Developing Qualitative and Quantitative Methodologies for the analysis of Political Discourse on Twitter: *An Irish Case Study*

Dónal Mulligan

Thesis presented for the award of PhD, June 2016 School of Communications, Dublin City University

Supervisor: Dr. Miriam Judge

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of *Doctor of Philosophy* is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:	
0	

ID No.:

Date:

Acknowledgements

I wish to express my sincere gratitude to everyone who has helped me in the course of this research. While the completion of a doctoral thesis is a long and solitary endeavour, it is not a lonely one when it is supported by the wisdom, guidance, humour, encouragement, and solidarity of colleagues, friends, and family. I was lucky to have received such support throughout and am glad, at the completion of writing up this work, to take time to say 'thank you'.

Firstly, to DCU School of Communications, from which a scholarship funded this research, and in which many members of staff offered support and guidance. Particular thanks are due to Professor Pat Brereton, Head of School, to my teaching colleagues on the Multimedia programs, and especially to Dr. Trish Morgan, whose advice and sharing of experience was invaluable at many stages. Of course, thanks must also go to Dr. Miriam Judge, the supervisor of this work, for her encouragement and patience throughout.

To the participants who took part in this research, offering their time and sharing their insights, I am deeply indebted. I could not have completed this project without their generous contributions of knowledge, and the stimulating ideas that emerged from speaking with them.

I am also sincerely grateful to the many friends who have offered support during the course of the long process of completing this work, and I am especially appreciative of the help of Eoin Wilson and Gráinne Walsh. To my partner, Gary O'Sullivan, who is concurrently working on his own PhD but who frequently and generously took the time to read, advise, and discuss mine - to its substantial betterment - I owe an enormous debt of gratitude and love; *Thank you, very much indeed*.

Finally, special and sincere thanks go to my family - in particular to my parents who have been an unwavering source of support and encouragement, and to my grandmother who has similarly always provided an encouraging word and a proud validation, through this and many previous efforts. *This work is dedicated to them, in recognition of their considerable dedication to me, over the course of my education.*

Abstract

Developing Qualitative and Quantitative Methodologies for the analysis of Political Discourse on Twitter: An Irish Case Study

This work aims to develop the analytical methodologies for the study of political discourse taking place on Twitter, by examining large-scale quantitative approaches and informing these with qualitative sources. The study of Twitter-mediated political communication is a relatively new field, lacking a definitive framework for data collection or analysis; methodologies used in the literature are often disparate and *ad hoc*, reducing their potential for comparative investigations. Whilst recent studies have begun to provide the foundations of a coherent analytical toolset, the field remains one of rich potential but without well-tested frameworks. The extent of bias in the dominant data collection approach and the suitability and effectiveness of analysis techniques are largely unclear, and are subject to poorly defined caveats in the literature.

This thesis takes a mixed-methods approach to the investigation and development of these analytics, utilising case studies from Ireland (where the medium is well established for political communication). In-depth interviewing of agents in the Irish context provides information on their conventions, behaviours, and communicative strategies, which informs large-scale quantitative analyses by the researcher. Concepts from Bourdieu's *Field Theory* are used to understand the power structures within this mediated space, and to relate it to other fields, and an extensive review of the interdisciplinary literature constituting the field of Twitter political discourse analysis is undertaken.

This review critiques established methodological approaches in the literature and argues that a framework built on clearer qualitative understanding of the participants involved improves the insights and comparative value of data derived from Twitter. Case studies demonstrate the efficacy of methodologies used and the thesis responds to the limitations of noncommercial data collection by academics. These studies thereby also provide the first largescale rigorous analyses of the structure of and participation in Irish political discourse on Twitter, providing both methodological model and comparative data for similar research elsewhere.

Table of Contents

1. Introduction	
1.1 Introduction to Twitter	
1.1.1 Twitter usage in Ireland6	
1.1.2 Twitter's use for political discussion7	
1.1.3 History of Twitter, and development of its established	
affordances & conventions8	
1.1.3.1 Addressing and re-broadcasting content	
1.1.3.1 Content-marking: Hashtags	
1.1.3.2 Notes on the use of Hashtags	
1.2 Extant research on Twitter	
1.2.1 The study of political discourse on Twitter	
1.2.2 Hashtags in data collection	
1.2.2 Implications for research15	
1.3 Theorising political discourse online and on Twitter	
1.3.1 A network-based model	
1.4 Primary Hypothesis and Research Themes	
1.4.1 Primary Hypothesis	
1.4.2 Research Themes	
1.4.2.1Addressing the networked structure of Twitter political discourse221.4.2.2Understanding the social context of Twitter political discourse221.4.2.3Establishing the extant methodologies for Twitter discourse analysis231.4.2.4Testing the effectiveness of collection and analytical methods231.4.2.5Informing methodology with qualitative investigation of actors241.4.2.6Developing the understanding of how hashtags are used in political discourse24	
1.4.3 Prioritised Research Questions	
1.4.4 Scope	
1.4.5 Levels of enquiry	
1.5 Methodology	
1.5.1 Quantitative Methods	
1.5.2 Qualitative Methods	
1.5.3 Mixed Methods	
1.5.4 Case Studies	
1.6 Layout of the thesis	
1.6.1 Thesis Guide	
1.6.1.1 Chapter Outlines	

2. Tech Situa	nology & Society: ting mediated social communication	
2.1 I	Introduction	35
2.2	Theories of Technological Determinism	36
2.2.1	The work of Marshall McLuhan	
2.2.2	Determinism in visions of the 'Information Age'	37
2.3	Theories of the Social Shaping of Technology	40
2.3.1	Raymond Williams' critique of McLuhan	40
2.3.2	The Social Shaping of Technology	41
2.3.3	The Social Construction of Technology	43
2.4 A	Actor-Network Theory	45
2.4.1	Technological Agency	45
2.4.2	The Network of Actor-Network Theory	47
2.4.3	Applicability of Actor Network Theory to Twitter	
2.5 I	Bourdieu's Theories of Social Interaction	50
2.5.1	Habitus	50
2.5.2	Field	51
2.5.3	Capital	52
2.5.4	Doxa	53
2.5.5	The field of a social media platform	53
2.6 (Conclusions	55
2.6.1	Implications for this research	57
3. Poli	tical Discourse on Twitter and its Analysis	
3.1 I	Introduction	59
3.2 I	Related Literature Reviews	62
3.3 I	Disciplines, Theories & Methods	63
3.3.1	Theories from Political Science	64
3.3.2	Theories from Communications & Media Studies	66
3.3.3	Theories from Psychology	67
3.3.4	Methods	68
3.4 I	Data Collection & Selection	70
3.4.1	Data Collection Methods for Twitter	73
3.4.2	Data Selection	76

3.5 E	thical Considerations	79
3.5.1	Ethical Practice for Internet Research	80
3.5.2	Ethical imperatives for this research	81
3.6 C	Conclusions	82
4. Met	hodology	
4.1 I	ntroduction	85
4.1.1	Big Data: Deep Data & Surface Data	86
4.1.2	Opening the "black box"	
4.1.3	Research Design and Pilot Study	
4.2 N	lixed Methods Research Design	90
4.2.1	Situating the Mixed Methods approach	90
4.2.2	Categorisation of research questions constituent to this thesis	92
4.2.3	The Explanatory Design model	93
4.2	3.1 The Explanatory Design: Participant Selection model	94
4.3 1	witter Data Collection	96
4.3.1	Ethical Considerations	96
4.3.2	Twitter Data Sources	97
4.3	2.1 Twitter Website	97
4.3	2.2 Twitter's Application Programming Interface (API)	97
4.3	2.3 Data Resellers	98
4.3	.2.4 US Library of Congress	99
4.3	2.5 Data source for this research: Twitter APIs	
4.3.3	Twitter Data Structure	100
4.3	.3.1 <i>Tweets</i>	100
4.3	.3.2 Users	106
4.3.4	Data Collection Software	108
4.3	4.1 Critique of Existing software	109
4.3	4.2 Software development for data collection	111
4.3	4.3 Emerging software concurrent to the research	114
4.4 Ç	Quantitative Methods: Twitter Data Analysis	116
4.4.1	Statistical Analysis	116
4.4.2	Network Analysis	117
4.4.3	Descriptive Analysis	118
4.5 Ç	Qualitative Methods	120

4.5.1	Interview Design
4.5.1	1.1 Thematisation
4.5.1	1.3 Interviews
4.5.1	1.4 Transcription
4.5.1	1.5 Analysis
4.5.1	1.6 Verification
4.5.1	1.7 <i>Reporting</i>
4.6 Pi	lot Study
4.6.1	Determining the research case
4.6.	1.1 Pilot Study: Hashtags & Keywords
4.6.1	1.2 Pilot Study: Interactions with identified users
4.6.1	1.3 Pilot Study: Results, Refinement and Findings
4.6.1	1.4 Selection of candidates for interview
4.6.	1.5 Caveats for the pilot study
4.7 C	onclusions
5. 2014	Irish Local Elections (#le14) Case Study 136
5.1 In	troduction
5.1.1	Research Design
5.2 D a	ata Collection
5.3 E	valuation
5.3.1	Manual Coding of Data141
5.3.2	Network Analytics144
5.3.3	Descriptive Analysis (DA)145
5.3.4	Framework145
5.4 R	esults & Analysis
5.4.1	Overall Network Structure
5.4.2	Network Metric Results148
5.4.3	Descriptive Metric Results
5.4.3 5.4 C a	Descriptive Metric Results
5.4.3 5.4 C a 5.5 C a	Descriptive Metric Results
5.4.3 5.4 C a 5.5 C a 5.6 C l	Descriptive Metric Results
5.4.3 5.4 Ca 5.5 Ca 5.6 Cl 5.6.1	Descriptive Metric Results

. 2015	Irish Marriage Referendum (#marref) Case Stud	1y 165
6.1 In	troduction	163
6.1.1	Research Design	164
6.1.2	Choice of hashtag	165
6.2	Data Collection	167
6.2.1	Comparison of data collection software used	
6.3 E	valuation	
631	Actor Coding	179
0.5.1		
6.3.2	Methods of analysis applied	173
6.3.	2.1 Network Analytics	173
6.3.	2.2 Descriptive Analytics	174
6.3.	2.3 Periods for this study	174
6.4 R	esults & Analysis	
641	Overall Network Structure	175
649	Analysis of the ten 50 contributors	177
0.4.2	Analysis of the top 50 contributors	1//
6.4.3	Mention Network Metrics	178
6.4.4	Descriptive Metrics	183
6.5	Caveats & Limitations	186
6.6 C	onclusions	188
6.6.1	Revisiting the research questions for this case study	
662	Opportunities for further study	191
Insigh in Iris 7.1 In	ts from interviews with highly active users h political discourse on Twitter htroduction	
7.2 P a	rticipant profiles and overviews	196
7.2.1	Participant NP1	198
7.2.2	Participants NP2a and NP2b	199
7.2.3	Participant NP3	200
7.2.4	Participant NP4	
7.2.5	Participant NP5	
7.2.6 7.9.7	Participant NP6	
798	Participant NP7	203 203
7.2.9	Participant LP1	
7.2.10	Participant LP2	
7.2.11	Participant LP3	
7.2.12	Participant LP4	205
7.2.13	Participant A1	
7.2.14	Participant A2	

7.2.15	Participant A3207
7.2.16	Participant PM1208
7.2.17	Participant PM2
7.2.18	Participant PM3
7.2.19	Participant OM1
7.2.20	Participant OM2
7.2.21	Participant C1
7.2.21	Participant C3 212
7.3 Un	derstanding the social context of Irish political
dis	scourse on Twitter
7.3.1	Habitus, Illusio & Doxa215
7.3.2	Capital in the field, and its accumulation and transfer
7.4 T	echnological agency investigated221
7.4 H	low hashtags are really being used224
7.4.1	Conversation 'hanging off' the hashtag
7.4.2	Factors influencing inclusion and exclusion225
7.4.	2.1 <i>Economic factors for curtailment</i>
7.5 Ir	nplications for research
7.5.1	Hashtags
7.5.2	Support for actor sub-categorisations
7.5.3	False positive "contributors"
7.6 Li	imitations of this qualitative research
7.6.1	Opportunities for future research
7.7 C	onclusions
8. Conc	235 235
8.1 S 1	ummary and Interpretation of Findings236
8.1.1	Results and conclusions of the case studies237
8.1.	1.1The #le14 and #marref case studies238
8.1.	1.2 Insights and conclusions based on interview responses
8.1.2	Interpretation
8.2 K e	ey Research Themes and Primary Hypothesis Addressed241
8.2.1	Addressing the networked structure of Twitter political discourse241
8.2.2	Understanding the social context of Twitter political discourse
8.2.3	Establishing the extant methodologies for Twitter discourse analysis242
8.2.4	Testing the effectiveness of collection and analytical methods
8.2.5	Informing methodology with qualitative investigation of actors

8.2.6	Developing the understanding of how hashtags are used in	
	political discourse	243
8.2.7	The Primary Hypothesis	244
8.3 M e	ethodological Imperatives for the Study of Political	
Dis	scourse on Twitter	244
8.3.1	Reporting	245
8.3.2	Digital Ethnographic Practices	245
8.3.4	Security of data	246
8.3.5	Actor coding	246
8.3.6	Specific Analytics	247
8.3.7	Periodic Analyses	247
8.4 C	ontributions to the Field	249
8.5 L	imitations of the Research	
8.6 P	ossibilities for Further Research	
D'1 1'	1 .	250
Dibliogi	rapny	
Append	lices	
A.3 A	ddenda to Chapter 4	275
A.3.1	Corpus of academic literature related to political discourse on	
	Twitter	275
A.4 A	ddenda to Chapter 4	
A.4.1	JSON Formatted Tweet Object Example	
A.4.2	Stars and Hearts – Implications of changing Twitter affordances	
A.4.3	Data collection software screenshots	
A.4	.3.1 Single archive default view in yourTwapperKeeper	
A.4	.3.2 Archive list view in yTK_DCU_v3	
A.4.4	Sample Participants' Guide	
A.4.5	Sample Consent Form	
A.4.6	Interview Questions	
A.4.7	Overview of Interviews Conducted	291
A.6 A	ddenda to Chapter 6	292
A.6.1	Example conversation with a user (via Twitter Direct Message,	
	04 February 2014), about the use of the #ssmref hashtag as an	
	alternative to #marref	

A.6.2	Tweets, tagged #marref, captured per minute by yTK_DCU_v3,
	during Referendum Count (23 May 2015)293
A.6.3	Table of the 50 most frequent contributors to #marref-tagged tweets, coded 294
A.6.4	Table of the 50 most frequently @-mentioned contributors in
	#marref-tagged tweets, coded
A.6.5	Table of the 50 hashtags most frequently co-occurring with
	#marref in the dataset

Chapter 1

Introduction

1. Introduction

The aim of this research is to explore how political discourse on Twitter is analysed, and to contribute to that analysis through investigation of existing disparate methodological approaches, and by the design of effective research strategies. A goal of the work, therefore, is to assemble a *digital methodology* for collecting and analysing data relevant to political discourse on Twitter. This is accomplished by collating appropriate methods from across disciplines, testing and combining these, and designing new approaches and tools based on identifying shortfalls and limitations in those currently available. These processes will be informed by qualitative investigation of the social context in which contributions to these discourses are made, and the work will address gaps in current understanding of how factors at the microlevel of Twitter interactions affect the data collected and analysed, and the scope of accuracy in resultant findings. Analysis of Twitter-mediated political discourse is an emergent field of academic enquiry across disciplines, and this thesis is presented in the context of computational social science, which emphasises the investigation of social and cultural phenomena using large sets of rich *digital trace data*¹ to uncover patterns of behaviour and interaction, especially in social networks (Lazer et al. 2009). This work will include case studies of Twitter discourses around major Irish political events in 2014 and 2015, and complementary qualitative data from highly active individuals within them - using these to present insight on the Irish context, and more focussed research imperatives for scholars using similar large sets of Twitter trace data to investigate political discourse elsewhere.

There are a number of reasons for this choice of research theme. The study of Twitter-based discourses, facilitated by access to enormous quantities of constituent 'big data', capable of being analysed by computational tools and methods, represents a new methodological frontier replete with intriguing opportunities and challenges for communication researchers. The field is at a very early stage of development, positioned at an interdisciplinary nexus of communications studies, political science, sociology and computer science, with theoretical approaches and methods from all of these being deployed. Studies from these various academic standpoints have already begun to demonstrate the rich potential of working with this data, but no sense of 'best practice' has emerged for the field. In this thesis, the researcher chose to engage with this topic because of the opportunities for contribution to this new field, the challenges of navigating the cross-disciplinary nature of the methodologies

¹ Digital trace data refers to that data, and especially metadata, which can be collected and interpreted to provide insight on the 'digital footprint' of a user, and on their activities and interactions in an online space; evidence of human and human-like activity that is logged and stored digitally (Howison, Wiggans and Crowston, 2011)

applied in it, and the potential to uncover insights particular to Irish political discourse and participation in it. Working with Twitter trace data and situating the behaviours that generate them in a networked social context of norms, conventions and interactions requires engagement both with computational approaches (programming, data storage and sorting, network analytics) and socio-cultural approaches for understanding technology-mediated interactions, political participation, and systems of power. The researcher, having a cross-disciplinary academic experience that maps well onto these skill requirements, chose to pursue the topic and goals outlined, with a view to contributing to the coalescence of methodologies in this new field, and the development of best practices for research. The choice of this subject, and the design of the research in response to it, also directly addresses recent calls for the derivation of *natively digital methodologies* in the computational social sciences that specifically respond to the challenges and opportunities of working with the big data of the social web (Bruns & Stieglitz 2012; Freelon 2014; Rogers 2009, 2010).

This thesis therefore conceptualises Twitter (and the specific political discourses taking place within it) as a network in which constituent interactions between users may be understood and analysed, individually and in aggregate, via their corresponding digital trace data. A thorough review of the interdisciplinary literature related to the topic is presented, in order to identify commonalties and interesting anomalies, and to focus on key characteristics of the extant research. Testing of appropriate methods for Twitter political discourse analysis follows, informed by qualitative data in a *Mixed Methods* research design. The goal of the thesis is to contribute to research design methodology in this emergent field, which is formed through disparate disciplinary approaches, and which currently lacks the consistency of defined frameworks of investigation, or well-understood parameters for research.

This chapter continues with an introduction to Twitter as a communications medium, briefly outlining some elements of its history and development, including relevant detail on the mechanisms and affordances within it that affect interactions and distribution of content. This information foregrounds more detailed discussion in later chapters of the technical detail of data related to tweets, and of the significance of this for analysis of the aggregate patterns of interaction in a given discourse. It is intended, therefore, to form an introductory primer for many of the core concepts of the work, and to begin to situate this study within the fledgling field of Twitter study and the analysis of discourses within it.

Some detail is provided in this chapter on the introduction, use, and value as an affordance of the *hashtag* - those topic or category markers included by users in their Twitter content. As will be made clear in this and subsequent chapters, hashtags are a central device of topic exploration, collaboration, and public-formation within Twitter discourse. Due additionally to their ability to act as anchors for data collection on a given topic, hashtags have become crucial to the methodological schemes for discourse analysis on Twitter. Hence some introductory detail of the centrality of hashtags to research, and the limitations they apply upon it, are provided here.

This chapter also introduces the theoretical setting of the thesis and sets out the research hypothesis of this work, indicating the thematic components of it, and their indicative sets of research questions, as well as the scope and levels of enquiry. This situating introduction concludes with information on the layout of the thesis and a guide to those chapters that follow.

1.1 Introduction to Twitter

The use of Twitter has become common in many nations, with its growth corresponding to the increased ubiquity of mobile devices in developed and developing countries, and with their associated modes of communication – particularly notions of immediacy and currency as important characteristics of daily mass communication, news gathering, crisis reaction, audience participation, etc. (Bruns & Burgess 2011; Bruns & Highfield 2013; D'heer & Verdegem 2014; Fuchs 2011; Maireder & Ausserhofer 2013).

Fig. 1.1Number of monthly active Twitter users worldwide2, by end of
quarter, from Q1 2010 to Q3 2015



Twitter's user base has continually grown world-wide since its introduction in 2006, and the service claims over 307 million active monthly users³ for the third quarter of 2015 (Statistica

 ² Sources: Twitter SEC-filing, 2013; Investor Relations annual reports, 2012, 2013, 2014; Q3 Report
 2015 (Twitter Inc. 2013; Twitter Investor Relations 2014, 2015)

³ It is worth noting that there is some vagueness to this term and that is frequently cited in the literature without contextual information on the specific activity required to include a user in this cohort. Twitter has continually published MAU figures in its quarterly and annual reports, but generally without an accompanying definition. The 2013 *Form 10-K* filing by Twitter to the US Securities and

2015; Twitter Investor Relations 2015) a number that has appeared to plateau over 2015, as growth comparatively slowed in the three quarterly statistics published (Fig. 1.1, above). Concurrent with the history of growth of the size of Twitter's user base, has been a steep growth within it of the proportion of users accessing the service using mobile devices; by mid-2013 this had reached 60% (Moore 2013), and it surpassed 80% in Q3 of 2015 (Twitter Investor Relations 2015).

1.1.1 Twitter usage in Ireland

Specific statistics for the use of Twitter in Ireland have not historically been published by the company, as they have for other larger nations (e.g. USA, UK), and since the inclusion of global positioning coordinates with tweets is disabled by default, and the 'location' metadata field is user-contributed and unreliable, any analysis of the proportion of all tweets sent that originated in Ireland would be subject to a likely very substantial error, which is difficult to define. Knowledge of the extent of use of the medium in the Republic of Ireland has therefore depended instead on representative national omnibus polling. The Ipsos MRBI Social Networking Quarterly report (e.g. Q2 2015 in Ipsos MRBI 2015) has published tracking of the percentage of the Irish adult (15+) population who have a Twitter account, and demonstrated a consistent rise since tracking began in 2010 to a level of 27% in mid-2013, at which it has since remained relatively stable (+/-2%) in subsequent quarters). It is important to note that this metric captures only the incident of a given person's registration of an account with Twitter, and not any indication of their level of use. Only in later tracking did Ipsos MRBI publish findings on the percentage of those users who use Twitter daily, amongst those who have an account. From 2013 onward, this has seen slight growth, from 35% to around 39%, where it has remained stable (+/- 1%) since 2014. At the end of Q2 2015, 28% of Irish adults were reported to have Twitter accounts, of which 39% made daily use (Ipsos MRBI 2015).

These metrics can be summarised to show that around 11% of Irish adults are using Twitter daily - a higher proportion than in most countries for which data is publically available; cf.

Exchange Commission does include a definition, however: "Twitter users who have logged in and access Twitter through our website, mobile website, desktop or mobile applications, SMS or registered third-party applications or websites in the 30-day period ending on the date of measurement" (Twitter Inc. 2013). Of these users, several million are "SMS fast follow" users, who opt to receive some updates from the service by SMS, but do not hold Twitter accounts and are not logged in. For the purposes of this research, these are not counted among reporting of Twitter users, following the example of metric archiving services (e.g. Statistica 2015)

8.2%⁴ in the United States, based on social media usage reporting by Pew Research in 2015⁵ (Duggan et al. 2015). These comparisons, as well as reports on the per-capita ownership of Twitter accounts by nation (e.g. Lipman 2014), have consistently placed Ireland in the top five to ten of countries, based on Twitter use.

1.1.2 Twitter's use for political discussion

Twitter's role in facilitating political discussion has also grown with it, and properties of its mode of *microblogging*, as well as the adopted conventions of its use, lend it a particularly central role as a medium for political discourse in contemporary societies (Bruns & Stieglitz 2012; Couldry 2013; Larsson & Moe 2011). In contrast to social networking sites (especially Facebook - the dominant member of this category), contributions (tweets) on Twitter are by default a public offering rather than one absolutely restricted to the discrete and specific audience of *followers* a user has⁶. This allows for the indexing, searching, discovery and tracking of commentary in tweets and hence the ability of individual users to initiate or participate in discussion and debate outside of the *followers/following* groups of others whom they have chosen as their immediate network. Contributions to the discourse are made to a public space accessible by any other user, or by any member of the public via the search features of the Twitter public timeline on its website7. An important affordance of Twitter's interfaces is in the presentation of hashtags as system hyperlinks (discussed in more detail in this chapter, and in chapter 4). Given that hashtags become links to all other recent public tweets that also include them, making them a primary point of breaking-out into a more public space within Twitter. This process of establishing topic-markers as points of departure from a closed follower/following network to a wider discourse on that topic is central to the formation of issue-specific discourse publics within the medium, therefore making Twitter hashtags an important focus for researchers (Bruns & Burgess 2011).

 $^{^4}$ 23% of adults, of which 36% report daily use. (Duggan et al. 2015 p. 4)

⁵ Note caveats that this survey was for the period of late 2014, though Irish usage was stable through that period, and that this survey defined adults as 18+, compared to 15+ in Ipsos MRBI. Pew found that among teens, social media usage was higher, though this was greatly skewed toward specific social platforms such as Facebook, Instagram and Snapchat. Nevertheless, the figure may be slightly higher for the US, when usage by those aged 15-17 is included.

⁶ Twitter's presentation of the tweets is important here, however, since while they are accessibly public, they are functionally served only in the 'timeline' of followers by default. Affordances allowing users to search for or display tweets outside their own timelines therefore become important to *breaking out* of a highly contained local discourse network.

⁷ The public search option on Twitter's website at https://twitter.com/search-home allows anyone with internet access to search the aggregated feed of all public tweets for a recent period (of approximately two weeks) without requiring membership of Twitter or login credentials.

1.1.3 History of Twitter, and development of its established affordances & conventions

Twitter began in early 2006 as a microblogging service, allowing its registered users to share short statuses with others. The length restriction of 140 characters that remains for tweets today arose in the initial design of the system to send this content via SMS, being later "renarrated as a feature" (Ellison & boyd 2013 p. 2) and incorporated into the brand concept of Twitter as a medium of concise, current communication⁸.

Twitter incorporated elements of other social networking sites, such as avatar images, profiles (as brief character-constrained biographies), and self-directed personal networks of other users to be *followed* – whose aggregated tweets appear in the user's timeline (*Twitter stream*) in reverse chronological order. This stream could be accessed initially via the Twitter website and later became the central feature of mobile and desktop clients, which diversified to allow users to split and manage the stream by topic or by self-created lists of users, and to track topics by keyword or tag from all users' public content.

1.1.3.1 Addressing and re-broadcasting content

Crucially, many of the methods of interaction and participation on Twitter today originate from organic user behaviours and conventions adopted during its early use. The syntax of @-replies (that of referring to a specific user by prefixing their username with an @ character, e.g. @stephenfry) derives from an older convention in Internet Relay Chat (IRC) where it allowed users to direct responses to an individual present among a chatroom group. Honeycutt and Herring (2009) discuss this development as a form of addressivity where a user indicates the specific intended target(s) of a comment within a publicly accessible space, inviting individual attention in a non-personal environment and prompting discrete conversational interaction within a many-to-many communication forum. Twitter adopted the @-reply into its core functionality, allowing it to become a system of notification to users that they are mentioned, and in doing so instituted the mechanism of seeking attention from and inviting conversation with individuals in the public space of the stream. Mentions of @tagged users in tweets in the public stream become hyperlinks to that user's profile and stream, and Twitter client applications prompt conversational interaction with a *reply* button, specifically establishing and promoting this mode of discussion in the space through these affordances.

⁸ See: Twitter's Creative Director, Biz Stone's doctrine that "creativity comes from constraint" (Zinko 2009).

Similarly, the practice and instituted method of *retweeting* was an emergent one, developed and officially adopted following user conventions after – and dependent on – the emergence of @-tagging. Initially, a convention developed in order that a tweet from one user could be re-broadcast to the followers and public of another; the original message was prefixed with "RT" (retweet) and the @-tagged username of the originator of the message (e.g. "RT @BarackObama Four more years"9). This initial convention was less uniformly adopted than that of @-replies, as it lacked the syntax precedent that the former had in IRC and because the attribution of the original author's username and the letters "RT" counted toward the 140 character limit imposed in re-posting the message. It was therefore the case that attribution was inconsistent and the "RT" was occasionally replaced with one quotationmark character (boyd et al. 2010). Twitter responded to this practice, and to the limitations imposed by necessarily including the originator's username in the subsequent retweet, by adopting a mechanism that allowed retweets as an in-built system function within the stream and client UI, and used meta-data to store the originator's details. This allowed the current format, wherein a tweet of any length (up to the 140 character limit) can be retweeted, with the full attribution of the original author included externally and so not counted toward the limit, but accessible in the stream and client interfaces.

1.1.3.1 Content-marking: Hashtags

The specific categorisation and tagging of tweets arose through the convention of *hashtags* – the combination of the *#* character with a single keyword identifying the topic to which the contribution belongs. Boyd et al. (2010) identify the development of the hashtag as topical categorization marker with the earlier practice of *social bookmarking* (Golder 2006) prevalent on blogs before the launch of Twitter, and possibly stemming from the conventions of many programming languages to set apart keywords and variables with unique characters, such as *#*. Following a widely supported call for their official institution led by Chris Messina one year after launch (2007a, 2007b), Twitter incorporated and established hashtagging of content by creating dynamic hyperlinks from any hashtags present in a tweet, providing a user with an aggregate stream of all other recent tweets that also include the given hashtag. This system - as well as the subsequent development of multiple split and tracked streams based on single or combinations of hashtags in Twitter clients - allowed for a highly topic-centric presentation of discussion among users. The mechanism of the hashtag, and its default function of pulling together all recent public contributions thus tagged, marks a

⁹ To date, the third most retweeted tweet of all time, at >740,000 system-tracked retweets. https://twitter.com/BarackObama/statuses/266031293945503744 (Accessed: December 2015)

profound 'breaking-out' from the user's immediate and self-selected user group and is an important difference in how Twitter handles topic-based searches compared to other dominant social networks, like Facebook, which tend toward in-group limitation of possible search results obtained.

1.1.3.2 Notes on the use of Hashtags

There is much discussion in the literature on the mechanism of the hashtag as a tool for the formation of emergent issue publics, centred on current or developing topics and transcending immediate user-selected networks to include wider public contributions. Bruns and Burgess's influential article (2011) on the formation of *ad hoc* publics establishes an initial methodology for the examination of such groups and offers examples of the formation, development and dissolution of publics responding to acute political events and crises. Techniques for the computational derivation of metrics and patterns useful in understanding hashtag-public participants are explored further in a later chapter of this thesis, and inform the methodology and critique thereof given in this work.

It should be noted that the intended mechanism of the hashtag – as topic or category indicator – has been diluted, subverted and evolved through use and convention, such that it is now as often used in satire, irony or for emphatic effect as it is to deliberately mark a given contribution as belonging to a particular topic. Bruns and Burgess note that "it is unlikely that significant, unified communities will exist around generic hashtags such as #Japan or #Australia, for example" (2011 p. 5); topics or categories which are so broad as to potentially contain many incidental discussions offer little to study the specific development of issue publics, outside of acute events which might occasionally overwhelm them. Likewise, a substantial variety of hashtags exist as emphatic markers, which may be useful only in supplying context in fields such as sentiment analysis (e.g. #rage, #fail, #happy) when they mimic the extra-textual intent of older markers like emoticons. Others have developed over time to become continuing tropes and memes among a Twitter community now literate to such conventions (e.g. #mysexlifeinmovietitles, #inappropriatefuneralsongs), or as standing arrangements to invite conversation (e.g. #edchatIE¹⁰), and these are tracked or explored for entertainment or discussion, rather than representing coherent issue-based publics. Others

¹⁰ An open discourse around Irish education, involving teachers, researchers and other education professionals

still exhibit a meta-function, describing modes or actions of users within the medium (e.g. #teamfollowback, #followback¹¹).

While none of these hashtag types offer direct insight to those studying public issue-based discourse, it is useful to establish the flexibility with which hashtags are used, and subverted, when studying groups unified by the use of issue-specific tags. We may conclude from these diverse usages that the inclusion of a hashtag in a tweet is likely to be the product of a complex set of understood social conventions for its particular use, and that the deliberative processes by a user choosing to include one are grounded in part in their ability to contextualise it from experience. This insight is an important motive for the inclusion in this study of depth analysis of active Twitter users.

¹¹ Used by Twitter users wishing to signal their reciprocity for following; i.e. if a user follows them, they will follow that user in return.

1.2 Extant research on Twitter

Despite occupying an emergent and evolving space in the study of media, Twitter (and the discourse taking place within it) is already the subject of a great deal of research in the literature. Such analysis represents an interdisciplinary endeavour, with contributions from Psychology, Computer Science and Mathematical Science particularly helping to inform its emergent methodologies in the social sciences. The ready accessibility to Twitter's large volumes of sortable 'big data' - rich, multidimensional content in the form of aggregated tweets - has interested scholars, who can use it to undertake large-scale analyses of online social interaction, in some cases with implications for a wider social context. Studies exist in areas as diverse as sentiment analysis, recommendation systems, election prediction, citizen journalism, or crisis reaction, with something of a common toolset of open-source collection and metadata-based analysis methodologies beginning to emerge. Such studies and methodologies prompt questions of the representative quality of communicative and organisational social interactions on Twitter to such interaction more widely in society. It is important neither to generalise from the specific demographic structure of Twitter users to the population at large, nor from any dataset collected by means of simple aggregation methods to all data comprising the totality of content on a topic; the latter caveat being less clearly expressed in much of the literature and a subject of enquiry in this research. Such caveats aside, the analyses of Twitter interactions still provide demonstrably valuable insight and detail for many fields specific to, and outside of, the medium itself.

1.2.1 The study of political discourse on Twitter

While a consistently increasing number of studies specifically address the role(s) of Twitter in politics, this research is fragmented and very often only partially or tangentially connected to available research on the same topic. A primary reason for this is that the research area is fairly new; a common cannon of study, identified as relevant by the community of researchers working on the topic and setting out a widely-agreed set of data-gathering, analytical, or explanatory methodologies, has not yet been established. Little effort has been devoted in the literature to the systematic identification of common empirical patterns in various studies addressing the topic of Twitter-mediated political discourse and the field therefore lacks even a reliable list of phenomena, events, behaviours, etc., that might be expected to be observable at a meta level in new studies undertaken. Additionally, and importantly, researchers from many different scientific fields are working on the role of Twitter in politics. Political scientists, communication scholars and computer scientists have published on this topic. These publications come from different theoretical perspectives, use

very different methods and – likely most relevant for the apparent lack of established practice in the field - are published in different journal categories. These differences make it difficult for researchers to productively engage with relevant work in their area of interest but stemming from other scientific backgrounds. This leads to a rich but loosely interconnected literature on the use of Twitter in politics, engagement with which may require searches through and methodological knowledge of unrelated academic disciplines.

This thesis includes, in its literature review processes, a large-scale review and analysis of the accessible corpus of academic publication on the topic of political discourse on Twitter (See: Chapter 3). This review is undertaken for several reasons: (1) to address the disciplinary disparity outlined above by drawing together the available studies across categories and outlining their commonalities and differences in methodological approach, (2) to investigate the theoretical framing for research questions and analyses, and (3) to establish the extent of reporting on methodology, and especially on data collection strategies.

It is clear from the literature that methods and conventions of the daily use of Twitter contribute to the medium's particular facilitation of highly-current public discourse on political issues; the ability to see and track public replies from other users to a tweet (@replies), the ability to retweet (RT) the contribution of another user as a mechanism of endorsement or to call attention and request comment, and the ability to categorise contribution as specific to a subject via the use of hashtag - all these offer significant advantages for Twitter as a network which prompts more open participatory discussion than other social media (Bruns & Stieglitz 2012). The fact that Twitter's dominant mode of access is from mobile devices, combined with the inherent capacity for the formation of *ad hoc* issue publics suggested by the abilities outlined, make it particularly well-suited to timely and reactive discussion of emergent issues. Indeed, the literature includes influential studies of the especially well-adapted roles it occupies in crisis-communication (Bruns & Burgess 2013), election tracking and prediction (Bermingham & Smeaton 2011; Larsson & Moe 2011; Tumasjan et al. 2010), and political unrest (Procter et al. 2013; Tonkin et al. 2012; Vis 2013), among many other examples of highly contemporary and fast-changing discussion by massive contributing groups of interested users. A unifying factor among these studies, despite their often-unrelated fields, is that the method of gathering sets of data for analysis, and to support conclusion, is via the hashtags used.

1.2.2 Hashtags in data collection

Aside from the advantages to researchers of using all tweets aggregated under a specific hashtag as a dataset in which the content has been deliberatively tagged as pertaining to a

certain topic by its contributors, the prevalence toward (and reliance on) hashtag based datasets is also significantly *technologically* determined. The tools and methods readily available to researchers for gathering such data tend to focus collection exclusively on selection processes based on identifying included text terms in tweets; here the hashtag, generally being communally adopted by an interested group, designed to be short and unique, and indicative of contributing users' judgement of the relevance of their contributions on a topic, is an obvious choice for researchers, and has come to dominate data collection strategy.

At present, Twitter's Application Programming Interface (API), discussed in some detail in Chapter 4, allows for the easy collation of recent tweet content by hashtag (Twitter 2015a). The API is used by all applications that access Twitter and the methods and limitations implied by it are universal and independent of the particular client or application used to gather data. The API in its most recent version (v 1.1, available since 2012) allows access to a greater amount of the meta-information accompanying individual tweets than did version 1.0 and can be used to draw result sets with more nuance than previous search requests (which were largely based on simple aggregate sets of recent tweets, by keyword, hashtag, or username). However, use of this more precise functionality in API v1.1 is not widespread, and tools used in data-gathering in the literature¹² do not draw on newer functionality to gather intricately derived sets of tweets, but rather focus on older large scale simple search approaches. (This includes the most common tool cited by researchers, *yourTwapperKeeper¹³*, designed in 2010 for API v1.0).

This shortfall has significant implications for research design; since the majority of social scientists do not have the specific programming skills to modify, update or replace the extant open-source tools, the particular abilities and limitations of these have come to define such data collection, and have consequences for the subsequent analysis available. The specific metadata fields recorded for every archived tweet (chosen from those available in v1.0 of Twitter's API, in the case of yTK) become the *de facto* standard variables upon which findings are based and comparative analyses are made by scholars. The historic lack of any other open-source tools offering greater functionality (based on more contemporary affordances of API v1.1) means that Twitter studies, while consistent in their collection method across the

¹² Where these are mentioned at all - the majority of studies fail to report their collection methodology; see the findings of the review on existing literature in chapter 3 for a discussion of this. ¹³ yourTwapperKeeper (yTK) is an open source tool based on the website functionally of the nowremoved www.twapperkeeper.com, which was previously used to aggregate Twitter content online, and provided users of its service with comprehensive archives of captured content in a variety of formats (O'Brien 2013). Greater analysis of yTK specifically, and of other available tools, their approaches, and the limitations of these, is offered in Chapter 4, Methodology.

social sciences and across geographic locales, operate on a necessarily smaller set of variables than are actually available, and gather data based on less nuanced filtration of Twitter content than is possible.

1.2.2 Implications for research

It is evident from the literature, as Chapter 3 will detail, that datasets, which have been used to draw conclusions on user participation in public discourse on Twitter, are taken almost exclusively from hashtag-based sets collected over defined periods and subject to the limitations outlined above. It is simply too difficult or time-consuming for researchers to collect other relevant tweets, outside those specifically hashtagged to a topic of interest. Data used is therefore subject to a defined and substantial selection bias; of all tweets that form the totality of Twitter discourse on a given topic, only those explicitly hashtagged as such will form part of the datasets subject to the vast majority of current analyses in the social sciences. The action of hashtag inclusion by users is also assumed to be a deliberative and rational behaviour, explicitly stemming from the user's wish to signal and contribute to a discourse topic, though factors modifying this behaviour are unexplored.

Furthermore, there exists a currently unresolved question as to the overall accuracy of inclusion in even this data, since it has historically been unclear whether particular query endpoints of the Twitter API furnish a complete set of all public tweets for a given hashtag, and in the case of applications tracking and collating tweets in real time, it is difficult to ensure uninterrupted collection over long periods since any network or server outage will result in a data gap.

