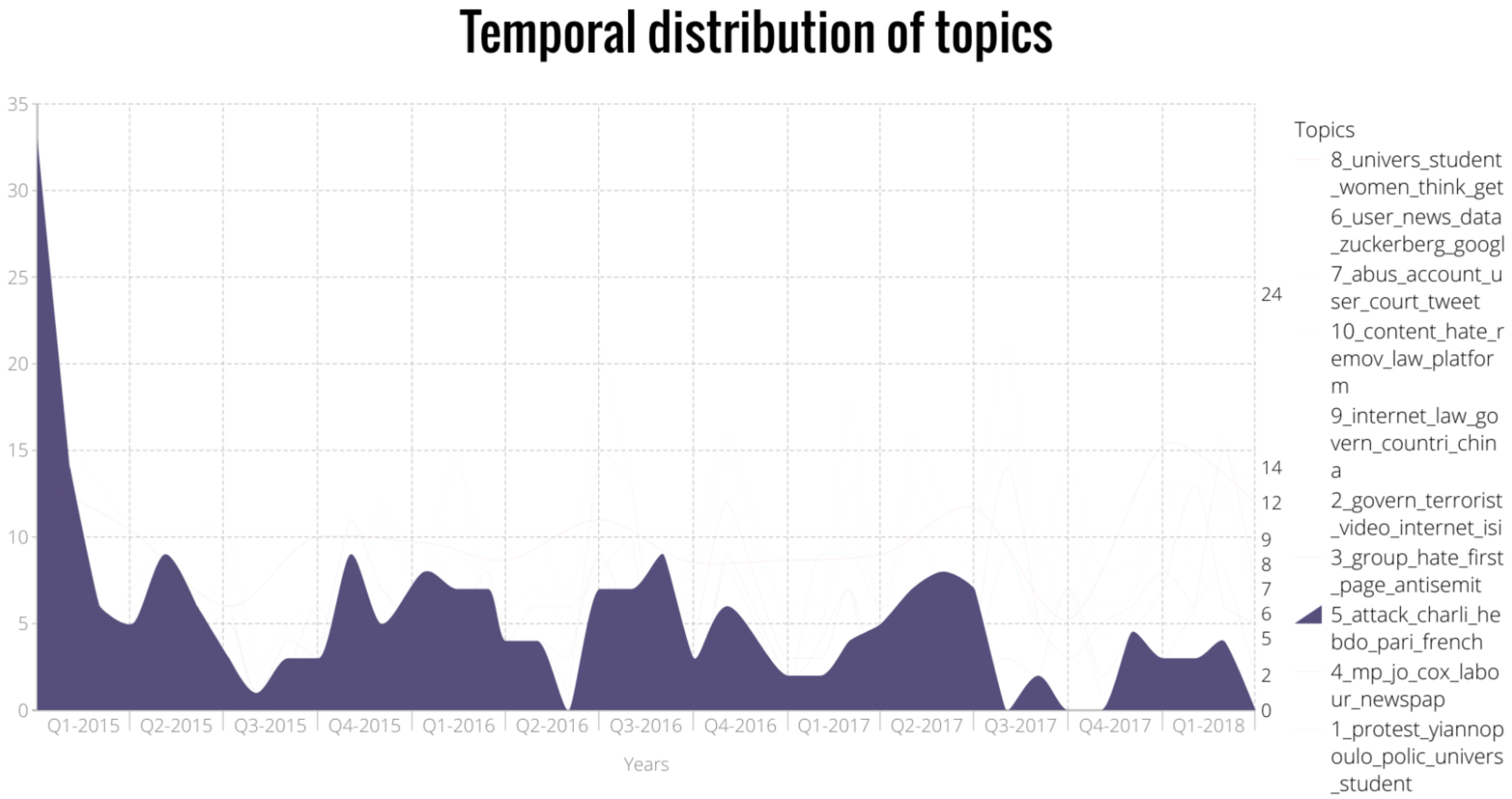
Temporal distribution - Topic 3



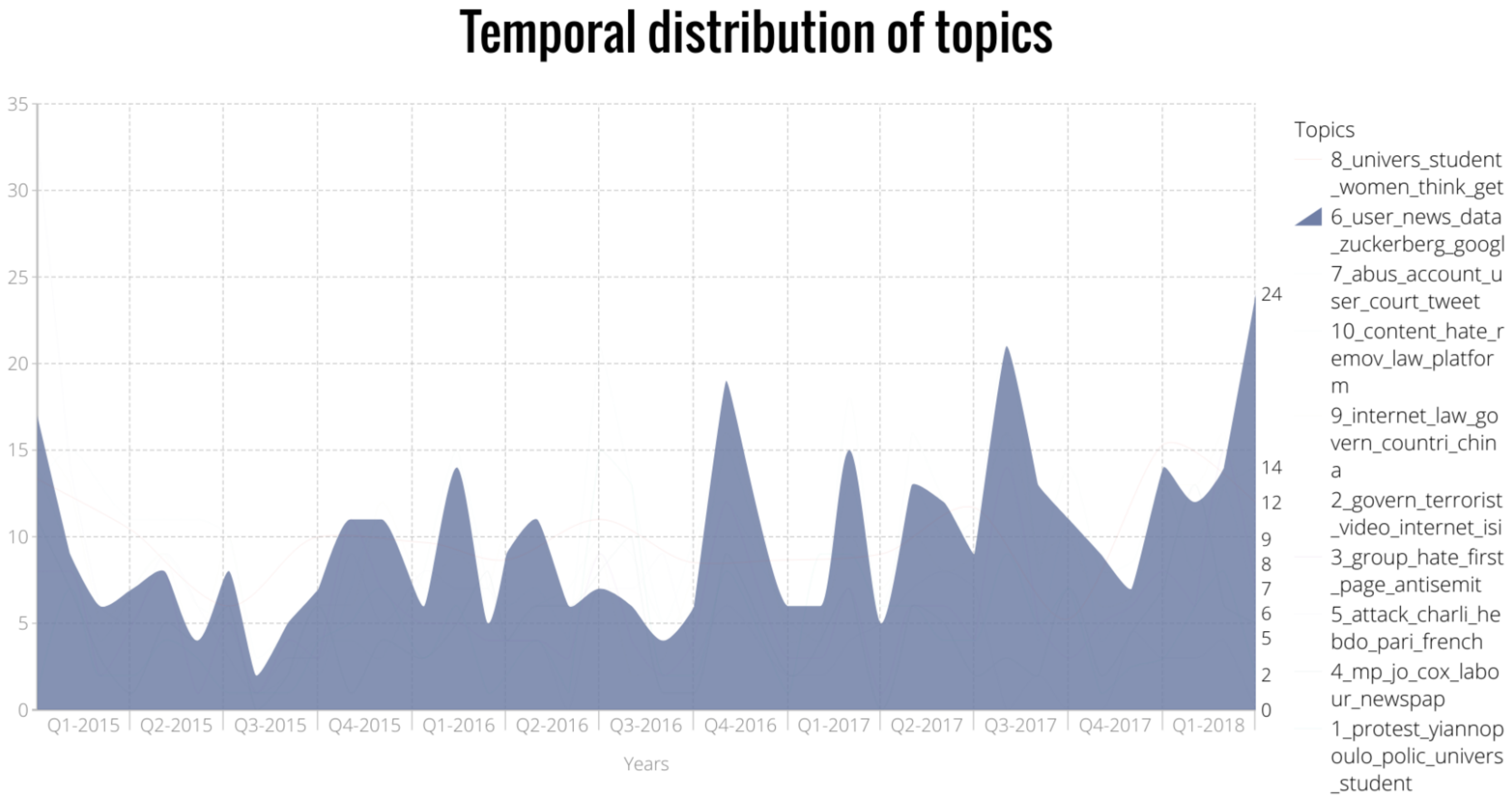
Temporal distribution - Topic 4



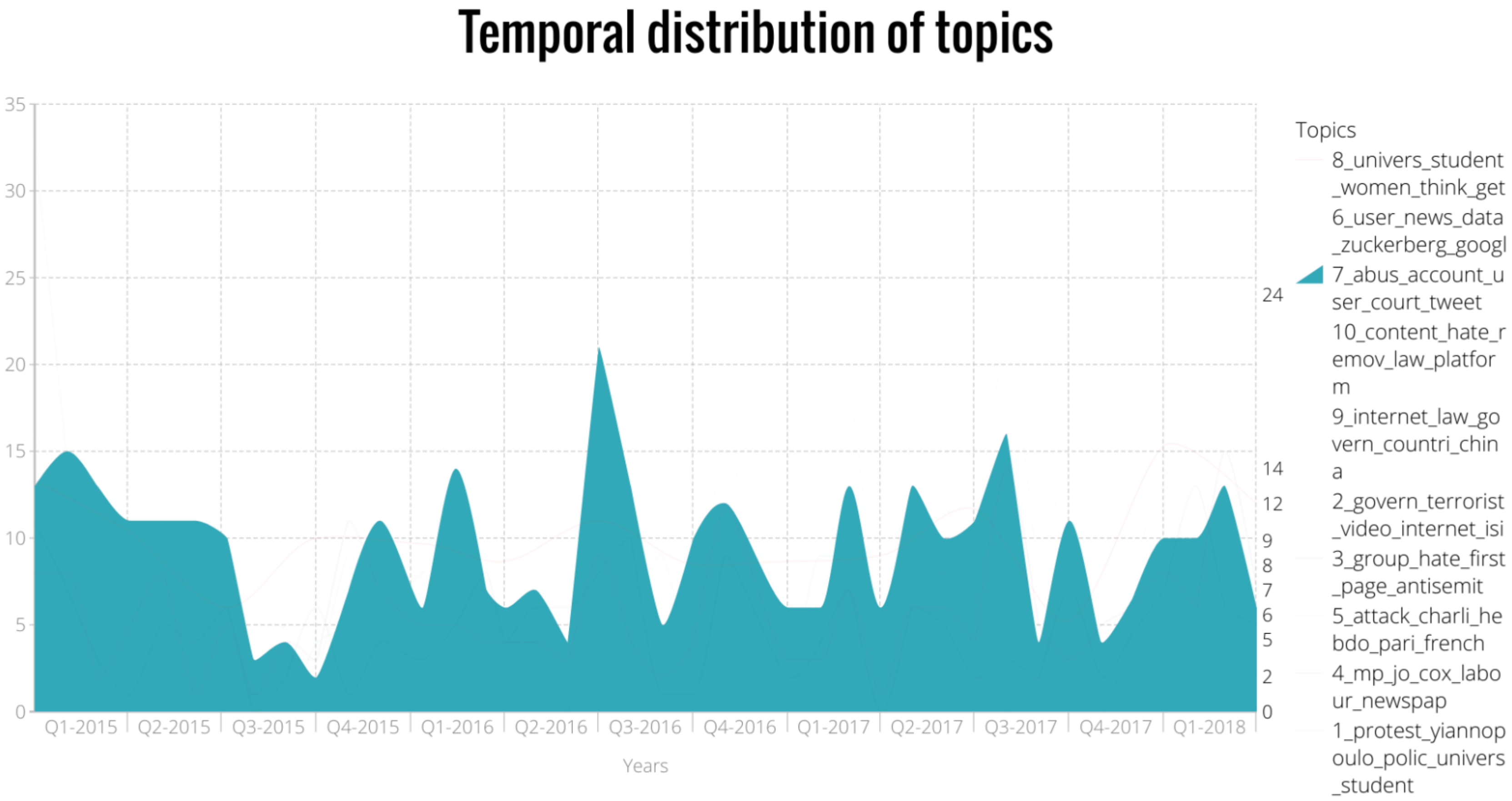
Temporal distribution - Topic 5



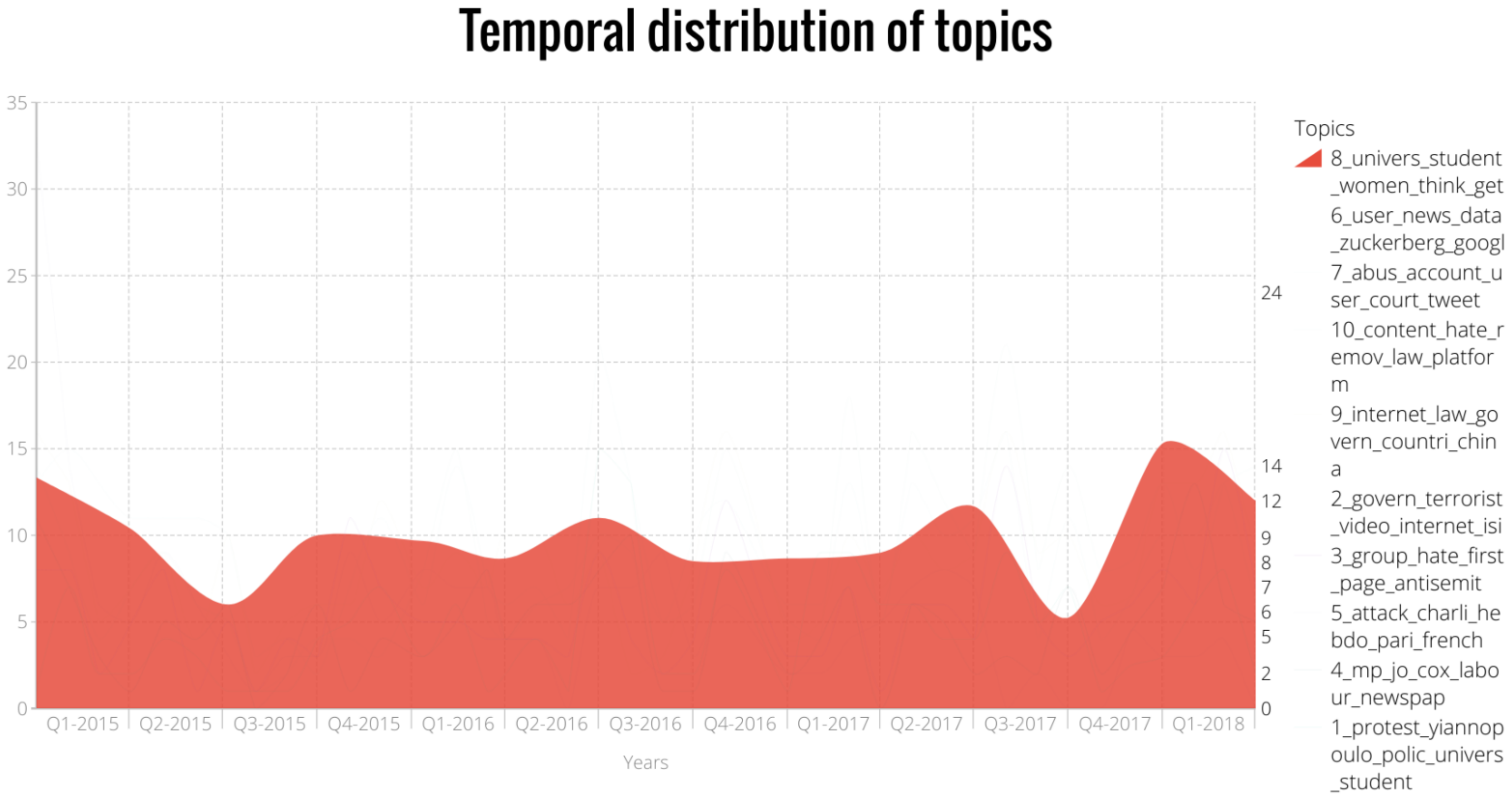