"Indeed, it is true that unless the Twitter API can be trusted to deliver all matching tweets without disruption, no study of Twitter that uses these processes can possibly achieve 100% accuracy. Further, as the API is the only access point to large-scale Twitter data which is available to researchers outside of Twitter itself, there is no opportunity to independently verify the quality of the data set." (Bruns & Stieglitz 2012 p. 179)

Grappling with Twitter's API and creating custom applications or scripts capable of pivoting out from a hashtag set to include untagged replies to tweets within it – the conversation prompted by or related to the content tagged and included in the set – is possible but difficult, generally requiring a level of programming expertise largely unavailable among Social Science researchers, and is more likely to be seen in Twitter studies within the field of Computer Science (e.g. Mohammad & Kiritchenko 2014). It is clear from acknowledged examples in the literature, as well as from examination of tweets collected in hashtag-based aggregations, that conversation directly related to a topic of interest is occurring around the tagged tweets, but is not being captured and included in analysis. Relevant contributions from users forming part of a thread of conversation may occur either before or after a single tweet from that discussion thread has been captured by virtue of being tagged. Large hashtag datasets are replete with examples of tweets which are themselves @-replies, and thus form part of a hitherto untagged conversation thread, marking the point where a user includes a tag to signal the relevance of the on-going conversation to a topic; the previous contributions to the conversation will not be included because they were not also tagged.

Similarly, a tagged (and therefore included) tweet may prompt one or more follow-up discussions in which users do not tag every subsequent contribution – indeed the reasons why it is unlikely that they would do so are explored in a later chapter on the performative elements of tweeting behaviour – and so all follow-up discourse, reaction, or analysis, would be absent from the collection. Bruns and Burgess's study on the formation of issue publics on Twitter acknowledges the absence of such content contributions, which "*hang off* the hashtag" (Bruns & Burgess 2011 p. 6), noting that: "not all such responses will themselves include the original hashtag, and lengthy conversations between two users who found each other through their shared use of a hashtag may follow on from that discovery but take place entirely outside of the hashtag stream itself" and thus not be included in any analysis of Twitter discussion on a topic, despite very clearly belonging to that discussion.

Notions of how representative the hashtagged set is of all discussion related to or 'hanging off' it must be in question, as the relationship between tagged and untagged contributions on a topic hinges on the behaviour of the user public involved; the degree to which users deliberatively include tags and understand a coherent shared grammar of their use is difficult to ascertain and may be wildly different in differing topic areas (e.g. conversation related to the 2014 Irish local elections, some of which will be tagged #LE14, versus conversation related to the television show *House of Cards*, some of which will be tagged #HouseOfCards; these distinct publics may have different priorities and conventions for the use of tags and understanding these is required to understand the degree of bias between the hashtagged sample and the whole for a given public). It is clear then, that investigations to begin the process of more clearly defining how hashtags are used with regard to certain topics, or within certain cohorts, are worthwhile contributions to the field.

Bruns' (2012a) article on mapping dynamic conversation networks within Twitter begins the process of building an accessible methodology for scholars to visualize and analyse the degree

to which a hashtag-derived data set contains @-replies, in an effort to account for and understand these snapshots of missing conversational threads. Bruns calls for further research to move toward greater sophistication in dealing with Twitter data and cites Rogers's (2010) methodological suggestion of *following the medium* - that Social Sciences research in newer and quickly-evolving media, like Twitter, ought to proceed with "emphasis on natively digital methods" (Rogers 2009 p. 5), responding to organically emerging opportunities for data collection and explanatory analysis.

Finally, it's worth noting that, much as the early convention of manually retweeting another user's contribution with the "RT" prefix was hampered by that syntax itself requiring some of the limited 140 characters available, the inclusion of any hashtag(s) in a tweet necessarily shortens the text space available for the content being tagged. It is therefore the case that there must be a subset of tweets (unmeasured in studies in the literature) that explicitly belong to a topic and might otherwise have been hashtagged as such, but have not been because of the constraint of the character limit. The degree to which this may be the case is a direct function of the priorities and behaviours of the given issue public generating them. Their personal or shared priority to include or exclude a hashtag at the expense of shortening a message, their notion of whether or not content is implicitly understood by others to belong to a topic when untagged, and their performance of tagging as an explicit cue to share some content beyond their own followers are all relevant factors. Social study of these factors among users who are actors in the discourse on Twitter is required to contribute to the understanding of what types of content are absent from the hashtag-based datasets.

This thesis seeks to respond to calls for research by establishing, through qualitative means, an understanding of the behaviours and conventions of actors and their interactions at a micro level, *for the specific case of political discourse in an Irish context*. Such findings can be used to directly address the question of representative validity in datasets for this context, but are also useful to support and advance theoretical frameworks to explain how political discourse is mediated on Twitter, and to develop methodologies in this emergent field which may better account for qualitative aspects.

1.3 Theorising political discourse online and on Twitter

Positioning Twitter-mediated political discourse with regard to social scientific understanding of communications and media requires a tracing through of the broad theoretical approaches of the interaction of technology and society. Our frameworks for the analysis of internet-mediated discourse in general are informed by concepts such as *technological determinism, social shaping* and *adoption of technologies*, and theories of their interplay with society at levels from the macro (cultural) to the micro (personal). These concepts and theories offer useful explanatory potential in the analysis of the genesis, nature, and limitations of public discourse on Twitter specifically (and mediated via the Internet more generally) and are explored in this thesis.

Whilst the reporting of explicit theoretical frameworks for studies of political discourse on Twitter in the literature is scant, it is useful to the research design of this thesis to explore and implement a model for how social interactions take place in the mediated social space of Twitter, as a guide for qualitative research, as an explanatory framework for analysis of discourse, and as a source of expected observations, which can be verified in the data.

In the early stages of this research, it was clear that a technological determinist stance might be applicable, since the innovations particular to Twitter's adoption and convention are rooted in technological progress, the ubiquity of mobile internet-connected devices, etc. The diffusion, and especially the specific appropriations and adoptions, of technology cannot be explained by a determinist stance in which technology is autonomous and exogenous, however, and the literature on the relationship of technology and society makes clear this critique.

The next chapter of this work goes on to examine the theoretical frameworks of the technology-society relationship and discusses the *social shaping* and *social construction* of technology, which posit a complex and multi-dimensional set of social, cultural, economic, political and personal factors influencing the adoption of technologies and the societal changes, which may follow. These approaches too are subject to criticism that focussing so deeply on the social context of a technology's adoption, usage and success, may deny its agency. *Actor-Network Theory* is subsequently explored as a 'third way' theoretical framing in which the role of such social factors is considered alongside the agency of the technology and of non-human constituents of its holistic conception as an *actor-network*. This model provides for lines of enquiry in the research as to the extent of non-human agency in Twitter, via its

affordances, hardware interfacing, software, etc., without isolating or fetishising technology. It also prompts consideration of social situatedness and emphasises the need for analysis at the micro level to build a richer understanding of agency and the domains affecting it at the nodal level of the network being studied. For this reason, it was deemed an appropriate model for framing the research design, and directly influenced qualitative enquiry and the emphasis on network structure in the quantitative methodologies explored.

In conceptualising the specific social settings in which constituent interactions in a Twitter discourse are made, the thesis employs the work of Pierre Bourdieu (1977, 1990, 2003), and specifically the concept of *field* to view the mediated social space of Twitter. The related concepts of *habitus, doxa*, and *capital* (discussed in detail in the chapter following) are thus contextualised for this field and together offer a paradigm for understanding both the interactions between actors in the field of Twitter political discourse, and the interaction between fields as struggles for power or attempts to transfer position and capital to and from settings outside of Twitter.

The conceptualisation of Twitter as a *Bourdieusian field* is a useful framing for the structure of qualitative investigations of the actors within it, as it provides a paradigm for how social interactions occur. Specifically, field theory suggests the necessity of examining how understanding of the 'rules of the game' (*illusio*) for the particular mediated space of interaction have come about, and to what end the agents within contribute and interact. These concepts, particularly the notion of the mapping onto Twitter - and reflexive examination - of pre-existing norms and expectations for social status and interaction (habitus), give rise to a set of interview questions for actors identified as highly active in the studied space. Bourdieu's framework therefore provides a foundation for exploring qualitatively the social interactions underlying patterns of communication identifiable by computational (quantitative) means, and is critical to the mixed methods imperative of this study to inform such 'big data' analyses with explanatory insight from archetypal actors involved.

Bourdieu's applicability extends beyond the qualitative component of the work, however, as the notion of Twitter as a distinct field of social practice is also supported by computational analyses, which demonstrate identifiable patterns of interaction between agents with common agendas (e.g. activists, journalists). Expectations set up in the conceptualisation of Twitter as a distinct field are borne out in detected patterns of communication between agents (Chapters 5, 6) and explained by subsequent interview-based investigation (Chapter 7). Thus, Bourdieu's wider views on the necessity to be holistic and methodical in the practice of social science, and to make use of complementary data sources in the

19

investigation of social phenomena (Swartz 1997), also inform the mixed methods research design of this work.

1.3.1 A network-based model

In keeping with the notion of an actor-network from ANT, and following the trend in the literature to conceptualise and analyse Twitter discourse as issue-based networks (having users as *nodes* and connections between them as *edges*) this thesis adopts a similar network-based model in examining the structure and dynamics of political discourses on Twitter.

The networks in question are those outlined in Bruns and Burgess's (2011) work on the formation of *ad hoc* issue publics, constituted of users who coalesce around a particular topic hashtag, using it to mark their own contributions and to find those of others whom they may not be linked to as a follower. These networks are dynamic, changing with time as a discourse develops, and specific methodological considerations are made for this, providing analysis of the developing patterns of discourse prior to, proximate to, and post political events. Attempts are made to identify and code the categories of actors participating in these networks and to investigate how these actor cohorts interact with one another and over time.

Notions from Actor Network Theory are important to this conception of Twitter discourse, as they emphasise the co-constitutive nature and associated agency of non-human elements in the network. The technologies that afford Twitter-based communication also act in and upon the network; rather than being 'dumb tools' of human intent and interaction, they (sometimes separately) shape and limit that interaction and must be considered when analysing the network. Technological agency is therefore examined at various points in this research design; hashtags are considered both as passive markers of topic and as active creators/enablers of conversant publics, the role of *bots* and aggregators is noted and used to screen data for irrelevant spam contributions, as well as to investigate behaviour of actors associated with them. In using a network-based model to analyse the field of Irish political discourse on Twitter, this thesis draws on the Actor Network Theory's reconciliation of technological determinism and social shaping in recognising that a network of interaction such as the studied Twitter publics is shaped and defined both by the action and affordances of technology and by the social structures and systems within it.

1.4 Primary Hypothesis and Research Themes

Having outlined the contemporary state of research into issue-based discourse on Twitter (which is a well-established interdisciplinary field, typically highly quantitative in nature, and of which political discourse is a key subset), and having summarised a theoretical framework from which to construct the research, a primary hypothesis for this work can be formulated.

This introduction has outlined how technological factors specific to Twitter as a medium have brought about social norms for its use, and vice versa, in an interplay between actors and network which must inform any analysis of the two. In responding to the evolving state of Twitter as a medium of societal discourse, researchers must test the biasing effect of the research methods employed, and the assumptions made of the actors involved. This thesis will explore these themes in the specific context of Irish political discourse, which may be shown to possess the properties of Twitter discussion common to other studies (modes of organising around hashtags to form issue-based publics, actors displaying archetypal behaviours in interaction, existing scholarly research that uses current methodologies for data collection, visualisation and analysis, etc.).

1.4.1 Primary Hypothesis

Political discourse on Twitter can be understood as mediated interactions between and within dynamic networked publics, analysis of which may be derived from combined computational methods, and meaningfully informed by contextual qualitative knowledge of the actors involved.

This hypothesis is systematically explored in the research presented here, through engagement with a set of research themes, from which follow questions that may be used to incrementally support it. The next section outlines these themes, relating them to the information introduced in this chapter, and foregrounding their development in the various chapters of this work.

1.4.2 Research Themes

This thesis explores the ways in which the primary hypothesis may be supported, by thematically addressing the foundations on which it is posited, through literature review and original investigation. The research themes are:

- (1) Addressing the networked structure of Twitter political discourse
- (2) Understanding the social context of Twitter political discourse

- (3) Establishing the extant methodologies for Twitter discourse analysis
- (4) Testing the effectiveness of collection and analytical methods
- (5) Informing methodology with qualitative investigation of actors
- (6) Developing the understanding of how hashtags are used in political discourse

These broad themes are necessarily interrelated, and it is difficult to separate some of these in the body of the research. However stating them thus offers a clarification of the routes of enquiry undertaken, which are both methodologically diverse, and stem from an interdisciplinary approach. The following summaries describe how these thematic research imperatives are operationalised in the work, setting out the specific research sub-questions implied by each.

1.4.2.1 Addressing the networked structure of Twitter political discourse

The primary hypothesis for this work posits that participants in political discourse on Twitter, or in discourses mediated by social media more generally, are structurally arranged in a network of interactions. This may seem self-evident from the common description of Twitter as 'social network', but it is necessary to establish the particulars of these structural arrangements, which may be multiple, ephemeral, and which can only be partially captured by the affordances of Twitter's API and the methods employed in data collection and analysis. This thesis views Twitter as an actor-network, in the sense outlined by Latour (2007) and others, and considerations are made of the role in the network of non-human agency (e.g. mediation by software used, communicative implications of applications or devices, etc.). Issue-centred *networked publics*, communicating on a particular topic, with dynamics in overall and internal cohort size and volume of interaction, are the model used in this work, and this conception is established in the literature and tested in case studies. Questions that follow from this theme are: *Who are the participants in specific political discourses on Twitter and how can they be functionally categorised and structured with regard to one another? What are the bounds of a given discourse network and to what extent is it captured in research? How do hashtag-based discourse networks overlap?*

1.4.2.2 Understanding the social context of Twitter political discourse

The theme of investigating social context for Twitter interactions around political discourse is central to this thesis, since it is an interrogation of the cultural or socio-cultural *situatedness* of the discourse under analysis. In addressing it, it is necessary to engage with established theories of the nature of the relationship between (media) technology and society generally, frameworks for understanding communicative social interaction, and the deployment or adaptation of these for the case of online social media, and the specific discussion of politics

on Twitter. This work draws on the theories of Bourdieu particularly, whose concepts of *field* and *habitus* offer a useful staging for examining both the interactions in the discourse as struggles for power and position, and the social contextualising factors that influence and limit those interactions. Thus, questions that emerge from this theme and are addressed in the thesis include: *How are social structures from related fields (broadcast and print media, professional politics) reflected in the field of Twitter? How is status (social capital) and prestige (symbolic capital) accumulated in Twitter discourses? How is this capital transformed or transferred between fields? How are interactions on Twitter affected by knowledge of the medium, its conventions and affordances?*

1.4.2.3 Establishing the extant methodologies for Twitter discourse analysis

In situating Twitter discourse and its study with regard to the literature, this work critically examines the theoretical framings, data collection strategies and modes of analysis in evidence in that corpus of literature pertaining to political discourse in Twitter. Of particular concern here is the seeking out of effective methods, capable of delivering useful descriptive metrics for users and their networks, especially where these have comparative value (i.e. are not overly onerous to undertake and result in metrics that may be meaningfully compared across studies, and especially across topics and geopolitical locales). Questions arising from this theme are: *From what disciplines are studies of Twitter political discourse emerging? How are methodologies reported in the literature? What are the dominant methodologies for data collection, and analysis? To what extent are theoretical frameworks explored for these analyses and how do these relate to the methods used? How are issues around the ethics of big data research explored?*

1.4.2.4 Testing the effectiveness of collection and analytical methods

Following on from preceding themes, this thesis seeks to undertake large-scale case studies using specific methods identified as appropriate and chosen from the literature - in some cases adapted, redeveloped, or combined. This theme implies pragmatic original research, in which issues identified in this chapter, and related to data selection in particular, are engaged with in an experimental attempt to build a useful toolset. This thematic research responds to the following questions: *How do chosen data collection tools affect the structure of the Twitter data they collate and store? Can tools be designed to maximise the available data-components for a given tweet object? How can the coding of actors in a studied discourse be most effectively accomplished? Which resultant metrics from analytical methods available offer most comparative value? From which disciplines can methods be combined?*

1.4.2.5 Informing methodology with qualitative investigation of actors

A sample of participants identified in the pilot study process, whose archived Twitter contributions forms the basis of macro- and meso-level political discourse are interviewed to establish a depth of qualitative information on the behaviours, motivations and internalised conventions (*habitus*) with which they operate. These interviews further contextualise the societal, group and personal factors underpinning their modes of engagement with Twitter and directly set out evidence for the validity of assumptions about how users operate. This theme is a contextualising one, as it necessarily uncovers information on contemporary Twitter discourse of particular political issues in Ireland. Questions posed are: *How have specific actors come to use Twitter for political discourse? What contextualising knowledge have they transferred from other fields, and from personal experience? To what extent is the co-agency of technology identifiable in actors' descriptions of their participation? How do actors use the affordances of the Twitter client or website in their participation?*

1.4.2.6 Developing the understanding of how hashtags are used in political discourse

A research theme that follows from those previous, and is central to the primary hypothesis of this work, must be the development of knowledge on the use of hashtags in political discourse on Twitter. Both the structural nature of the networked issue-based publics on Twitter generally, and the methods for gathering data derived from these for analysis, rely on the hashtag: issue publics use this affordance to deliberately mark their contributions as related to a topic, and to find other conversation on a topic outside of their follower networks (Bruns & Burgess 2011); scholars select tweet objects for collection and subsequent analysis based on the presence of the hashtag in the tweet and the associated assumption that this specifically marks it as pertinent (E.g. Messina 2007b; Boynton et al. 2014; D'heer & Verdegem 2014; Mohammad & Kiritchenko 2014 - See also: Chapter 3). Thus, a useful contribution of this work is to address the gaps outlined in this chapter in knowledge around how hashtags may be used, and whether selection upon them biases data collection toward representation of certain actors, or segments of conversational interaction. Questions posed include: How do actors deploy hashtags and what factors modify or limit their use? Do the behaviours associated with hashtag inclusion vary with actor type? Can qualitative insight, derived from interviewing of key actors, provide explanatory value for patterns apparent in quantitative hashtag-based data collections?
1.4.3 Prioritised Research Questions

While detailed separation of the research themes may offer a framing of the intersecting sets of questions underpinning the research, and deriving from the Primary Hypothesis, it is also helpful to briefly state the summarising research questions at work in this thesis, in order of their priority:

- Can a methodological framework be devised for the study of political discourse on Twitter, which combines and extends existing collection and analysis methods and results in useful explanatory and comparable metrics?
- 2) What new information on the extent of the biases and limitations of collection and analysis methods can be uncovered?
- 3) Specifically, how do user norms and conventions for the inclusion of hashtags shape the data resultant from selecting upon them for data collection?
- 4) In the testing of the methodological toolset devised, what results and insights for the particular case of Irish political discourses may be uncovered?

Thus, the priority for this research is to contribute to the field through the development of demonstrably useful methodologies currently absent in the literature, and to address the limitations of existing studies, in which poorly defined caveats and biases exist. In doing so, the methodological tools explored will provide new data for the Irish case, which may be directly compared with existing studies and used as a basis for further research.

1.4.4 Scope

Conducting research, especially mixed methods research, with a goal to improve methodological toolsets in an emergent field necessarily constitutes a multi-faceted and complex approach to myriad research opportunities. The process of defining the goal of this thesis - as must be the case with any doctoral study - involved iterative processes of focused inclusion and exclusion of potential pathways of investigation. As outlined in this section, the study aims to provide useful contributions to the field in the form of carefully reviewed and selected methods, used in combination to produce meaningful, reproducible, descriptive and analytic metrics, informed by qualitative investigations of the social factors involved, for the specific case of political discourse on Twitter in Ireland.

While the findings on effective research design strategies, derived in this work, are very likely to be directly applicable to Twitter political discourse analysis in other locales, and potentially to analysis of discourse on other subjects, the findings on the social context and conventions of actors involved are necessarily narrowly focused on Ireland, and even there cannot be fully representative of the actors captured in the large case studies. It is beyond the scope of this work to undertake qualitative investigation at a larger scale; this branch of the research being novel and almost wholly unexplored in the literature, will represent only the beginning of a process of qualitative situation of these discourse analyses.

During the course of this research, two closely overlapping areas of enquiry were identified, which could provide additional situatedness for this study and imply routes of further research, but which are outside the scope of this work. These are qualitative investigation of the nature of *trust and reputation* in this space, and its effect on social interactions within Twitter, as well as investigation of the *performative aspects of self-presentation* in the medium. While both are interesting and potentially rich avenues of investigation, elements of which were uncovered in the interviews for this thesis, both require situation in a separate research literatures, differently directed interview design, and methods for deep content analysis that were deemed to be outside the remit of this thesis.

Finally, it is important to note that this thesis does not discuss Twitter text - the micro-level text content of constituent tweets - in the captured discourses, but rather is focussed on the collection and analysis of the trace data related to these, and the aggregate patterns of influence and interaction between and amongst groups that can be illustrated by it. The applications of content analysis methods to digitally mediated social spaces is a separate academic sub-discipline to that of the social computing setting for this thesis; one longer established and with many publications devoted to its study (e.g. McMillan 2000; Weare & Lin 2000).

The concluding chapter of this work offers a greater discussion of the particular opportunities for further research identified in the course of this thesis.

1.4.5 Levels of enquiry

This research is broadly structured as a tripartite set of enquiries, exploring the complex networked publics of specific political discourses on Twitter at three levels of analysis and evaluation (1) the individual social context of particular actors, whose *habitus* grounds and limits their discrete interactions in the mediated space, (2) the broader domain of particular discourses in which actors interact, and of which aggregate analyses may be made and insights derived on practices among cohorts of users, up to (3) the collection of discourses on Twitter comprising the political generally, which may be analysed as a *Bourdieusian field*, and in which concepts of *doxa*, *illusio* and *capital* may be applied in an explanatory design model.

1.5 Methodology

This study, in addressing both quantitative 'big data' analytics for social media discourses, and depth-focussed research at the level of individual actors, takes a *Mixed Methods* approach to the combination of data collected and evaluated. This methodological perspective is essentially *post-positivist* in nature, concerned with positivist *and* phenomenological processes of enquiry for both selected social discourses online, and their individual, political, cultural, temporal and technological contexts. While positivist approaches are evident in some Twitter studies in the literature, and seek to apply associations of 'scientific method' and goals of strict objectivity in their quantitative analyses, this methodological stance has limitations in the study of social interactions, in attesting to a singular knowable world. A post-positivist framework, in contrast, can take account of a multiplicity of individual subjective realities, and concerns itself with issues of the *situatedness* of knowledge within a holistic worldview with which the researcher is inherently connected and interdependent.

Post-positivism also moves the methods of data collection and analysis from a system of objective verification and proof of hypotheses, to one with the scope for inductive theory construction based on observations and on the identification of themes within rich qualitative findings. The mixed methods approach of this thesis is outlined in more detail in chapter 4, and is chosen for its *pragmatic* worldview in facilitating the complementary use of quantitative and qualitative methods, deductive and inductive reasoning, toward the goals of understanding both the network structure of Twitter political discourses more generally and the socio-cultural factors at the meso- and micro- level which led to their creation.

1.5.1 Quantitative Methods

A variety of quantitative approaches are investigated in this study, ranging from the computational derivation of simple descriptive metrics for studies datasets, to more complex *Network Analytics* methods borrowed from other disciplines and carried out on data that has been coded.

A focus of this study, as outlined in earlier sections of this chapter, is the development of data collection techniques to respond to and improve upon the dated but dominant tools in use in the research literature. Some detailed attention, therefore, is paid to the structure and sources of Twitter data available, analysing the viable and relevant variables in its constituent metadata and developing new approaches to collect this. The iterative development of a tool

to greatly increase the efficiency of coding users in very large datasets is an outcome and useful contribution to the field.

Quantitative methods for the analysis of discourse networks are tested in this thesis via two large-scale case studies on discussion of national political issues (Chapters 5, 6), in which the datasets were collated using the new software methods described, and the analytics result in metrics which may be usefully compared with prominent studies in the literature.

1.5.2 Qualitative Methods

Methodological techniques for interviewing, following the seven-staged approach of Kvale (2007), were employed in this research to investigate themes (especially themes 2 and 5) identified in the research design. These interviews provide insight on the micro-level of specific actors' social contexts (habitus), as well as at a macro-level in addressing themes across actors around the use of hashtags and the implications of selecting on them for Irish political discourse analysis. Meso-level interrogations of specific patterns uncovered in communication between actor cohorts were also made here, with interview participants offering relevant explanatory insight. This phased approach, in which qualitative data is collected subsequent to, and related to, a quantitative study, is a feature of the mixed methods approach taken overall.

1.5.3 Mixed Methods

This thesis makes use of the *Explanatory Design* model for mixed methods research outlined by Creswell and Plano Clark (2011), which is intended as a framework in which "qualitative data helps to *explain or build upon* initial quantitative results" (ibid., p.71, emphasis added).

Figure 1.2 The Mixed Methods Explanatory Design¹⁴



The general structure of this research design is a two-phased approach (Fig. 1.2) beginning with the collection and analysis of quantitative data, followed by the subsequent collection and analysis of related and complementary qualitative data, as described in the preceding sections. The overall interpretation of the data takes place in a subsequent concluding phase and typically for this approach, greater emphasis is placed on the quantitative phase of the

¹⁴ Adapted from the Explanatory Design models given in *Designing and Conducting Mixed Methods Research* (Fig 4.3a in Creswell & Plano Clark 2011 p. 73)

study. Creswell and Plano Clark specifically cite research designs such as this one as those to which this model is most suited, highlighting its efficacy in providing a staged approach, suitable to a single researcher, in which qualitative data is a source of situating and contextualising insight for quantitative methods (ibid. p.73-74).

The authors also describe an adaptation of the Explanatory Design model for the specific sub-process of using quantitative data to select participants for qualitative inquiry, such as interviewing. This process being crucial to the thesis, the *Explanatory Design: Participant Selection Model* (Creswell & Plano Clark 2011 p. 74) is also used in the research design of this work, and is discussed in greater detail in chapter 4.

1.5.4 Case Studies

Central to this research are the large-scale case studies of Irish political discourse, in which methods for data collection and analysis are demonstrated, and from which insights on the dynamics of participation and influence within Twitter discourse of major Irish political events are presented. These case studies develop from a pilot study, outlined in the methodology of this thesis (chapter 4), and are each detailed in the separate chapters constituting the major quantitative and methodological investigation of this thesis (chapter 5 and 6).

The first case study examines the Twitter discourse around the Irish local elections of 2014, collected from the *#le14* hashtag, and is significant for its comparative value with an existing influential study in the literature - D'heer and Verdegem's (2014) analysis of Twitter discussion of the 2012 local elections in Flanders, Belgium. In addition to providing a useful comparison of the patterns of interaction and influence among actor groups between the two political locales, this case study also advances the methodologies of its smaller Belgian counterpart and derives effective new coding schemata for actors and period specific staged analysis of the discourse dynamics.

Following from the 2014 study, a second larger dataset is gathered on the 2015 Irish marriage referendum #marref, and is used to demonstrate the consistency of effectiveness in methods from the first, the continued development of per-period analyses, and the inductive utility of such datasets for prompting investigations of social contextual events external to Twitter. This case study also broadens the overall scope of deployment of these methods, in offering an analysis of a discourse event centred on a *binary issue*, rather than the multiplicity of political issues within the preceding local election discussion, and within much of the literature.

1.6 Layout of the thesis

This thesis consists of eight chapters in 5 sections (Fig. 1.3)

Figure 1.3 Thesis Outline



1.6.1 Thesis Guide

Having introduced the research area, foregrounded the approach and stated the primary hypothesis and major themes in this chapter, this document will next examine existing literature to situate the theoretical framework of the approach, and the relevant situation of thesis with regard to existing work. This introduction is followed by a chapter outlining the methodologies and research design (chapter 4), two chapters in which computational methodologies for analysis of Twitter discourse are applied to large-scale datasets on case study topics from Ireland, (chapters 5 & 6), a chapter on the qualitative findings of this work (chapter 7) and a concluding chapter (8).

A more granular outlining of the specific inclusions of each chapter is presented to conclude this introduction.

1.6.1.1 Chapter Outlines

Chapter 2 opens with a discussion of the theories of technology and society that focus on their interrelationship. The chapter outlines the concept of technological determinism, citing the work of McLuhan (1964) and describing its inherent causal directionality, before moving on to the development of its themes in the work of more recent theorists such as Castells (2000, 2006, 2009) and Shirky (2008), who have situated it for new media technologies such as social networks. A counterpoint is then offered in the discussion of the theories of social shaping of technology and social construction of technology, beginning with Williams' (1974) critique of McLuhan's work and developing through influential examples in the works of MacKenzie and Wajcman (1985, 1999) on the social contexts of technological introduction, adoption, and success. The chapter then introduces the work of Latour (1996, 2007) on Actor-Network Theory, as a reconciliation of the 'hard' directionality in some aspects of the previous approaches, and discusses the fit of this theory for the particular case of Twitter. Finally, the work of Pierre Bourdieu is introduced, briefly discussing his imperatives for empirical work in sociology, and then exploring his field theory of social interaction in some detail, applying its various concepts to the case of Twitter. Discussion here is also informed by the work of Papacharissi and Easton (2013), who have attempted to deploy Bourdieu's concept of *habitus* in particular, as a framework for understanding social convention and adaptation in new media platforms like Twitter and Facebook.

Chapter 3 moves on to a structured literature review of a corpus of academic publications on the topic of political discourse on Twitter, outlining the method of their collation and then sorting and discussing these across a number of categories, based on their review. This chapter addresses the disparate nature of publications on the topic, stemming as they do from separate disciplines, and draws conclusions on the commonalities among them, as well as on the implications for further research, which this thesis responds to in part.

Chapter 4 is a lengthy exploration of methodology, functioning both as an outline of the research design of subsequent components of this thesis, and a required primer on the data structure, collection methods, and modes of analysis available for Twitter. The chapter attempts jointly to set out and justify the mixed methods approach of this work as a whole and to equip the reader with an understanding of the methodologies deployed in Twitter analysis in the literature, as well as its specific development in this work. The former consists of a discussion of the philosophical standpoint of the research, the implementation of a phased pilot study process and the staged design of interviewing. The latter consists of a detailed overview of the rich data affordances of the Twitter API, a summary of methodologies for its use in data-collection, critiques thereof, and a foregrounding of the specific approaches taken in the major case studies that follow.

Chapter 5 centres on the dataset of tweets gathered on the Irish local elections of 2014 using the established hashtag for that political event, *#le14*. The chapter provides descriptive methods for the set, details the development of the system of actor coding introduced in chapter 4, its specific application in this case, and the results of various Network Analytics approaches in examining the resultant data. Comparison is made in the findings with two other studies of a similar nature, widely cited in the existing literature, and discussion is presented on the development of those methodologies and the effectiveness of the choices made.

Chapter 6, similarly to the preceding chapter, focuses on a hashtag-derived dataset - in this case #marref, the dominant tag for discussion of the 2015 Irish marriage referendum. Similar actor coding processes and network and descriptive analytics are deployed for this larger dataset, with additional findings presented on the effectiveness of the updated collection methods in use.

Chapter 7 introduces the qualitative data derived from the interview processes with those 24 participants whose contributions may be analysed in this work. The chapter examines pertinent themes that emerged across the interviews, including insights on actor knowledge and behaviours within the discourse generally and the specific case studies analysed in the preceding chapters, trends on the use of hashtags and thus relevant considerations when selecting data upon them, and imperatives for further research.

Chapter 8 summarises the main findings of the various studies and offers conclusions on the research by reference to the primary hypothesis and themes. The thesis concludes with an identification of some of the limitations of this research, and suggestions for future research possibilities.

Chapter 2

Technology & Society: Situating mediated social communication

2. Technology & Society: Situating mediated social communication

2.1 Introduction

This chapter reviews the literature pertaining to those theoretical frameworks for conceptualising and understanding the relationship of technology and society, and seeks to identify salient concepts that can inform this work. Specifically, this chapter will outline the development of thought around the effect of media technologies on human communication, expression, organization and therefore discourses at a societal level, as well as the effect of that society – its norms, conventions, power structures, etc. – on the technologies' introduction and use. An examination of the social theory of *fields*, by Pierre Bourdieu, follows and develops a framework for understanding the social interactions mediated by a new technology. The chapter thus investigates central themes of the research questions identified in the introduction to this work; how can we understand the adoption and social context of online discourse, and how are these social interactions related to wider structures and systems of power in society?

The chapter begins with the establishment of the *technological determinism* viewpoint for the relationship of media technologies to society, and traces the criticisms of this stance through the development of theories of *social shaping* and *social construction* of technology. The interplay between the technological and the social examined here prompts consideration of the multidimensional context of the adoption and use of a media technology, such as Twitter - factors influencing this may be political, cultural, and economic.

Finally, the work of Pierre Bourdieu is outlined - specifically discussing his theory of fields, which forms a useful explanatory framework for understanding the relative social positions and struggle for power of agents within a given setting, such as the mediated space of Twitter. Bourdieu offers concepts that help to position this given field in the broader social setting of those conventions, beliefs, and predominant social structures that govern it. The chapter concludes with an analysis of recent work by social scientists to relate Bourdieu's theories to social media particularly and to identify specific affordances therein that modify how social structure and agency are maintained and adapted by participants.

2.2 Theories of Technological Determinism

Central to any discussion of the media-society relationship as a *technologically determined* one is the work of Canadian communications theorist, Marshall McLuhan. Lister, Dovey, Giddings, Grant and Kelly (2009) demonstrate, through extensive examples of the on-going applicability of McLuhan's theories to contemporary discourse on this relationship, the continued powerful relevance of McLuhan decades after his 1960s publications on 'mass media'. Castells, who has prominently published on the sociology of the 'information age', also points to McLuhan as "the great visionary who revolutionised thinking in communications", through his framing of mass media communication systems as "a distinctive mode of cognitive expression" (Castells 2000 p. 357).

2.2.1 The work of Marshall McLuhan

McLuhan's feted aphorism that "the medium is the message" (1964 p. 7) is predicated upon his contention, in Understanding Media: The Extensions of Man, that the content of a mediated communication is secondary to the medium that is shaping it. He further underscores this point in his 1967 work, The medium is the Message: An Inventory of Effects, asserting more strongly still that content is unimportant compared to the medium of its communication; "societies have always been shaped more by the nature of the media by which men communicate than by the content of the communication" (McLuhan (1967) in Lister et al. 2009). For McLuhan, the content of a transmissive medium may be thought of as "ineffectual in shaping the form of human association" (1964 p. 9) - a redundant distraction from a more relevant need to understand the *character* of the encapsulating medium or technology, and its inherent capacity to extend communication beyond the corporeal and proximal. McLuhan's determinist viewpoint therefore sees these media (technologies) as an "extension of ourselves" (1964 p. 7), redrawing the temporal and spatial boundaries in which our physical bodies can reach and communicate - firstly in the mechanical age through the extension of our bodies in space, and later in the advanced technological ages to a state where "we have extended our central nervous system itself in a global embrace, abolishing both space and time" (ibid.).

McLuhan's contention is that these processes of media and technology-based extension have profound societal implications and consequences, which are distinct from any content transmitted by them. Proposing the example of the electric light bulb as a content-less electric information carrier – "a medium without a message" – McLuhan suggests that regardless of the purpose of the light's transmission or use, the fact of the light itself has effects. Similarly, he asserts that media and technology in themselves act upon society, and (like the light) "shape and control the scale and form of human association and action" (McLuhan 1964 p. 9).

McLuhan's reduction of societal actors to passive transmitters and receivers of communication enabled by - but incidental to - the media in use, is an inherently strong technologically determinist viewpoint, in which it is clear that technology must drive social change through its capacity for communicative extension. This paradigm places media, and the technologies that drive them, outside of society and culture - acting upon these but autonomously independent of their effect; technological progress, here, is the causal agent of social and cultural change. This view of the centrality of an inevitable technological progress exogenously acting upon and reshaping media, politics, art, culture, etc., is widely to be found in journalistic accounts of web and social media communications (Castells 2009; Morozov 2013; Preston 2001; Wu 2011). MacKenzie and Wajcman (1999), as well as Lister et al. (2009) also discuss this determinist tendency in the specific context of contemporary news media's elevation and celebration of technological development, particularly of new social media, as a source of radical social change. In these cases, it is the determinist viewpoint first elucidated by McLuhan that has come to dominate the discourse on the relationship of cultural change to media technologies – a reductionist model in which the implications of a medium's usage, for individuals, groups, or society at large, arise from the presence, adoption and development of underlying technologies, irrespective of content produced or communicated.

2.2.2 Determinism in visions of the 'Information Age'

Although Manuel Castells rejects the label of technological determinist in his multi-volume work *The Information Age: Economy, Society and Culture*, it is clear from its first volume (*The Rise of the Network Society* (2000)) that Castells' claims about the development of broadcast media are strongly deterministic in nature. In chapter 5 of the work, Castells contends that society is undergoing a seminal change because of the *convergence* of communication forms enabled and shaped by technology. For him, the emergence of an 'interactive network' represents a unity developed between machines in which the textual, aural, oral, and visual are integrated. The creation and adoption of this 'information superhighway' has fundamentally changed communication, and thereby culture and society, and is derived from and directly dependent on the emergent technologies of 'new media'. In Castell's view, reality is linguistically structured, and the media shaping, transmitting and extending that language are thus crucial building blocks of culture; communication, he asserts, is the decisive shaper of culture, and media the shapers of communication. Though he would distance himself from it, this view of culture mediated through communications technologies, and therefore dependent on their

limitations and affordances, transformed by their developments, etc., is fundamentally a technological determinism.

Castells' vision for the on-going development of the 'information age' - and with it the transformation of societies - is often uncritical and utopian. No reference is made in this work (2000) to the existence of a 'digital divide' in access to the transformative enabling technologies, apart from an acknowledgement that their systems are neither fully developed, not equally geographically spread at the time of writing – though inevitably to be expected to achieve universality with time; "it is a certainty that it will develop and embrace at least the dominant activities and the core segments of the population in the whole planet" (Castells 2000 chap. 5). No discussion of the timeline for such a global extension of media technologies is offered by Castells, nor does he disentangle the requirements of functional and technological literacies – absent in the majority of the world's population, but necessary dependencies for such media – from the processes of radical change he asserts will result.

A similarly utopian and highly deterministic narrative for the contemporary co-development of technology and society is offered in the works of Clay Shirky. His (2008) Here Comes *Everybody* more narrowly focuses on the recent development of the social internet than does Castells' cumulative review of decades of converging media technologies, though the posited outcome for human society is the same - radical cultural change enabled (indeed compelled) by the communicative and organisational affordances of new media platforms and their underlying technologies. While grounded in concepts and terminology from the 1930s economics paper The Nature of the Firm (Coase 1937), Shirky's discussion of the shifting transaction costs that limit group organisation in societies is nonetheless in line with the works of McLuhan and Castells; it is not the specific content discussed, nor agenda espoused, that creates the net societal and cultural changes, according to Shirky, but the aggregate *capacity* for these, brought about by the media platforms (and technologies) of the internet. He provides as example, the 'mass amateurisation' of journalism and photography afforded by online tools, specifically citing Twitter as 'threatening to change the way news is spread'. Significantly for political discourse, action, and mobilisation, Shirky posits that the 'social tools' of the recent internet have fundamentally altered the processes of group formation, dynamics and organisation such that individuals are empowered to challenge societal conventions of organisation and governmental power. For Shirky, the lower time, effort, and experience costs of effective collaboration, allow groups to quickly mobile and campaign, "without needing traditional organisational structures" (Shirky 2008 p. 21).

A critique may be made of the extent of universality - in terms of access and required ability - of the enabling tools that determine the social change in Shirky's paradigm, similarly to that

made of Castells' vision for the inevitability of the information age. Reviewers of the work, such as Brabazon (2008) and Morozov (2011), criticised the exclusion of older, poorer, illiterate, or infrastructurally limited citizens from Shirky's collection of individuals capable of harnessing the transformative organisational potential of the new media social tools. This determinist viewpoint, maintained without a social context of the technologies' adoption and usage, has remained a critique of Shirky, as it has with Castells; "[the] assumption that 'we' can learn about technology from technology - without attention to user-generated contexts rather than content - is the gaping, stunning silence of Shirky's argument" (Brabazon 2008).