Temporal distribution - Topic 6



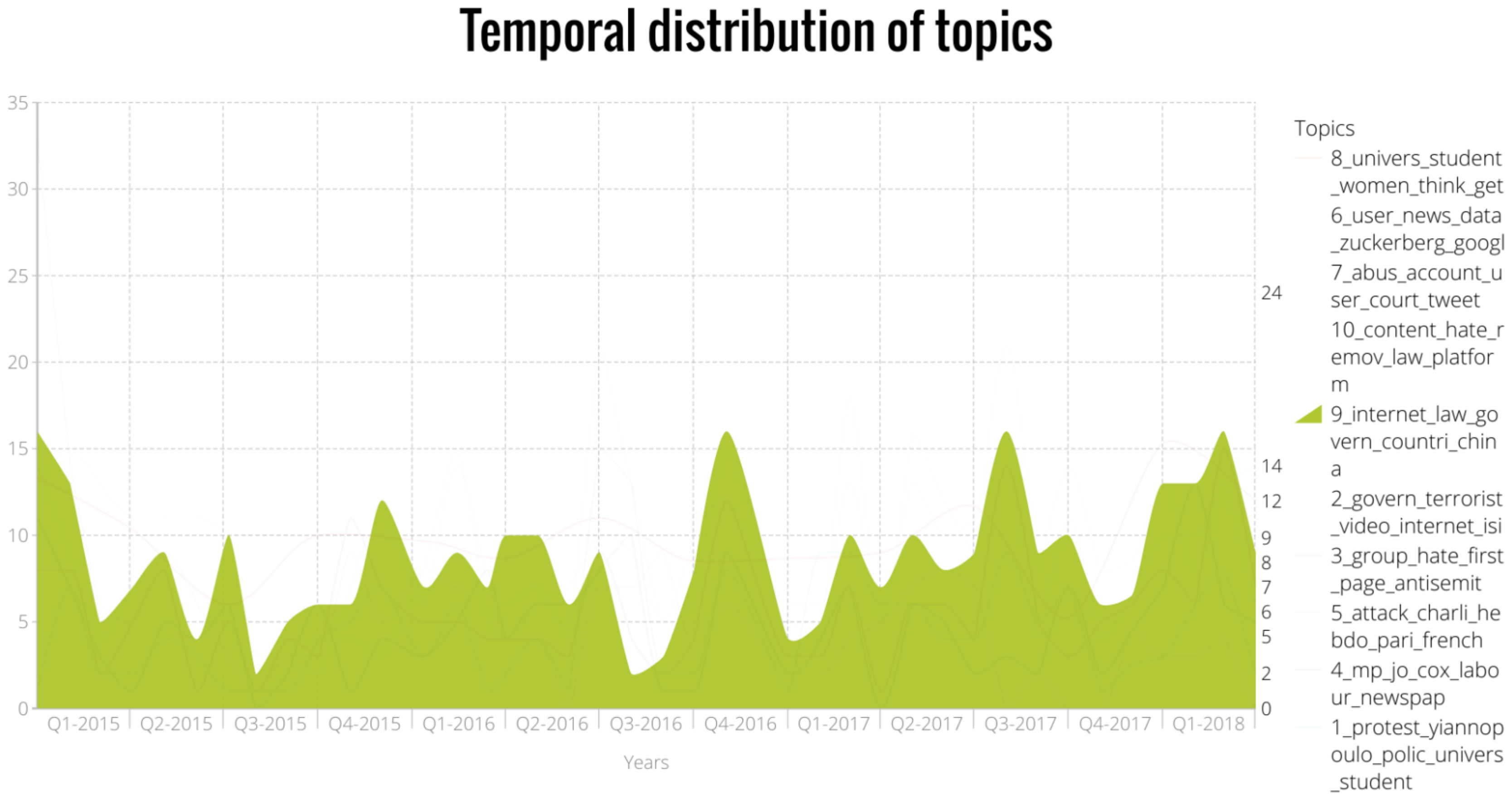
Temporal distribution - Topic 7



Temporal distribution - Topic 8

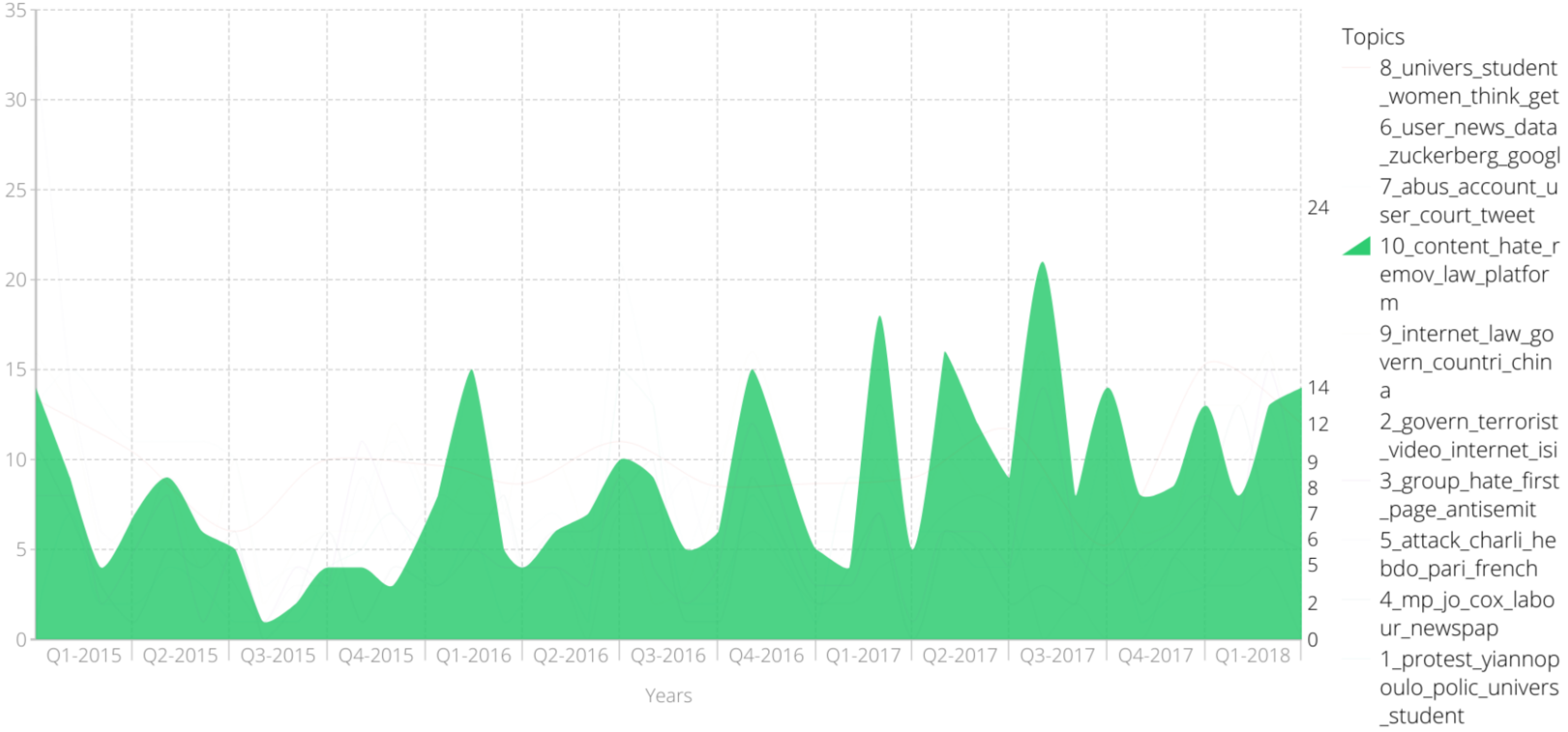


Temporal distribution - Topic 9



Temporal distribution - Topic 10

Temporal distribution of topics



6) Newspapers sampling

Newspaper sample: original publications in the sample and in the subset for qualitative analysis

2015

Large sample		**Subset**	
The Guardian	234	The Guardian	57
MailOnline * sig	217	Independent.co.uk	32 / 29
Independent.co.uk	168	MailOnline * sig	30 /33
telegraph.co.uk	110	mirror.co.uk	21
mirror.co.uk	66	telegraph.co.uk	20
The Times (London)	44	The Times (London)	15
Express Online	29	Express Online	9
The Sunday Times (London)	24	i-Independent Print Ltd	6
The Daily Telegraph (London)	20	The Sunday Times (London)	5
The Independent (London)	17	The Sun (England)	4
The Sun (England)	17	DAILY MAIL (London)	2
i-Independent Print Ltd	14	The Daily Telegraph (London)	2
DAILY MAIL (London)	13	The Independent (London)	2
The Observer (London)	10	The Observer (London)	2
Daily Mirror*	4	The Express	1
The Sunday Telegraph (London)	4	The Independent on Sunday	1
The Express	3	The Sunday Telegraph (London)	1
MAIL ON SUNDAY (London) *	2		
Daily Star *	1		
Sunday Express *	1		
The Independent on Sunday	1		
The People *	1		