2.3 Theories of the Social Shaping of Technology

While the determinist approaches of the theorists outlined in the previous section assert technology as the primary actor in a unilateral and inevitable progression-based causality, critiques of these ideologies have sought to reposition the technology-society relationship as a more complex interplay. This section opens with an exploration of the early critique, by Raymond Williams, of McLuhan's determinist stance, and follows this with an introduction to theories of the social shaping of technologies that suggest that social and cultural contextual factors, unacknowledged by determinist theorists, play a central role in defining the co-evolution of technology and society.

2.3.1 Raymond Williams' critique of McLuhan

Williams was amongst the earliest cultural theorists to stand as critic of McLuhan's views on technology and society. His 1974 work, *Television: Technology and Cultural Form*, expresses concern at what Williams viewed as the linear causality and directionality implied in McLuhan's statements about 'effects' and 'consequences' brought about by technologies. Williams rejected these *a priori* 'effects' as simplistic and vague, hiding much more complex philosophical questions about the social dimension of technologies' use, and suggested that a closer examination of the appropriateness of a *cause - effect* paradigm was required. For Williams, the discussion of 'effects' brings about the implication of a cause or causes, and sets up a simple (and possibly spurious) model of determined causality in which the introduction and adoption of new technologies themselves may be viewed as exogenous starting points for the changes documented. The absence of analysis of a holistic investigation of the societal and individual context, he contended, masked questions such as "whether it is reasonable to describe any technology as a cause, or, if we think of it as a cause, as what kind of cause, and in what relations with other kinds of causes" (R. Williams 1974 p. 2).

Williams argued for a pragmatic approach to the questions of the technology - society relationship, leveraging practical and empirical research in social and cultural studies to assess the role of human agency in the success or failure of a particular technology¹⁵. He extends the *cause - effect* paradigm implied in McLuhan's work to lead from technological to societal change, to a questioning of the causes leading to the observable effect of a technology's adoption or rejection within a society. Thus, Williams broadens a simple unidirectional determinism that would position technology as the major actor, to a more

¹⁵ Television, in the case of his (1974) work, though these arguments are equally applicable in any communications medium arising from new technologies.

complex viewpoint that requires us to question the iterative interactions between both the technology and its adopters.

Where McLuhan (and indeed Castells and Shirky later) propounded theories in which *mankind* is a largely homogenous group entity, subject to uniform causality for the societal changes within it, Williams' theories prompted enquiry centred on the complexity and specificity of social, cultural, scientific and political dimensions of technology.

Williams early responses to McLuhan have therefore come to bring about wider questions on the role of society in the specific fields of study of cultural production, path determinism, the relationship between science and technology, between technology and politics, etc., leading Lister *et al.* to declare his work to "have been a major, shaping contribution to the constitution of an academic discipline [cultural studies]" (2009 p. 73).

2.3.2 The Social Shaping of Technology

Following Williams' work, MacKenzie and Wajcman further developed the study of technology within the context of specific social dynamics in their essay collection exploring the technology - society relationship, *The Social Shaping of Technology* (1985). The authors, echoing the principal sentiments of Williams, assert that determinist accounts of technology reduce its complex relationship with society to too simple a cause and effect sequence, in which no consideration is given to how social actors and the interrelationships between them affect the actions they take using the technology¹⁶.

MacKenzie and Wajcman observe that the social sciences are behind disciplines such as engineering in their discourses of technology, in that engineers have long accepted that the technological superiority of an object in terms of utility alone cannot counter more 'social' perceptions of shortfall in cost, aesthetics, fitness for purpose, politically acceptability, etc., which may cause a technologically superior object to be rejected by society at large or by particular groups within it. Thus, the existing sociological discourses of technology were considered by the authors to be 'a naive technological determinism' that was simplistic, both politically and intellectually. This position encouraged political passivity and reduced critical discourse to positions that adopted an uncritical embracing of technology, a defensive adaptation to it, or an unexplored rejection of it.

¹⁶ It is worth noting, however, that by the major second edition of their work, (MacKenzie & Wajcman 1999), the authors acknowledge that between editions it has become generally accepted by academics that there does not exist a unidirectional relationship between technological and social/cultural change, but that this is a 'two-way relationship'.

The essays collated by MacKenzie and Wajcman explored the relationship between technology and society in a series of case studies that sought to unpick the apparent 'common sense' fit of technological determinism, by investigating the social dimensions of the cases.

One such case, often cited in subsequent summaries of MacKenzie and Wajcman's work, is that of Robert Moses – a senior New York City planner at work in the mid-20th Century. The case study demonstrated that apparently insignificant details from a technologically determinist point of view, such as the clearance height for bridges - analogous with limitations of affordance in media technologies – were later demonstrated to have had specific and intentional social intent; Moses' design of the clearance height was a deliberate discouragement of the passage of taller public transport vehicles through certain wealthy areas, directly reflecting his own class and racial biases (Winner in MacKenzie & Wajcman 1999). Thus, the author contends that it was not the civil infrastructural technology, in isolation, that determined the pattern and extent of travel, but the particular shaping of the technology by a socially biased designer. Analysis of the case study develops the point that complex social factors of politics, prejudice and taste determine how a technology is implemented, which may in turn determine further societal change afforded by it.

A similar case study, which underscores the economic factors in the technology - society relationship is that offered by Bijker (1995), who discusses the introduction in the United States of the fluorescent light bulb by *General Electric* in the late 1930s. The electric companies, supplying power to the homes and businesses that might make use of the new bulb, argued against its efficiency on the grounds that it would hinder their profits by reducing demand for electricity. Their lobbying resulted in the re-design of the bulb and a sequence of litigation by government antitrust investigators and GE, culminating in the latter's successful defence of their altered design based on interference to it potentially hindering the on-going war effort of the time. Thus the introduction of this particular technology was shown to be subject to factors such as the economic concerns of pre-existing powerful actors, established commercial law, national and supranational politics, and the temporal state of war; the relationship between technology and society can therefore be shown to be affected by the convergence of highly complex social, political and economic factors not inherent to the technology itself.

These case studies, and others, combine to call into question any notion of the inevitability of adoption (either initial, or continued after success in one social or geographic space) of a 'best-fitting' technology. Indeed, any conception of a technology being a 'best-fit' for delivering some efficiency, affordance, or change-making innovation, would itself be highly subjective and grounded in social, economic and cultural dynamics that vary widely.

42

MacKenzie and Wajcman observe that where technologies are successful, their adoption and usage often begins a process of iterative improvement (as user experience is relayed to designers who respond to adapt the technology). These processes are capable of incrementally strengthening the technology, but are themselves an obvious social-shaping of it based on its contextualised use, and are likely to direct its continued design toward the specific requirements of the group where it is first well-adopted. Thus, the early adopters of a technology may shape its subsequent development, subjecting it to refinement based on the social context of their usage of it (MacKenzie & Wajcman 1999).

2.3.3 The Social Construction of Technology

Distinct from, and subsequent to, the theories of the social *shaping* already outlined, is the concept of social *construction* of technology (SCOT), developed by Bijker and Pinch. Central to SCOT is the notion of *interpretative flexibility* – that a given technology (technological *artefact*) may have different meanings and interpretations among various groups it is relevant to. Taking account of these relevant social groups, and observing how the shared and differing meanings they attribute to an artefact may affect its development and adoption, is central to the theory (Wiebe E. Bijker et al. 1987; Kline & Pinch in MacKenzie & Wajcman 1999). The groups include obviously influential agents, such as engineers, designers, and users, but also include advertisers, journalists, politicians, civil leaders, and salespeople, who are neither producers nor users of a technology, but who influence the context of its presentation and its situation in society (Pinch in Oudshoorn & Pinch 2003). These groups may be delineated by their shared or diverging interpretations of the artefact and the meanings they ascribe to it can explain the complex system of its implementation in society.

Kline and Pinch suggest an observable endpoint to the processes of interpretative flexibility, where 'closure' and 'stability' occurs for the technology, usually coincident with its achieving some dominance in its particular field of purpose (Kline & Pinch in MacKenzie & Wajcman 1999). They note, however, that it is possible for new issues to arise in the usage of a closed and stable technology, after it has become established, leading to a new iterative cycle of interpretative flexibility and possibly a subsequent restabilisation. Thus, from the SCOT viewpoint, it is obvious that a technology cannot be assumed to have the same meaning for all groups, in all spaces, or at all times, leading to a segmented co-construction of a technology's stable place within a given set of social, spatial and temporal factors, and a significant question over any sense of inevitable universality.

The *social shaping* and *social construction* theories of technology have both countered the determinist stance by posing questions for the researcher on the complexity of the technology

- society relationship, the multidimensional contextual factors for a technology's introduction, and the interactions affecting and bounding its development, adoption, use, and success within a particular social, cultural, political and economic domain. These theories too, are subject to critique, particularly that in excessively deconstructing the social dimensions of a technology's introduction, the sense of its effect is diminished or negated. An attempt to derive a more balanced model of the technology - society relationship is evident in the emergence of Actor-Network Theory, in which technology holds a significant position of agency, but is still subject to examination, contextualisation, and criticism. An overview of this theory is presented in the next section.

2.4 Actor-Network Theory

The work of Latour and others in the development of Actor-Network Theory (ANT) has been particularly influential for the conceptualisation, within the social sciences, of the relationship between technology (especially media technologies) and society. Latour and Strum's investigations in primatology (in MacKenzie & Wajcman 1999), demonstrated a system of continuously negotiated and renegotiated roles constituting the social structures within which communities of baboons lived. The overall baboon society, they observed, was the result of individual organisational roles, rather than the other way around. Latour and Strum contended that this understanding of baboon society by primatologists was close to the model held by sociologists studying human society, and that the latter disregarded the added complexity of human interactions, which are not subject to the same spatial and temporary limitations. Whilst interacting primates were bounded geographically and to the present, human actors draw on shared and borrowed histories, complex systems of communicated intent, and the sort of extensions of their corporeal bodies in space and time, alluded to by McLuhan. Latour and Strum also discuss the centrality of the use of symbols and material resources to human interaction, making the technological means by which these are created and exchanged inextricably part of the society. Thus, they conclude that technology, being crucial to the individual co-creation of structure, is constitutive of human society rather than apart from it; they suggest that any attempts either to set technology aside as independent determiner, or to theorise social relationships without it, are erroneous and amount to 'baboon theory', focussing on primitive behaviours and disregarding embedded technological considerations (ibid.).

2.4.1 Technological Agency

For Latour, then, it follows that technology must itself possess *agency* within the sociological context - an attribute ascribed in previous theories only to the human actors in the interactions from which structure arises. In *Aramis: Or a Love of Technology*, Latour (1996) provides a *sociotechnical* study of the failure of the Parisian personal rapid transit (PRT) system developed from 1970 until its abandonment in 1987, and dubbed 'Aramis'. The described intention of the Aramis PRT was to provide a hybrid transport system, something between train and car¹⁷, offering the flexibility of the latter in a point-to-point network of connections

¹⁷ It was conceived as a 'platooning' system of individual-use vehicles, similar in capacity to a small car, which could operate together in the manner of a train at busy sections of transit, and split apart to provide the specific and personalised transport needed for their occupants.

between existing hubs, intending to provide convenience while also limiting congestion and pollution.

Latour offers an analysis of its failure in a fictional narrative form, in which a professor of sociology and his assistant gather intelligence on the specifics of the failure of the system, in the hope of reaching a firm conclusion on who or what 'killed' Aramis. This form of writing is a hybrid of novel and sociological investigation, in which the fictional figures gather and cite real data on the project and interviews with actual personnel involved in its development. Latour explains that the intention of the work was to show technicians, engineers, and those responsible for the creation and introduction of technologies that they cannot conceive of a technological artefact "without taking into account the mass of human beings with all their passions and politics and pitiful calculations, and that by becoming good sociologists and good humanists they can become better engineers" (Latour 1996 p. viii). His dual goal is to demonstrate to technologists that the creation of artefacts is not viable without consideration of human influence, and to humanists that technology and machines are worthy of their respect and study, as integral constituents of society.

At the end of the work, Latour has his fictional professor come to the conclusion that there was no definable 'crime', no single fatal cause or agent responsible for the 'killing' of the Aramis project. Instead, he concludes, it was the cumulative and inherent weaknesses in many dimensions - financial, political, design, planning-related, and interactional - that brought about the demise. Despite these 'fragilities', the creators and developers of the project believed that the inherent value of the project's contribution to the lives of the Parisian people must drive it to a successful conclusion; "You believed in the autonomy of the technology" (Latour 1996 p. 292). While certainly a direct critique of technological determinism, the work also documents a shortfall in social shaping theories, which are similarly unable to account for the totality of the complex relationship between the social and the technological in the Aramis case. There is no clear deterministic path, either technological or social, for the evolution of the PRT project, which many times changed direction and aim and was ultimately kept going by 'the love of the project' until its eventual intractable failure.

Central to the Latour's view of the broad web of society, then, is the fact of human and technological agency and the implications of both; human actors are free to act and communicate, while simultaneously residing within social structures, the norms of which minimise their individuality. The technologies included in, and necessary to, these social structures and continuous interactions are also ascribed agency - here this is not a marker of intentionality (another factor of human agency), but an acknowledgement of the role of the

technology as mediator rather than intermediary¹⁸. Actor-Network Theory attempts to dismantle the dualistic distinction between agency and structure, and between technology and society, and offers something of a 'third way' to the theories of technological determinism and social shaping outlined earlier in this chapter.

2.4.2 The Network of Actor-Network Theory

Intrinsic to Latour's concept of ANT is the inclusion of a networked dimension, in which agency (human and non-human) occurs. This aspect of the theory is controversial for its principle of *generalised symmetry* - weighting technological agency equally with human agency. It is somewhat easier to conceive of society as intrinsically networked where those networks composing it are social and have as their *nodes* only people, rather than people and materials and machines.

Law, in his *Notes on the Theory of the Actor-Network* (1992) contextualises this less intuitive conception of network well. He argues that knowledge is social, the product of a network, and necessarily also material, arising from the combination of non-human entities (computers, laboratory equipment, etc.), as well as human ones (researchers, scientists, and their specific skills derived from social processes of education). Scientific knowledge, for Law, is a *heterogeneous engineering* of necessarily interlinked discrete entities - human, non-human, social, technical, and textual. In extending this network understanding to all social interactions, he points out that almost all are mediated through non-human entities, none of which are passive intermediaries. All aspects of society, therefore, arise from myriad actornetworks in which symbolic information is routed and modified by its transmission.

Such a conception of societal networks denies a special position to human agents within them, by excluding a hierarchy of humans and objects. Law contends that the ethical questions this raises for some, are answered by the necessity to distinguish ethical enquiry from sociology, treating them as separate fields capable of informing one another (Law 1992). For him, the fact of generalised symmetry, and its lack of affording a special place to humans, does not lead to a de-humanising reduction of those actors to the status of machines, nor to a consequence that human rights need not be considered and applied. He sees value in considering the actors (human and non-human) as equal and necessary components to the overall network, and posits that this model of investigation, when separate

¹⁸ The distinction here lies in the quality of the entity to either create or multiply difference in a given transaction (mediator) or to passively transmit or transport without limitation or transformation occurring (intermediary). ANT proponents content that the latter is the position given to too much of the studied world, and especially to technology, by sociologists traditionally. (Latour 2007)

to an ethical one, can provide greater insight into a given circumstance, citing the example of machine-assisted-living, in which the human-technology relationship is complex and benefits both from ANT analysis of its structure and development, and ethical analysis of the human morality involved. Bijker, too, has responded to criticism of ANT as amoral, by similarly pointing out that it does not necessitate amorality and that moral (or political) judgments may be separately made on a network once it has been described. (W. E. Bijker 2010).

2.4.3 Applicability of Actor Network Theory to Twitter

A structural definition of the persistent or ephemeral 'links' between users who follow or message one another in the mediated space of Twitter requires the conception of an interlinked network, where the users are nodes and their various connections links (*edges*, in network terminology) between them. This simplified model is widely used, especially for visualisations of connectedness and conversational interaction in Twitter (E.g. Bruns and Burgess, 2011; Ausserhofer and Maireder, 2013; Highfield, Harrington and Bruns, 2013), but a more useful conceptual framework when considering the shaping of Twitter communication by technological and social effects is the ANT network. Actor Network Theory conceives of communication networks not as the simplified structural arrangements of interacting humans outlined, but as complex webs of agency and effect, in which technological actors (e.g. hardware, algorithms, etc.) are not denigrated as agenda-less tools of human interaction, but as equally co-constitutive components of the overall network.

Thus ANT provides for a reconciliation of the disparate focuses on technological or social determinism in the earlier theoretical frameworks outlined, and instead treats technological and human social agency as mutually constitutive elements of the mediated interactions. Such a view allows for research to equally value technological affordances and shaping actions in considerations of how observed patterns of communication might have emerged. In the case of Twitter, factors such as the 140-character constraint, the topic-based organisational potential of hashtags, the algorithmic determination of observable content from followed users, and the automation of content sorting and aggregation are all clear technological shaping factors on any communication taking place in the space. Similarly, the reflection of social status based on actors' relationship to the wider field of politics, the evidence of user-led innovation and change in how critical elements of the social network have developed, and the existence of norms and expectations around tweeting and mentioning other users, point to social shaping of the technology and of the communication facilitated by it. ANT's imperative to consider both of these sets of shaping influences as strongly interlinked in a complex web of interaction usefully suggest both lines of enquiry and modes of interpretation for analytical investigations of Twitter. No understanding of how and why communication on a particular political topic or event unfolds as it does is complete when viewed as a solely technologically- or solely socially determined phenomenon. The agency of human and non-human components of the *Actor-Network* must be considered, and in doing so, useful insight that might be ignored by earlier approaches may be obtained. Chapter 7 of this work particularly outlines how consideration of non-human agency in the Twitter networks studied gave rise to insights about the shaping of contributions by and for technological effect.

2.5 Bourdieu's Theories of Social Interaction

Similarly to works in ANT, Pierre Bourdieu has attempted to replace the dualistic nature of debate about structure and agency with an understanding of society as a complex symbiotic relationship, the subjective and objective parts of which are greatly interlinked. His investigation of the nature of culture in society was concurrent with an influential epistemology for sociology, centred on reflexivity on the part of the researcher in an effort to derive relative objectivity. Bourdieu termed this *participant objectivation* (Bourdieu 2003).

Jenkins' critical introduction to Bourdieu's work (2007) stresses the importance of an engagement on the part of researchers with the global zeitgeist; "he asserted the right, the duty indeed, of the public intellectual to engage with politics and the issues of the day, whether they be poverty, immigration, or globalisation" (Jenkins 2007 p. x). For Bourdieu, any sociological enquiry, just as the phenomenon that was its subject, must be situated in a cultural, political and social domain that was local, national and supra-national. He asserted, and demonstrated by example in his own work, that theory and research are necessarily mutual and that systematic empirical work was necessary to frame understanding; 'theory without empirical research is empty, empirical research without theory is blind' (Bourdieu in Jenkins 2007 p. 10). While these assertions by Bourdieu concern research methodology and the role of the researcher in society, rather than constituents of his theories of societal interaction, they are relevant to this thesis more broadly - the pragmatic mixed methods approach of which is detailed in Chapter 4. Bourdieu's key concepts, as they related to cultural theory and the interactions constituting human society, are outlined in the following sections. In conceiving of Twitter as a Bourdieusian field of social practice, these concepts together offer a useful framework of understanding how Twitter might operate as a distinct mediated space, reflecting structures and norms from outside it, and developing practices and conventions within it, all of which shape its communicative interations.

2.5.1 Habitus

In building toward Bourdieu's concepts of *field* and *capital*, that of *habitus* is important to establish. In his framework, habitus constitutes a system of pervasive, lasting dispositions, formed through unconscious appropriation in childhood. It is initially formed from parents and built upon early sets of logical distinctions (hot / cold, dark / light, etc.), which are subsequently applied with increasing complexity as analogical schemes for the solution of problems, based on past experiences. Bourdieu summarises this as a 'practical sense' and 'feel for the game', stemming from our internally developed logic (Bourdieu 1990). Though internalised by us, habitus is rooted in family, class, and social contexts; "individual practice

50

as regulated by the logic of practice is always a structural variant of group and especially class practice" (Garnham & Williams in Collins 1986 p. 120). Bourdieu suggests that the unified practice of a class group is in part determined by the habitus, positing the formula: [(habitus) (capital)] + field = practice (Bourdieu 1990 p. 101)

Habitus is a useful construct, in that it does not separate social structure from agency; "embodied dispositions [...] are generated by structural features of that same social world" and "agents' dispositions to act are themselves formed out of pre-existing social contexts" (Couldry (2004, p358) in Papacharissi & Easton 2013). Habitus, though the product of socialisation processes begun in childhood that impart structural practices that become taken for granted, is not fixed at a point and thereafter a rigid set of dispositions to be acted out. Rather, it is necessarily reflexive as its implications for social position and agency are brought to and tested in new fields, where specific practices may develop to extend it. Papacharissi and Easton's (2013) work on the *Habitus of the New* discuss this reflexivity in the context of new social media, such as Twitter and Facebook, positing that a permanent state of novelty inherent to these allows for a sustained and accelerated reflexivity, prompting actors within them to examine and adjust their general tendencies to reproduce certain patterns of action, and conform to roles transferred from other social settings, or *fields*. Innovations in technology giving rise to new fields of social interaction, therefore, may be shown to give rise to *hybrid habituses* that incorporate new values with old ones.

2.5.2 Field

Field is a core concept of Bourdieu's sociology and comprises the setting and social positions of agents, governed by the interaction of the agents' *habitus, capital* (anything which is held to be significant for the social agents, e.g. money) and the rules of the field. These are specific social arenas, in which struggles take place over particular resources and access to them (Jenkins 2007). Power and position are important organisational attributes of the hierarchical, relational, concept of field, "within which human agents are engaged in specific struggles to maximise their control over the social resources specific to that field - the intellectual field, the educational field, the economic field, etc.' (Garnham & Williams in Collins 1986 p. 122). Bourdieu asserts that while fields may exist independently of one another, some fields dominate others and their struggles are reflected in them. He cites the example of the predominant *field of class struggle*, whose ultimate goal of accumulating capital will result in the development of skills to cope with or advance this aim, detectable in the struggles of other fields it imbues. The *field of power* horizontally permeates all others, controlling the exchange of capital (social, symbolic, or physical) within them, and influencing the struggles and relative positions of agents. Such fields are not vague

conceptions, for Bourdieu, but concretely describable and bounded settings - "a veritable social universe where, in accordance with its particular laws, there accumulates a particular form of capital and where relations of force of a particular type are exerted" (Bourdieu 1993 p. 164). People may be expected therefore, to experience power differently depending on which field they are operating in at a given time.

Activity, particular to each field, develops as a set of market-like interactions (struggles) between actors, who are competing for specific benefits associated with it. This struggle defines the objectively observable behaviours of actors, which relate to their volume of capital, trajectory of position (and power) within the field, or ability to influence the rules governing it. The extent to which participants may be able to make use of their accumulated capital is a function of their adaptation of habitus to the given field - their ability to apply the internalised systems of logic and understanding to the power relationships and capital struggles of a given field. Participants in a particular field subscribe to it by implicit acknowledgement of the stakes and the rules for their acquisition, rather than by any explicit mean. Bourdieu terms this acknowledgement the 'social illusion', *illusio*, of the field; it is the understanding, via the adaptation of habitus, of 'how the game is played'.

Setting the mediated communication space of the *Twittersphere* – specifically the Actor-Network comprising such political discussions as form the basis of the case studies in this work – as a distinct *Bourdieusian field*, recognises that status, structure, and power are experienced differently here than in the related fields of traditional media, politics, etc. Twitter has a separate *illusio*, shaping the reflexive adaptation of the habitus of its actors; struggles for status within Twitter take unique forms, though are often linked to the retention or accumulation of external status and power.

2.5.3 Capital

Bourdieu posits the existence of a class struggle for domination over the modes of production, and therefore over capital. Capital, in this particular sense, is physical and stems from the historic, pre-industrial fusion of material and symbolic fields of production. As society developed, a corresponding economic field grew and separated from the material and symbolic fields. Thus, in post-industrial societies, capital takes many forms and is essentially that which a particular field values and struggles to obtain or to control access to, often economic. Shifts in the definition, value, and possession of differing forms of capital can lead to profound changes in a field's makeup. Bourdieu differentiates between multiple types of capital, including economic (monetary), cultural (class-based knowledge, tastes, and resources), social (the potential for opportunity based upon relations among actors), and symbolic (honour and prestige). In general, the more capital one possesses, the more power they may wield. (Bourdieu 1993; Swartz 1997). Of those forms of capital outlined, social and symbolic capital are most relevant to the power dynamics of the mediated space of Twitter, as actors within it struggle for prestige of position in the form of markers of influence (followers, retweets, etc.) and associated opportunities to reflect or transcend a position in a related field (e.g. politicians wishing to establish similar hierarchical dominance within the field of Twitter as within the related field of politics, or amateur media agents, such as bloggers, seeking to attain position not afforded to them in the media field, respectively). It should be noted that Bourdieu acknowledges processes both of transfer of capital between fields, such as is outlined in these examples, as well as the conversion between different forms of capital. The latter is readily evident within the field of Twitter, where certain actors have sought to accumulate social and symbolic capital associated with influential position, so as to leverage this to attain employment as professional pundits in the related media field, this acquiring economic capital.

2.5.4 Doxa

Bourdieu (in Benson & Neveu 2005) defines *doxa* as "the universe of tacit presuppositions that we accept as the natives of a certain society" (ibid., p.37). It is therefore a field's *taken-for-granted* state that goes largely undisputed and constitutes the experience that "the natural and social world appears as self-evident" (Bourdieu 1977 p. 164). Doxa stands apart from *opinion* - the sphere of values and truths that may be openly discussed, questioned and contested - where Bourdieu offers the concepts of *heterodoxy* and *orthodoxy*; orthodox values are those which fit the *status quo* of the field, while heterodox values are those which depart from accepted norms. Together, these concepts establish a model for how the social space is created and limited. Thus, as actors are socialised into a field, they start acquiring field-specific capital, forming a situated habitus, and eventually become accustomed to the doxa of that field. The end result of this process is *practice*. (ibid.)

2.5.5 The field of a social media platform

Authors, such as D'heer and Verdegem (2014) have recognised the *Twittersphere* as a mediated social space to which the concept of field is applicable. Within the setting of Twitter, as with any social media platform, a unique set of interaction rules govern the struggle for social and symbolic capital among those present, and the adaptation of habitus by actors creates the power dynamics observable within it. While the specific affordances of

Twitter, especially the limitation of expression to discrete character-limited utterances, and the mediation of that expression toward follower groups or publics, provide the necessary rule specificity for a discrete Bourdieusian field in itself, Papacharissi and Easton (2013) identify a set of affordances of social media platforms generally that affect how habitus is modified within them. Extending boyd's (2010) four structural affordances of networked publics (persistence, replicability, scalability and searchability), they add shareability as a necessary affordance of the social media platform; persistence, here, refers to the automated archiving processes that preserve the interactions within the space, *replicability* the capacity for duplication of content within it, *scalability* the potential visibility for that content via systems of following or publication to defined publics, and *searchability* the access to given content via affordances to tag or search for it. Papacharissi ((2011) in Papacharissi & Easton 2013) argues for the addition of *shareability* referring to the tendency within social media spaces to explicitly encourage sharing over withholding; "What renders networks lively is the flow of information between individual network nodes. Without information flowing between individuals, the network becomes a static, asocial environment" (ibid., p10). Combined, the authors argue, these form a habitus of the new, explaining how individuals reconcile physical and digital interaction practices, maintaining or adapting dominant social practices; participants perform online similarly to how they perform face-to-face interactions within cultural norms constructed for other fields, while seeking opportunities in the affordances of the new field to acquire position and power.

Thus, the social architecture of online spaces like Twitter compels actors within them to use opportunities to publically offer narratives about themselves and to share opinions. What may be mistaken for narcissism, Papacharissi and Easton argue, "may also be understood as an expression of agency that conforms to and seeks to relate the individual to the habitus" (2013 p. 14). Discourse specific to the platform is constructed by actors within it, who engage in a performative authorship that is self-reflexive and acknowledges the affordances, limitations, and tropes of the medium itself. Digital literacy is therefor a crucial requirement to participation in the field of a mediated social space such as Twitter, since knowledge of and fluency with its affordances are central to success and the accumulation of capital.

Findings on the digital divide cited in Papacharissi and Easton (ibid., p19) show that it is not just *access* to social media technologies that blocks certain actors from using them, but fundamentally different disposition toward them, in which irreconcilable habitus impedes fluent participation in the platforms. Here again, the authors argue for the dominance of a field of class power over the fields of social media platforms, since deeply engrained practices and familiarities produced by class and education are at the heart of the required fluency; "those without fluency lose agency and fluency is a product of class" (ibid, p20).

2.6 Conclusions

This chapter has reviewed certain key theoretical frameworks for conceptualising the complex relationship between technology and society, governing the development, implementation and introduction of new technologies within the specifics of their social, political, cultural and economic contexts. The utility of Actor Network Theory as a viable 'third way' framework necessitating equal consideration of technological and social shaping effects was discussed and its implications for directing the qualitative research outlined. A distinction was drawn between the notion of *network* as a simplified structural arrangement of linked communicating actors, and that of *Actor-Network*, as a conception of mutually co-constitutive human and non-human agency. While the former is a useful device in providing systematic computational analysis, and thus *Network Analytics* form and important part of the methodology being developed by this thesis, the latter is a crucial guiding framework for what these methodologies attempt to describe – complex communication spaces shaped by multiple technological and social factors.

The chapter also outlined the key insights of Pierre Bourdieu related to social structure, position and interaction, providing a conceptual framework for understanding these within the specific context of a given field - in this case, Twitter - and implying necessary lines of investigative inquiry for the social study of that field. The particular nature of interaction in the space of the *Twittersphere*, the nature and struggle for status within it and related externally to it, and the evidence of a set of norms and conventions for behaviour there were outlined in justifying the apt treatment of Twitter as a distinct field of social practice in the Bourdieusian sense.

In responding to the indicative questions that this chapter posed, the material covered may be summarised in the following particular conclusions:

RQ2.1 How can we understand the adoption and social context of online discourse spaces such as Twitter?

The earlier sections of this chapter outlined a narrative development of those theories of technology and society that have come to influence the understanding of their relationship in the social sciences. Whilst a strongly determinist stance around digital technologies was shown to be problematic in the simplicity of its assumptions and the isolation of the technology from any social context of its use, likewise shortfalls were identified for a 'hard' social shaping approach. The chapter outlined Actor-Network Theory as one that accounted for mutual agency between human and non-human actors and better situated technology,

without isolating or fetishising it. This is especially relevant for the technological case of a communications medium such as Twitter, where issues of convergence are important, where actors are engaged in processes of innovative use of the technology that is in turn responsively adapted to these, and where actors knowledge of the affordances, limitations and tropes of the medium results in highly reflexive use of it that emphasises the non-human components of the network.

Twitter's adoption may be understood then, as the development of a responsive actornetwork, capable of facilitating highly current and discrete packages of information exchange, and within which external social, political, cultural and economic domains are translated and transforming. A holistic investigation of any particular discourse within Twitter, therefore requires the researcher to engage with both the social context of its use and the existence within it of non-human agency in the mediation of its interactions.

RQ2.2 What is the nature of the interactions constituting online discourse, and how are these social interactions related to wider structures and systems of power in society?

Armed with the concepts outlined from Bourdieu, we can understand Twitter (or any other social platform for online discourse) as a discrete *field* in which participants are engaged in a struggle for social and symbolic capital, and concerned with position (and transfer of position from other related fields). We can use these concepts both to explain observed aggregate trends in the interaction between certain kinds of actors, as well as to predict potential patterns of interaction. The processes of *capital* accumulation and transfer, in particular, can shed light on observable behaviours among those actors most actively interacting in a discourse and struggling for specific position within it.

Concepts of *habitus*, *illusio* and *doxa* are also useful in informing any analysis of Twitter discourse, since they prompt questions as to the pre-existence, mapping-on, and novel development of norms, conventions and values for the field, relating it to wider social contexts and structures. Habitus, in particular, carries with it implied questions as to the historic social interactions and involvements of an actor, especially one who demonstrates fluency of action in the Twitter field; these may guide actor-centric investigations, such as interviewing. Similarly, the development of illusio - an actor's understanding of the rules of the field - is an important quality of this fluency and a necessary point of investigation when assessing whether successful actors share some particular insights about how a field operates, or indeed whether assumptions made by researchers about the systems governing their participation are even valid.

2.6.1 Implications for this research

Bourdieu's theory of fields implies observable patterns of behaviour within a given field, corresponding to the power dynamics involved. In particular, the struggles of participants to attain position, both by accumulating capital and seeking to transfer capital from other fields should lead to aggregate trends of interaction at the macro level within a studied Twitter discourse, as well as to specific behaviours and understanding of the rules, norms, tropes and affordances by individual actors at the micro level.

Following Bourdieu's sociological methodologies, urging the support of theory with empirical research, this work seeks to establish corresponding data for these macro and micro phenomena, through the development of specific discourse-based case studies, and the interviewing of key actors. The design of the analysis for the former will focus on the identification of actors' professional association with the field of politics, therefore relating their relative presence and position in other relevant fields to that of Twitter. Consideration of aggregate interactions by categorised actor groupings will be informed by a network approach, derived from an understanding of the relationship of Twitter to wider social domains as an actor-network, the structure of which is co-constructed by participants and by non-human agencies of the technology.

Design of the interview component of the research, outlined in detail in Chapter 4 and its appendices, will take into account the need to establish the habitus of participants, the derived social norms of the medium and the veracity of assumptions made about how actors understand the rules and affordances (*illusio*) of Twitter in their on-going use of it. Acknowledgement of the agency of the technology itself will inform some of the questioning around actor's use (or not) of particular clients and tools that materially affect the presentation of parts of the discourse to them.

Detail of the methodological approaches in this work continues in Chapter 4, while this literature review section moves in the next chapter to an aggregate analysis of the state of the literature concerning political discourse on Twitter and its analysis.

Chapter 3

Political Discourse on Twitter and its Analysis

3. Political Discourse on Twitter and its Analysis

3.1 Introduction

Since its launch in 2006, Twitter has rapidly gained worldwide popularity, with over 300 million *monthly active users* as of Autumn 2015, generating over 500 million tweets daily (Twitter Investor Relations 2015). The simplicity of its character-limited format has led to its widespread use, particularly for highly current conversational and status-update based interaction, and it has become an important source of study for scholars, particularly because of the possibilities afforded in accessing geographic, thematic, or event based discourse and communications patterns. Twitter's 140-character messages (tweets) are relatively easy to process and store, and access to this stream of data (and user account metadata) is made available through Twitter's own application programming interfaces (APIs) and related thirdparty services. Given Twitter's default public setting for messages sent - fewer than 10% of users opt to "lock" accounts, restricting access to selected followers only¹⁹ - researchers have increasingly recognised the value of studying Twitter data to gain a better understanding of its mediation of communication, its users, and their patterns of interaction. Such studies have increased in number and academic field over time, and Twitter data analysis has been shown to have powerful explanatory usage in disparate investigations of social discourse. (boyd et al. 2010; boyd & Crawford 2012; Weber et al. 2013)

As a platform, its use has been well documented for communication, reportage and mass coordination during significant global political and social movements (e.g. 2008-2009 Iranian protests (Grossman 2009), 2011 Arab Spring protests (Huang 2011), the 2011 London riots (Procter et al. 2013; Tonkin et al. 2012; Vis 2013), the Occupy movement (Juris 2012), etc.). Similarly, it has seen widespread study during national and international crises and disasters (e.g. earthquake events and detection (Earle et al. 2011), tsunamis (Bruns & Burgess 2013) and floods (Bruns et al. 2012)). More recently, Twitter has been used for the study of electoral discourse and prediction (Bermingham & Smeaton 2011; Bruns & Highfield 2013; D'heer & Verdegem 2014; Larsson & Moe 2011; Lilleker & Vedel 2013), and such studies have increased in frequency and geographic spread. In each of these cases, as well as in

¹⁹ A current or recent proportion of private Twitter accounts is difficult to assess, given that this information has not been directly published in recent years by Twitter, as it was previously (e.g. Meeder *et al.*, 2010; Moore, 2009). However, verifiable analyses showed that it continued to fall through 2012 and 2013, and that fewer new accounts were being created as private (Adrianus 2013; Beevolve.com 2015)

literature related to finance, logistics, marketing, epidemiology, and many other fields, Twitter has emerged as a valuable data source for the analysis of complex real-time events, user-interactions, discussions and opinions.

A goal of this chapter, which forms a second arm of the literature review component of this research, is therefore to attempt to address the problems of the disparate publications and lack of canon within the field, by identifying commonalities and divergences among the available studies. This literature review attempts to include as many studies as possible that provide and assess empirical evidence on the use of Twitter in political discourse. Studies are presented here with regard to their research questions, their research design, and their mode(s) of data collection, data selection and findings. The aim is to compile a collection of empirical evidence that might allow for the detection both of patterns common across different studies, and of interesting observations, divergences or anomalies that were only evident in specific cases.

Since Twitter may be mentioned in the introductions or broad setting of many scholarly works on politics, the only included studies are those that directly or indirectly referred to (1) the use of Twitter by politicians and campaigns, (2) the use of Twitter by publics during election or *ad hoc* issue campaigns and (3) the use of Twitter by various users to comment on mediated campaign events - such as televised candidate debates, party conventions or election day coverage. This has the effect of excluding studies addressing the use of Twitter in some other areas of politics, e.g. studies on political marketing and political journalism are excluded. The corpus of literature included in this review was derived from keyword searches of scholarly journal indexes (matching terms like "twitter", "election" "politics" etc., in various combinations), which were then checked against the topic focus of this analysis. The cited sources in these relevant publications then provided a widening network of related literature, in which the published work of all included authors was examined for additional studies fitting the criteria but missed by the initial searches. The process yielded 116 studies, including case studies in 19 nations (some with multiple sub-national regional foci), from 2008 to 2015²⁰. In this chapter, these studies are sorted and grouped based on categorical characteristics: disciplines in which they are published; theories in which research questions are framed; methodological categories, including mode(s) of data collection, selection, and analysis; and geographic specificity of the research.

²⁰ A full list of constituent studies which form this corpus, ordered by year, is presented in the addenda for this chapter (Appendix A.3)
In attempting to set out a coherent account of existing literature, relevant to political discourse on Twitter, this chapter will better situate the case studies and methodological developments of this research, and provide insight into the developments and current state of the field. Additionally, such a review allows for some examination of - and the posing of questions related to - the ethics of Twitter-based research in this and other fields. While this is not a significant focus of this research work, ethical considerations of the use of publicly available social network data is also an underdeveloped theme in the field generally, and a review of the context and extent to which it is mentioned or discussed in this corpus may provide useful insights.

3.2 Related Literature Reviews

There is a small, but notable, set of related scholarly work that tracks and maps research on Twitter and which this contribution complements. For several years, internet researcher danah boyd [sic] has curated a growing Bibliography of Research on Twitter & Microblogging²¹ that has been published within disciplines such as "communications, information science, anthropology, sociology, economics, political science, cultural studies, computer science, etc." (boyd 2015). While this excellent resource contains over 260 scholarly works on Twitter, it remains an informal collection, spanning disciplines, and presented without review, meta-analysis or commentary on either the state of research, or its development over the period of the collection. In their article classifying Twitter-related academic papers, Williams, Terras and Warwick (2013) take an initial step to map the terrain of research regarding Twitter. The authors classified over 500 papers on Twitter and related microblogging research, published between 2007 and 2011. Through a content analysis of the papers' abstracts, Williams et al. found that the analysis of tweets, more than Twitter users or the technology itself, were the most common focus of these papers. Further, based on their analysis of abstracts, the authors were able to map the domains in which Twitter research was taking place, were able to generalise about the (often multiple) methods used for data analysis, and provided frequency counts of the most commonly used words within the abstracts.

These works provide a set of inroads to holistically mapping the state of academic research on Twitter as a whole. However, significant questions about the state of research generally, and about the type and extent of that which relates to politics, remain to be addressed. Building on the work of boyd and Williams *et al.*, this review provides new depth to these mappings by focussing more closely on the collection, use, and analysis of Twitter data by researchers, and on the inherent ethical dimensions of this line of research. By including an analysis of the full-text of the identified Twitter research corpus, this chapter provides greater detail on topics such as the exact volumes of Twitter data being gathered, the number of Twitter users potentially impacted by this collection, and the means by which researchers are acquiring their datasets – aspects which have not been addressed in other work to date.

²¹ http://www.danah.org/researchBibs/twitter.php

3.3 Disciplines, Theories & Methods

A review of the collected literature pertaining to the use of Twitter in politics (Appendix A: Chapter 3 Addenda) readily demonstrates that the substantial majority of research is empirical, descriptive and focussed on data collected directly by the researcher(s). A very small minority of studies explicitly seek to situate or contextualise their analyses or findings with regard to theoretical discussion at the level of the field as a whole, or even to reference theories or themes from political science as explanatory frameworks in their investigations. It may, of course, be argued that some works in this corpus contribute to or derive an explanatory basis from certain theories *implicitly*, but for clarity of this categorisation, only those studies where a theory is explicitly mentioned are included in Table 3.1 below, which is ordered by the academic discipline from which the theory originates.

Theory	Studies
Political Science	
Polarisation / Partisanship	(Bode et al. 2011; Feller et al. 2011; Gainous & Wagner 2014; Godnov & Redek 2014; Hanna et al. 2011; Kim & Park 2012; Mustafaraj et al. 2011; Plotkowiak & Stanoevska- Slabeva 2013; Smith et al. 2014)
Normalisation vs. Mobilisation	(Ausserhofer & Maireder 2013; D'heer & Verdegem 2014; Jürgens et al. 2011; Klinger 2013; Mirer & Bode 2015)
Social / Symbolic Capital	(D'heer & Verdegem 2014; Dimitrova et al. 2014; Gainous & Wagner 2014)
Deliberation	(Bruns & Highfield 2013; Kim & Park 2012)
Communications & Media Studies	
Media Framing	(Groshek & Al-Rawi 2013; Eun-Ju Lee & Oh 2013)
Agenda Setting	(Broersma & Graham 2012; J. H. Parmelee 2014; Wallsten 2014)
Limited Exposure	(Gainous & Wagner 2014; John H. Parmelee & Bichard 2013)
Transmediality / Intertextuality	(EJ. Lee & Shin 2014)

 Table 3.1:
 Theoretical framing in studies of Twitter in politics

Psychology	
Effect of exposure on behaviour / learning	(Dimitrova et al. 2014; Gainous & Wagner 2014; Hong 2013; EJ. Lee & Jang 2013; EJ. Lee & Shin 2014; John H. Parmelee & Bichard 2013; Cristian Vaccari et al. 2013)
Gratification	(Eun-Ju Lee & Oh 2013; John H. Parmelee & Bichard 2013)

 Table 3.1:
 Theoretical framing of studies on Twitter in politics (cont.)

3.3.1 Theories from Political Science

As evidenced above, contributions to a number of political theories can be found in the corpus, and this discipline, perhaps unsurprisingly, accounts for the majority of explicit theoretical framing in this literature. These include normalising or transformative mobilising aspects of Twitter for political systems and locales, political polarisation and partisanship, social and symbolic capital generation and transformation, and processes of deliberation among political actors.

The role of the internet as an arena fostering or hindering deliberative discourses on politics has been discussed in many studies and for many stages of the internet's technological development (Castells 2006; Andrew Chadwick 2006; Davis 1999; Rash 1997; Wilhelm 2000). Some researchers in the literature on Twitter have analysed the interactions of politicians and the public from this perspective; assessing whether and how politicians interacted with citizens within the medium. Since these interactions were both rare and largely one-way, the authors tend to conclude that Twitter is a tool not used by politicians for deliberative discourses (e.g. Kim & Park 2012).

Early discussion on the effects of the internet for political systems and interactions focused on the question of whether the internet would lead to a transformation of politics, in which new political actors rise to power, enabled through new opportunities of information distribution, political coordination, participation, representation, and new communication practices, or if the internet would become a normalised information environment, with the same communication modes and practices, where the same political actors would dominate the political sphere on- as off-line (Davis 1999; Rash 1997). Analyses of whether prominence for political actors in the traditional political environment would also lead to prominence on Twitter, reference this *normalisation vs. mobilisation* debate, and form the most numerous category in this sorting (D'heer & Verdegem 2014; e.g. Jürgens et al. 2011). These studies tend to show that, in most countries, the actors who are powerful offline tend also to dominate the political discourse on Twitter. A notable exception is in Germany where the *Pirate Party*, a party with no seats in the national parliament, is by far the most dominant party online (Andreas Jungherr 2014). With regard to communication practices, the findings are more mixed. Here, it appears that across countries studied, professional political actors do not radically break with their offline communication practices, but adapt these for Twitter, especially in response to crisis events (e.g. Ausserhofer & Maireder 2013).

Particularly in the United States, political scientists have discussed whether the political system has become more polarised over time. Researchers cite a number of reasons that might lead to an increasingly polarised electorate: highly partisan cues from political and social elites, an increasingly atomized social structure that might lead to people predominantly interacting with others like them, and a media environment with dedicated channels for partisan coverage (Prior 2013). Most studies addressing political polarisation on Twitter map the interactions of politically vocal Twitter users and try to determine whether such users tend to interact more frequently with users who share their political conviction. These studies are in the tradition of similar approaches in which researchers mapped linking practices between political blogs and found that political blogs tended to link to others sharing the same partisan convictions (Adamic & Glance 2005). The evidence with regard to Twitter appears to be more mixed than in previous studies of blogging. While there is evidence that users in the USA do indeed tend predominantly to retweet the messages of users who appear to share their political convictions, users also tend to interact through @mentions and conversational threads across partisan lines (Conover et al. 2012; Hanna et al. 2011; Mustafaraj et al. 2011; Smith et al. 2014). These findings lack useful comparative studies in other locales, and for other types of political system (proportional representation plurality vs. majoritarian binary party systems). Also, it is not at all clear if these findings true for specific interactions of users on a particular online service - do indeed speak to the wider polarisation debate as understood in political science.

Political uses of Twitter have also been discussed in the context of *social or symbolic capital* as understood by Robert Putnam (2001) and Pierre Bourdieu (1991), and in which social networks between people enable and foster political engagement. Although this concept and its use in political science has been disputed (e.g. Woolcock 2010), it has remained a useful explanatory framework for researchers focusing on the power of the internet to foster civic and political engagement (e.g. Rainie et al. 2012), or on the transformation of power through *fields* or mediated spaces (D'heer & Verdegem 2014). The view that digital tools enable new and meaningful interactions between people, which in turn might lead to new or increased civic or political participation, causes some researchers to expect Twitter to show similar effects (Gainous & Wagner 2014). The small number of studies which use social and symbolic capital theories do so for framing purposes, and do not engage in substantial detail with the theories, or establish any evidence at the micro level of individual actors that suspected demonstrations of capital generation or transfer are taking place.

3.3.2 Theories from Communications & Media Studies

Following from those studies that explicitly cite theory from political science, those that leverage communications studies concepts are the next most numerous. Here, researchers use concepts dealing with the agenda-setting and agenda-building processes of media coverage, the media framing of messages referring to political candidates, and the selective exposure of audience members to opinions concurring with their own.

When investigating Twitter's use in politics, some researchers have sought to examine the way Twitter messages might influence, and be influenced by, a wider media agenda. The studies included in this review that explicitly referred to *agenda-setting* research revealed that journalists increasingly incorporate Twitter messages by politicians and sometimes other users in their coverage of political topics, with the tweets either themselves the topic of a story, or the source of quotes or background information in on-going reporting (Broersma & Graham 2012; John H. Parmelee & Bichard 2013; Wallsten 2014).

Media *framing* refers to the different contexts in which information can be presented to an audience, and the potentially transformative effects on interpretation for the recipient, based on these presentations (Iyengar & Kinder 1987). Studies using this concept examined, for example, the specific words used in messages commenting on election results for candidates in the 2012 American presidential election and demonstrated an absence of partisan framing (Groshek & Al-Rawi 2013). Another study showed that messages by politicians framed with personalised information, led to different effects in political learning among recipients, when compared with similar but depersonalised messages (Eun-Ju Lee & Oh 2013).

One of the oldest theories of political contextualisation in communication research is that of *selective exposure*; the tendency of people to expose themselves to information in accordance with opinions already held by them. Media environments that allow audiences wider choice in what stories they may be exposed to seem to foster this tendency (Prior 2013). In this theoretical context, researchers have discussed the tendency of Twitter users to follow accounts of politicians belonging to parties they support, and generally to use social networking services to selectively find information in accordance with their existing political views (Gainous & Wagner 2014; John H. Parmelee & Bichard 2013). Those studies which have explicitly framed investigation of follower / following networks within Twitter using a

theory of selective or limited exposure, have based their findings on quantitative data derived from the Twitter API. This research necessarily provides a macro-level analysis and is not contextualised by any micro level or qualitative investigations of what the effective partisan exposure is for studied users.

3.3.3 Theories from Psychology

Another set of studies addresses the use of Twitter from a psychological perspective, though these form the smallest set of research in the corpus with an explicitly identified theoretical basis. Here, the studies principally relate to the motivations that lead people to interact with politicians online and their observable or quantifiable behaviour in doing so. These studies connect to a literature in Psychology on *Uses & Gratification* inherent in media usage (Rubin, 2009) as well as recent work on psychological mechanisms in the emergence of public opinion and political cognition (Lodge and Taber, 2013; Zaller, 1992).

Studies using theory from this discipline also focus on the effects exposure to tweets has on audiences with regard to political learning and the evaluation of political candidates and issue (Dimitrova et al. 2014; Gainous & Wagner 2014; Hong 2013; E.-J. Lee & Jang 2013; E.-J. Lee & Shin 2014; Eun-Ju Lee & Oh 2013; Cristian Vaccari et al. 2013).

3.3.4 Methods

While the studies discussed above share the theoretical discussion they address in their research, there are also studies that approach the use of Twitter in politics with specific shared methods using rich data collected on Twitter to perform quantitative analyses. Here the two key approaches evident are Network Analysis (NA) and automated Sentiment Analysis (SA).

Studies in which a method of Twitter data analysis is explicitly stated are sorted into these two category groups in Table 3.2, below. A small number of studies utilise both methods.

Method	Studies
Network Analysis	(Ausserhofer & Maireder 2013, 2013; Bode et al. 2011; Boutet et
	al. 2012; Bruns et al. 2012; Bruns & Burgess 2011; Bruns &
	Highfield 2013; Conover et al. 2012; D'heer & Verdegem 2014;
	Feller et al. 2011; Godnov & Redek 2014; Groshek & Al-Rawi
	2013; Hanna et al. 2011; A. Jungherr & Jurgens 2014 p. 2015;
	Jürgens et al. 2011; Kim & Park 2012; Larsson & Moe 2011; Lin
	et al. 2014; Mascaro & Goggins 2012; Metaxas & Mustafaraj
	2012; Mustafaraj et al. 2011; Plotkowiak & Stanoevska-Slabeva
	2013; Shamma et al. 2010; Smith et al. 2014)
Sentiment Analysis	(Dang-Xuan et al. 2013; Diakopoulos & Shamma 2010; Lin et al.
	2014; Plotkowiak & Stanoevska-Slabeva 2013; Stieglitz & Dang-
	Xuan 2012)

 Table 3.2:
 Methods utilised in studies of Twitter in politics

In studies interested in the analysis of social networks, the use of Twitter conventions - such as follower / followee relationships, and interactions between users through @-mentions, favourites or retweets - are used to construct a social network in which the users form *nodes* and their interactions *edges* in a network graph. Studies, using Network Analysis methods usually focus on questions of who interacts with whom, grouping actors according to type; professional politicians, citizens, media, etc. (Ausserhofer & Maireder 2013; e.g. Bruns & Highfield 2013; D'heer & Verdegem 2014). This research uses a disparate set of NA methods, from simple quantification of nodes/edges in a set, to complex in- and out-group interaction measurement. Those studies that make most detailed use of Network Analyses in explanatory investigations of political discourse on Twitter tend to be the work of academics and institutions with a strong Computer Science focus, which is the origin of this method set. Research based on applying NA techniques to large scale data sets tends to have as its investigative focus a set of broad descriptive questions about the overall patterns of participation in political discourse related to a locale (e.g. Slovenia in Godnov & Redek 2014), event (e.g. Queensland floods in Bruns et al. 2012), or election (e.g. Belgian local elections of 2014 in D'heer & Verdegem 2014).

Studies using automated sentiment analysis also use textual data provided by Twitter, which undergoes content analysis in the form of attributing positive or negative score to each tweet based on the occurrence of sentiment-indicating keywords (e.g. "hate", "like", "approve", etc.) in each tweet. Various approaches are used in defining the keyword set and attempting to account for conversational context, modifying phrase combinations, etc. (Bermingham & Smeaton 2011). Researchers in this section of the corpus have focused on the sentiment expressed in messages referring to specific political candidates (@-mentioning them), to interactions by specific users (coding them and examining all tweets by user category), and referring to events (by selection and analysis of tweets containing specific keywords or hashtags). While SA methods are acknowledged in the literature to be imprecise and to have difficulty accounting for certain user intents (e.g. sarcasm, irony), these methods have been found to be useful in indicating aggregate sentiment, especially per actor group for specific electoral events (Bermingham & Smeaton 2011; Groshek & Al-Rawi 2013; Stieglitz & Dang-Xuan 2012).

3.4 Data Collection & Selection

Grouping the corpus of studies included in this review by their research design results in four categories: studies conducting specific experiments, studies using quantitative data not derived from Twitter itself (e.g. surveys of users), studies using qualitative data (e.g. interviews or ethnographies), and studies using quantitative metadata, derived from Twitter and demonstrating user activity (e.g. retweets, @-mentions, URL inclusions). Of these, the last is by far the largest category, to the extent that almost every study necessarily includes Twitter metadata. Since the other categories contain a small minority of the studies, only these are detailed in Table 3.3, below, which outlines those research approaches. It is important to note that while these studies in this table utilise the research approaches shown, almost all also utilise Twitter trace data.

Research Design	Studies
Qualitative (Interviews, Case Studies, Ethnographies)	(Abroms & Craig Lefebvre 2009; Anstead & O'Loughlin 2012; A. Chadwick 2011; Andrew Chadwick 2013; Enli & Skogerbø 2013; Grusell & Nord 2012; Andreas Jungherr 2014; Nilsson & Carlsson 2014; J. H. Parmelee 2014; John H. Parmelee & Bichard 2013)
Quantitative <i>(Surveys)</i>	(Bekafigo & McBride 2013; Dimitrova et al. 2014; Gainous & Wagner 2014; Eun-Ju Lee & Oh 2013; John H. Parmelee & Bichard 2013; Pew Research Center 2013; Rainie et al. 2012; Smith et al. 2014; Strandberg 2013; Cristian Vaccari et al. 2013)
Experimental	(EJ. Lee & Jang 2013; EJ. Lee & Shin 2014; Eun-Ju Lee & Oh 2013)

 Table 3.3:
 Research approaches utilised in studies of Twitter in politics

Some studies used surveys to investigate specific quantifiable factors, such as assessments of how many people were using Twitter politically (Rainie et al. 2012; Smith et al. 2014) or for the identification of characteristic variables showing a statistical relationship with political Twitter use (e.g. Bekafigo & McBride 2013; Dimitrova et al. 2014; Cristian Vaccari et al. 2013).

Studies based on surveys allow researchers to compare their findings on political Twitter use with established research on political participation elsewhere, such as models of voluntarism (Rainie et al. 2012). This potentially allows the identification of characteristics that Twitter use for political discourse shares with other technologies for political participation, as well as identifying factors that are solely characteristic for the political use of Twitter. Some important challenges are posed by this research approach, however. As Twitter is still used by a minority of citizens in any given country, varying widely from one to another, a true sample of a population would have to include a very large number of respondents to allow for any meaningful discussion of political uses of Twitter. In the present literature, there are only two studies which employ very large survey samples to attempt to draw meaningful conclusions for a national population's participation: (Gainous & Wagner 2014; Smith et al. 2014).

Certain studies try to address this issue by selecting Twitter users based on their use of political hashtags, or their decision to follow specific identified political actors. Once selected in one of these ways, the researchers invite this politically active population through Twitter messages to participate in a survey (John H. Parmelee & Bichard 2013; Cristian Vaccari et al. 2013 both use this method.) While this selection adjustment might give some indication of highly active Twitter users for the studied discourse, this practice is of course heavily dependent on self-selection from the respondents, and might thus lead to indeterminably biased results.

A small but recently growing number of studies, developed by Lee *et al.*, and others, are using controlled experiments to measure effects on recipients following the exposure to Twitter feeds containing political information (E.-J. Lee & Jang 2013; E.-J. Lee & Shin 2014; Eun-Ju Lee & Oh 2013). Results from these studies to date shows the considerable potential of these experimental methods to provide data capable of identifying mechanisms common to political discourse on Twitter and in other media. These research designs then situate Twitter in the theoretical context of established political psychology.

The overwhelming majority of studies, of course, focus on digital metadata from Twitter. These metadata document the activities of users on digital services, by identifying known interaction types with the service (e.g. retweeting). Analyses based on this data type are thus analyses of readily available variables provided by Twitter (via APIs), and stand in clear contrast to research in which data was externally collected to measure and control for specific phenomena, such as in the surveys or experiments. The studies vary in complexity from simple manual transcription of the number of messages and followers of politicians' Twitter account (e.g. Golbeck et al. 2010) to the extensive collection of structured data and metadata through Twitter's APIs (e.g. Ausserhofer & Maireder 2013). This metadata (occasionally termed digital trace data for its description of human behaviour mediated through digital technologies), has been at the centre of most research on the political usage of Twitter, since its inception. In principle, every interaction of users with Twitter is documented and, within limits of a specific API access level, is accessible for researchers. This opens up interesting areas of potential for researchers interested in the use of digital tools, the effects of exposure to certain types of messages and the flow of information through networks constructed through manifest interactions between Twitter users. Twitter offers access to a data type that is representative of the type of social network interaction data collected by other online services like Google or Facebook, but in contrast to these services, Twitter allows researchers comparably comprehensive access to its data²². Thus, research on political behaviour on Twitter, using metadata, may show patterns generalisable beyond it, offering opportunities for the development of general digital methods for use in research on mediated political activity online (Lazer et al. 2009; Rogers 2009).

The research potential in such relatively easily accessed, rich, structured data is likely the root cause of the readily identifiable bias toward highly quantitative data collection and analysis in the corpus, with only a small minority of studies taking qualitative approaches. Among these, detailed ethnographic case studies (e.g. Andrew Chadwick 2013), content analyses of sets of tweets (Ausserhofer & Maireder 2013; Graham et al. 2013; Jackson & Lilleker 2011), and interviews with political actors and journalists on their use of Twitter and their assessment of its effect on politics (Nilsson & Carlsson 2014; J. H. Parmelee 2014), have proven the value of informing the insights of quantitative methods with qualitative detail. The latter has been especially useful in providing verification of suspected user behaviours and conventions underpinning observed patterns at a meso and macro level in quantitative studies, particularly for the case of electoral events.

²² A discussion of the data structure, sources and limitations inherent to Twitter's API is presented in the Methodology of this work (Chapter 4)

3.4.1 Data Collection Methods for Twitter

As evidenced in the corpus for review, studies on the use of Twitter in politics rely greatly on metadata variables, parsed from vast volumes of data retrieved automatically from Twitter's APIs, usually with various third-party software solutions. It is important to characterise these collection approaches along two categorical divisions, therefore; the specific API being used (which limits the structural nature and temporal window of data available), and the software employed to query it (which may further limit the data and metadata received, parsed and stored, based on its programming).

Tables 3.4 and 3.5, below, set out those studies in the corpus that specify the API or software used in the collection of their working dataset, respectively. It is important to note here that though almost every study (114 of 116 reviewed) necessarily had to collect data from Twitter, *the majority do not report how this was done*, and some that do report only the API used and not the particular software or study-specific scripts that were used. This highlights a crucial (and problematic) attribute of the state of Twitter research, at least for the study of political discourse – that researchers, editors and journal reviewers are indifferent to how data was collected and are perhaps unfamiliar with the important connotations the various collection methods have for the completeness and biases within the resulting dataset.

Twitter API used:	Study
REST	(Anstead & O'Loughlin 2012; Boynton et al. 2014; Dang- Xuan et al. 2013; Diakopoulos & Shamma 2010; A. Jungherr & Jurgens 2014; Andreas Jungherr 2014; Mascaro & Goggins 2012; Shamma et al. 2010)
Streaming	(Ampofo et al. 2011; Bode et al. 2011; Conover et al. 2012; Groshek & Al-Rawi 2013; Hanna et al. 2011; Hemphill et al. 2013; Lin et al. 2014; Metaxas & Mustafaraj 2012; Mustafaraj et al. 2011; Otterbacher et al. 2012; Roback & Hemphill 2013; Cristian Vaccari et al. 2013)

 Table 3.4:
 Data collection API used in studies of Twitter in politics

Table 3.4: Data collection API used in studies of Twitter in politics (cont.)

API mentioned, but not	(Ausserhofer & Maireder 2013; Elmer 2013; Gilmore 2012;
specified	Golbeck et al. 2010; Plotkowiak & Stanoevska-Slabeva 2013;

Cristian Vaccari et al. 2013; M. Vergeer et al. 2013; Maurice
Vergeer et al. 2011; Maurice Vergeer & Hermans 2013)

As briefly detailed in the introduction to this work, and explored in greater detail in the next chapter, the Twitter REST API provides structured data results for searches based on a particular tweet, user, timeline, locale, hashtag or keyword, but is limited temporally and in completeness of the result retrieved; only data from the last 6-9 days are available via this API, and an unspecified portion of possible results for a given search are withheld²³. The REST API is therefore suitable for post-hoc searches based on known data fragments (e.g. retrieving a recent tweet based on an ID) or post-hoc updates of existing data (e.g. retrieving favourites metadata for a stored recent tweet to assess the number of favourites it received since it was first collected).

The Streaming API, by contrast, provides Twitter data in real time, requiring a constant monitoring, collection and storage solution to gather relevant tweets and metadata as they occur. This API is limited in not allowing the same post-hoc updating afforded by the REST API, but does not have a content-limiting factor and in theory allows for the collection of all public real-time data from Twitter, so long as the criteria for what's being collected mandates a data stream capable of being collected and stored fast enough with given hardware. The total output of the Streaming API is often referred to in the literature and documentation as the "Twitter Firehouse". Real-time collection via this method requires suitable hardware and software and a defined set of collection parameters, set up in advance.

While many of the studies did not specify which of the two APIs were used in their data collection, some – seen in the "Unspecified" section of Table 3.4 – indicated that an API was used but did not say which one. In all cases, except Vaccari *et al.*, the type of data described in the study, or the software mentioned, implies that the Streaming API was used, but for clarity and in keeping with the convention of earlier classifications to use only explicitly stated information, these are separately categorised.

Similarly, the affordances (such as data storage, customisability of search criteria, etc.) of a given software option for data collection will necessarily limit the data available and may bias the sample. In many cases, this too was not specified, but those studies which did so are

²³ Developer documentation of the APIs states that for some query endpoints (e.g. search/statuses) which are very likely to be used for retrieving tweets, the result sets are restricted to a 10% fixed sample of the total matching set (Twitter Developers 2015). Other endpoints are not explicitly described as having a set limitation, though authors in the research literature speculate that this restriction is more widely in place (Bruns & Stieglitz 2012)

categorised in Table 3.5, below. In the majority of cases that did not make their data collection software explicitly clear, details of the metadata available, limitations on the study, and suggestion that outages in collection created gaps in the dataset, point to the use of *yourTwapperKeeper* (an open-source tool well-established in the early literature). However, for clarity only those studies with an explicit identification of tool are presented here.

yourTwapperKeeper	(Bruns et al. 2012; Bruns & Burgess 2011; Bruns & Highfield 2013; Bruns & Stieglitz 2012; Larsson & Moe 2011, 2014; Moe & Larsson 2013)
NodeXL	(Kim & Park 2012; Smith et al. 2014)
DiscoverText	(Groshek & Al-Rawi 2013)
Topsy	(Bekafigo & McBride 2013)
Tweet Archivist	(Boynton et al. 2014)
DataSift	(Hawthorne et al. 2013)

 Table 3.5:
 Data collection software used in studies of Twitter in politics

The very slight extent to which researchers make explicit the software in use to collect their data demonstrates how far the field has yet to go in developing even agreed standards of reporting on data collection. It is possible, though unlikely, that some researchers are coding scripts for their data collection needs, specific to their project and not sharing these in the public domain. Even where this may be the case, such *ad hoc* bespoke approaches to data gathering leaves little ground for either development of standards or completeness of insight for other researchers wishing to undertake comparative work.

An agreed specific method for gathering data is then very far off, with the most popular software cited, yourTwapperKeeper²⁴ (yTK), functionally obsolete having not been updated for several years and being based on a now deprecated version of the Twitter API, which did not facilitate the same degree of metadata available currently. The use of yTK and other open-source solutions for Twitter data collection has been bolstered by extensive discussion of these in the early literature, and by manuals and provision of accompanying scripts by some researchers (e.g. Bruns & Stieglitz 2012). Commercial software services, such as *Topsy* and *DataSift* have acted as resellers of Twitter data, leveraging a direct relationship with the

²⁴ https://github.com/540co/yourtwapperkeeper - last commit: 24 May 2013

service to obtain and sell datasets based on more elaborate search criteria than would be possible via API access. Aside from the prohibitive cost likely being a substantial limiting factor for the use of these services in collection of data by academics, recent moves by Twitter to end their relationship with certain resellers, and to directly sell datasets has made many commercial providers increasingly obsolete in this field (Halstead 2015). It is very interesting to note that, despite indications by Twitter that research data may be provided to academics free of charge (Scola 2015), no study in the literature shows such collaboration. Neither does any study source Gnip²⁵, the remaining professional Twitter data seller affiliate. It is very likely the case that, while either of these channels might guarantee a substantial and complete archive of tweets and metadata on a topic, the terms and conditions of the publication of such data, and especially the requirement by Twitter to approve any results and analyses before publication, would make research based on these sources unfit for the academic peer review processes.

There remains then, a distinct lack of a reliable, affordable and usable tool for researchers in the field (or related fields where Twitter data collection is an important component of research).

3.4.2 Data Selection

Not only do the studies reviewed show considerable variance in their methods for collection of data, but so too for their criteria for selecting which data to collect. Broadly, there are two approaches in use: the first – and very substantially the more prevalent – is to collect all tweets containing one or more relevant hashtags, usually in real time, and to subsequently analyse the resulting data set as an archive of discussion on that topic. The second, less-used approach is to identify a cohort of users to track – either by their function (e.g. professional politicians, media actors), their "bio" description (e.g. women declaring an interest in current affairs), or an action based on detectable trace data (e.g. all users who followed a person of interest to the study, or who used a particular hashtag with a certain frequency). The total tweet output from this cohort then forms the dataset for analysis.

Both approaches carry problems of bias, the extent of which may be undetermined. Data selection based on the inclusion of a hashtag (or other keyword) remains the principal method of most studies, especially around electoral events (e.g. Bruns & Burgess 2011; Larsson & Moe 2011), and its utility is easy to appreciate; real-time data collection via the Twitter Steaming API (dubbed the *firehose*) requires a set of search markers to gather by, and

²⁵ https://gnip.com

hashtags and keywords suit this method where they can be decided on or predicted in advance. The obvious – and substantial – issue is, of course, the extent to which hashtags or keywords actually appear in the totality of conversation on a topic. It is likely that user behaviours and conventions underpin the usage of hashtags especially, and that certain tweet contexts may be less likely to contain them – replies to a tweet that already contained a given hashtag, for example, may be unlikely to repeat it, yet likely will form part of the Twitter conversation on that topic and be missed by this selection process.

Keywords are likely to exhibit similar limitations and patters of usage in conversation, though their nature – as part of regular text communication, rather than as deliberative markers somewhat set apart from it, as hashtags are – make them less likely to be affected by social convention within the medium. Of these two searched-for inclusions in tweets to select, it is hashtags that form the significantly greater bulk of the corpus selection methods, likely because the notion of their inclusion representing a deliberative, rational marking of context by the user is a powerful one when considering how to classify data for inclusion. Evidence is seen in the literature of researchers' knowledge that hashtags are imperfect as a selection criteria and reliance on the explicit tagging aspect of their use to offset those caveats (Ausserhofer & Maireder 2013; e.g. Bruns & Burgess 2011). Such attribution of the deliberative function of the hashtag as a positive trait for data collection belies the associated bias inherent in a feature that requires actors to understand these usage conventions; if hashtags are predominantly and most appropriately used by those who are best acquainted with the norms of the medium, selecting on them biases the sample toward savvy, longerestablished actors.

While hashtag sets may therefore skew toward too-specific data which misses contributions from less experienced users, or those in the latter stages of a conversational thread, selecting on keywords is likely to inflate and dilute a sample as an initially indeterminable volume of false-positives will be accumulated.

Both of these content-based selection methods must necessarily be set up in advance for collecting real-time data, which may account for their widespread presence in the Twitter politics literature where terms and hashtags (especially around electoral events) are generally well-agreed and predictable in advance (e.g. Bruns et al. 2011; D'heer & Verdegem 2014). This is less the case for unplanned events, such as disasters, scandals or other breaking news, analysis of which is often retrospective (Bruns et al. 2012; Tonkin et al. 2012; Vis 2013).

The second broad approach is the collection of all messages posted by a preselected cohort of users. The criteria for the selection of relevant users vary in the studies reviewed. Some

collect all messages by users identified by their official function – for example politicians, activists or journalists (Ausserhofer & Maireder 2013; e.g. Bruns & Highfield 2013). Other studies collect all messages by users who follow or mention specific accounts (Bekafigo & McBride 2013; C. Vaccari & Valeriani 2015), or who use specific hashtags or keywords (Ausserhofer & Maireder 2013; Jürgens et al. 2011; Lin et al. 2014).

These selection decisions are accompanied by similar trade-offs to those discussed above, and as in the cases of hashtag or keyword based selections, there is no systematic analysis method, or comparative study basis available if different selection criteria produce datasets that differ to an unknown degree. It seems obvious then, that researchers need, at the very least, to more fully understand the biases involved in order to work toward producing a set of best practices for collection and selection.

3.5 Ethical Considerations

In 2010, the United States Library of Congress announced the creation of an archive of Twitter's public content, recognising its importance as a source of mediated public discourse and *zeitgeist* record; "*Every public tweet, ever, since Twitter's inception in March 2006, will be archived digitally*..." (Raymond 2010). This announcement clearly validated the research importance of Twitter, but also prompted concerns about creating a permanent archive of tweets, and whether such a proposal was properly aligned with users' understanding of how the platform worked and their privacy expectations (Scola 2015; e.g. Zimmer 2010).

Particularly relevant are questions regarding how academic research on Twitter has proceeded thus far, such as:

- What disciplines are engaging in Twitter research and with what amount of scrutiny of ethics in research design and collection methodologies?
- What is the nature of research questions being investigated, and what specific data and metadata is collected in these processes?
- How does (or can) consent operate in these studies and are users given any opportunity to opt-out?

In the reviewed corpus, only a small number of studies even acknowledged the existence of ethical issues around using Twitter data for analyses (e.g. Ausserhofer & Maireder 2013), opting only to mention this area as one of a series of caveats to consider for this and future work. Only one study tackled the issue with any depth, attempting to engage with a wider literature around ethics in big data and metadata research; Anstead and O'Loughin (2012) question the assumptions made by researchers that user agreement with Twitter's terms and conditions (which state that tweeted content may be public and searchable) indicates a level of consent and may be considered to extend to user knowledge of processes involved in academic study that may capture their Twitter communication, archive it, and perform analyses for publication.

It is clear from the current state of research that the ethical considerations for the field are as yet in a very early stage of development and that difficulties exist around informed consent in particular, because of the fragmented nature of data collection and selection, the quantitative focus on large-scale data-sets and the comparative lack of qualitative personalised research contexts in the field.

3.5.1 Ethical Practice for Internet Research

While the literature immediately pertinent to the field of Twitter political discourse research shows the aforementioned lack of focus on ethical questions, a wider review of literature related to Internet research more generally might be useful in deriving principles of best practice. In *Digital Media Ethics*, Ess (2010) provides an overview of the issues around the protection of individuals' privacy in research that uses online (and especially social networking) data. Ess suggests that an imperative exists for researchers to carefully assess the degree to which a studied site or service may be considered by its users to afford privacy, whether or not that expectation is well-founded. Thus, it remains a possibility, particularly for communicative media that promote content and status sharing, that users may perceive their interaction to be limited to only friends/followers, when this is not the case.

This notion is present in the earlier work of Moor ('Towards a theory of privacy for the information age' in Spinello & Tavani 2004), which highlights the difference between *normative privacy* and *natural privacy*. The former is a well-founded expectation of privacy based on law or custom (e.g. one can expect privacy in their own home, or of personal information held by a hospital), while the latter is a reasonable expectation based on determination that privacy is likely, despite explicit laws or the presence of others to enforce it. Such expectations of natural privacy may exist based on perceptions about the location or conventions for an interaction or event – e.g. an activity might be expected to go unobserved by virtue of taking place in an extremely remote location; no law guarantees this outcome, though this is reasonable to assume based on likelihood. In cases of natural privacy were breached – despite the remote location of the previous example, another person may stumble upon the activity. Moor therefore terms the (usually accidental) failure of natural privacy a '*violation*', emphasising the degree of difference in seriousness.

Thelwall (2010) develops the notions of normative and natural privacy from Moor's work and applies it to specific cases of online interaction and activity. In the case of the public web, he contends, it is unreasonable for users to expect contributions made to be totally protected; there is no normative privacy to be violated in this space. Thelwall sets out a case that "researching the public web should *not* be subjected to ethical scrutiny for privacy concerns", though he allows that natural privacy breaches may occur when contextual integrity is broken- e.g. information posted with the intention of being read by friends is used in a different context, such as academic research. Like Ess, he suggests that researchers be cognisant of the consensus view among users of how their data for a particular service may be accessed and whether breaches may be felt by its inclusion in research. Both authors agree that activity by the researcher to disseminate results and contribute to a public awareness that academic research related to online data exists, is an important practice for informing users.

Zimmer (2010) similarly calls for greater engagement from scholars with ethical considerations of their work in this field, in his blog post addressing 'big data harvesting' in general, by academics. He suggests that mechanisms of consent-seeking, that aim to make the public aware of the nature of this research and offer the possibility of withdrawing content, are necessary, and that qualitative research can help to better frame the individual and their privacy in studies that often operate at depersonalised macro levels.

3.5.2 Ethical imperatives for this research

Drawing on the concepts and principles around data privacy outlined here, two imperatives in particular are clear: (1) minimising the extent to which user data at the level of the individual is re-published within academic work where possible, and (2) engaging with users to promote awareness that research is occurring and that their data may be used. Both imperatives act to minimise the possibility of contextual privacy breaches by limiting the incidences with which personal data is published in altered contexts from the users' original intentions, and limiting the extent to which any academic use of the data would constitute a surprise and perceived breach to a user who had expectations of natural privacy.

As the next chapter will outline, the methodology for this study seeks to preserve user anonymity in so far as is possible, including specific mention of profiles, particular tweets, or other identifiable information only with the express consent of the users involved. The selection method for data collection in these case studies operates on hashtags, the use of which likely already indicate an extra action of deliberate marking of contributions as 'more public' than a user's immediate network of followers, and so the imperative to preserve privacy for these may be less than a tweet that does not contain them. However, since part of the data necessarily contains conversational interactions, which are directly addressed to particular users, an ambiguity still exists as to the privacy expectation for at least part of any such set. The qualitative elements of this research contribute to a greater awareness among users of the nature of studies being conducted, as they specifically contain formalised systems of obtaining informed consent, during which the research and data used is discussed in detail.

3.6 Conclusions

This chapter offers a comprehensive overview of the literature across several academic disciplines, as it pertains to political discourse on Twitter, and its analysis. Its contribution to that literature lies in the identification of commonalities in theme, theory, and method across the corpus of work, and the following may be concluded from this analysis:

Theoretical framing for research questions and analyses are explicit only in a minority of works, and amongst these theories from political science are the most prevalent – especially theory related to the normalisation vs. mobilisation debate, which has been established as a useful frame of reference for studies of participation in the discourse across many countries and regions. Theory from Communications and from Psychology is also evident in the literature, though narrowly centred on certain sub-topics and specific researchers. Many works in the literature operate without a wider reference to theory in the field or beyond.

Methods of analysis for Twitter-mediated political discourse fall broadly between Network Analysis techniques – in some places taken from highly specific computer science approaches – that focus on the interaction and interconnectivity of users in a given conversational network on Twitter, and Sentiment Analysis that seeks to provide a useful, quantifiable measure of a qualitative aspect of the content being studied. A small number of studies employ both methods, but the majority focus on network analytics to describe cohort interaction, communicativity and reciprocity in geographic locales, around events and around elections.

The literature shows little development of good practice around the disclosure and detailing of how data used in the studies is collected or selected. Collection methods are often not documented, or are dependent on demonstrably outmoded or obsolete software approaches for retrieval via Twitter's APIs. Considerable work is required to develop a standard of reporting and possibly a standard collection method for the field, and there is no agreement evident on a particular approach to gathering data, save that data provided directly by Twitter does not seem to form any part of the literature, likely because of attached conditions.

Data selection via hashtag is the dominant method for gathering datasets for analysis, though the sources of bias implied by this method are poorly understood and not investigated in any detail in the literature. There is no best practice evident either for content-based (hashtag / keyword) searches in real time, or for user cohort based collection, both of which operate in an *ad hoc* form, despite some acknowledgement of caveats required for the completeness and bias of the resulting data.

Finally, ethical considerations are almost wholly unexplored in the literature and calls for their development have not yet been met with meaningful analysis or discussion. Rogers' (2009, 2010) calls for the development of digital methodology specific to new fields such as Twitter is occasionally referred to among some studies seeking to begin to build a foundation for Twitter analysis, but the questions around informed vs. implied consent remain unanswered in the field's specific literature, though crucial to its maturing as a social science discipline. Examination of the wider literature concerning the ethics of Internet research points to the possibility of user expectations of privacy for their online data, including contributions made to an ostensibly public medium such as Twitter. This research acts, based on principles of ethical practice suggested for the field, to minimise the likelihood of breaches of user privacy and to fully inform participants of how their data may be used and how research such as these case studies take place more generally, contributing to greater public understanding of an academic context for their online data.

Chapter 4

Methodology

4. Methodology

4.1 Introduction

This chapter introduces the methodological framework of this research – a *Mixed Methods* approach based on dominant practices in the social sciences for the staging and structuring of such work. The chapter is necessarily lengthy, as this research is itself concerned with responding to the disparate and often *ad hoc* methodology in the specific field of Twitter discourse analysis, proposing and testing frameworks in an attempt to assess their suitability and efficacy, and contributing to knowledge both by this development and by a focus on complementary qualitative insight. It is required then, to outline several aspects of the methodological structure of this research:

- Digital methodologies related to the selection, capture and storage of large amounts of Twitter data, via automated means
- Quantitative methodologies for the statistical and (network) structural analysis of Twitter datasets
- Qualitative methodologies for the selection, capture and analysis of information from identified actors in the case study
- The design of a pilot study to establish the research case and to define key variables and coding schemata
- A Mixed Methods design appropriate to the useful combination of the insights gained from the practices above, toward informing effective methodological combinations

The affordances of Twitter's Application Programming Interface (API) and the beginnings of Twitter research in the field of Computer Science has tended to greatly skew research of the medium toward large-scale quantitative analyses (See: Chapter 3), and this study employs, combines and evaluates many of these existing approaches. However, researchers in the social sciences have established the benefit of informing the conclusions of quantitative studies - especially those descriptive of phenomena at a macro level - with the depth of insight obtained from qualitative investigation (Patton 1990). Studies have existed for many decades, in which qualitative and quantitative data is gathered simultaneously, on the same research question, but a formalised approach to mixed methods as a distinct research design is more recent (Creswell 2014; Creswell & Plano Clark 2011) and is considered highly suitable for the specific goals of this study.

4.1.1 Big Data: Deep Data & Surface Data

In framing the specific context of computational methods within the Social Sciences, it is helpful to have a broad typology for the kinds of data that social science researchers collect using automated or computational means. Manovich (2011) proposes two types of such data, built on the paradigms of quantitative and qualitative categorisations. One is *surface data*, which entails data collected from many actors/users, and where the collection method explicitly limits it to surface-level information. The other is *deep data*, which is more detailed and nuanced, often derived from a small subset of studied participants, and may require subjective qualitative evaluation to categorise.