2016

Large sample		**Subset**	
MailOnline * sig	254	The Guardian	50 * sig /45
The Guardian * sig	183	MailOnline	35 * sig /40
telegraph.co.uk	109	The Independent (United Kingdom)	25
The Independent (United Kingdom)	104	telegraph.co.uk	18
mirror.co.uk	73	Express Online	15
Express Online	56	The Times (London)	13
The Times (London)	44	mirror.co.uk	11
The Daily Telegraph (London)	33	Independent.co.uk	9
Independent.co.uk	29	The Daily Telegraph (London)	7
The Sun (England)	27	The Independent - Daily Edition	3
The Sunday Times (London)	25	The Sun (England)	3
i-Independent Print Ltd	14	i-Independent Print Ltd	3
The Independent - Daily Edition	10	DAILY MAIL (London)	2
DAILY MAIL (London)	9	The Express	1
Daily Mirror	8	Daily Mirror	1
The Observer (London)	6	The Observer (London)	1
The Express	4	The Sunday Times (London)	1
The Sunday Telegraph (London) *	3		
Sunday Express *	2		
MAIL ON SUNDAY (London) *	2		
Daily Star *	2		
The Independent (London) *	2		
Daily Star Sunday *	1		

2017

Large sample		**Subset**	
MailOnline	299	MailOnline	46
The Independent (United Kingdom)	128	The Guardian(London)	34
The Guardian(London)	123	telegraph.co.uk *sig	33/ 28
telegraph.co.uk *sig	98	The Independent (United Kingdom)	30/35
The Times (London)	69	The Times (London)	15
Express Online	53	Express Online	9
mirror.co.uk	51	The Sun (England)	7
The Sun (England)	30	mirror.co.uk	5
DAILY MAIL (London)	21	The Observer(London)	4
The Sunday Times (London)	20	The Daily Telegraph (London)	3
i-Independent Print Ltd	20	The Sunday Times (London)	2
The Daily Telegraph (London)	19	i-Independent Print Ltd	2
The Guardian	14	DAILY MAIL (London)	2
The Observer(London)	12	The Sunday Telegraph (London)	2
The Independent - Daily Edition	11	The Guardian	2
Daily Mirror	10	The Independent - Daily Edition	2
The Sunday Telegraph (London)	7	Daily Mirror	1
The Express *	5		
MAIL ON SUNDAY (London) *	4		
The Observer (London) *	3		
Sunday Express *	3		

2018

Large sample		**Subset**	
MailOnline	86	The Guardian(London)	24
The Guardian(London)	80	MailOnline	21
The Independent (United Kingdom)	55	The Independent (United Kingdom)	13
The Times (London)	39	telegraph.co.uk	11
telegraph.co.uk	33	Express Online	6
Express Online	33	The Times (London)	4
mirror.co.uk	17	mirror.co.uk	4
The Sunday Times (London) *	15	DAILY MAIL (London)	3
DAILY MAIL (London)	15	The Daily Telegraph (London)	3
The Daily Telegraph (London)	15	The Sun (England)	3
The Sun (England)	14	The Independent - Daily Edition	2
The Observer(London)	9	The Observer(London)	2
i-Independent Print Ltd	9	The Sunday Times (London)	2
The Independent - Daily Edition	9	i-Independent Print Ltd	2
Daily Mirror	5		
The Express	4		
The Sunday Telegraph (London)	3		
MAIL ON SUNDAY (London)	2		
The People	1		
Sunday Express	1		

*sig = statistically significant difference

* Publications with very low representation in the original set are not present in the subset.