Deep data collection and analysis has the goal of gaining a specific understanding of a very small sample of data, with a tendency toward *'why'* questions shaping its research usage. This type of investigation tends to use qualitative methods such as interviews, participant observation and focus groups to gather data on an event.

Surface data, in contrast, seeks to establish or verify patterns across large datasets and tends to be employed widely in (and derive methods from) Computer Science, and related fields such as economics. This research posits *'what'*, rather than *'why'* questions and results in generalisable conclusions based on analyses of simplified variables. Frequently, the methods related to surface data are greatly less time-intensive than those of deep data, and may allow for greater agility and reactivity during extended research endeavours (Manovich 2011).

Big Data, though widely used in the literature and in journalistic and other extra-academic sources, is a poorly defined term that broadly combines both surface and deep data. A widely cited article by boyd and Crawford (2012) defines big data, somewhat vaguely, as "a cultural, technological, and scholarly phenomenon", proposing it as a type of data that moves beyond the scope of traditional methods of capture, storage and analysis, necessarily requiring new approaches to collecting, indexing, retrieving and managing it. Mayer-Schönberger and Cukier (2014) assert a characteristic *Three Vs*: volume, velocity and variety – the last of which is an indication of the combinational nature of some sources, where surface quantitative data is combined or structured with other complex information that requires qualitative methods to categorise, parse, or assess. These authors specifically cite Twitter data as an example of this.

Though big data is considered by some authors to consist only of surface data (Kitchin 2014; e.g. Uprichard 2013a), the literature review (Chapter 3) of this work shows that computational researchers in the social sciences have been engaging with Twitter's big data outputs at more than a surface level and using qualitative means in many cases. The deep data aspect of these studies is focused on the qualitative analyses of the content of collected tweets. These methods usually consist of content coding, but computational methods (e.g. sentiment analysis) are also widely evident, where qualitative factors are identified via language markers (keywords, phrasings, etc.) and very large datasets are automatically coded. Such an approach is exemplary of the new methodologies emergent in response to the research possibilities of this data type.

There is also evidence in the literature of big data use by social scientists to provide a starting point to identify meaningful patterns, in the bottom-up generation of new hypotheses. Using abductive reasoning on Twitter datasets, researchers have identified patterns in order to form their new hypothesis before formally proceeding with deductive theory construction and inductive empirical testing (e.g. Ausserhofer & Maireder 2013; See also: Dixon 2012). Despite the problems associated with capture, storage and accessibility, and requirements for new ways of dealing with the data, Twitter-derived big data has clearly been changing the way researchers understand and experience knowledge (boyd & Crawford 2012; LaValle et al. 2010).

While contextualisation of collected Twitter data is achieved in some studies through the use of selection methods (e.g. choosing only those tweets from a specified geo-locale or containing a specific hashtag or keyword), or post-hoc content analysis (e.g. coding tweets based on language used or media attached), there is little evidence in the literature of qualitative methods being used to inform understanding of the social context of the creation of the tweets. For example, many studies acknowledge the limitations of selection based on hashtags (e.g. D'heer & Verdegem 2014) but do not take action to collect contextualising information about how hashtags are used, from outside of the Twitter dataset itself. The focus, therefore, is almost uniformly on the information available in the Twitter dataset(s) in use, and not on the collection and analysis of complementary data from other sources external to it. This research seeks to broaden that scope, by setting not only to collect and analyse big data from Twitter, but qualitative data derived from a case study of those actors who created it. The use of this complementary, but structurally very different, strand of data in this research necessitates a research design that allows for the concurrent collection and analysis of qualitative and quantitative data, seeking insight in their combined explanatory potential. A mixed methods approach was therefore deemed appropriate.

4.1.2 **Opening the "black box"**

Researchers in the literature (Borra & Rieder 2014; boyd & Crawford 2012; Burgess & Crawford 2011) have drawn attention to issues arising from the use of software to extract and study data from (mostly proprietary) social media platforms that enable and shape communicative expressions and interactions of hundreds of millions of their users. These issues include the methodological, epistemological, ethical, political, and legal dimensions of research on what is increasingly referred to as Big Data, amongst which Twitter research is often the exemplary case. Critiques from these scholars often centre on the "black box" nature of both a medium and its associated analytical tools. The extent to which unseen algorithmic rules in social media service like Twitter determine what is seen, and by whom, is a concern for researchers seeking to understand the discourse within these, its context(s) and its participants. Another "black box" arises with the tools used to acquire data for analysis; many commercial solutions (and even some open-source projects) that focus on the collection of Twitter data for research purposes, are deliberately or unintentionally vague in their description of how the data was collected, what (if any) changes have been made to its structure, and how complete it is. It is perhaps little surprise then, that trends across the literature (Chapter 3) show a correspondingly poor state of reporting and discussion of Twitter data collection strategies, their limitations, and their effect on the completeness of the set for a topic it pertains to represent. It is likely in many cases that the extent of technical knowledge required to engage with the actions related to the acquisition, storage, structuring and retrieval of Twitter data prevents researchers in the Humanities and Social Sciences from fully understanding the "black box" of collection, limiting their reporting of process and causing their reliance on simplified methods and tools already prevent in older literature.

Scholars such as Rogers (2009, 2010) urge researchers to pay close attention to how data objects are defined, and to how they are affected by the processes involved in their capture and analysis. Rogers calls for the advancement of *specifically digital methodologies*, which seek to produce and test tools and processes for the capture, store and analysis of the large-scale complex datasets in use in computational social science. As Winner (1980) asserts that tools - like the research artefacts they operate on - have politics, researchers must consider the method of digital data *collection* to have particular epistemic orientations that will have repercussions for the production of academic knowledge resulting from and dependent on them. Rather than glossing over them, they need to be understood in sufficient detail to make informed choices on the use and design of tools that engage with these objects of data. This chapter seeks to provide some of this unpacking, by establishing a detailed overview of

the structure, sources and affordances of available Twitter data, and using this information to demonstrate an informed process of tool and method selection for its analysis.

4.1.3 Research Design and Pilot Study

This chapter outlines the methodological approach undertaken for the project, including decisions related to the design of specific components falling with the paradigm of quantitative and qualitative methods and the justification for these choices. As discussed in the introductory chapter and developed in the choice of Bourdieu's Field Theory as a useful explanatory framework for understanding the how and why of the political discourse taking place on Twitter (Chapter 2), a goal of this thesis is the collection and analysis of qualitative information from identified active users involved. The decision to complement design and assessment of quantitative Twitter data analysis in large-scale case studies (Chapters 5, 6) with interview-derived deep and subjective data from a selection of participants in those discourse datasets stems from lack of knowledge, at the micro level, of how and why users behave as they do. The qualitative component of this research begins to address this absence of information in the literature. Sometimes (if mentioned at all) this is dealt with in cursory caveats made by researchers, that intention ascribed to particular user actions, or assumptions made about the level of user knowledge or social convention, are difficult to assess and currently unknown. Few studies have turned to qualitative methods to access these insights, and those that have tend to be from the Psychology or Marketing literature in the field (e.g. Meier et al. 2014).

An initial step required to both establish the case for this research (what is "Irish political discourse on Twitter" and who is participating in it), was to design an exploratory pilot study, discussed later in this chapter. This study uses a combination of the knowledge about the nature of the data, the digital methods for that data's collection, and the methods of statistical and descriptive analysis that can be performed on it, to arrive at a useful coding schema for actors and a pool from which to select highly relevant interview candidates.

4.2 Mixed Methods Research Design

4.2.1 Situating the Mixed Methods approach

Maykut and Morehouse's influential guide to undertaking qualitative investigation outlines a perception, within the natural sciences in particular, that such research is comparatively unscientific, unfocused or deficient in methodological rigour (1994 p. 3). Their assertion is that because qualitative research has its basis in the phenomenological branch of philosophy, rather than positivism, this has tended to relegate it in the eyes of many researchers, who have criticised it for its perceived lack of "objective inquiry based on measurable variables and provable propositions" (ibid.). This assertion is widely echoed in the research design literature of the late 20th Century, positing a dichotomy between the two approaches, identifying qualitative investigation as "a minority voice [in need of] strong defenders" (ibid. p2). That defence has been offered by authors such as Patton (1990) who contend that phenomenological research focuses on contextualisation, specificity of understanding and personal meaning, thereby offering fundamentally different but equally important research value, compared with the positivist position. Similarly, in supporting the need for qualitative work, Lincoln and Guba (1985) directly contrast the positivist paradigm of a single objectively-knowable world with the multiplicity of socio-psychological individualised subjective realities inherent to phenomenology. So too, the position of researcher as disinterested and independent scientific investigator is contrasted with the inherently connected and interdependent phenomenologist, and procedural linear unidirectional causality with complex event interplay and multi-dimensionality. Lincoln and Guba's conclusion is that the resulting levels of abstractness and specificity in these contrasted approaches highlights a central tension in the separate goals of each; while quantitative research tends toward verification and proof of hypotheses, qualitative research seeks discovery of new insights, offering deep and specific rather than broad and generalisable findings.

Though the separation and perception of relative merit of the two approaches has been changing as the natural sciences take on qualitative approaches, much as the social sciences earlier did with quantitative methods, contemporary literature maintains these distinctions (E.g. Creswell 2013). Attempts to mix methodologies and to integrate both forms of data in a general procedural framework are relatively new, and mixed methods strategies emphasise the explanatory power of the contrasting data types and the distinct phasing of the study or studies to appropriately address the research question involved (Creswell & Plano Clark 2011).

Creswell and Plano Clark posit a superior understanding of research problems via integration of both methodologies than can be attained from either alone (ibid., p5). They contend that the human capability to approach investigation and to solve problems using both inductive and deductive logic, numbers and words, observation and abstraction, make mixed methods a "natural" framework, especially where research is concerned with social investigation or situations where multiple worldviews may be relevant. The separation and later combination of methods from quantitative and qualitative approaches allows the researcher to seek expert assistance in the established techniques of either, and to approach a research question more holistically in the design phase. This makes such research potentially more difficult, as the investigator requires training in approaches from both (historically mutually exclusive) areas. The formalisation of investigative procedures and research design involved in mixed methods studies, as well as the demonstrated value of insights gained, has led to its consideration as a legitimate *third way*, and its increased acceptance, especially in Social Science (Denzin & Lincoln 2005).

This study adopts a mixed methods approach specifically grounded in a pragmatist worldview, *Pragmatism* being identified in the literature as the philosophy offering the best-fit worldview for such research (Creswell & Plano Clark 2011; Tashakkori & Teddlie 2009, 2010). This approach values both objective and subjective findings, and emphasises the centrality of the research question and its practical, observable consequences. In this study, a pragmatist approach is most suitable, since the research question is premised both on known and unknown information; Twitter discourse methodologies are somewhat established, studies are acknowledging shortfalls and caveats, but the extent of validity in many foundational assumptions is not clear. Addressing this research question involves complementary quantitative and qualitative approaches: to investigate and understand specifics of the agency of actors involved; to analyse their mediated interaction at a macro level, informed by this; and to propose and test combinations of methods from the field in an attempt to develop an effective set for this analysis.

4.2.2 Categorisation of research questions constituent to this thesis

In identifying the potential suitability of a mixed methods approach, the particulars of the research were divided into thematic questions that suited quantitative, qualitative, or mixed approaches to their design and implementation:

Quantitative Questions:

- *i.* What data can be meaningfully captured from Twitter, and to what extent does the method of capture limit this?
- *ii.* For a given Twitter dataset, what patterns of communication, conversation, reciprocity and influence may be computationally discerned?
- *iii.* For a Twitter dataset centred on a specific political event, how do patterns of communication and interaction vary over its duration?
- *iv.* For the particular case(s) of Irish political discourse on Twitter, which users are most involved and how do their modes of interaction correlate to their identified association(s) with the field of politics?

Qualitative Questions

- *i.* How do identified Twitter users understand the conventions and affordances of the medium and to what extent might this understanding affect their specific interactions?
- *ii.* How are such conventions negotiated and established in the specific context of Irish political discussion on Twitter?
- *iii.* How are social, cultural and political factors transferred from outside the sphere of Twitter and to what extent do these influence actors' behaviour and view of their role(s) in the medium?

Mixed Questions:

- *i.* Are behaviours among archetypal actors described in the qualitative findings borne out at macro level over the course of quantitatively analysed political events?
- *ii.* Are there implications for the explanatory efficacy of quantitative approaches stemming from the particular qualitative insights obtained?
- *iii.* To what extent do these findings affect our understanding of Twitter political discourse based on the existing literature, and how can their combined insights inform further research design?

These questions, taken as a constituent set, combine to provide more complete knowledge of the effectiveness of any proposed methodological framework for the field, than would either qualitative or quantitative approach taken alone. The research design for the mixed methods approach to staging this research follows the outline of Creswell and Plano Clark (2011) and is detailed in this chapter.

4.2.3 The *Explanatory Design* model

In *Designing and Conducting Mixed Methods Research*, Creswell and Plano Clark (2011) set out a number of designs for the integration of quantitative and qualitative data in a study, citing examples of their use, and the strengths and challenges of each. Of these designs, the *Triangulation Design*, which is predicated on giving equal weight to both qualitative and quantitative data on a single issue, is identified by the authors as the most common and well-known. The Triangulation Design is a single-phase design, in which both types of data gathering occur simultaneously - a factor cited by the authors as contributing to their recommendation of it as an efficient strategy for new researchers seeking to blend data on a topic. This design was deemed unsuitable to this study, however, since it requires complementary and equally weighted sources of data to be gathered concurrently, rather than emphasising the interpretative ability of qualitative data as a secondary input toward the goal of situating and better understanding the context of a largely qualitative computational approach.

Thus, the alternative *Explanatory Design* model (ibid. p. 71)²⁶ was chosen as the 'best fit' research design strategy for this thesis. This design emphasises the primacy of an initial quantitative data collection and analysis phase, which precedes and is later informed by a separately staged qualitative data gathering and analysis. The staged approach allowed the researcher to deploy methods from the literature to collect and analyse Twitter trace data on the major case study topics of this thesis, and to use subsequent interviewing to aid in interpreting these. Specifically, this research integrates Creswell and Plano Clark's model for the *Explanatory Design: Follow-up Explanation* structure to phasing the research (See Fig 4.1).

Analysis of the data from the pilot study and major case studies, as well as findings from the review of literature in the field, was used to construct a set of questions to follow-up in qualitative research, informing the design of the interviewing stage. The selection of candidates for interview required a system of computational quantitative selection, based on metrics of activity in the Irish political discourse taking place on Twitter, and this sub-process of the research was itself modelled on another Explanatory Design structure identified by Creswell and Plano Clark, and set out in the next subsection.

²⁶ See also: Fig. 1.2 in the introductory chapter of this work.

Figure 4.1 Phases of the Explanatory Design: Follow-up Explanations Model²⁷



4.2.3.1 The Explanatory Design: Participant Selection model

Participants were selected for interview during the qualitative stage of the research using a system of computational selection based on the tracking of their contributions to a wide variety of relevant hashtag and keyword identified topics on Twitter, as well as on their interaction with particular agents (national politicians, political press, etc. See: section 4.6).

Fig. 4.2, overleaf, sets out the stages of this model, which uses a small-scale pilot study to collect and analyse quantitative Twitter trace data, the results of which form the parameters for a set of highly active contributors to the Twitter political discourse, from which actors identified with particular professional associations to politics could be chosen. The interviews that followed constitute the qualitative stage of the work, and thus this selection process is a structured component of the wider mixed methods strategy used in the thesis.

²⁷ This figure depicts a model adapted from Creswell and Plano Clark's Fig. 4.3 (b) (2011 p. 73)



Figure 4.2 Phases of the Explanatory Design: Participant Selection Model²⁸

²⁸ Adapted from Creswell and Plano Clark's Fig. 4.3 (c) (2011 p. 73)

4.3 Twitter Data Collection

4.3.1 Ethical Considerations

Though little discussed in the substantial corpus of literature reviewed in preparing this work (See: Chapter 3), a brief outline of the ethics of gathering Twitter data is useful in initially framing the method. Tweets gathered via Twitters APIs are necessarily public tweets, unless specifically set up otherwise²⁹. Therefore, from a purely technical point of view it might seem that there are no ethical issues in gathering Twitter data, since the process amounts to the recording of messages explicitly created by users who have indicated their understanding of the public nature of their output. However, since tweets are an expression of users' opinions, are regarded as highly current, often reactive, conversational, and possibly ephemeral in nature, it is difficult to contend that true informed consent has taken place, and that users understand that they may become part of a permanent archive or academic study.

In the case of the datasets used in this research, as in the vast majority of those in the literature that employ this same selection criteria, the use of hashtags in the tweets may indicate an additional acknowledgement of the public nature of such contributions by users. It may be justifiably asserted that those users, who included a hashtag, deliberately intended their tweets to be viewable by others searching for that categorising tag and – crucially – intentionally extended the scope of their message beyond their follower network, into an explicitly public context. Investigation of the degree to which such understanding exists among the users involved is necessary, though absent from the literature. Questions exploring this issue form part of the interviewing portion of this research.

Finally, since this research is aimed at finding patterns in the tweets that comprise collected datasets, rather than focusing on the content of individual contributions, ethical distance is maintained. Therefore the potential is minimal for any violation of perceived privacy, or harm done to any individuals, by the computational gathering and analysis present in this study, or tested as part of the development of a toolset. Additionally, all phases of the project

²⁹ Since the Twitter API requires authentication as a registered Twitter user, it is possible (though poor practice) to collect the tweets while authenticated as a specific established user with a given network of accounts followed. Some among that network may have opted for a "locked timeline", ostensibly making their tweets private to their followers, and it is possible that such users have previously allowed the user now collecting to access their tweets. This is not the case with this research, but this possibility is impossible to discern in any study in the literature due to the poor reporting of how collection was managed.
received *Research Ethics Committee* clearance from Dublin City University, before commencement (See: Appendix A.4 - Chapter 4 Addenda).

4.3.2 Twitter Data Sources

Before examining the type and structure of data that can potentially be captured from Twitter, it is necessary to briefly distinguish the points of access to this data, in order to understand why sources used in the literature and in this study have particular limitations and implications for completeness of data and for analytic potential.

4.3.2.1 Twitter Website

The standard source of access to tweets for ordinary users of the service is the Twitter website (or Twitter's client applications for mobile or desktop platforms, which are extremely functionally similar). This source is also used by researchers in a small number of studies where manual reading of content (either tweets or user profile pages) was an element of their methodology. The website is free to access, and for a logged-in user can provided as much historical data as the user can manually browse back through, making it a useful source where ease of access to a small sample of data is relevant.

4.3.2.2 Twitter's Application Programming Interface (API)

The Twitter API (in version 1.1³⁰ at time of writing) provides access to data objects, such as the *tweet* and *user* objects outlined in Section 4.3.3, in response to HTTP queries sent to specific endpoints (target URLs). For example, a GET query to the *search/tweets*³¹ endpoint will return a set of JSON formatted tweet objects matching specific search parameters passed in the request; i.e. a researcher could request the most recent 5 tweets containing the keyword "Ireland". The API for Twitter is separated into a Streaming API and *Representational State Transfer* (REST) API. These setups are designed to enhance Twitter's commercial value by strategically limiting much of the returned data in order to drive developers and marketing analysts toward licensed data-resellers (See: 4.3.2.2), rather than with the needs of researchers in mind.

The Streaming API is the more complex to use, requiring an authentication procedure and a persistent connection to Twitter's servers from which data is returned. Two principal modes are available; (1) a 1% sample of all public Twitter statuses (tweets) in real time, and (2) a

³⁰ https://dev.twitter.com/overview/api

³¹ The resource URL for this endpoint is: https://api.twitter.com/1.1/search/tweets.json

filtered stream from the *statuses/filter*³² endpoint, which returns all tweets, as they are created, which match a set of keywords, users, or geo-locations defined in the query that originated the stream. While the filtered stream is subject to a limit on the number of keywords, user ids, or location definitions, it constitutes the only API source where completeness of results is possible; provided that Twitter data of interest can be identified by a set of such parameters in advance, it is therefore possible to collect the totality of that data as it occurs in real time.

The REST API points use the eponymous information transfer paradigm (Richardson & Ruby 2007) to specify the available types of requests, types of data returned, and endpoints to use. Researchers can use the REST API to send requests, which return results in the form of tweets or users. This API set is subject to two important limitations, however: (1) searches for tweets made will return results only for the last 6-9 days, and (2) an unknown percentage of the total results are intentionally omitted³³. Queries to certain endpoints are not subject to these limitations; while general search type queries (e.g. show all tweets mentioning "Ireland") will be limited in these ways, specific look-up queries (e.g. show the tweet with a specific given id) will return complete information for that item. This makes the REST API useful to researchers in two ways - it can be used to partially fill gaps in data from the very the recent past, and it can be used for very specific post-hoc analysis by looking up single data points for users or tweets (e.g. a researcher could access the full data available for a specific user, or the data and metadata for a particular tweet of known id substantially in the past, to check how many times it has been since liked.

4.3.2.3 Data Resellers

Twitter's utility as a source of consumer and marketing insight has led to its commercialisation via the sale of sections of data through licenced resellers. Without an easy source of access to the data, purchase of ready-formatted datasets matching specific bounding or search parameters has become one mode of acquisition for researchers also. Paid platforms have developed which combine data acquisition and simple computational analysis, broadly divided into those catering to commercial interests (e.g. Topsy, Twitonomy) and those oriented toward academic enquiry (e.g. DiscoverText, Truthy). Some of these

³² This endpoint is described at: https://dev.twitter.com/streaming/reference/post/statuses/filter and its resource URL is https://stream.twitter.com/1.1/statuses/filter.json

³³ "... it's important to know that the [REST] Search API is focused on relevance and not completeness. This means that some Tweets and users may be missing from search results. If you want to match for completeness you should consider using a Streaming API instead." https://dev.twitter.com/rest/public/search (accessed September 2015).

platforms allow the free collection of limited-sized datasets, but without payment of oftensubstantial fees, these data are difficult or impossible to export or extend.

Borra and Reider (2014) note difficulties for researchers in the use of purchased data aside from cost alone; the data collection and analysis methods involved are often "black-boxed", subject to specific paradigmatic approaches, assumptions and premises, and metrics offered are derivative of and highly focussed on marketing analysis (ibid., p265).

4.3.2.4 US Library of Congress

A potential future data source, which is currently inaccessible but which may eventually provide access to the totality of historical Twitter data is the US Library of Congress archive. In April 2010, the Library of Congress announced that it would acquire the entire archive of tweets and make this available to researchers (Raymond 2010). While initially celebrated, authors such as Scola (2015) have noted the lack of any progress in provision of the data since the announcement; "[M] ore than five years later, the project is in limbo. The library is still grappling with how to manage an archive that amounts to something like half a trillion tweets. And the researchers are still waiting."

The management of the data by the Library presents an enormous challenge in scale alone; the archive from Twitter's foundation to the announcement of this acquisition in 2010 already contained 20 billion tweets – at the end of 2015, this number is in the trillions. Creating workable access and retrieval methods for this enormous dataset, as well as a system of authentication of "*bona fide* researchers" remains a focus for the Library, but no announcements have been made since the most recent status report described the importance of, and ongoing work on, the archive (Osterburg 2013).

4.3.2.5 Data source for this research: Twitter APIs

Due to the limitations of the cost and method involved with using paid sources of Twitter data, and the unfeasibly time-intensive nature of manual analysis from the Twitter website, this research, like the vast majority of studies in the literature, opts for a methodology of data collection via the Twitter API. Whereas interfacing with the API is possible via a number of proprietary and open-source software applications, variously used in other studies, this work examines the nature and research utility of the data itself, the options available and their limitations on that data, in order to inform the choice of collection model.

4.3.3 Twitter Data Structure

Having outlined those sources of Twitter data available, and situated the source for this research, it is helpful to understand the resultant data's structured nature and constituent components. Though Twitter data objects retrieved via APIs have dozens of structured subcomponents, the text (message part) of each tweet and the user who sent it are the key values visible to any Twitter user negotiating the medium in its conventional forms (via the Twitter website or client applications). This section therefore first discusses these components, before providing detail on other relevant data obtained. An understanding of the totality of API-retrieval data is necessary to the informed consideration of how best to operate a data gathering strategy for Twitter, as well as what points of analysis will be possible on the results returned.

4.3.3.1 Tweets

The term *tweet* is used throughout this work, and elsewhere in the literature, to mean the total data-point for a Twitter message, including all structural elements and metadata, though everyday users of the medium may understand the term as just that text portion of the tweet that entails the 140-character-limited message. (Some users, and some authors in the literature, make use of the term *"status update"* to indicate a tweet). Retrieval of a tweet object is possible via the APIs, and the structure is well documented on Twitter's Developer website (Twitter Developers 2015). These multidimensional JSON³⁴ objects comprise around 40 fields (key-value pairs) of data and metadata, many of which are themselves complex objects with other fields, and only one of these (the field called "text") comprises the actual message of the tweet that might be familiar to a user. The remaining data is a series of objects, strings, integers, and Boolean flags, much of which may be very useful to researchers seeking means by which to computationally categorise or analyse a given tweet. An example of a full tweet object, randomly chosen from the studies comprising this work, is given in the chapter addenda (Appendix A.4.1: *JSON formatted tweet object example*).

Research-relevant fields of a tweet object include: *text* (the 140-character limited message), *user* (an object describing the user who sent it), *id* (a unique identifying number), *created_at* (a UTC formatted timestamp for the creation of the object), *source* (the utility used to create the tweet – e.g. the Twitter website, a mobile client, etc.), and *entities* (an object containing a

³⁴ JavaScript Object Notation – a lightweight data exchange format comprising key-value pairs. For more see: http://www.json.org

structured array of components identifiable in the tweet – e.g. URLs, hashtags) (Twitter Developers 2015).



Fig 4.1 Annotated example of a tweet, shown on Twitter's website (November 2015)

Fig 4.1 Legend

- 1 User (the sender of the tweet), displaying real name, username, and profile image
- 2 A hashtag included in the tweet
- 3 *Text* of the tweet
- 4 Attached *media*: an image
- 5 *Retweet* and *Like* counts
- 6 Links to users most recently interacting with the tweet (via retweets, likes)
- 7 Time that the tweet was *created at*
- 8 Interface buttons for actions related to the tweet (reply, retweet, like, more)
- 9 *Replies* to this tweet in chronological order of their creation
- 10 An *@-mention* in a tweet sent in reply

The tweet object fields therefore contain much of the information available to Twitter users who are accessing the object as part of their view in a Twitter client application, or via the Twitter website. Figure 4.1, above, illustrates the extent to which this data is mediated to the user on the Twitter website "detailed view" for a single tweet³⁵. Many of the points of information secondary to the message text can be seen here, while some metadata is not reflected to ordinary users and is available only via the API.

Metadata components comprise the majority of the tweet object's fields and many of these are only situationally useful (e.g. *retweet_count* provides an integer value for the number of retweets a tweet has received, but this value will always be zero if the tweet was captured in real time, as in this study, since it only later accrues these – *post-hoc* analysis of this field may be useful in establishing the reach and influence of a tweet with time). Some fields, however, are very applicable in analytical methods seeking to explore the networked nature of Twitter discourse; *in_reply_to_status_id* and *in_reply_to_user_id*, for example, give the id of the particular tweet object and user, respectively, that a reply originated from. Such meta-information can usefully be employed in building a conversation thread comprising subsequent tweets and replies over time, though this capability is seldom explored in the literature.

A tweet necessarily contains some text, up to 140 characters in length in any supported language, and may contain embedded media (images, video), URLs (links), hashtags (category markers), @-replies (indicators of addressivity, see section below), and non-text special encoded characters (e.g. emoji). Only the user who created a tweet may delete it from Twitter (though this action will also necessarily occur after the point that the tweet has been captured in real time, such as occurred in this study, and though relevant to discourse analysis may not be perceptible to researchers).

Actions that may operate on a tweet (and therefore will be recorded in its meta-data, after the tweets initial creation) include processes related to liking (previously called "favouriting"³⁶), retweeting, quoting, or replying. Data concerning these actions for a given tweet, as well as metadata pertaining to the inclusions in the tweet, are important selection and analysis variables.

³⁵ https://twitter.com/EndaKennyTD/status/588611261287399424

³⁶ This change occurred in November 2015 (https://blog.twitter.com/2015/hearts-on-twitter), and its implications for user expression, intention and research related to same are discussed briefly in the chapter addenda (*Appendix A.4.2*)

4.3.3.1.1 Retweets

Redistributing another user's tweet to one's own following is known as retweeting, and Twitter affords this ability via its interface and API. The practice of manually retweeting content developed organically among users from the earliest days of the service (boyd et al. 2010; Bruns 2012b), via a process of copying a given status and prefixing it with "RT:" (retweet) before sending it again. This manual practice remains today, despite Twitter's facilitation of retweets via the interface since 2009.

Significantly for this research, and related studies of political discourse on Twitter, retweets account for a substantial volume of all activity on a topic, especially where topics intersect with highly current news issues. This is because the retweet may act as a form of information sharing (passing on an important message from another user during a time of relevance), endorsement (demonstrating agreement with the stated opinion of someone else via rebroadcast of their message), or driving attention (providing a tweet for others to see, critique, react to, etc.). Studies of Twitter's use in crisis communication demonstrate a steep increase in retweeting during crisis events (Bruns et al. 2012), and studies of political discourse show substantial retweeting correlated with breaking news in campaigns, announcement of election results, etc. (D'heer & Verdegem 2014; Larsson & Moe 2011; Maireder & Ausserhofer 2013), bearing out these usages.

Screening of captured tweets for both the metadata flag indicating a retweet via Twitter's systematic affordances, as well as content markers within the text ("RT:") indicating a manual retweet, is necessary since either approach alone will lead to under-counting their number. It is unclear in much of the literature whether this is accounted for in other studies.

4.3.3.1.2 Favorites (Likes)

With retweets, the number of likes (formerly "favorites") a tweet receives may act as an indicator of its relative influence. This process is complicated by (1) the various repurposings of the feature by users, who may like a tweet for many more reasons than agreement, approval or endorsement, (2) iterative changes by Twitter to how the affordance has been labelled and mediated, and (3) the necessity that likes accumulate after the initial publication of the tweet –the point when it is captured by real time collection software. Further discussion of the changes to this feature, as an example of the necessity for researchers to remain critical and aware of changing patterns in affordances on Twitter, is presented in Appendix A.4.2 of this work.

4.3.3.1.3 @-Replies (@-mentions)

The use of an @-reply (the inclusion of the username of another Twitter user prefixed with the @ symbol) is an indicator of *addressivity*; the deliberate addressing of content toward an identified actor (Honeycutt & Herring 2009). Though apparently simple, this requires some detailed understanding of usage for effective deployment as a selector for research.

An @-reply occurring at the beginning of a tweet results automatically when users opt to use Twitter's in-built "reply" button on any tweet and is therefore substantially the most popular manifestation of these among tweets. Twitter's default mediation of @-replies in this beginning position is to hide them from the followers of the sender who are not @-replied; i.e. only the intended recipient will automatically see such a tweet in their default view, and moreover this mentioned person will be alerted via notification systems in their Twitter client. This leads to two important (and likely initially unintended) consequences for their usage:

- Users who want to reply to someone, but explicitly want their replies to be seen by their followers often prefix the @-reply with a dot, ".", to override Twitter's default policy of hiding it. This action is readily identifiable in political discourse in particular, where a public performative aspect to conversation is often desirable and users wish to draw the attention of others to their on-going replies. Analytics that search for and identify this behaviour may be useful for tracking this performative and consciously public aspect of Twitter conversation.
- The ability of an @-reply to draw that user's attention, via default notification systems, leads to its widespread secondary use as a driver of attention rather than as part of an organic reply process. Many users deliberately include the @-prefixed username of those whom they wish to see a tweet. This practice is, of course, especially common in political discourse, where users may wish to draw the attention of political and media actors to messages they create. This practice is more accurately referred to as @-mentions, since no reply is taking place, although the mechanism derives from the same reply addressivity. The term @-mention is more often applied in the literature where the addressing occurs in a position other than at the beginning of a tweet.

4.3.3.1.4 Hashtags (#)

Discussed in the introduction (Chapter 1) of this work, hashtags are a user-generated communicative feature, used for categorisation and tagging of tweets, and included in the message body. The inclusion of these tags is considered by many scholars to indicate a deliberative practice, on the part of users, to explicitly mark their contributions as belonging to a specific category and public stream. Selection of Twitter datasets upon hashtags is the dominant data collection mode in the literature.

4.3.3.1.5 URLs

A list of any and all URLs included in a tweet is separately provided as an array as part of the tweet object, simplifying access to and processing of links shared by users. The provision of URLs by users constitutes a sharing of additional information, external to the message of the tweet, and increases in the presence of URLs have been shown to correlate with periods of intense information-sharing, such as crises, natural disasters, election results, and breaking news (Broersma & Graham 2012; Bruns et al. 2012; D'heer & Verdegem 2014; Eun-Ju Lee & Oh 2013; Papacharissi & de Fatima Oliveira 2012; Cristian Vaccari et al. 2013). A measurement of the proportion of tweets containing URLs has been used by researchers in automated computed analysis of Twitter behaviour; Bruns and Steiglitz (2012), for example, examine cycles of information sharing (via URL inclusion) and subsequent retweeting.

4.3.3.1.6 Media

Inclusion of media in tweets (termed "embedded", by some researchers, but more accurately "attached") is a well-established user practice, and image and video files shared with a tweet are accessible via the API. Twitter strips EXIF data from images uploaded to the service (Twitter 2015b), in order to avoid collection of the location and other metadata that users may not be aware of submitting in their image files. This feature, while helpful for user privacy, also greatly impedes the detection of fake or altered images and makes their authenticity an area of considerable question for researchers interested in examining the media attached to tweets (e.g. Gupta & Kumaraguru 2012). The studies in this research do not examine the specific content of media inclusions, which would be subject to these caveats; though some of the metrics investigated here rely on establishing whether media was attached at all, as an indicator of certain types of Twitter behaviour and usage.

4.3.3.2 Users

The Twitter *user* object returned by certain API requests is a similar JSON-formatted data object to the tweet object³⁷. It consists of around 40 fields as key-value pairs and a simplified version of this object is the data returned for the *user* field of the tweet object detailed above. Similarly to the tweet object, much of the content of a user object is available via client applications and the Twitter website and will be familiar to ordinary users of the service. Figure 4.2, below, shows a section from the user profile page³⁸ of a participant in this study's quantitative and qualitative analyses, demonstrating which fields of the user object are represented in this view.





Fig 4.2 Legend

- 1 Profile image (a picture selected by the user to represent themselves; usually a portrait)
- 2 *Count of total tweets* sent by this user
- 3 Count of following and followers
- 4 Real *name* (set by user) and *screenname* (Twitter username)
- 5 Sorting options for the history of this user's tweet output
- 6 *Verified* flag, indicating a Twitter-verified account, where shown
- 7 The user's *description* of themselves; "Twitter bio"
- 8 Self-declared *location* and *link*, and date the user joined Twitter
- 9 Interface buttons for interacting with this user (Tweeting to, Directly messaging)

While most approaches to data collection in the literature focus exclusively on the capture of tweet objects as part of a real-time collection of Twitter content, based on some search

³⁷ https://dev.twitter.com/overview/api/users

³⁸ https://twitter.com/CathMurphyTD

criteria for the topic of interest, it may be useful to construct queries to the Twitter API for user data specifically. In this study, post-collection analysis of a dataset of tweet objects on political discourse topics involved the coding of the users involved, to categorise them by their professional association with the field of politics. This was achieved using custom software (detailed briefly in this chapter), which queried the users/lookup Twitter API endpoint³⁹ and returned, parsed, and presented a matching user object for the relevant user's id. The result of this process was similar in effect to visiting the user profile page of every user to make an assessment of their category - a process described in several studies in the literature that used such coding (e.g. Maireder & Ausserhofer 2013). This study's customised approach, however, allowed the researcher to explore any relevant fields from amongst the totality of information available on a given Twitter user via the API, rather than being fixed to the conventional web view. It also allowed for the integration of the coding process directly; rather than opening each user's profile page in turn, and in a second application assigning that user a category code, a single web-app was programmed to take a user id from the dataset, query the API for the corresponding user object, parse and display relevant fields, and allow the researcher examining this output to press a key which set the code in the researcher's database. The approach greatly increased the rate at which user coding could take place.

Particularly useful fields to retrieve when attempting to categorise a user include *name* (the name of the user, as they chose to define it), *screen_name* (their Twitter username, which often in the case of politicians includes prefixes or suffixes related to their party or role), *profile_image_url* (a URL for the avatar image chosen by the user), *description* (the 140-character "bio" text afforded to Twitter users to describe themselves), *followers_count* (the number of Twitter accounts following this user), *friends_count* (the number of Twitter users this user follows), and *location* (a user-defined indication of the user's geographic location⁴⁰).

4.3.2.2.1 Description

The *description* field of a Twitter user object, which contains the short text that the user has provided to describe themselves, generally contains a brief explanation of what they do. The "bios" for professional political and media figures, official organisations (such as political parties and NGOs), and non-professional media (blogs, bloggers, etc.) tend to provide a

³⁹ https://dev.twitter.com/rest/reference/get/users/lookup

⁴⁰ Though often used in studied in literature, this field is deeply unreliable as a valid indicator of actual location of a user, as was demonstrated in the trend for users to set their location to "Tehran, Iran" during the Arab Spring revolutions, as a mark of solidarity and perceived help to disguise the accounts of users actually located there and potentially in danger.

concise self-description useful for category coding. Indeed, researchers have noted this field as the key determinant for identification of the topic expertise or association of a user (Maireder & Ausserhofer 2013; E.g. Wagner et al. 2012). The self-description by ordinary users, in contrast, tends to be highly variant, and entirely absent in some cases.

4.3.2.2.2 Follower / Following counts

The *followers_count* and *friends_count* field may be useful indicators of user influence generally, though perhaps not specifically for the topic network under study. A large number of followers indicates that the tweets sent by that user will be seen in the timeline of many other users, affording greater exposure for those messages. Typically, an influential user, especially one with status in the fields of media or politics, will have both a very high follower count and an asymmetrically lower following (*friends_count*), indicating that their messages are widely received but that they are not correspondingly connected with those users who receive them (Wagner et al. 2012). For most active users without a public or celebrity profile, their mediated communication with others like themselves tends to create to a broad symmetry of their follower/following counts, as both members of any communicative pair add one another to the accounts they follow.

4.3.2.2.3 Verified

Twitter user objects also contain a *verified* field, which stores a Boolean (true / false) value indicating whether the user has verified their identity with Twitter officially. Since this service is not open to the public, but rather to a fuzzily defined set of *"highly sought users in music, acting, fashion, government, politics, religion, journalism, media, sports, business and other key interest areas"* (Twitter 2015c), it cannot be widely used for verification of the identity of most users in a cohort to be categorised, but where present (usually for senior politicians and national media agents) can be a valuable indicator of authenticity.

4.3.4 Data Collection Software

Having established the best-fit of a collection strategy leveraging the Twitter API – specifically the streaming API for the real-time collection of identified filtered data, and the REST API for post-hoc classification look-ups – a software application could now be chosen for its ability to effectively select on and collect the particular data objects understood above.

The *ad-hoc* programming of API queries and the creation of parsing and storing algorithms for the resultant data objects requires a significant level of technical skill, acknowledged to be a major barrier to research in many studies in the literature (Borra & Rieder 2014; e.g. Bruns

& Stieglitz 2012; D'heer & Verdegem 2014). It was therefore an initial goal of this work to respond to the methodological gap between the potential richness of data collection afforded by the API and the existing software solutions which failed to fully meet that potential. This aspect of the research constituted an analysis of the collection software in use in the literature (see also: Chapter 3), and the development of new webserver-based applications capable of utilising the API to gather more complete and better-structured datasets.

Since the architecture of the structured information objects (described in the previous section) constitute what Uprichard (2013b) describes as the "*ontology of the case*" – the defined and related data points that together construct the information available to us – it follows that any strategy for collecting that data ought to preserve its structure as completely as possible. Metadata (or specific fields of a data object) lost at the collection stage are therefore unavailable at the analysis stage, limiting both the contemporary analytics available for the set, as well as any future methodology that may become relevant to the field and seek comparative findings from historical archives to utilise. As data can be captured, stored, and retrieved in many different ways, this research sought to maximise the completeness and primary integrity of the information, while providing an end-product archive which would be fast, responsive and easily integrated into both proprietary and open-source analytics applications.

4.3.4.1 Critique of Existing software

During the early research period of this work, two open-source collection applications were evidently well established in the literature; yourTwapperKeeper⁴¹ and 140kit⁴². Both had begun as web-based services, accessible to the public, and both needed to be removed in response to changes in Twitter's *Terms of Use*⁴³ in 2011 that limited the publishing of its data. In both cases, the source code for the sites were published as open-source projects; scholars could therefore take and install this code on a private webserver, utilising it to collect a dataset for their own research use (rather than direct public publishing of raw data)⁴⁴.

⁴¹ https://github.com/540co/yourTwapperKeeper (Accessed: November 2015)

⁴² https://github.com/WebEcologyProject/140kit (Accessed: November 2015)

⁴³ https://dev.twitter.com/overview/terms/agreement-and-policy (Accessed September 2015)

⁴⁴ It should be noted that this practice is neither endorsed as acceptable, nor expressly forbidden, by Twitter's Terms of Use, and that this model for researcher-specific collection has operated in something of a loop-hole since 2011, being potentially subject to shut-down by any future moves by Twitter to explicitly address this. This is noted by some authors in the literature (Borra & Rieder 2014; e.g. Bruns & Stieglitz 2012) but may be poorly understood more widely, certainly not being well discussed or recognised in caveats or limitations outlined.

Of the two applications, yourTwapperKeeper (yTK) was substantially the more widely used in those studies that specified how their Twitter data was collected. This is likely to derive, in large part, from the detailed discussion of its utility and the subsequent publication of dataprocessing scripts designed for with it, by Axel Bruns and other researchers from the Queensland University of Technology (QUT), Australia (Bruns & Burgess 2013, 2011; Bruns & Stieglitz 2012). The yTK project derived from www.twapperkeeper.com - a (now removed) website that allowed users to set up a search-specific collection of tweets in real time, exporting this as a comma separated values (.CSV) file of the fields stored for each tweet captured. Neither the original site, nor the subsequent open-source project offered any analytical component, though Bruns and his colleagues made available a number of GAWK⁴⁵ scripts which simplified the process of parsing the resultant CSV files for direct use with mapping and quantitative analysis software (such as GEPHI or R). Even with the provision of these scripts, processing of Twitter data compiled this way is difficult and requires a high standard of working knowledge in object-oriented programming and the use of regular expressions - skills likely to be found more readily among Computer Science researchers than among their counterparts in the Humanities and Social Sciences.

				archive yo	ur own twe	ets	
Archive ID	Keyword / Hashtag	Description	Tags	Screen Name	Count	Create Time	
22	#bigdata	gen research	research	dm_phd	1040644	Tue, 05 Nov 2013 12:39:44 +0000	147
23	#analytics	gen research	research	dm_phd	403748	Tue, 05 Nov 2013 13:17:30 +0000	147
24	#vinb	Vincent Brown	test, tv, pilot	dm_phd	84515	Mon, 11 Jan 2014 11:17:47 +0000	10
25	oireachtas	pilot	pilot	dm_phd	961	Tue, 12 Jan 2014 23:08:14 +0000	147
26	#dubw	pilot	pilot, phd	dm_phd	1215	Tue, 12 Jan 2014 13:18:47 +0000	147
27	#ep2014	euro elections	test, pilot	dm_phd	46725	Fri, 24 Jan 2014 09:19:06 +0000	147
28	#ep14	euro elections	test, pilot	dm_phd	19813	Fri, 24 Jan 2014 19:03:25 +0000	10
29	#ccven	const. conv.	pilot, test	dm_phd	16557	Fri, 24 Jan 2014 19:57:39 +0000	147
Your TwapperKeeper - version 0.6.0							

Figure 4.3 Interface of yourTwapperKeeper software: Archive list

The yTK project was developed by John O'Brien, an American computer programmer and consultant, who moved on from the project in January 2012, leaving the archive without updates from that point. In part because of this lack of continued development, but more substantially because of technical shortcomings of the software (grounded in its development for an older and less feature-rich version of the Twitter API), its direct use for this research

⁴⁵ GNU Awk (https://www.gnu.org/software/gawk/manual/ Accessed September 2015), is an implementation of Awk – a text programming language designed for data extraction.

was discounted. Important limitations of yTK in the current (2012) implementation available online include: (1) that data is not stored using a UTF-8 character set, excluding the possibility of encoding and correctly analysing any accented or special characters, including all non-Latin alphabets, (2) that specific users' total outputs cannot be collected – only tweets containing keywords can be captured, (3) that storage methods used out-dated and comparatively slow functions for interfacing with the database, and (4) – most crucially – that it omitted the majority of fields returned by API queries and stored only 12 of the 40 or so fields potentially returned. The latter factor – a product of the software's design at a time with the Twitter API (v1.0) simply returned fewer fields – leads to very substantial limitations on the analysis possible, especially since the fields omitted are chiefly metadata related to a tweets position within a conversation network⁴⁶.

The alternative software candidate, 140kit, was subject to the same limitations, especially of data fields recorded, but had the additional issue of lower stability over long term use; several researchers report its comparative likelihood to crash when collecting large volumes of data. This issue is a particularly serious one for real-time data collection, since any outage in active collection causes gaps in the dataset.

It became necessary, therefore, in order to proceed to the analytical stage of investigating Twitter discourse, to acknowledge that the collection software in use in studies in the literature - largely glossed over and taken for granted – was a crucial limiting factor on efficacy. The researcher therefore undertook development of software to address identified shortfalls in the open-source software available. This practice begun as a staged updating of the yTK collection software to address issues of its stability over long periods of use, its efficiency in database transactions and its limited collection of available fields of trace data. The building of ad hoc software solutions for processes involved in the coding of research data was also undertaken, concurrent with some of this work, and both of these projects of tool-building ultimately contributed to this thesis and, in part, to open-source collaborations in the field.

4.3.4.2 Software development for data collection

An early focus of this research was the improvement of the yTK project, through updating of its programming, architecture, and storage paradigm. To achieve this, the researcher made contact with O'Brien, the project's author, and communicated regularly during March 2014 on issues related to the implementation and efficiency of the project. In conjunction with

⁴⁶ E.g. *in_reply_to_status_id*, identifying which other tweet a given one was sent as a reply to

O'Brien, an updated version (yTK_DCU_v1) was created and installed on a server in Dublin City University⁴⁷ and was used for the real-time capture of Twitter data pertaining to the first case study in this work (the Irish local elections of 2014, #le14). This version was cloned onto a second server and iteratively improved by the researcher throughout 2014, specifically addressing the speed of data storage and retrieval, the encoding of non-Latin characters and the capture in their entirety of the fields of a tweet object. Thus, by the commencement of the second case study (#marref, in 2015), a substantially updated version of yTK was in use.

The most recent version (*yTK_DCU_v3*, January 2015) also included a number of new features, chiefly related to usability⁴⁸:

- A responsive HTML interface, which allowed the application's views to scale and fit a mobile device, extending the potential for a researcher to check on their on-going data collection more regularly from a mobile or tablet.
- A re-design of the workflow organisation, to segment functionality into dedicated areas (setting up new keywords, exporting existing data, at-a-glance displays for current tracking, etc.)
- Removal of unnecessarily query-intensive displays of data in individual archive views

 the default yTK software displays a large amount of recent tweets from any query viewed, formatting it to look equivalent to the Twitter website timeline, and including profile images of the users involved and links to the Twitter website. This view, which creates a loading delay caused by the retrieval and formatting every time it loads, is not of substantial use to a researcher, who very likely needs only to see the current status of the archives, the number of tweets it contains, and links to export it.

It should be noted that while the servers in DCU, on which these two software iterations ran, remained stable over the course of data collection for the first case study (#le14), one of the two was hacked in the summer of 2014. The server, which was 'rooted' (i.e. taken over maliciously by hackers) was used for short period to carry out attacks on servers in Italy, and was therefore quarantined and removed from the DCU network. Since the researcher was unable to guarantee the future security of webservers deployed *ad hoc* to the DCU network for data collection, a decision was made to use Amazon Web Services (AWS) from September 2015 onwards. AWS is a paid service, in which scalable server instances can be

⁴⁷ http://twarchive.dcu.ie

⁴⁸ Comparison between various views in the default yTK package and the yTK_DCU_v3 version developed over the course of this research may be made with reference to the screenshots supplied in the addenda to this chapter (Appendix A.4.3)

deployed with a particular specification and a monthly charge is owed based on the volume of data transferred. The service is intended for the deployment of large-scale web applications and services and the costs are low for the case of Twitter data collection. Additionally, unlike the case of adding an *ad hoc* collection server to the DCU network and potentially exposing it to security risks, AWS has the advantage of on-going security expertise and system administration, substantially relieving this workload from the researcher.

Concurrent with the improvement of the yTK software, was the development of a new web application designed to work in conjunction with the structured database containing the archived tweets. This software, given the working title *yourUserCoder* (yUC), was created by the researcher to simplify and speed up the categorisation of users responsible for the production of tweets in a given dataset.

The yUC software – most recent version yUC_DCU_v2 , July 2015 – retrieved and parsed the yTK database table for a given query-based dataset, generating a list of all unique users recorded, by username and user id. This new list – substantially shorter than the list of tweet objects, since a single user is usually responsible for many tweets that occur on a given topic, especially in a bounded political discourse - was then systematically displayed in a web view (See: Section 4.3.3.2). This view (Fig. 4.4) displayed all relevant fields for a given user (e.g. name, description, image, verified flag, etc.) and facilitated the user's categorisation by providing the researcher with the option to immediately choose a coding category. This process stored the code on that user's record, and loaded the next user for categorisation. The software displays a progress bar for the cohort currently being coded, and following the submission of a coding decision, that choice is confirmed in the subsequent loading, offering the researcher a chance to review and revert if necessary. The yUC form loads quickly and can be operated either by clicking the buttons corresponding to dynamically-loaded coding categories, or by using the relevant number on the keyboard. Thus, the software achieves a workflow for Twitter user coding that is multiples faster than the scheme of copying a username, looking up that user's profile on the Twitter website in a browser, and recording the determined code in a separate application for quantitative analysis. This latter process, attested in some studies in the literature, requires that the coding of users for a dataset commence when the set is complete and exported to another application for analysis (e.g. nVivo, R, Tableau). Crucially, this workflow with yUC may be carried out on a currently active dataset, since it retrieves and stores the coding data alongside the on-going collection, and allows a concurrent rather than staged approach to coding, with the export at the end of the process already containing information about the users determined by the researcher.

Figure 4.4 yourUserCoder (yUC) software; view of coding in progress



(Note: This image is altered to remove reference to another user's screenname)

The yUC software was used for the user coding of both of the major studies (#le14, and #marref) in this work (See: Chapters 5 and 6, respectively.)

4.3.4.3 Emerging software concurrent to the research

In mid-2015, attention was drawn to the development in late 2014 of Twitter data collection software termed DMI-TCAT by Eric Borra and Bernhard Rieder (2014), researchers at the Digital Methods Initiative⁴⁹, at the University of Amsterdam. DMI-TCAT *(Digital Methods Initiative – Twitter Capture & Analysis Toolset)* is a software application developed along similar lines to those of this research. The toolset seeks to respond to the same shortfalls in collection methods, but also addresses some features of analysis by providing the capability not just to export data for use in analysis and visualisation applications, but to also pre-process the data and output the results of simple computable analyses.

⁴⁹ https://wiki.digitalmethods.net/Dmi/DmiAbout

While initially launched in 2014, DMI-TCAT was brought to the attention of many researchers through the endorsement of its use by Dr. Axel Bruns (2015), whose previous manuals and scripts for yourTwapperKeeper had helped centralise it in the Twitter research community. Unlike yTK, however, DMI-TCAT is an actively updated open-source project, with several contributing researchers and developers. Its iterations through 2015 have led to its position as a stable, feature-rich collection tool, which exhibits the same completeness of data collected, and utility of interfaces, as the yTK variant improved for this research. Moreover, the developers of DMI-TCAT have surpassed the database interaction methodology underpinning yTK and the software derived from it, giving the former a performance advantage when dealing with large amounts of incoming data, and a greater long-term stability when tracking keywords over many months. For this reason, the improvements to yTK in this research were suspended and a DMI-TCAT install was deployed on an additional server, and switched to for on-going data collection concerns in late 2015. Contact was made with the development team for DMI-TCAT, in order to share knowledge on the creation of the tools for this research, in the spirit of open-source software creation.

Absent from DMI-TCAT, however, is any tool that approximates the functionality of the yUC software in coding users in a dataset. A current endeavour of this research, beyond the work of this thesis specifically, is to continue to improve and develop yUC with a view to its incorporation into the DMI-TCAT set, and the extension of its improved coding utility to many other researchers in the field.

4.4 Quantitative Methods: Twitter Data Analysis

The overwhelming extent to which the study of political discourse on Twitter relies on quantitative metrics and methods is readily evidenced in the literature (cf. Chapter 3). This may seem an obvious consequence of the data involved; the presence of numerical and Boolean metadata, the highly structured and segmented objects involved, and the flexibility of API query design (at least for those researchers with the technical skill to deploy these), all point toward a fit with computational quantitative methods that might be difficult to apply for any other widely-used medium of communication. It is important to note, however, that while 'traditional' quantitative methods for the substantiation of deduced patterns of correlation and production of evidence are important, it is necessary to supplement these with methodologies capable of investigating Twitter datasets as interlinked data points. The sense of a network to be analysed, rather than a collection of discrete data objects, is central to any attempt to examine and meaningfully discern a political discourse on Twitter, so it follows that methods from Computer Science suitable to these topological investigations should be deployed here.

4.4.1 Statistical Analysis

The obvious first steps in the analysis of a dataset pertaining to Twitter discourse on a topic under study are the derivation of simple quantitative metrics and statistical descriptors for the set; these include the total number of tweets captured in a given timeframe, the number of these that are replies, the number that contain URLs, the number of unique users, the number of users, the mean / minimum / maximum number of tweets per user, etc. An extensive list of these numerical factors are easily derived; indeed both the software developed from yTK for this research and the DMI-TCAT software sought to continuously display much of this information while a dataset was still under collection, so as to give researchers an on-going sense of the conversation being captured.

The segmentation of these statistical measures into customisable intervals (hourly, daily, weekly, monthly, etc.), and the capability to easily derive and analyse sub-sets of a total archive as delineated event-related periods, make these powerful explanatory variables for temporal analysis on a topic. It is possible to provide time-specific insights from simple volume (e.g. *how many tweets were sent on the day before an election event, compared to the average or to a specific other period?*), to complex examinations of correlation of behavior distinguishable in the metadata with events from the study's duration (e.g. *to what extent are increases in information-sharing indicators such as retweeting and URL inclusion seen in conjunction with national televised debates?*). Determinations of such correlations may be subjected to the methods of probability

determination and investigations of degree of significance, which are central to the quantitative toolset.

These quantitative analyses on a dataset can provide not only confirmation for deduced hypotheses about how discourse on a topic should be expected to unfold, but also observational support for inductive reasoning about the cause of patterns detected. Analysis of the latter type might form the basis for sets of questions to frame qualitative inquiries (e.g. interviewing archetypal users from a particular cohort on the behaviours that might underpin a detected pattern in communicative strategy). In this regard, the deductive or inductive directions of potential theory development, and the quantitative and qualitative methodological outcomes, show the applicability of a mixed methods research model to Twitter discourse analysis.

4.4.2 Network Analysis

As already illustrated, Twitter users engage with one another through the affordances of @mention and retweet structures in the medium, and that these can be algorithmically detected in a dataset through analysis of certain fields of the tweet objects collected. As a result, it is possible to extract network structural information from Twitter data using the techniques and metrics of *Network Theory*, which is increasingly used in many academic disciplines outside of Computer Science (Bruns & Stieglitz 2013; Burt et al. 2013) and has recently been shown to have potent explanatory potential for the inter-group dynamics in electoral discourse (e.g. D'heer & Verdegem 2014). Network Theory utilises analyses based on the structured nature of communication networks, taking a particular cohort of discourse participants as *nodes* (individual Twitter accounts) and examining the conversation as a set of *edges* (connection relationships between nodes). Network *topology* refers to a layout of the nodes and edges based on the reply and retweet relationships captured in a recorded period of Twitter discourse on a subject. This network model uncovers patterns in interactions among users and usefully provides metrics on the degree of in-group and out-group connectivity.

Network Theory metrics (e.g. *E-I index* – a measurement of network *cohesion* indicating the degree of communication internal to and external from subgroups within the network⁵⁰) can then be used to provide a detailed description of the studied network(s). Using data gathered from the Twitter API, two topological networks are easily generated for a dataset: the *connection network* and the *@-mention network*.

⁵⁰ This metric is described in some detail, and given specific context, at the introduction of its use in examining the network of #le14 discussion in Chapter 5. (See also: Krackhardt & Stern 1988)

A connection network is constructed based on follower and following data for the participant nodes – the overall number and specific ids of all accounts that follow a particular user, and all accounts that the user follows are available via the API, as discussed earlier. This network is largely a passive structural arrangement which may pre-exist a period of studied discourse and change little during it, reflecting the *status quo* of interconnectivity among the studied users and illustrating power structures and sources of influence within the discussion space and wider than the specific studied hashtag event that created the set.

The mention network, however, is a dynamic and active set of interconnections representing the edges created between nodes in communication (directed replies sent and messages retweeted) during the period of study. This network is specific to a given bounded dataset (and therefore to the event being tracked by hashtag) and may provide researchers with highly pertinent information about the discourse; patterns of and changes in influence, reciprocity, information-sharing, and communicative directionality are all discernable.

To these two networks, the theory offers methods for *Centrality Analysis*; metrics at the node level for *degree centrality* – the number of ties, or degrees of connection to adjacent nodes - and *between-ness centrality* – the distance in connections to distant nodes. Centrality Analysis provides us with a computable method to understand the influential players within a studied network and to quantitatively compare these. Total edges (connections) between these nodes (individual users of Twitter) can be measured as a proportion of the maximum possible edges for the given network to provide a network density, useful for illustrating the extent of absolute cohesion within the group. The methods of Network Theory also provide for analysis of grouped nodes as modules. The extent to which modules, as cohesive sub-groups of a network, exist can provide important information about communities within a studied dataset and their degree of interaction (Wasserman & Faust 1994). These methods, in particular, provide a framework of quantitative investigations from Computer Science that may be readily applied to the analysis of Twitter discourse among cohorts of categorised users in a dataset.

4.4.3 Descriptive Analysis

As outlined in Section 4.3.3.1, individual tweet objects contain a large amount of information, as metadata, accompanying the message text of a given tweet. *Descriptive Analytics* focus on quantitative statistics derived from this data, such as the number of tweets, distribution of different types of tweets, and the number of hashtags. Descriptive statistics are widely used elsewhere in political discourse analysis, as well as in other studies of Twitter communication (e.g. Bermingham & Smeaton 2011) and are a standard inclusion in research

based on other types of data (e.g. surveys, questionnaires). While a small number of metrics (e.g., sample size, response rate, responder profile) are used for the survey data, the enriched nature of Twitter's 'big data' enables intelligence extraction, using a large set of metrics regarding tweets, users, hashtags, URL inclusions, etc.

Descriptive tweet metrics aim to present a simplistic, but broad and essential, picture of the data using several metrics (number of tweets, mentions per tweet, number of hashtags, etc.). Understanding who tweets, and in what form, who replies, and who retweets is important when attempting to uncover explanatory patterns which might reflect group or individual power, relationships, or interdependence within the cohort. Descriptive metrics for user objects and categorised groups of user objects may also be pertinent. Information such as the size and symmetry of follower and following counts, for example, can reflect the influence level of a user, and means for these values can give researchers some sense of the comparative influence levels of cohort groups. Such user information is closely related to the outcome of centrality analysis (Bruns & Stieglitz 2013).

Finally, the number of tweets containing one or more URLs in their text is easily computed and this subset may be a useful source of further investigation. URLs could be news releases, reports, articles, and more, and are a direct example of information provision by users to their followers. Thus, analysing URLs can reveal subjects of interest at the micro level of individual tweets, periods of lesser of greater information provision in the macro level of the discussion period, and individuals or modules within the network where greater or lesser linked information sharing takes place.

4.5 Qualitative Methods

The qualitative potion of this research took the form of interviews with key participants in Irish political discourse on Twitter, selected via the pilot study (outlined in the next section) and structured according to Creswell and Plano Clark's *Explanatory Design: Participant Selection* model (2011 p. 74).

4.5.1 Interview Design

This research adopts the 7-stage model (*thematising*, *designing*, *interviewing*, *transcribing*, *analysing*, *verifying* and *reporting*) outlined by Steiner Kvale (2007). These stages establish the interviewing process as a holistic one, emphasising the relatedness of the interviewing to the thematic subquestions of the research hypothesis and weighting aspects of the design phase and postinterview practices of the analysis and generation of findings.

4.5.1.1 Thematisation

The research questions were divided into key sub-questions and thematically grouped, as outlined in Kvale's model. These themes were then used to generate both *thematic* and *dynamic* questions; the former eliciting information from the participant related to the research agenda and the latter maintaining a lively and engaging flow for the participant (ibid.). The questions were organised in a staged set of groupings, introducing and producing knowledge on each theme, with follow-up or prompt questions identified to further develop responses where necessary. Questions were posed which related to the themes first abstractly and then via grounding in the context of specific examples and event. The questions, with information on their themes, are given in Appendix A.4.5 of this work.

4.5.1.2 Design

The interviewing process was designed next; this process involved the creation of informative material, the selection of the participants, communication with them on the topic of the research, two pilot interviews, and associated iterative developments to the questions and related materials.

These design processes also necessitated an ethical dimension and prompted analysis of the use and storage of participants' data and the inclusion of a clear policy for this. A consent form was created to reflect this and, used in conjunction with the Participants' Guide, to verify informed consent was given. A particular ethical concern was the anonymity preferences of participants, some of whom would be sitting professional politicians and media

agents discussing details of their work, their parties, candid information on their estimation of and their relationship with others, etc. It was decided to offer three levels of anonymity to participants; full credit, partial anonymity (identifying them by their professional association with the field) and full anonymity. These options were made clear on the consent form and discussed, with examples, in the Participants' Guide.

Selection of participants was made via the automatic generation of a shortlist of (initially 40) candidates, who demonstrated the greatest frequency of contribution to the discourse identified in the quantitative pilot study (See: Section 4.4). These candidates were contacted, provided with the guide, and interviews were requested. These interviews were organised in phases, corresponding to the two major case studies (#le14 and #marref) and scheduled in accordance with the overall project plan timeline.

4.5.1.3 Interviews

Having scheduled the interviews, these were conducted at a series of locations chosen by the participants (where a quiet space within Dublin City University was offered as a choice, and used by about 25%). Some participants were unable to schedule physical meetings for interview and three interviews were conducted using Skype⁵¹ Voice Over IP (VOIP) software. Despite the facility for video chat using this software, only one of the three interviewees opted to enable a video feed, and this feed was subsequently of such low quality and so prone to image-freezing that all three of these interviews were functionally identical to telephone interviews in that no visual cues were afforded. These interviews tended to be considerably shorter than interviews carried out in person; the mean length for Skype interviews in this study was 20m 00s, compared with 35m 22s⁵² for interviews in person. The Skype interviews provided less detail and showed more perfunctory responses from participants, who rarely diverged from very direct answers to the questions and often required follow-up and prompt questions. This is unsurprising, as Kvale (2007) and Seidman (2013) both suggest that telephone interviewing (which Skype approximates) provides diminished capacity to the researcher to build rapport, react to provide a comfortable flow, read gestures and body language, etc.

In keeping with Kvale's methodology, a short *briefing* took place before each interview, in which the purpose of the research was once again explained, the use of the recording

⁵¹ http://www.skype.com

⁵² This is the mean of 21 of 22 interviews in person, discounting the outlying very short interview NP7, which was a separate addendum interview to NP6. (See: Appendix A.4.6)

equipment was noted, an opportunity for questions was provided, and the participant made it clear that they were comfortable to begin.

Recording took place using the voice-memo app for iPhone, which provides a simple interface, and file compression specifically designed to provide small audio files that preserve the frequencies of the human voice with high fidelity. The iPhone was placed face-down during the recording period so as not to divert attention (as the display, when active, shows a stopwatch style constant output of the time that might have served to distract, call attention to duration, or otherwise interrupt a comfortable flow). Recording of interviews conducted by Skype utilised the open source audio recording and editing package *Audacity*⁵³.

Note-taking during the interview was minimal, in order to avoid distraction, and was limited to reminders for the transcription process, follow-up points for verification or investigation, and indications of extra-textual sources of relevance to the meaning of responses (body language changes, gestures).

The interviews followed a semi-structured format, which allowed for extensive flexibility between participants in the length and detail of responses, degree of follow-up, and in some cases question order. Kvale (2007) and others (King & Horrocks 2010; Seidman 2013) assert that drawing on personal experience of the area of research can be important to building rapport, phrasing questions, and responding dynamically during the interview. In some cases, previous acquaintance (via Twitter) with some of the participants could have potentially extended this past useful empathy to a situation of likely bias in responses or tailoring of answers toward mutually known subject areas, events, or issues. Some candidates were discounted in the design phase for this reason and care was taken during interviewing of participants to provide comfortable affability and demonstration of area knowledge on behalf of the researcher, whilst maintaining sufficient professional distance to ensure that interviews could reasonably be carried out by another independent researcher and show intersubjective similarity in findings.

As suggested in Seidman's *Interviewing as Qualitative Research* (2013), a short *debriefing* process was included after each interview. This process was begun in the final question of the set, which was designed to broaden the focus to a more openly discursive topic of the philosophical underpinning of the research and allowed the interviewer to move toward a more balanced conversational mode. After concluding the recording, debriefing continued with the participant able to ask questions about the process and about the research generally.

⁵³ http://audacityteam.org

A list of anonymised categorised interviewees, and interview information for each (including date, location, and duration) appears in the addenda for this chapter (See: Appendix A.4.6).

Signed consent forms, which necessarily identify participants, cannot be included with this work and are stored at Dublin City University in accordance with the stated ethical policy for data collection and storage⁵⁴.

4.5.1.4 Transcription

Each interview was transcribed shortly after being conducted, in two phases. Initially, using software (Audacity) that slowed down the recording to a tempo which could be comfortably typed at without excessive pausing, then a second pass comparing the completed transcription to the recording in real time to verify the completeness and add any notes required related to tone that were not perceivable when listing at the lower speed. The researcher completed all transcriptions directly.

4.5.1.5 Analysis

Printed hard copies of each transcribed interview were used in the analysis phase, during which physical markers (coloured sticky notes) identified coded themes within each. These themes among the responses were related to the thematic sets of research questions, which underpinned the initial design of the interviewing. This process consisted of the identification of commonalities and differences across responses to particular themed questions, rather than of a deep linguistic analysis of the content. Specific quotations that evidence these themes were marked in the transcripts for potential use in the subsequent reporting of findings.

4.5.1.6 Verification

The thematic findings were then drafted into a first-stage framework which could be examined for *validity*, *objectivity* and *relevance* (Kvale 2007). Kvale, in *Ten Standard Objections to Qualitative Research Interviews* (1994), identifies the difficulty in attributing objectivity, in particular, to interview material. Objectivity here is related to minimisation of bias and to *intersubjective consensus*; "In phenomenological philosophy, objectivity is reached through the intentional acts of consciousness and is an expression of fidelity to the phenomena investigated" (Kvale 1994 p. 151). The

⁵⁴ Interview recordings and transcripts to be held for 24 months subsequent to the submission of the study for examination. Consent forms and other identifying notes in hard copy to be held by the thesis supervisor following the ethical policy of the university. See also Appendix A.4.3 *Sample Participants' Guide*

interviews were thus checked for intersubjective consensus, identifying where insights on a theme emerged from knowledge co-constructed across participants, and between participant and interviewer, allowing knowledge production specific to each thematic question via what Kvale terms "*the voice of the interview*".

The verification stage for each interview consisted also of the post-hoc re-examination of the transcriptions and analysis to ensure that each was in line with the level of anonymity requested by participants. This proved, in some cases, to be especially difficult where those participants who had requested full anonymity made key contributions that necessarily identified them by their role. This was a particular problem among professional politicians, many of whom requested partial or full anonymity but whose useful and relevant discussion of self-representation and the transfer of power between fields were laden with references to their party, position in cabinet, or involvement in policy and campaigns, which would readily identify them. Care was required to mitigate any possibility of breaching the anonymity of participants and a policy was adopted of applying the highest possible level of anonymity across all contributions as a default. Inclusion of full transcriptions as an appendix to this work was also abandoned following an analysis of the level of redaction required to have the content meet the ethical standards; instead copies of the recording, transcriptions and notes will be kept securely at Dublin City University in accordance with the storage and destruction policy laid out in the Participants' Guide and approved by the Research Ethics Committee.

An additional complication to the verification procedures occurred when some participants invoked their right to withdraw entirely from the research substantially after they had been interviewed and these phases had been completed for their contributions. Though frustrating, removal of reference to their contributions and destruction of their recordings took place, as requested. It is likely that this change of mind was prompted by news developments late in the period of this research, which examined usage of social media by some prominent politicians and their relationships with their party in managing this communication. Those who withdrew from this study were all national politicians, and one other within that category retroactively changed their anonymity level to full at this time. In one case, a withdrawing participant stated that their reason for no longer remaining part of the research concerned their perception that their contributions (specifically comments related to their relationship with their political party and its control of their Twitter output) might become public and "show some sort of disloyalty" to party on their behalf. I confirmed permission to provide this reason as part of this study, which the withdrawing participant allowed provided that neither they, nor their party, could be identified in doing so. These events provide useful insight into the volatility of research both on current political issues,

124

and on the use of a new medium within which reflected power relationships and modes of organisation from other fields have yet to become firmly established among relevant actors.

4.5.1.7 Reporting

Continuing to draw on Kvale's methodology, data compiled and verified from the interviews was written up as part of a discursive chapter (Chapter 7) emphasising readable, concise reporting of the themes explored, to provide insight on a number of research sub-questions related to actor knowledge and behaviour underpinning Twitter communication. Reporting of the interviews required the use of indicative quotes from participants to evidence analysis.

4.6 Pilot Study

4.6.1 Determining the research case

In order to begin to understand and bound any concept of "Irish political discourse" on Twitter, it was necessary to establish foundational insights around what this might entail, who might be involved, and to what extent, etc. A pilot study was designed in February-March 2014, intended to constitute the kind of *digital ethnography* espoused by researchers such as Murthy (2011) and others, who seek to apply computational variants of ethnographic techniques for the study of group interactions in localised space, to those of a computermediated social space. A similar undertaking is attested to in some studies in the literature, which seek to build, from first principles, a working set of active users who discuss specific political topics (e.g. Maireder & Ausserhofer 2013). Following this broad paradigm, the pilot study first determined a set of indicative search terms that might identify contributions to political topics of interest and the users who made them. The intent was to create a list of hashtags and keywords related to political discourse in Ireland specifically, drawing on researcher knowledge of the area; it was important to exclude from the set those terms or tags which might capture a more geographically diverse discourse, or an entirely irrelevant one, so hashtags and terms were initially individually screened for a one-week period, with the resultant data examined manually for trends that might warrant exclusions. Terms sufficiently common to transcend an Irish context - words like "politics", "government", or "referendum" - were immediately excluded. Some terms, especially some hashtags that were commonly used within an Irish context to indicate a relevant topic, overlapped other topics and uses elsewhere and could not be included; a good example is the hashtag #ff, used in political tweets by many Irish users to indicate the Fianna Fáil party, but also used very widely by Twitter users on Fridays for the "Follow Friday" phenomenon, wherein users recommend other accounts worth following on that day, using the tag. The suitability analysis process resulted in two final lists, of 51 hashtags and 44 keywords respectively, which demonstrated effectiveness in identifying aspects of Irish political discourse with a high degree of topicality.

4.6.1.1 Pilot Study: Hashtags & Keywords

The initial set of hashtags related to known topics within Irish political discourse was identified by the researcher, based on relationship of the topic to politics in a variety of ways:

1. Hashtags used to indicate *Irish political parties* (e.g. #fiannafail, #finegael, etc.) or derived from *Irish government terms* (e.g. #seanad, #dail, etc.)

- Established hashtags for current *electoral events* (e.g. #ee14, #ep2014, #seanref), *political events* (e.g. #ccven), or *constituencies* (e.g. #dubw)
- Hashtags promoted in the reportage of politics and current affairs (e.g. #VinB, #rtept, #rtedrivetime, #twip, etc.)
- Hashtags related to breaking political news, scandals and crisis events (e.g. #gsoc, #shatter, #anglotapes, #pantigate)

The hashtag set was then supplemented by keywords likely to occur in tweets related to Irish politics. Some of these keywords were identical to hashtags identified (e.g. seanad, pantigate), but most were peculiarly Irish political terms unlikely to occur as hashtags but likely to be mentioned in text (e.g. *oireachtas, taoiseach, tanaiste, ceann comhairle*, etc.)

4.6.1.2 Pilot Study: Interactions with identified users

A second source of data collected for the pilot study was selected by the capture of tweets directed at identified actors from the fields of politics and media. Tweets containing @-mentions of these actors were likely to constitute either replies to them, or attempts to attract their attention, and analysis of a sample of these showed a very high proportion (96% for 250 randomly selected tweets) showed a text topic that was identifiably related to political discourse. Maireder and Ausserhofer's (2013) study of Austrian political discourse used a similar selection method.

The set of users (n=252), with whom Twitter correspondence was captured, included all members of the *Oireachtas* with a Twitter account, the official accounts of all Irish political parties, accounts related to politics reporting in Ireland, specific accounts of journalists identified as political correspondents working in Ireland, public policy institutes, university politics departments & their identified scholars.

4.6.1.3 Pilot Study: Results, Refinement and Findings

Over an initial 60 day period, (12 January – 13 March 2014) these tracking sets together gathered a combined dataset (n=238,660) of unique twitter interactions⁵⁵ (tweet objects) from >6000 users.

Users archived in this set were then ranked by frequency of interaction, number of hashtags and keywords used, and by appearance in the set of interactions with established users by

⁵⁵ The total set was based on the individual tracking of many of the keywords, hashtags, etc., and a substantial overlap was found, constituting tweets which were recorded in more than one query simultaneously; e.g. a tweet might contain both the keyword 'seanad' and the hashtag '#ccven'

role. A shortlist of users was selected from the total (n=311), the cohort of which met criteria of contribution frequency (average of >3 daily tweets for at least one hashtag or keyword over the period), and communicative scope (at least 3 hashtags or keywords used during the period OR at least 3 identified political actors interacted with); most users in the set demonstrated substantially greater metric values than these minima. This interim working set could therefore reasonably be shown to be representative of active participants in the on-going discourse around specifically Irish political issues in early 2014. From this, the set of the top 100 users by frequency of contribution was deemed the pilot study set of most active contributors, and it was from this set that interviewees would later be selected for the qualitative phase of this research. The total Twitter output for the 100 users in this pilot study set was followed, beginning 21 March 2014, and was analysed concurrent with the collection of the local election tweets (tagged #le14) during early to mid-2014.

The pilot study set was subject to repeated monthly analysis, recalculating based on the criteria above for each 30-day period until June 2014, in an effort to minimise any unintentional selection bias grounded in the particular temporality of contributions to the discourse made during the initial pilot phase. In fact, the set of 100 most active users only changed by the addition of 9 new users over that period, based on the initial criteria for inclusion. The nine users displaced during the period of the adjustments continued to be tracked, though where detail is provided in this research of the extent of inclusion of pilot set users in a particular case study, this is a measurement of the inclusion of the top 100 users at the end of the study.

The same re-assessment of the pilot study set was carried out in 2015, during the early phase of the second major case study – the 2015 Marriage Referendum - and demonstrated considerably greater change; by April 2015, 36 new users now appeared in the top 100, based on these metrics. This is discussed in Chapter 6 of this work, which examines the #marref case study in detail. Overall, the working set first obtained via the selection methods outlined here yielded a largely stable representation of highly active users within this discourse over the two-year period.

Examination of the users in the pilot study set took place to assess the actor categories to which each might belong, based on their association to the field of politics. This categorisation followed that of several studies in the literature where actors were classified by role (Bruns & Stieglitz 2013; D'heer & Verdegem 2014; Maireder & Ausserhofer 2013), in initially proposing three categories; (1) professional politicians, (2) media, and (3) citizens.

Examination of the interaction between users in the working set, over the period of the pilot study, showed broadly similar communicative trends to studies in the literature (especially Maireder and Ausserhofer's work, which was also most similar in method, See Chapter 5). The process of evaluating the user data for this three-category coding led to some important insights, however. It became clear that (1) professional politicians at both national and local level were in the set, and (2) both professional media and bloggers were in the set, and it was difficult to classify these according to the somewhat vague definitions of "media" given in some studies. In this tripartite classification, the remaining "citizen" category became a default for any user not easily fitting the first two.

4.6.1.3.1 Separation of professional and other media actors

A determination was made to split the media category into *professional media* and *other media*, such that these two sub-groups could be taken together if an exactly similar category comparison was desired with the existing studies, but the possibility to independently analyse either was preserved. This decision was motived by the substantial number of *other media* users in the set - a contingent of highly active bloggers, who seemed to exhibit different communicative strategies with politicians than did the *professional media* actors, and who had a substantially higher frequency of contributions.

The category split for media prompted consideration of whether it might be methodologically useful to also split politicians into local and national sub-classifications. It was less immediately clear, than in the case of media actors, whether communicative behaviors strongly varied between these two, so a quantitative investigation was designed to assess this.

4.6.1.3.2 Analysis of national vs local professional politicians

As explored earlier in this chapter, metadata representing certain actions related to a tweet, or inclusions in it, are useful quantitative variables for analysis of Twitter behaviour. Following the example of Bruns and Burgess's (2013) work on identification of communication patterns in Twitter discourse around crises, the pilot study used the metrics of *percentage of user's tweets which were retweets of others* and *percentage of user's tweets which contained an* @-mention to assess factors of information sharing (the users retweeting others to share messages, show agreement, etc.) and conversational directionality (the user is deliberatively addressing another user, indicating conversation, replies, reciprocity, etc). Plotting these percentage metrics along y and x axes of a graph, a scatter-plot of all tweets containing the #le14 hashtag⁵⁶ during the period of study was created (Figure 4.4, below).

Fig. 4.4Case study set contributors who mention #le14 (01 - 30 March
2014)



This graph shows three clusters for users who exhibit similar scores in both metrics, indicating popular modes where a particular balance of content retweeting and addressed interaction with others occurs. The same graph was then reproduced, highlighting those users who were identified as professional politicians. This new graph (Figure 4.5) indicates a strong presence of professional political actors among two of the user clusters.

Fig. 4.5 Graph from Fig. 4.4, highlighting professional politicians



The clusters substantially occupied by professional politicians correspond to two communicative modes: (1) very low levels of retweeting combined with low levels of

⁵⁶ The hashtag for the Irish Local Elections of 2014; this was the most-used hashtag of those tracked in the pilot study, and the selector for one of the two subsequent major case studies in this work.

addressed interaction - the users are broadcasting original information to their Twitter followers but showing little reciprocity, and (2) a medium level of retweeting combined with low and varying levels of addressed interaction -20-40% of total content tweeted by these users came from others and was retweeted to their followers, while only a slightly more directed conversational behaviour is seen.





A final pass of the data highlighted the national and local politicians separately (Figure 4.6) and demonstrated that there is a clear behavioural divide between the two, particularly in regard to their tendency to retweet content related to the #le14 hashtag.

Inductive reasoning might offer a likely postulate to explain the observed tendency; during the early period leading up to local elections, national politicians tweeting press releases or otherwise "broadcasting" party news or information about the timing of the elections, may be being retweeted by local politicians seeking contextualisation in a national party or to attract followers by useful information provision. This pattern could have been explained more precisely by additional analyses, such as a measure of the extent of URL inclusion, for example, but for the needs of the pilot study, these findings were sufficient to justify the initial postulate that separate coding of local and national political actors might usefully capture differences in their communicative behaviour.

4.6.1.3.3 Separation of "Activists" from the "Citizen" category

A final departure from the three-category actor coding of the literature involved the separation out of "activists", who were otherwise accounted for among the citizen default cohort. This action arose from the notion that "citizen" was, in practice, a very broad term within this model, when it might be interpreted by researchers to mean the ordinary users of Twitter who do not have any professional association with the field of politics. Including

those who work with NGOs, lobby groups, issue and awareness organisations, etc., seemed at odds with this. It also seemed likely (and was borne out in initial examination of a number of such identified users in the working set) that activists might exhibit unique communicative strategies in response to their need to build cause awareness, target specific politicians, interact with media and the public, etc.

4.6.1.3.4 Creation of an actor coding schema for users

Code	Category	Criteria
1	Politicians: National	Current or former professional political representatives at national level (e.g. TDs, Senators) and their identified employees and agents;
		National Political Parties.
2	Politicians: Local	Current, former and candidate professional politicians at local
		(council) level and their identified employees and agents; local party political groups
3	Activists	Groups or individuals who are professionally associated with politics but are neither representative politicians nor media agents, e.g. NGOs, local issue groups, lobbyists.
4	Professional Media	Outlets and journalists that publish reportage of politics via mass media
5	Other Media	Groups or individuals who publish blogs, opinions, or private reportage of politics via online means, but who are not professional journalists
6	Citizens	Actors commenting on or engaging in discussion of politics, who are not identified as professionally active in the field.

Table 4.1:Actor coding schema for Twitter users engaged in Irish political
discourse

A useful result of the pilot study was the creation of a demonstrably valuable categorisation system for actors within Irish political discourse on Twitter. Code numbers for actor type were set, along with definitions developed in the processes outlined here, resulting in the table (4.1) above.

4.6.1.4 Selection of candidates for interview

The pilot study for this work formed part of the exploratory quantitative phase of the mixed methods design for participant selection outlined, and as such became both an investigative endeavour for the determination and first stage testing of quantitative variables and coding, as well as supplying a useful selection method for qualitative research. The design phase of Kvale's interview methodology (See: section 4.5) calls for the selection of candidates who can be identified as pertinent to the themes of the research and likely to offer deep insight from a worldview under investigation. The pilot study's role, both in generating a shortlist of highly active Twitter users from the discourse under examination, as well as refining a typology of those users, provided the selection mechanism for a series of interviews with archetypal participants likely to offer highly relevant insight.
Users from the various actor categories (Table 4.1), could be ranked by overall frequency of contribution during the pilot period, by frequency of contribution on a particular issue identified by a hashtag or hashtag/keyword set, or by other metrics applied to the pilot data.

The working set of highly active users thus provided a source of participants for interview, who were contacted (via Twitter) in the design phase of the interview portion of this work. A pilot interview took place with the first participant contacted; a 'citizen' designated C1 in the research interview participant (see chapter addenda; Appendix A.4.6).

4.6.1.5 Caveats for the pilot study

The pilot study undertaken in this work is subject to some important limitations and caveats:

- Choice of accounts by role is more subjective, outside of government and party members, and interactions with any identified are weighted equally in determining the inclusion criteria for active users
- 2. Choice of hashtags/keywords provide spectra of Irish-centric relevance but are treated equally, potentially resulting in the capture of highly niche subsets of a broader political discourse
- 3. Sporadic, high-frequency, single issue contributors may engulf regular multi-issue contributors, making a recognition of the temporality of the study key to assessing its legitimacy. E.g. "*Pantigate*", an Irish political controversy from February 2014, was especially widely discussed on social media at the time of the initial pilot study, and many of the top contributors overall related to that topic only. Steps to require multiple topic contributions, or interactions with identified key actors, as well as repeated monthly re-assessments of the cohort, were taken to mitigate this bias, but a tension exists between a desire to reflect the broad inter-topic nature of a political discourse, without excluding actors, such as single-issue activists, whose contributions and position in the network may be very important and persist over time.
- 4. Original, retweeted and @-tagged content was treated equally, removing any conversational context from the contributions captured. The pilot study identified active users by the topic identification, frequency and addressivity of their contributions alone, and did not make any exclusions based on niche communicative strategies; the selection mechanism used would not screen out a user who only retweeted the content of others, but did so frequently and across multiple topics, for instance. While this was not the case for any user in the study, it remains a concern to be noted for any development of these pilot methods.

4.7 Conclusions

This chapter has outlined a *post-positivist*, specifically mixed methods approach to the research questions of this work, describing available research methods and their suitability to the segmented goals constituting the overall aim of improvement of methodologies for Twitter discourse analysis. The chapter argued that the combination of understanding the structure of *Big Data* available from Twitter, developing tools to effectively collect it, testing qualitative methods for their efficacy in interpreting and analysing that data, and informing these processes with insights from users who created the data, constitutes a multi-staged and mixed approach that directly addresses many gaps in the literature.

Examination of the structured nature and sources of *big data* offered by Twitter's APIs made clear the scope of complex analysis possible at each data point, through programmed selection upon these various fields. While it has been shown that much of the literature focuses narrowly on a small number of tweet object fields, and fails to adequately examine the suitability and effect of data collection modalities, it is clear from this chapter's analysis that scope exists to establish improved practices for working with Twitter data, contingent on the capture and analysis of richer information in response to the needs of specific questions.

Having justified this overall approach, the chapter set out details of the methods of implementation of the various stages of research, including digital, quantitative and qualitative approaches chosen for their appropriateness of use with Twitter datasets, and for their investigative potential.

This information concludes the section of the thesis that deals with theory, data and method. The next chapters move on to the empirical aspects of the research begun in the pilot study outlined at the end of this section, and describe the findings on effectiveness of various methods when applied to specific case studies. These case studies also provide insight into the nature, structure and participation in Irish political discourse on Twitter.

Chapter 5

2014 Irish Local Elections (#le14) Case Study

5. 2014 Irish Local Elections (#le14) Case Study

5.1 Introduction

This thesis has established that Twitter has emerged as an important sub-network within internet-mediated spaces where political issues are publically negotiated, through its increasing use by professional politicians, journalists, activists, and citizens. The number and range of investigations into the structure of - and participation in - political communication on Twitter supports this and form a basis for developing an investigative framework.

This chapter applies a methodological framework of investigation to Twitter discussion of the 2014 Local Elections in Ireland, and provides the findings and analysis of the data collected. It outlines related literature, both in methodological approach and in findings derived from similar data and setting, and offers an analysis of the structure and participation of an Irish network of Twitter participants.

Chapter 3 set out those aspects of a quantitative analytical framework for political discourse on Twitter which have emerged in the literature (See especially: Bruns & Burgess 2011; D'heer & Verdegem 2014; Larsson & Moe 2011). These tools and methods have been used effectively in event-centred political discourse analysis on Twitter, focusing mostly on investigating the networked structure of the mediated conversation around specific political events. Discussion and follower networks in Twitter are highly geographically localised (Quercia et al. 2012) and a number of studies have sought to investigate the structure of electoral and national discourses in various locales: e.g. Sweden (Larsson & Moe 2011); Queensland, Australia (Bruns & Highfield 2013); Flanders, Belgium (D'heer & Verdegem 2014); Austria (Ausserhofer & Maireder 2013). These studies have demonstrated that the mediated space of Twitter is one that affords interaction between citizens and professional politicians, and that while social structures from outside Twitter (e.g. political party power structures, relationships between politicians and media agents) are represented or transposed to it, new engagements – especially from citizen agents – are a feature.

These methods are seen as a recent departure from traditionally content-based analyses of campaigns and elections in Political Science (Lilleker & Vedel 2013), moving instead toward consideration of the *networked nature* of the Internet (and specifically Twitter) and the application of associated tools and methods from Computer Science (especially Network Analytics; topology and distribution analyses, etc.). These studies form the basis for a set of tools and approaches designed for Twitter analysis – the beginning of a new method called

for by Rogers (2010) and others to respond to the growth in use of Twitter for public discourse with a toolset and approach specifically developed for this mediated space, rather than attempting to apply older methods designed for incompatible discourse spaces.

5.1.1 Research Design

This focus of this case study is the network of interactions created by those users who tweeted about the 2014 Irish Local Elections and explicitly marked their messages as pertaining to this topic by the use of the #le14 tag; it seeks to examine that network by answering a series of research questions:

- RQ.5A: Who are the participants in the #le14 Twitter discussion and how can they be meaningfully categorised?
- RQ.5B: How do these participants interact with one another and do their modes of interaction vary according to role (media agents, professional politicians, activists, etc.)?
- RQ.5C: How can analytical methodologies for networks be applied to investigate the patterns of interaction in Twitter discourse, using this case study as an example?

To answer these questions, an initial method was developed based on the approach used by Queensland University of Technology researchers (Bruns & Burgess 2011; T. Highfield et al. 2011; Tim Highfield 2012), which analyses a hashtag-centred network through examination of the indicators of actor influence within it. Specifically, this approach utilises the addressivity (Honeycutt & Herring 2009) of the @-mention function in Twitter to derive an overall pattern of discourse directionality and a metric of actor influence based on the number of tweets sent to them. These researchers used this metric to demonstrate patterns of interaction as well as phases of discussion linked to specific users and groups.

This method was augmented by the inclusion of other forms of analysis of the network, including examination of its *centrality* and *density*, metrics detailed and examined later in this chapter, and descriptive analysis based on algorithmic examination of the content and inclusions of the tweets collected.

5.2 Data Collection

Collecting Twitter data (tweets and metadata) begins with identifying the topic of interest using a keyword(s) or hashtag(s), and requires the use of Twitter's Application Programming Interfaces (APIs), as previously outlined. This case study follows the literature's wellestablished methodology⁵⁷ of using the open-source "yourTwapperKeeper" (yTK) software (O'Brien 2013) for the capture and storage of Twitter data. The origin and usage of this software is discussed in the previous chapter.

A slight upgrade to the standard version of yTK was used for the collation of the #le14 data, and iterations of new data collection software were piloted alongside it for comparison of data archived⁵⁸. The version used (*yTK_DCU_v1*) was developed in conjunction with yTK's author, John O'Brien, and consisted of changes only to an issue of database interfacing during storage and retrieval owing to the software having been built, several years previously, on PHP programming structures now deprecated. Thus, the software used for #le14 collection approximates the yTK version available on open-source sharing site GitHub in all aspects of functionality save that it had been updated to function with the latest version of PHP at the time of the study. Concurrent improvements to yTK undertaken by the researcher (and discussed in more detail in the previous chapter) led to a more significantly altered version of it being deployed for the second case study, though in this case the collection ability of the software approximates exactly its historical use and meaningful comparison may be made to other studies in the literature that have used it.

The yTK software ran on a webserver based on the Irish Higher Education Authority's network backbone⁵⁹ and collected data via Twitter's Streaming API. Various approaches to collection are possible and the degree of data capture achieved is subject to limits imposed by the particular API used (e.g. Twitter's Streaming API is *rate limited*, allowing only a fixed amount of results per minute to be returned). The ideal data capture framework is one that collects the entire Twitter output for a certain period and extracts all relevant data points. In

⁵⁷ Bruns' (2012a) paper on mapping conversation patterns on Twitter has become something of a manual for the process, as it offers both an endorsement of the open-source tool approach as well as a guide to how the data might be meaningfully explored. It is frequently cited in the subsequent literature and its methods are a foundation for the studies cited in the introduction of this chapter. ⁵⁸ Part of this research involved the development of this software to respond to data collection shortcomings identified in the literature and in testing such as this case study. (See: Chapter 4) ⁵⁹ In the case of this study, yTK ran for data collection on a webserver physically located on Dublin City University's network (http://multimedia.dcu.ie/twina). This server ceased operation at the conclusion of this study, as data collection for this research was moved to an Amazon Web Services (AWS) hosted server (http://twarchive.dcu.ie/comms)

practice, however, the vast amount of tweets produced (about 500 million per day), means that all studies using Twitter data need a sampling process that relies on some search criteria (hashtags and/or keywords included) to identify a meaningful subset of tweets for archiving and subsequent examination. yTK uses the *Twitter Firehose* (Streaming API), which connects in real time to the totality of output from Twitter, sorting and saving tweets and their metadata according to such specified search criteria. Some limits must therefore necessarily exist in situations where the current rate of relevant Twitter contributions for a given search term exceed the rate of results allowed. Similarly, the database transaction (parsing and storing the incoming JSON data object) takes time in each cycle and this process (a factor of the efficiency with which it is coded, the hardware of the server, etc.) will also limit the overall rate of collection.

For this study, only tweets containing the established hashtag for the 2014 Local Elections in Ireland - #le14 - were captured⁶⁰. This hashtag arose from a conventional development among Irish Twitter users to apply the format "le" followed by the year to local election discussion. The same convention has developed around general elections (e.g. #ge11, #ge16) and was seen in an attempt to use #ee14 for the European Parliament elections, which ran concurrently with #le14⁶¹. The #le14 tag saw widespread adoption among users and was promoted in broadcast and printed media (e.g. appearing onscreen as a suggested way for users to take part in conversation around the election arising from current affairs programming on RTÉ television).

Formal data collection was conducted between 21 March 2014 and 21 June 2014, with the election concerned taking place on 23 May 2014, allowing for the capture of discourse leading up to and including election, as well as subsequent discussion of the aftermath.

The final dataset includes 112,689 tweets containing the hashtag #le14, and their associated metadata. The frequency of #le14 tweeting increases sharply approximately one month prior

⁶⁰ The extent to which a corpus of tweets gathered via hashtag selection is representative of the totality of conversation on a topic is a substantial question of this thesis and is discussed elsewhere. See Section 5.3.5 for caveats specific to this study and its methods.

⁶¹ The comparatively divergent use of #ep2014, #eu14, #ee14, and #ep14 all to mark conversation around the European elections is a by-product of the supra-national nature of the European Parliament, causing Twitter discussions – and therefore associated hashtags – to appear in many countries and to be cross-used. This forms an interesting contrast with the relative consistency of the #le14 hashtag, which was established from Irish Twitter convention rather than simultaneously imported from multiple sources. Clear (and widespread) discussions can be observed in the Irish *Twittersphere*, in which the suggestion to use #ee14 is proposed in order to maintain the established convention (and low character count), but ultimately the greater international use of #ep2014 won out, resulting in its widespread adoption in Ireland too. Such cases point to the importance of understanding and tracking the conventions and usage of hashtags in studies such as these.

to the election, increases slowly in the week prior to voting day, peaks on the subsequent day (when the counts take place and results are announced) and falls sharply.



Fig. 5.1 Tweets per day during the #le14 studied period

5.3 Evaluation

While Twitter data collection can rely on tools interfacing with structured API services yielding a standardised format of data, its subsequent analysis becomes challenging because the data are both less consistent (e.g., text constraints, informal expressions, in-group communicative conventions) and more enriched (e.g., accompanying user profiles, follower-following-dynamics, hashtag use, URL inclusion) than traditional data found in electoral and political discourse studies (e.g. survey data). The use of diverse research methods and metrics is necessary to extract intelligence from the highly enriched and unstructured social media data (Fan & Gordon 2014; Zeng et al. 2010), though any well-established methodological framework of analysis is absent for the field. With this in mind, this chapter also focuses on developing an analytical approach encompassing relevant but disparate research methods and metrics in the literature, in order to answer the outlined RQs. It applies computational methods of Network Analysis (NA) and Descriptive Analysis (DA) to the #le14 dataset, yielding insights into the patterns of participation by individuals and groups that also informs detailed interviewing in the qualitative data collection elsewhere in this thesis.

5.3.1 Manual Coding of Data

The set of tweets comprising the archive for #le14 was sent by a total of 17,994 unique Twitter users. Metadata fetched for each user, comprising their specified full name, description ("bio"), location (if offered), etc., was used to manually code them according to actor categories. This process was initially a slow iterative one, following the conventional manner outlined in some studies in the literature, where the Twitter profile page for each user is examined and the determination of their category made in an accompanying application. As chapter 4 has detailed, the development by the researcher of the yourUserCoder (yUC) software greatly increased the rate of coding, and allowed user codes to be stored in and retrieved from the same SQL database in which the yTK-collected Twitter data resided.

Analyses of political discourse on Twitter commonly use a simplified coding structure for identifying participants for the purpose of grouping and examining conversation dynamics apparent in their messaging, though scholars are careful to acknowledge that the nature of identity online, and specifically on Twitter, makes it problematic to code in any traditional categorisation scheme (Lewis et al. 2013). D'heer and Verdegem (2014) adopt a threecategory model in their comparable analysis of the 2012 Belgian (Flanders) local election, designating participants as either political-, media-, or citizen-agent based on their username and description. While identifying the inherent problems in using the self-definition of these user-created data as identifiers for the purposes of the study, their work is nonetheless in line with previous studies of this type (e.g. Larsson & Moe 2011), which have similarly relied on these identifiers for simplified coding of discourse participants.

The threefold coding structure of politician / media / citizen is problematic however - not only because of the identifiers used, but because of the difficulty in fitting these categories to some cases where users overlap them, where single real-world agents may be represented by multiple Twitter accounts, or where agents' self-identified professional identity does not clearly fit with the simple scheme.

D'heer and Verdegem identify the first two categories, political and media agents, with salient profession identities (2014 p. 7) and include Twitter accounts representing political parties, as well as those used by individual professional politicians at national and local levels in the same grouping. Similarly, the official Twitter accounts of media outlets (e.g. newspapers) are categorised together with those of individual journalists. Nuance based on the type of agent within the categorising cohort is sacrificed in favour of providing the relative simplicity of the three-category analysis. For the third category (citizen agents), however, their choice to exclude those users and accounts publicly affiliated with NGOs, political movements, events, unions, or municipalities, represents a substantial alteration to the network structure.

A similar, though extended model appears in Ausserhofer & Maireder's (2013) study of Austrian national politics. Here the politician / media / citizen model is extended to include a fourth category; "experts", who are professional actors not fitting the journalist or media categories (e.g. lobbyists, political scientists). This model is useful, in not requiring the removal of a potentially large number of actors who do not fit the earlier three-category approach, though some problems remain in the separation of actors between "expert" and "citizen"⁶².

In the coding of the #le14 data, the researcher opted for a further extension of these models into a total of six categories; (1) National Politicians, (2) Local Politicians, (3) Activists, (4) Professional Media Agents, (5) Other Media Agents, (6) Citizens. Further detail of the criteria for each category is given in the table (5.1) below. A further coding of (0) was applied to Twitter accounts among the participant cohort which were not relevant to the discourse network, despite tweeting using the #le14 hashtag; these largely belong to *spam-bots*, which

⁶² Definitions in the work (2013 p. 298) leave the coding of citizen activists and political bloggers somewhat ambiguous, based on a subjective assessment their professional association with politics.

frequently tweet advertising or other linked information using popular hashtags. A total of 802 actors were coded thus, and 1591 tweets originating from them were removed from the analysis.

Code	Category	Criteria
0	Irrelevant	Spam-bots or other unidentifiable and suspicious accounts
1	Politicians: National	Current or former professional political representatives at national level (e.g. TDs, Senators) and their identified employees and agents; National Political Parties.
2	Politicians: Local	Current, former and candidate professional politicians at local (council) level and their identified employees and agents; local party political groups
3	Activists	Groups or individuals who are professionally associated with politics but are neither representative politicians nor media agents, e.g. NGOs, local issue groups, lobbyists.
4	Professional Media	Outlets and journalists that publish reportage of politics via mass media
5	Other Media	Groups or individuals who publish blogs, opinions, or private reportage of politics via online means, but who are not professional journalists
6	Citizens	Actors commenting on or engaging in discussion of politics, who are not identified as professionally active in the field.

 Table 5.1:
 Definitions of annotated actor categories

The separation of professional politicians into two sub-groups arose from the findings in the pilot study for this research (see chapter 4) that national and local politicians behaved very differently on Twitter in their modes of communication and follower interaction. The studied event, being a local election, also prompted an increased focus on this potential difference. Similarly, the structure and ownership of large professional media outlets in Ireland has prompted a substantial national discourse in recent years and the development of a large cohort of non-professional media agents (bloggers, online publications, satire sites, etc.). The separation of these from the established professional media provides a means to discern whether some of the notions of Dutton's (2009) Fifth Estate can be demonstrated in the Irish context. These separations, together with the 'Activists' category comprising NGOs, lobby groups, etc., allow for a more meaningful remaining 'Citizens' category in which users who are active in discussing politics, though not professionally associated with it, are captured. Notably too, this categorisation system removed the requirement to exclude any agents from the analysis, while still maintaining an ability to group categories together to form comparative analyses with existing studies (e.g. Categories 1-4 and categories 5 and 6 may be taken as two groups representing the professional / non-professional actor categories seen in previous analyses).

5.3.2 Network Analytics

The methodology chapter of this thesis (see section 4.4.2) outlines how the tools of Network Analysis have come to be used by computational social science researchers investigating discourse networks on Twitter. Such networks are constructed (as elsewhere in computer science from where this framework is borrowed), as a collection of *nodes* (points – in this case Twitter users) and *edges* (connections between them, also termed *ties*).

In the construction of a *mention network*, these ties between users are determined through examination of that portion of the total dataset of #le14 tweets sent that contain @-mentions of users. As outlined in chapter 4, @-mentions (also termed @-replies when they specifically target a user whom a tweet is being sent in response to) are addressivity affordances in Twitter, which allow for the specific direction of a tweet toward a named user. By computational selection of each @-mentions, a tie can be recorded between the user who sent the tweet containing it, and the user the tweet was directed to using it. In some cases a single tweet may contain multiple @-mentions, resulting in the recording of several ties between the sender and those mentioned users. The total set of users connected by such ties is the *mention network* derived from the dataset, and represents a map of conversational directed interactions recorded. For the purposes of most network analyses, the direction of the tie is unimportant (Scott & Carrington 2014), however in the case of measuring reciprocity – the extent to which replies were forthcoming from users who had a message directed to them – the recording of ties as directional is important.

As chapter 4 described, a second *connection network* of users and the ties between them may be constructed by examination of the followers of each user, recording a tie between a user and each other that follows them. Since the total Twitter output of any given user is theoretically possible to be seen by their followers⁶³, the user-follower relationship is important to consider. Influential users, with many followers, are more likely to have their content seen, and therefore potentially replied to or retweeted. A computational determination of all the follower ties between the users in the dataset provides a much more interconnected structural

⁶³ In practice, of course, the timeline of any given user is filled with the latest tweets of every other they follow, meaning that even when constantly updated, the user will only see what tweets they are capable of discerning in real time during the time they log in. This is further complicated by a change made by Twitter, beginning in Autumn 2014, to how the timeline operates. This change, which approximates features of the Facebook timeline, uses unrevealed algorithmic determinations of the 'popularity' of tweets from amongst those followed, 'holding' these in the timeline for when the user next logs in, resulting in a 'While you were away' collection of popular missed tweets. Thus a user will see recent algorithmically selected tweets from the last 48 hours, as well as the real time output of all users they follow.

network, which (unlike the dynamic and ephemeral *mention network* composed of directed tweets) represents the more permanent status quo of how users are connected to one another.

In this study, meta-data collected for the captured #le14 tweets was used to form the connection and mention networks for the set, which were subject to the centrality and network density analysis discussed in the methodology of this work, to determine patterns of communication among the coded participants.

5.3.3 Descriptive Analysis (DA)

In this study, the corpus of 111,098 tweets remaining after the exclusion of those tweets associated with category 0 actors was algorithmically coded using GAWK to parse the structured data and add fields classifying certain properties of each tweet: (1) whether a retweet (RT) of an existing message, (2) whether containing an @-mention, (3) whether containing a URL, (4) whether containing other hashtags. Of those tweets that contained @-mentions, a secondary algorithmic analysis took place to mark those which contained the @-mention at the start of the message, indicating that it is a reply to another user, or a message explicitly intended for another user, rather than one which incidentally mentions them. A further pass of this cohort of tweets marked those where an initial @-mention is preceded by a full stop; this indicates a message intended for a particular user, perhaps in reply, being intentionally made public⁶⁴. Of those tweets that contained other hashtags, the specific tags were archived to analyse trends in cross-topic categorisation by actors.

5.3.4 Framework

This approach seeks to blend methods employed in previous studies to achieve a comprehensive set of reproducible results for other events or locales, allowing meaningful comparative investigation with greater explanatory nuance than has been previously achieved. The manual coding of the actors involved, and algorithmically derived coding of the tweets sent, facilitates network and descriptive analyses to provide useful insight for this Irish electoral discussion case study.

While only these types of analyses are introduced, other descriptive analyses and metrics are certainly possible and facilitated by available metadata (e.g. geolocation information enables geographic description of messaging and modules, timestamps enable chronological analysis of when discourse peaks occur in daily or weekly cycles). Scholars note, however that "*the use*

⁶⁴ For discussion of the conventions associated with addressivity in tweets, see Section 1.1.1

of too many metrics is likely to cause information overload and confusion, rather than intelligence" (Chae 2015 p. 3). The DA metrics chosen for this study complement the NA methods in uncovering patterns of interaction at the macro level, and assist in generating an understanding of the discourse captured as a set of communicative relationships between actor groups.

The methodological framework may then be outlined as follows:

- Capture and archiving of all tweets containing a given hashtag for a defined period.
- Algorithmic coding of these tweets according to descriptive criteria (e.g. contains URL, contains hashtag.)
- Human coding of all actors responsible for the sending of these tweets, according to the category system outlined.
- Descriptive Analytics based on tweet content, overall and per actor category
- Network Analytics to investigate inter-category dynamics and group cohesion

Finally, several of the analytics used may yield additional insight when conducted per-period for divisions of the overall duration of discourse captured; e.g. D'heer and Verdegem demonstrate the efficacy of performing centrality analysis over defined periods of the Belgian local election campaign - pre-election, prior week, election day, post-election (2014 p. 10) – rather than as overall values for the dataset. The same periodic divisions are used in this study and correspond to the following dates:

- (P1) Pre-Election: 21 March 2014 15 May 2014;
- (P2) Prior Week: 16 May 2014 22 May 2014;
- (P3) Election Day: 23 May 2014;
- (P4) Post-Election: 24 May 2014 21 June 2014.

It should be noted that while the peak in frequency of tweeting #le14 occurs not on Election Day (P3), but on the first day of P4, when the counts take place (Fig 5.1). These periods then divide the discourse into pre-event, proximate pre-event, event, and post-event, and should capture discourse related to early campaigning, peak campaigning, voting itself, and discussion of results, respectively.

5.4 Results & Analysis

5.4.1 Overall Network Structure

The approach undertaken allows us to begin with an overview of the relative numbers by category of participant actors overall and per-period. Specifically, Table 5.2 provides the number of active users during the specific periods (P1 - P4) of the study, as well as in total (P*). Since the periods used are not equal (66 days, 7 days, 1 day, 29 days) and therefore not directly comparable to one another in absolute number of agents participating, the percentage participation by category for each period is a more appropriate metric of the dynamics of involvement at each stage. (e.g. Local politicians are substantially more represented as a percentage of overall participation during P2, than during P1).

Table 5.2	#le14 agents by coded category, by discourse period	Ĺ
-----------	---	---

	Nati Politi	ional icians	Local Po	oliticians	Acti	vists	Profes Me	ssional edia	Other	Media	Citi	zens	
	1	(%)	2	(%)	3	(%)	4	(%)	5	(%)	6	(%)	Total
P*	409	2.4	1914	11.1	121	0.7	131	0.8	635	3.7	13985	81.3	17195
P1	408	2.9	1366	9.7	87	0.6	123	0.9	621	4.4	11546	81.6	14151
P2	216	2.6	1909	22.7	121	1.4	126	1.5	337	4.0	5699	67.8	8408
P3	399	7.0	322	5.7	56	1.0	130	2.3	289	5.1	4495	79.0	5692
P4	286	1.9	1469	9.9	90	0.6	131	0.9	596	4.0	12318	82.7	14891

Some insights are immediately apparent in this breakdown: citizens (6), followed by local politicians (2), dominate the conversation at all stages; users participate in substantial numbers during the post-election period (P4), which includes the vote count; and a divergence in periodic participation by national versus local politicians is evident.

The objective structure of the network can be seen to change during each period, and the very high percentage of citizen actors supports the notion of Twitter as a medium of public discourse participation in the Irish context.

Similarly to other studies, the participation by politicians is seen to be greater in advance of the election; "It seems that politicians mobilize or inform (or are mobilized or informed) more before Election Day, or when the stakes are higher, than after Election Day" (D'heer & Verdegem 2014 p. 9). Notable in this analysis however, is the marked difference between how national and local politicians participate, as predicted by similar anomalies noted in the pilot study. The percentage presence of national politicians in the discourse of each period shows a sharp spike for Election Day (P3). This is likely the effect of a switch to mobilisation and support of linked local party candidates on voting day. Among local politicians, the sharp change in

participation comes in the week prior to the election, when candidates are motivated to engage voters and campaigning is most fervent. Local politicians are comparatively underrepresented on voting day itself, re-emerging in the post-election period, likely to announce count results and make statements.

Professional media presence is consistent throughout the periods, but achieves a greater weight in the discourse on Election Day and in the week prior. Other media's presence follows a similar trend, but exhibits a greater weight in P1, indicating a more substantial presence for bloggers and other non-professional agents earlier in the discourse. Finally, presence by activist agents peaks in the week prior to voting, likely corresponding to an agenda of lobbying politicians and mobilising voters.

This initial quantitative picture of the actor structure with the network, during key divisions of the period, forms a useful set of descriptions of the comparative presence of categorised participants, which is enhanced by examination of the connections between them and the type of messaging involved.

5.4.2 Network Metric Results

Network *density* is a simple measure of the communicative connectivity between users in the mention network, given as the number of @-mention links between them divided by the total number possible. For each period, this metric remained relatively stable and very low; (P1) 0.0009, (P2) 0.0021 (P3) 0.0019, (P4) 0.0011. These results show a network of very little interaction between agents, for a given total, but this is unsurprising for such a large data set; the very substantial number of citizens in the network makes the total possible connections extremely large. These figures are useful, however, in demonstrating (a) that the network is not tightly knit, and (b) that the network becomes denser immediately prior to and during the voting. Network centrality is a further useful measure, indicating the extent to which the network is centralised around particular actors; i.e. whether users tend to direct tweets toward the same actors, or distribute their directed messages widely within the overall cohort. Overall, for the network, this remains similarly low and stable during the four periods, becoming more centralised on and after the Election Day, likely as increasing numbers of users @-mention winning candidates⁶⁵.

A more nuanced measure is the degree to which each category cohort directs messages at actors in other categories, forming a picture of the *directionality* of interaction over the whole

⁶⁵ Overall centrality by period: (P1) 0.046; (P2) 0.048; (P3) 0.051; (P4) 0.053

#le14 period. Table 5.3 provides the percentage of each category's overall directed tweets that were sent to each other category, with X denoting those tweets which @-mention a user outside the sample.

Source :	1	2	3	4	5	6	X
1	10.3	22.6	7.3	16.1	2.2	2.2	39.3
2	14.7	33.1	3	9.1	12.2	16.2	11.7
3	22.9	28.2	2.1	13.6	0.7	3.6	28.9
4	10.6	16	0.9	20.5	0.2	0.7	51.1
5	16.5	26.6	3.6	11.6	10.3	9.1	22.3
6	11.2	13.9	0.3	9.2	0.3	19.3	45.8

Table 5.3Percentage of @-mentions by actor category, directed to actorswithin other categories (1-6), or outside the sample (X).

This analysis is useful in describing several aspects of the shape of conversation around #le14. Percentage of tweets with @-mentions of users not captured in the sample (i.e. messages directed to those outside the total cohort who used this hashtag) is lowest for local politicians. This is likely a function of their interaction being based strongly around responses to questions from others in the sample and the comparative jump in presence (See Table 5.2) immediately prior to voting, when they are responding to these mentions.

Excepting tweets directed outside the sample, it is the local politicians who are most often the ones that messages from others are directed to. (For all groups except citizens, they receive the most @-mentions of any category, and for citizens they receive the second-highest number). This is an interesting demonstration of the centrality of the local politicians to the discourse and might be expected, given the topic. It's worth noting, however, that 33.1% of @-mentions from local politicians was directed at other local politicians (i.e. in-group). This finding is less expected and may be the result of mentions of party colleagues by candidates, or local (constituency-level) discussions in which candidates mention one another or directly ask one another questions, performing debates on Twitter.

A further finding is that messages directed at non-professional media agents, such as bloggers (category 5), represent a very small percentage of those sent by most other groups, with significant exceptions for local politicians and group 5 itself. Here, it is likely (given that the reverse directionality is also evident) that local politicians are responding to questions arising from these media agents. It is interesting to note that it appears that local politicians exhibit significantly more willingness to engage with these users than do politicians at national level. The interactions for activists (3) provide some interesting insight, given that the percentage of tweets directed at them by other categories is very low, despite substantial proportions in the opposite direction. This likely describes the condition of using @-mentions to passively "tag" users rather than to specifically address a question to them; this usage can be seen in an examination of the tweets from the category 3 cohort, where many are links to statements which @-mention a political or professional media agent in order to get their attention, rather than posing a question demanding a directed reply in return⁶⁶.

When analysed across the periods, this matrix of addressivity between groups shows some slight variation, in line with what might be expected from the centrality and density measurements; total tweets directed to those outside the sample drop in number at each stage, as the #le14 discourse coheres, citizen mentions of local politicians grow at each stage, and national politician mentions of local politicians peaks strongly in P2, as endorsements and mobilisation become central to their interaction.

Because of the relatively large proportion of the total sample cohort in the citizen category, it is necessary to further elaborate on them, since this mode of analysis can disguise a larger reality of who among each category is actually addressing others. In the case of citizen actors, the mean value for other-users-mentioned is very small (0.002), while the total number of tweets with such mentions is high; this means that among the total citizens, a relatively small number are sending directed tweets, but those who are are sending many. It is important to note, then, that this overall picture captured in Table 5.3 is demonstrative of swathes of interaction between actor categories, but not of the level of participation in that interaction by the totality of the category.

An investigation of participation in directed messaging was undertaken in which the proportion of total users for each category who sent a tweet containing an @-mention was calculated: (1) 64.2%; (2) 31.9%; (3) 84.2%; (4) 71.4%; (5) 88.2%; (6) 5.9%. It is clear that though overall participation by citizens in the #le14 discourse is very high, the number who engage in directed tweeting (and therefore display conversational behaviour) is small and concentrated. The proportion of citizen users who sent more than 10 tweets containing @-mentions, indicating a sustained practice of addressed messaging rather than an incidental one, is smaller still at 0.6%.

⁶⁶ E.g "Hi @labour! Make sure your candidates don't miss our Citizen Engagement conference in Dublin next Thurs http://t.co/tm9LnQTzjl #le14", sent by user @DisabilityFed (Disability Federation of Ireland)

https://twitter.com/DisabilityFed/status/448073075184394000 (Accessed April 2015)

Finally, the @-mention network for #le14 was subjected to an External/Internal ratio analysis (E-I Index) of the sort used by D'heer and Verdegem in their investigation of the 2012 Belgian local elections, and based on standard Social Network Analysis (SNA) adoption (Scott & Carrington 2014) of the E-I index arising from organisational Network Theory (Krackhardt & Stern 1988). In this investigation type for Twitter, the ratio of directed tweets ("ties" in SNA terminology) from users within each category to others in their group (internal) versus to those outside their group (external) is presented as a value between -1 and 1.

[(Number of externally directed messages) – (Number of internally directed messages)] [Total Number of directed messages]

This calculation yields a ratio value where -1 represents all internal ties and +1 represents all external; i.e. higher values for the mean E-I index for a Twitter group/actor indicate that it is outwardly communicative (Hanneman 2005).

Table 5.4, below, shows the E-I index values for each category, during each period. The E-I index for the overall #le14 discourse period was positive for each category⁶⁷, indicating that none was disposed toward mostly in-group communication. However, the E-I value for local politicians was the lowest (0.338) – substantially lower than national politicians – indicating a tendency for this group toward greater internal directing of messages.

	1	2	3	4	5	6
P1	0.489	0.094	0.901	0.612	0.823	0.216
P2	0.507	0.053	0.922	0.589	0.927	0.963
P3	0.784	0.446	0.823	0.578	0.632	0.704
P4	0.801	0.487	0.956	0.436	0.793	0.451

Table 5.4 Group-level E-I indexes by period

These results reveal high levels of communication between groups; the discourse is an open one at the level of actor categorisation used in this coding model. Notable findings include: (a) the very high values for activists, during all periods, indicating highly externalised directionality in their tweeting; (b) sustained increase in values for both politician cohorts over the periods, indicating increasing interaction outside their groups with time; and, most significantly, (c) the very substantial rise and subsequent fall in E-I index value for citizens, who appear to go from the lowest levels of out-group messaging to the highest, in the week prior to the election, falling back to a low level post-voting. This trend for citizens may be

⁶⁷ Overall E-I indexes, by actor category: (1): 0.794; (2): 0.338; (3): 0.958; (4): 0.590; (5): 0.794; (6): 0.614

explained by a deviation from discourse behaviour centred on more in-group discussion before and after the event, to one of high out-group contact during the final stages of campaigning, when citizens engage more with politicians and media. Again, this result set demonstrates the additional insight gained by performing this analysis per period (cf. overall E-I values, which smooth the very significant changes in citizen behaviour).

5.4.3 Descriptive Metric Results

Of the total set of 111,098 tweets analysed, 61,845 (55.6%) were found to be retweets (RTs), and 10,199 (9.2%) contained @-mentions and represent directed/addressed messaging⁶⁸. 34,996 (31.5%) contained a second or further hashtag, the most popular of these being: #VinB, #ep2014, #ee14, #ep14, and #VoteNI2014 respectively. These additional markers indicate intersection of the discussion with other hashtag-defined topics, where the most popular (roughly equivalent in number to the next four combined) corresponds to a television current affairs program, which offered extensive coverage of the candidates and election. The remaining four of the hashtags most popularly intersecting #le14 correspond to the concurrent European Parliament elections (#ep2014, #ee14 and #ep14) and Northern Irish local elections (#VoteNI2014). These tweets, co-tagged with topic markers for the other elections, demonstrate that a substantial portion of the captured discourse involves commentary generalised to a wider political issue. Examination of the cohort of tweets cotagged with the European Parliament election hashtags, compared with those co-tagged #VinB shows a significantly higher degree of URL inclusion: 61% versus 9%. It is very likely therefore, that these tweets represent information-sharing among users, related to the election period generally. This component of the discourse - users directing others to linked resources (e.g. articles, count results) - is seen in other studies and is markedly different in nature to that part of the discourse related to television events. Moreover, those tweets cotagged #VinB exhibit a significantly higher degree of @-mention inclusion, than either those tagged for the EP election, or the set generally. Here, it is likely that the discourse is more directed because users are responding to one another in discussion of live events on the Vincent Brown television program represented. This is an example of the 'second screen' use of Twitter for "audiencing" described by Queensland University of Technology researchers, where users discuss a current television event in real time (Harrington et al. 2012; 2013).

⁶⁸ Of those tweets containing @-mentions, 7168 (70.3%) started with an @-mention, indicating a reply or addressed message. 518 of these used an initial .@, indicating an addressed message intended for public viewing. The majority (414) of these expressly public directed messages came from a small cohort of 8 highly active citizen users, indicating that this convention is likely a performative aspect of the Twitter presence of status-seeking agents.

Examining those tweets that were found to be RTs, an additional approximation of influence within the discourse may be devised based on the most-retweeted actors. At the overall level⁶⁹, professional media and national politicians are the most retweeted, reflecting established power / authority dynamics from outside the Twitter field. Activists are the least retweeted, and perhaps surprisingly, local politicians are substantially less retweeted than national, despite being the nominal focus of the #le14 discourse. This may reflect a greater weight of social and symbolic capital accrued to their national counterparts, making RTs of the national figure more desirable / likely. (See: Chapter 7).

	1	2	3	4	5	6
P1	22.4	1.1	0.2	32.2	26.1	18
P2	29.1	5.4	1.8	45.3	9.3	9.1
P3	30.4	5.2	0.8	48.9	13.1	1.6
P4	16.5	9.6	0.7	38.4	24.9	9.9

Table 5.5% of RTs originating in each actor category, by period

Again, the per-period findings of this analysis provide more specific and nuanced insight into the dynamics of influence within the #le14 discourse. While the overall dominance of established professional media as the source of retweeted content is clear, some additional trends are worth noting: (a) RTs of local politicians' messages rise throughout, from a very small proportion in the initial several weeks of campaigning - possibly indicating a growing recognition of candidates culminating in a desire by users to RT the count-related messages in the post-election phase; (b) non-professional media agents exhibit most influence in the earliest period - likely correlating with the early publication of information by bloggers in the period before the established media strongly enters the discourse - and return strongly postelection, likely because many of these agents publish count information for specific wards, and (c) citizens are a substantial source of retweeted material in the early part of the discourse but this dwindles starkly around voting day.

Combining these descriptive statistics, we can examine how many of the RTs contain URLs and additional hashtags and can see that the majority (71%) of RTs of professional media actors' messages contain a URL, demonstrating a sharing behaviour, likely of news stories related to the election published by a media outlet whose agents then tweet a link. We can see also that among the citizen messages retweeted, secondary hashtag inclusion is markedly higher than for other groups, and that these are a particularly substantial source of the #VinB hashtag, indicating that citizen agents are commenting on live television coverage of

⁶⁹ Total RTs in the #le14 corpus originate from the actor categories as follows: (1): 26.4%; (2): 5.3%; (3): 0.9%; (4): 41.2%; (5): 18.3%; (6): 9.6%

the election/candidates and are being retweeted for doing so. The same trend is present within the non-professional media cohort, which contains satire websites, amateur pundits responding to television events and aggregators of election content; the most retweeted user within category 5 is also the most prolific actor within the sample generally and is a blog tracking election literature (@electionlit).

It is clear from these findings that while citizens dominate the discourse in absolute number and in presence by period, the metrics associated with influence in the sample demonstrate that it lies chiefly with established actors and moves more toward these as the discourse develops. Citizens switch from a dynamic of greater in-group conversation directionality, where they are a source of RTs and engaging in information sharing in the earliest period of the discourse, toward radically out-group communication where they interact with established political and media agents and afford these increased influence in the network proximate to the election event.

5.4 Caveats and Limitations

Consistent with similar studies, the limitations of this investigation into Irish political discourse on Twitter chiefly relate to the data collection process; the use of hashtag-based datasets is a standard for such studies, though also acknowledged to have substantial bias. The issues related to the use (or not) of hashtags in content tweeted by actors in Irish political discourse networks on Twitter is dealt with in some detail elsewhere in this work (See: Chapters 1, 4, 7). The utility of the hashtag as selection criterion for inclusion in the set stems from the deliberative topic-marking action implied by its use; researchers may be unable to systematically track which other tweets from a user were related to the local elections, but understand that those containing #le14 are explicitly considered so by the users who tag them thus. The behaviours affecting and limiting these deliberative processes are explored in this thesis. Basing this study on the #le14 tag, however, situates it as a meaningful comparative case for similar studies, detailed below, that have used an established electoral hashtag to study political discourse around events in other locales.

It should be noted that with hashtag-based data collection, the issue of multiple topic associations converging on a single hashtag is possible; i.e. #le14 may come to mean something else in another locale, community, event network, etc., concurrent with this local election study and tweets unrelated to the election may be captured and dilute the analysis. While this does not seem to be the case here, and while "spam-tweets" (usually containing advertising links) that hijacked the hashtag for visibility to users were removed, it is possible in a dataset this large that some tweets are present which are irrelevant to the discourse intended for study. Moreover, tweets unrelated to the local elections, but containing the #le14 hashtag will likely arise from users who will then not fit categories 1-5 of the scheme (although the exclusion category, 0, will have removed some of these). The remaining category, 6: Citizens, will therefore become a default into which actors who have tweeted irrelevant messages will be sorted. This will likely inflate that category and have some distorting effect on the network metrics that describe that group's interaction with others (e.g. its E-I index will tend to be higher, since directed tweets irrelevant to this discourse are likely to be out-group). The methods of the study have sought to minimise this distortion, but without content scanning of every tweet in the corpus and subjective judgement of whether it meaningfully represents a contribution to local election discourse, it is not possible to completely mitigate this bias.

5.5 Comparison to Related Work

This study may be most closely compared to the work of D'heer and Verdegem (2014) on the Flanders (Belgian) local elections of 2012. This study utilised a similar data collection methodology and was, in part, the basis for some of the network analytics used here. Their study diverges from this one in its more simplified actor coding, and in its substantial reduction of the discourse network based on the exclusion of users not fitting its threecategory system. It is important to note, also, that the Belgian study centred only on the @mention network for the actor selection made; i.e. it supplies no comparable metrics for the entirety of the hashtag discourse captured, but focuses instead only on the interaction dynamics of addressed messaging extracted from the total corpus. Nonetheless, comparisons are possible with the similar component investigations in this study and yield insight into the differences between the Irish and Belgian cases.

Since the #le14 study coded actors to more subgroups than the Belgian study, density and in/out-group metrics are not immediately comparable; e.g. local politicians' tweets mentioning national politicians count as an external (out-group) tie in this study, driving up the E-I index value, whereas they would be classified as an internal (in-group, among politicians as a general cohort) tie in the Belgian model, lowering E-I index. This is demonstrable across the range of E-I values obtained in this study, which are higher than the corresponding group's values in D'heer and Verdegem's (2014 p. 11). Recalculating, such that categories 1 and 2 (national and local politicians) are combined, yields a value from a group more directly comparable. Comparing, using this re-grouping, we see that the political agents exhibit very similar E-I indexes for pre-election periods in both locales (e.g. compare Pl_{IRE}: 0.113 with Pl_{BEL}: 0.089). Only during the week-prior period do the values vary by more than 0.2 (a >10% fluctuation); $P2_{IRE}$: 0.269, $P2_{BEL}$: 0.103. This greater out-group discourse directionality among the Irish political actors may be accounted for by that period's marked increase in local politician presence; it is very likely that many local candidates took to Twitter for the first time in the discourse, and subsequently began addressing questions, in the week before voting. It is not clear from the Belgian study whether a corresponding usage spike occurs for politicians in this period, and this could prompt a comparative investigation of media conditions in both countries at the time, to suggest explanations.

Since the Belgian study only counted professional agents in the media category, this group's metrics may be compared with group 4 in this study. The groups are of similar size and the periodic E-I indexes are consistent (variance < 0.2) between locales, indicating that, like

156

political agents, professional media in both countries exhibit similar patterns of communication with agents of other types.

Among citizens, however, the E-I indexes diverge significantly after P1. This comparison involves the least similar groups, since the citizen category for this study is larger (n_{IRE} =6,556; n_{BEL} =3,393) than the Belgian case, though both exhibit an initially similar level of out-group communicative tendency in P1. It is the sharp rise in E-I index prior to and on Election Day in Ireland, described earlier in this chapter, which diverges from the comparatively stable Belgian levels throughout.

Regardless of this interaction anomaly among citizen actors, a very similar overall structure is evident in both local election cases; citizens form a large outwardly communicative cohort, interacting with media and politicians, who demonstrate lower reciprocity toward them than toward one another. D'heer and Verdegem's observation on the power relationships among the three groupings seems as valid for the Irish case: "*Taking part in the discussion is one thing but taking position is another. Whereas the conversation networks are dominated by citizens in terms of presence, the central positions are related to political and media agents.*" (2014 p. 10). A similar conclusion is reached by Ausserhofer and Maireder (2013) for national political discourse in Austria, where established political actors (professional politicians, media, activists) occupy more densely linked networks and message reciprocity is greater among these.

5.6 Chapter Conclusions

This chapter presents a number of insights related to the discussion on Twitter of the Irish local elections of 2014, and how the networks formed by the actors involved in those discussions are structured, and change through the period. Some of these findings may be summarised by returning to the foci of the research questions guiding this case study.

5.6.1 Revisiting the research questions for this chapter

Who are the participants in the #le14 Twitter discussion and how can they be meaningfully categorised?

It is clear that the majority of actors in this discourse network are not professionally affiliated with politics, nor occupying roles traditionally associated with political discourse in broadcast and print media; the #le14 discussion is dominated by citizen actors. Of those actors who are professionally occupied with politics, we have found it useful to separate the categorisation of professional politicians into local and national, following indications observed in the pilot study and supported in these data; local and national politicians behave differently in the discourse overall and exhibit marked divergences in their interactions with one another and with other actors during the various periods of the study.

Amongst media actors, we have seen that a substantial non-professional media group exists, in the form of amateur commentators, bloggers, etc., and that these actors have greatest presence earlier in the discourse before being displaced by established professional media proximal to the voting event. They are highly communicative and exhibit high in-group directionality and centrality, demonstrating an internal dynamic of debate and struggle for visibility consistent with our expectation of how symbolic capital might be generated in the space (See: Chapters 3, 7). Similar trends are observed among activist actors, who are distinguished by a very low reciprocity in messaging, corresponding to their communicative strategies.

Finally, the chapter has outlined that the citizen majority in the discourse is chiefly composed of incidentally contributing actors, with a very small but highly communicative cohort engaging in high frequency tweeting and extensive directed messaging. Among citizen actors, we can observe patterns of interaction, via the @-mention and retweet affordances of Twitter, that demonstrate information sharing, generalisation of the discourse to a wider political context and shared audience behaviours in response to real time broadcast media coverage of the election.

158

Derivation of these actor categories and analysis of the conversational networks based upon them is informed by the conception of the networks as part of a *Bourdieusian field* of social practice. It should be expected that such a field would reflect patterns of status and struggle reflected from other fields (e.g. media, politics) and that structures of power should derive from and respond to norms in those intersecting fields. Analyses based on these actor categories support this and the observed patterns are further examined in the explanatory qualitative research detailed in Chapter 7.

How do these participants interact with one another and do their modes of interaction vary according to role?

This chapter has shown that this discourse may be understood as a very large, de-centralised, and sparsely connected network of actors, itself composed of sub-networks organised by ingroup and out-group communicative patterns. By categorising actors according to aspects of their professional association with the field of politics, patterns of interaction are identifiable between them and between periods of the discourse.

It may be concluded that modes of interaction vary between groups and that combining descriptive analytics with metrics of the network structure can identify group-specific communicative strategies.

Through comparison of relevant metrics from this study with related work, this chapter demonstrates that the Irish context bears strong similarities to another Western European political setting, Belgium, as the very similar communicative trends for professional political actors and established media support this. It may be reasonably concluded that the substantial presence, distributed and decentralised network, and small highly active subgroup of actors associated with citizen interaction in the discourse is common to both settings and is not a unique pattern to the Irish context. This may be a feature of citizen presence in Twitter political discourse more generally, but this conclusion would require additional supporting studies from elsewhere.

How can analytical methodologies for networks be applied to investigate the patterns of interaction in Twitter discourse?

This study has demonstrated the investigative potential of an approach combining analytical methodologies from the literature, and augmenting these with a nuanced coding model. Specifically, this study has shown that a multi-phase approach to understanding the network of discourse is useful.

Quantitative metrics describing the overall network structure give a broad understanding of the essential nature of the discourse, and relative presence of actors within it. Coding of actors by specific professional association to the discourse field allows for iterations of analyses for groups and sub-groups, affording us greater investigative insight for communication patterns while still allowing simplified cohort models for comparison with other studies. We may conclude from comparison of the findings of this study with that of the 2012 Belgian local elections, that a more specific categorisation model offers a useful extension to explanatory power for the patterns of discourse, without a substantial additional requirement in coding.

Network Analytics provide findings on inter-group communication and patterns of influence in the discourse and metrics such as E-I index offer comparative quantitative measures for interaction among groups within the study and with similar groups in other studies. Descriptive Analytics allow us to derive associations between this discourse and other topics and to understand how patterns of messaging behaviour, such as URL sharing, are distinct to events in the discourse period or to groups of actors. These metrics can also illustrate the relative influence of actor groups, which may be compared to their relative presence and provide an understanding of the power relationships in the discourse network.

Crucially, this case study has demonstrated the explanatory value of per-period analysis of the overall discourse, modelled on the D'heer and Verdegem structure of subdividing the study so as to investigate changing patterns of interaction prior to, proximate to, and post electoral event.

As this research field develops, investigators must recognise the imperative to develop analytical toolsets that provide results that inform scholars on the specific discourse investigated, but crucially also offer comparative value for existing studies involving similar subject matter.

5.6.2 **Opportunities for further study**

This case study forms part of the directed research effort of this thesis and its findings inform the subsequent case study (#marref, Chapter 6) and qualitative investigations (Chapter 7). These case studies refine the methodological insights, test these approaches in a different type of political discourse, and unpick some of the findings of network analytics through interviewing of the actors involved.

It is worth noting, however, that other opportunities for study are implied by the findings of the chapter, which might be undertaken at a later time. An investigation using this framework during a future local election in Ireland, or in another electoral locale would provide interesting comparative data and demonstrate developments in communicative strategy or changes in particular group cohesion with time. A larger qualitative study of local and national political actors' use of Twitter would provide interesting insight, and might offer explanatory potential for some of the observed dynamics for these groups. This is particularly relevant for the case of local politicians' use of Twitter, which can be observed in this context to have more sudden adoption and greater openness of communication. The social and behavioural factors underpinning these patterns are outside the scope of this thesis, but provide a potentially rich area for investigation.

Chapter 6

2015 Irish Marriage Referendum (#marref) Case Study

6. 2015 Irish Marriage Referendum (#marref) Case Study

6.1 Introduction

This chapter outlines the second major quantitative case study of this research, based on the Twitter discussion of Ireland's 2015 marriage referendum. The referendum, which followed as an imperative of the recommendations of the Irish Constitutional Convention, held between December 2012 and March 2014, was widely discussed in traditional and online media. The specific question of the referendum was whether to enact the proposed *Thirty Fourth Amendment to the Constitution of Ireland*, which permits marriage to be contracted by two persons without distinction as to their sex. The referendum passed (by 62% of voters) on 22 May 2015, and a bill was subsequently signed into law, making Ireland the first country to legalise same-sex marriage by popular vote. Discussion of issues related to this historic social change in Ireland, was thus a lengthy and wide-ranging affair.

This subject was chosen for the second case study of this work for several reasons: (1) the nature of the referendum implied a discourse at national level; (2) the issue was a binary one, with actors likely to be positioned in favour of or against the passage of the referendum, which offered a contrast to the multiple and disparate foci of the discussions of local elections; (3) Ireland's particular legislative arrangement for the discussion of electoral and referendum events on broadcast media has previously been shown to drive discussion online instead (Barrett 2011); and (4) the referendum was signalled by the government substantially in advance and allowed time for initial debate around it on Twitter to coalesce around a particular hashtag. The last point is, of course, a substantial consideration for this type of research since the dominant data collections methods (already outlined in some detail in this work) require advance knowledge of a particular hashtag or keyword to track in real time. Some detail is provided later in this introduction on the selection of the particular tag (#marref) used to track the referendum discussion.

Similarly to the previous chapter, this one begins by discussing the statistical and descriptive metrics for the set as a whole, providing an overall sense of the #marref discourse, its dynamics, and its participants. The chapter then moves on to the specific network and descriptive analysis methods outlined for the #le14 case study. These methods were again deployed in the analysis of the marriage referendum dataset and allow for direct comparison between the two political discourse events, providing insight on the commonalities and differences in the network structure and participation involved. It is important to note that

while the #le14 case study was comparable with existing studies in the literature (Ausserhofer & Maireder 2013; and more particularly D'heer & Verdegem 2014), albeit using a considerably larger scale of dataset, there are no studies in the literature which apply these or similar methods to discourse around a referendum event in which participants are likely to split along a single binary cleavage of political opinion. Thus, as well as testing the previous methods in this new discourse setting, some simple analysis based on the coding of participants as pro- or anti- referendum proposition is also piloted. Other significant changes and additions made to the methods used in this second case study include its approach to data collection, additional analyses of the user cohort, and adaptation of the actor coding methods applied to them.

The chapter concludes with reference to the specific research questions underpinning it, and offers evaluation of the combined methods of collection and analysis deployed for this case – a larger dataset than that of the first case study, and one focussed on a binary national issue over a longer period.

6.1.1 Research Design

In common with the previous chapter, the research questions for this case study seek to examine the particular network of discourse around the #marref event with reference to similar research questions. This case study necessarily expands the focus of the questions set out for chapter 5, in order to add comparative evaluation between the two datasets as a component of this study. Thus, the #marref research is designed around these investigations:

- RQ.6A: Who are the participants in the #marref Twitter discussion and how can they be meaningfully categorised?
- RQ.6B: How can analytical methodologies for networks be applied to investigate the patterns of interaction in Twitter discourse, using this case study as an example?
- RQ.6C: How do the findings of this case study compare to that of the previous study of #le14 discourse, and what research insights may be drawn to inform the design of further studies?

6.1.2 Choice of hashtag

As in the case of the selection of #le14 for the study of the 2014 local elections, a hashtag needed to be identified as the selection variable for this case study. The possibility of using keywords alone or in combination was observed to be problematic during early testing; terms such as "gay marriage", "same-sex marriage", etc., were often used in widely varying geographically and politically distinct locales (e.g. the United States was concurrently engaged in regular debate on the issue following developments within its Supreme Court). Selection of a tag specific to the Irish context for Twitter discourse on this issue was important, and similarly to the discussion of election tags, a number of 'competing' hashtag suggestions were proposed by users in mid-2014. Many terms explicitly related to one of the two stances on the referendum (e.g. #right2marry) or had implicit connotations of a proposing or opposing view (e.g. #marriageequality, which despite appearing in the name of the bill moved by government as part of the referendum, was rejected by no-campaigners as implying a stance in favour of the motion through the use of the word equality). The most sustained 'alternative' tag was #ssmref (same sex marriage referendum), which saw roughly equivalent use to #marref during pilot tracking in early 2014, and was regularly included as a contextual tag in reporting by Irish online news site *The Journal*⁷⁰. This tag subsequently dropped from use as #marref became more dominant by frequency through mid-2014, leading to some users objecting to the use of #ssmref⁷¹ as redundant and secondary. By the beginning of 2015, use of #marref as the topic indicator for the referendum was dominantly established and observable in the tweets of both pro- and anti-referendum campaigners.

The process of observation of the 'settling on' a hashtag was once again shown in this case study to be a critical process for this type of research. Digital ethnographic methods participating in Twitter discussion, engaging users on their attitudes toward the use of particular tags, etc. - are an important consideration during the piloting of a study such as this, in order to understand the connotations of the use of a specific tag, and any implications for selection bias associated with it. The rejection of the #marriageequality tag by nocampaigners did not curtail its widespread use and co-occurrence with #marref, but selecting on this tag alone would have biased the sample strongly towards pro-referendum users. Similarly, many pro-referendum users expressed a non-specifically stated concern about the use of #ssmref as a tag for the referendum. When interviewed privately by the researcher

⁷⁰ http://www.thejournal.ie/ssmref/news/ (Accessed December 2015)

⁷¹ E.g. https://twitter.com/daithigorman/status/560724499239559168 (Accessed December 2015)

about this, using Twitter Direct Messages⁷², several described a discomfort with the closeness of "ssm" to "S&M" and wished to curtail its use and shift to a "*more neutral*" term. (See e.g. Appendix A.6.1)

Separately to issues related to users' (or groups of users') inclinations to include or exclude hashtags in general - a subject investigated in detail in chapter 7 of this work - this case study demonstrates the research value of detailed interrogation of the process of hashtag adoption by users, and suggests that further study on this issue would be of value to the field.

⁷² Direct Messages (DMs) on Twitter are a mechanism of private conversation with another user, the content of which is separate to the timeline, or any public presentations. An example of one of the DM conversations exploring the use of #ssmref with an anonymised user is presented in the addenda to this chapter (Appendix A.6.1)

6.2 Data Collection

This second case study diverges from the data collection methods of the previous through the use of alternative software for data collection. Chapter 4 has outlined the processes of software development by the researcher that occurred over the period of this research, and by the point of commencement of data collection for #marref, in January 2015, a substantially altered version of yourTwapperKeeper (yTK) was available for use, dubbed yTK_DCU_v3 . This version used newer PHP libraries for interfacing with the MySQL database⁷³, on which the software depends for structured storage of data, and was therefore more efficient in each cycle of real-time query, response, parsing and storing.

In order for a comparison to be made between this version and the more standard implementation of the yTK software (used on the #le14 case study, and in studies in the literature), yTK_DCU_v1 was also deployed on a second server concurrent with the main data collection, and was subject to 30-day checks on the comparative volume of #marref tweets collected by each. Chapter 4 has also noted that a further data collection option emerged for researchers during the early period of 2015, in the form of the Digital Methods Initiative's Twitter Capture and Analysis Tool (DMI-TCAT), produced by researchers at the University of Amsterdam (Borra & Rieder 2014). A third server was deployed on which this software was installed, and this ran concurrently with the two yTK versions during the phase of greatest frequency of #marref contributions in the final weeks leading up to the referendum. In total then, data was collected for #marref on three Amazon Web Service (AWS) server instances, each running a different collection software; (1) $\gamma TK_DCU_v 1$ – the standard yTK version, used in the first case study in this thesis and directly comparable with yTK's use in previous studies in the literature, (2) yTK_DCU_v3 – an updated version of yTK used as the principal source for this case study, and (3) DMI-TCAT, used for comparative reasons during the last phase of the study only. Figure 6.1 shows this distribution of collection modes, alongside the frequency of tweeting by day during the study period.

Data collection began on 30 January 2015 and ended on 26 June 2015. During this period, 267,223 tweets containing the #marref tag were collected, following a very similar overall frequency curve to that of #le14, with a very large spike in frequency occurring on the day of the referendum count, 23 May 2015. While the volume of tweets captured for #marref is more than double that of #le14, the same trends are observable of slow growth in daily use

⁷³ The *mysqli* library was used to replace the *mysql* library (upon which yTK is built), which was deprecated in PHP version 5.5 and removed entirely in version 7.

See: http://ie2.php.net/manual/en/mysqlinfo.api.choosing.php (Accessed December 2015)

of the tag in the long lead-in period, switching to greater usage in the month before the voting event (with more dramatic spikes notable in this dataset, corresponding to televised debates and specific related news coverage in other media), and a pronounced spike beginning on voting day, maximised on the day of the count, and dropping quickly to a final rate of incidence comparable with the start of the period (See Fig. 6.1).





6.2.1 Comparison of data collection software used

For that period of the study when all three data collection software packages were simultaneously deployed (21 April 2015 - 01 June 2015, incorporating the busiest stage of data collection, prior to and including voting and count days), comparison may be made between the effectiveness of all three. A simple measure of the average tweets collected per day shows little difference across the 42-day period when the three collections were in
session, with the average daily count slightly lower for the standard version of yTK than the researcher's developed version, and higher for DMI TCAT. However, particular analysis of the busiest day of tweeting (referendum count on 23 May 2015) shows very substantial differences between the volumes of tweets collected by each strategy. Fig. 6.2 shows the tweets collected by each software approach, by hourly segment for 23 May 2015, and the number of tweets collected can be observed to be substantially lower for yTK_DCU_v3, especially during hours 9, 10, 15, 17, and 18. In each of these hours, the sample collected by DMI TCAT is largest, showing collection of hundreds more tweets than yTK_DCU_v3 (on which this case study is based), in some cases.

Figure 6.2 Overlapping graphs of #marref tweets collected by three Twitter data collection software applications, by hour, for 23 May 2015



Further analysis of the tweets collected on 23 May 2015 provides an explanation for the hourly variance in software performance. At a frequency resolution of minutes, it is clear that DMI TCAT performs best (and yTK_DCU_v1 performs worst) at tweet collection during periods of short bursts of high frequency. A graph of the rate of tweets captured per minute by yTK_DCU_v3, throughout the day, shows that the hours corresponding to the greatest variance in collection shown in Fig 6.2 are those with the most spikes of >90 tweets per minute (see chapter 6 addenda; A.6.1). During the count day, these spikes likely correspond

to media reporting of the results, occurring at the opening of the count centres (hour 9), during the first media projections of a national result from initial tallies (hour 10), during live televised reporting of speeches from the national count coordination at Dublin Castle (in the early afternoon), and most notably during the minutes prior to and following the beginning of the main evening news at 18:00.

This micro-level analysis of the performance of the various software options deployed demonstrates that the choice of collection application, and its inherent dependencies on particular modes of API query, database transaction efficiencies, etc., can directly influence the *completeness* of the resulting dataset. It seems reasonable to conclude that the tracking of Twitter discourse, especially where that discourse is prone to exhibit short sharp spikes in frequency of tweets sent, is affected by limitations directly resulting from the choice of software in use - a variable not discussed in studies in the literature. While it appears that DMI TCAT is the best performing open-source collection software currently available⁷⁴, these observed variances also leave open a question of the extent of data capture possible using even it.

⁷⁴ The discovery of the superior tracking performance of DMI TCAT led the researcher to cease the iterative developments to yTK, adopting DMI TCAT instead for continuing projects.

6.3 Evaluation

Due to the scale of the dataset collected for #marref (n=267,223), and the 79,609 users represented within it, new strategies are required for meaningful sub-sampling of the user set overall. The coding of very large sets of users is prohibitively time-consuming for researchers, and while not explicitly stated as such, likely accounts for strategies evidenced in the literature to diminish the size of large datasets, or to disregard actor cohorts via computational selection and removal (e.g. Bruns & Highfield 2013; D'heer & Verdegem 2014; Gainous & Wagner 2014; Godnov & Redek 2014).

In this case study, two methods to reduce the user set under investigation were considered. The first was a simple elimination based on a criterion of frequency; users who were incidentally involved in the discourse, appearing only once in the #marref dataset, were computationally removed, yielding a smaller derived set of users for coding. The second strategy, which follows Bruns and Stieglitz's (2013) article on Twitter analysis, suggests a 90-9-1 percentage split, using computational selection to rank users by their activity in the discourse set, segmenting them into cohorts of the least active 90%, next least active 9% and remaining most active 1% of users.

Examination of the corpus of tweets showed that 69,938 users (88% of the total in the set) sent just one tweet with the #marref tag. This is not an unexpected finding, since studies in the literature which offer this breakdown (such as Bruns & Stieglitz 2013) show a similar 'long tail' in the graph of users to tweets sent. The remaining 9,671 users appear more than once in the corpus.

Examining the users in the set by 90-9-1 percentile split also offers a useful insight for quickly understanding the variance in participation among users recorded. The least active 90% of users were those who sent fewer than 5 tweets each, and they accounted for 31% of all tweets sent. The next 9% varied more widely in their participation, sending between 6 and 78 tweets each, and accounted for 22% of tweets sent. The top 1% by activity is constituted of those users who sent 79 or more tweets each during the period (some sending in excess of 1000) and these users accounted for 47% of all the tweets gathered. Fig. 6.3, below, summarises this information by showing the whole of the #marref-tagged tweet dataset as a pie-chart, segmented by the percentile bands of users are responsible for the majority of the discourse and that there exists a disproportionately engaged top tier of highly active users. Further analysis showed that 12.8% of all tweets (34,374 in total) originated from just the top 50 users in the sample. Additional attention was given to these top 50 users (a list of whom

appears in the addenda to this chapter, Appendix A.6.3); these were subject to additional coding, based on their pro- or anti-referendum position, and findings based on this coding are discussed in a subsequent section of this chapter.



Figure 6.3 #marref tweets sent, by 90-9-1 percentile split of active users

While this analysis makes clear that a viable strategy may be to select and code only the top 1% of actors, concerns at disregarding over half the collected tweets in those analytic methods that depend on actor coding motivated the researcher to instead choose the option to code all users who sent more than one #marref tweet. This resulted in a sub-set of the total 79,609 users, amounting to 9,671 actors who had each contributed at least twice to this studied discourse. Though still a large cohort, this coding was deemed manageable and allowed for a greater extent of comparison between the metrics derived from it for #marref, and those already stated in chapter 5 for #le14.

6.3.1 Actor Coding

Following the coding schema set out in the methodology of this work (chapter 4) and deployed in the analysis of the #le14 dataset (chapter 5), this case study uses the same 6-category system for the manual evaluation and coding of the 9,671 users in the set who sent more than one tweet each. This focus on actor coding necessarily reduced the collected dataset by 69,938 tweets, leaving a corpus of 197,285 tweets that had been contributed by actors who had participated in this hashtagged discourse at least twice each. Coding of the entire user set was considered too onerous and thus the findings based on the analysis of interactions between actor cohorts in this study are derived only from this specific subsample.

In keeping with the method established in this work, the yourUserCoder (yUC) application, developed by the researcher to aid efficient actor coding, was deployed on this subset and the following coding was applied: (1) National Politician, (2) Local Politician, (3) Activist, (4)

Professional Media Agent, (5) Other Media Agent, and (6) Citizen. Again, an additional coding of (0) was used to indicate users determined by the researcher to be irrelevant to the discourse but appearing in the set through their use of the hashtag. These were again found to consist mostly of obvious spam-bots or spam accounts sharing links using several currently popular hashtags for visibility⁷⁵. A total of 161 users were coded θ - *irrelevant* and their corresponding 6,957 tweets were removed from analyses of the data subset. This number may appear to indicate a much lower prevalence of spam accounts in this case study, compared with the 802 identified in that of #le14, but it is important to note that in this case only users having sent more than one tweet were being coded. This may point to a greater incidence of irrelevant users among the cohort that contribute only once to a hashtag set.

6.3.2 Methods of analysis applied

As in the previous case study, Network Analytics and Descriptive Analytics were applied to the dataset. An important methodological departure to note, however, is the effect of the reduced actor coding in this case study. Therefore, while computationally derived metrics are available for the entire dataset (267,223 tweets), some analytics necessitating coded actors are available only for the subset, already identified, of tweets sent by users who have contributed more than once (197,285 tweets).

6.3.2.1 Network Analytics

A *mention network* (as outlined in chapters 4, 5) may be computationally derived for the entire dataset, since it is constructed by examining the links between users in conversation, who @-mention (@-reply) one another. An impression of this network is useful for analysing those individual users who are most often mentioned, or for measuring the net internal or external communicative directionality for the whole group (i.e. are most mentions made of others in the set, or of users outside the captured set).

However, to perform analyses related to actors by category (such as those outlined in the previous chapter for #le14), users must be manually coded. Since this process only occurred for those 9,671 users who sent more than one #marref tweet, a second mention network is constructed for these. These two mention networks, termed *mention network 0* and *mention*

⁷⁵ A tweet captured in the set, from user @6_skinny, advertises a treatment for gout using the unrelated set of hashtags: #11m #gangtour #lrt #afc #marref #monday. This user account had been deleted by the time it was examined for coding - likely a forced deletion by Twitter for breach of Terms of Service - but was determined to be a spam-bot by examination of its 91 contributions to #marref, all of which were similar marketing links. Tweet source:

https://twitter.com/6_skinny/status/598091180702760000 (Accessed 30 August 2015)

network 1 represent the directed communication between all users in the set, and that of just those users who contributed twice or more, respectively. These networks are used to derive findings set out in section 6.4 of this chapter.

6.3.2.2 Descriptive Analytics

Similarly to the previous case study, metrics are derived for the full corpus of tweets via computational analysis of descriptive attributes of each. Tweets were algorithmically parsed and certain properties were recorded for each: (1) whether the tweet is a retweet of an existing message, (2) whether the tweet contains an @-mention, (3) whether the tweet contains a URL, and (4) whether other hashtags co-occur in the tweet with #marref.

6.3.2.3 Periods for this study

Following the insights yielded in the per-period analyses of the #le14 dataset, a similar 4period division is provided for the #marref tweets:

P1 :	Prior period (pre-referendum, somewhat distant from the event).					
	30 January 2015 - 14 May 2015					
P2 :	Proximate prior period (one week pre-referendum).					
	15 May 2015 - 21 May 2015					
P3 :	Event (referendum voting day)					
	22 May 2015					
P4 :	Post-event period (Count day, and subsequent weeks)					
	23 May 2015 - 26 June 2015					

These periods mirror the periodic structure of the #le14 analyses, as well as that of D'heer and Verdegem's (2014) Belgian election study. It should be noted however, that the prior period (P1) and post-event period (P4) are longer for this study than for either of these previous cases; this longer study period was intentionally set out to facilitate detection of when changes in the frequency of tagged discourse occur. In this case study, it can be observed that, as in the previous case, a marked increase in daily tagged tweeting occurs approximately one month in advance of the electoral event (Fig 6.1), corresponding to the period of most broadcast media coverage, and greatest campaigning. Further studies are required to establish whether this is a consistent trend outside these two examples in the Irish context, potentially suggesting a more fixed definition of these 4 periods as set durations with respect to a studied event. While utilising standard period lengths would offer additional uniform comparative metrics across events in different studies, this is not yet a concern in the field, as research remains insufficient to fix a demonstrably effective period for these.

6.4 Results & Analysis

6.4.1 Overall Network Structure

As in the previous case study, the coding of the users (in this case, those who have contributed more than once to the hashtagged discourse collected) may used to present an initial overview of the relative participation by actor cohorts, per period. P* gives the overall total number of participants and percentage presence for each actor category across the duration of the study, and P1 - P4 shows the absolute and relative participation by each during its various periods. As with the previous presentation of such data, it is important to note that the length of the periods vary greatly, so the percentage presence during each period is a more useful comparable indicator of group dynamics. Table 6.1 sets out this data, in the same manner as Table 5.2 did for the #le14 dataset in chapter 5. Figure 6.2 offers a visualisation of this data, indicating the changing presence overall and by actor through the four periods.

Table 6.1#marref selected users by coded category, by discourse period

	National Politicians		Local Politicians		Activists Me		essional Iedia Other Media		Media	Citizens			
	1	(%)	2	(%)	3	(%)	4	(%)	5	(%)	6	(%)	Total
P*	126	1.3	98	1.0	302	3.2	417	4.4	227	2.4	8340	87.7	9510
P1	74	1.3	56	0.9	199	3.4	140	2.4	121	2.0	5325	90.0	5915
P 2	111	1.3	71	0.8	302	3.5	389	4.4	198	2.3	7678	87.8	8749
P 3	89	1.1	32	0.4	228	2.9	360	4.6	46	0.6	7129	90.4	7884
P4	121	1.3	77	0.8	248	2.7	378	4.1	166	1.8	8225	89.3	9215





As in the previous study, some insights are readily evident in this breakdown: once again, citizens (coded 6) dominate the discourse during all periods, and in a substantially higher proportion than was the case in #le14; activists (3) similarly appear in greater numbers and

with a much greater percentage presence than in the previous case study; and the overall 'shape' of participation has changed with more users taking part in the discourse during the week prior to referendum day.

The structure of this coded subset of the #marref data shows that in the case of this event, activists, rather than national or local politicians, are the more prominent political actors at all stages, and are an early and sustained presence in the discourse. This may be unsurprising, given the nature of the political event, where leadership of the conversation shifts from established political figures to campaigning and mobilising groups on both sides of the referendum issue.

The prominence of professional media (4), especially from P2 onward, is partially accounted for by the interest among international media outlets immediately before, during and subsequent to the voting - around 40% of the users coded as professional media agents were observed to be journalists and outlets from other countries, with UK agents having particular prominence among these. While the result of the referendum was internationally significant and very widely reported, use of the specifically Irish #marref tag to mark Twitter reporting of it was not very widespread beyond Ireland, so the number of international professional media agents discussing the referendum is likely to be greatly larger than what is captured using this tag.

Similarly, other media (category 5), composed of bloggers, casual online journalists, etc., showed an international component, compared with that of the #le14 dataset, where this was wholly absent. Like the previous case study, these non-professional media agents became more established in the discourse during P1, showing an earlier interest in the upcoming referendum than professional media actors.

Most notable, however, is the extent to which citizen agents dominate this discourse. The categorisation of actors as 'citizen' versus 'activist' was more difficult in the case of this dataset than the previous, since the line between outspoken campaigning citizen and activist was blurred, especially where coding took place in the period close to referendum day, when many users changed their Twitter bio and profile image to campaign slogans, logos, etc. Only users whose Twitter information made clear that they were professionally associated with political activism, describing themselves as such, or indicating a role with a particular group were coded (3), while any users who had adopted campaigning stances in support of either side were deemed 'citizens'. This issue of interpretation is further discussed in the section on the limitations of this study.

6.4.2 Analysis of the top 50 contributors

An additional analysis took place of the most active 50 users in the dataset. As well as their already present actor category coding, these users were categorised as either in favour of the referendum (coded: 1), opposed to the referendum (coded: 2), or neutral (coded: 0), as in the case of media outlets, most of which maintained a neutral stance throughout, or until hours before voting began.

This coding was based on examination of each user's profile, as well as of a random selection of 100 tweets sent by them in the set. In every case, it was very clear what the user's stance was, and in some cases this was immediately evidenced in their Twitter profile alone, with many users having changed their bio and/or profile image during the campaign period. A list of these users, the coding for their referendum stance and actor category, and the number of tweets from each, appears in the addenda for this chapter (Appendix A.6.3).





Amongst the 50 most active contributors using #marref, 16 (32%) were identified as opposing the referendum. These contributors were responsible for a disproportionate number of the tweets from amongst these top 50, however (accounting for 40% of the total), corresponding to the general trend for those opposing the referendum to tweet more frequently. It is a reasonable assumption that this asymmetry of contribution frequency is a factor of the overall lack of representation of the opposing view within the cohort generally, as evidenced by the hashtag intersection (outlined in the nest section); if there are fewer nocampaigners, they likely involved themselves in more conversational and debate interactions on average, as well as tweeting more in general to represent their position on the issues. Fig. 6.5, above, sets out the composition of the top 50 users by referendum stance, and by actor category.

It is clear from this analysis, that citizens dominated not only the discourse generally, but the section of most active contributors to it. Indeed, of the top 10 contributors to the dataset, 9 are citizens (Appendix A.6.3). This finding, as well as wider picture of the discourse periods already set out, show that the #marref discussion was one in which citizen users had a greater presence than for #le14, and where they kept that presence at all stages.

6.4.3 Mention Network Metrics

Following the methodology of the #le14 case study, the mention network density was calculated for the entire dataset (*mention network 0*). This is a measurement of the communicative interconnectivity of actors and is computed from the given number of @-mentions between all users in the dataset, divided by the total number possible. In common with the previous case study, these values are very small for each period, indicating a loosely knit network; (P1) 0.0007, (P2) 0.0023, (P3) 0.0017, (P4) 0.0024. These results show a similar, (though initially lower) level of communicative cohesion throughout the study, to that of #le14. Notably, however, the network density fluctuates differently over the four periods in this case, showing a slight drop on referendum day (P3), likely as users made more declarative contributions and fewer directed ones, and an increase to the densest cohesion in P4 (greatly higher than the same period for #le14), likely corresponding to a substantial increase in conversational interactions associated with seeking result information on count day, and subsequent discussion of the outcome.

Table 6.2	Percentage of @-mentions by actor category, directed to actors
	within other categories $(1-6)$, or outside the sample (X) .

Source:	1	2	3	4	5	6	X
1	14.2	1.1	26.2	8.2	0.8	28.6	25.5
2	28.1	2.6	10.9	4.4	0.6	36.5	16.9
3	27.1	0.9	36	10.8	2.3	20.8	2.1
4	11.2	0.6	24.9	18.5	0.0^{76}	12.9	31.9
5	23.3	1.2	28.4	13.03	3.2	11.0	19.9
6	13.3	0.6	16.1	19.1	0.3	38.5	12.1

⁷⁶ Note that it is not the case that absolutely no content from professional media was directed to other media, but that at this level of rounding the small amount that did occur is negligible.

For the case of the sub-network of mentions between coded actors (*mention network 1*), a measure of the percentage of each cohort's total @-mentions that was directed to other actor groups, or outside the sample, offers a more detailed picture than the simple measure of mention density. Table 6.2 provides this overview.

This analysis offers a very distinct description of the directionality of the examined discourse, compared to the previous case study. Despite the "external" sample (X) containing not only all Twitter users not recorded by the study, but also those users in the dataset who only sent a single #marref-tagged tweet, actors directed messages to these at generally much lower levels than for #le14; i.e. this group was much more inwardly communicative. The highest proportion of external directing of messages occurs for professional media (4), which may be accounted for by the international cohort among that number, who likely directed messages to users outside of Ireland, and largely outside of this sub-network. A marked anomaly between externally directed messages in this case study, compared to the previous, occurs for activists (3), whose messages were mostly directed within this user set. This demonstrates an interesting departure from the local elections discourse, since the table shows that activists are the recipients of a large proportion of messages from all other users, and are likely responding heavily inside this set. Similarly to the strategies seen for #le14, the activists direct most of their messages to professional politicians, again likely @-mentioning figures whose attention they wish to draw to issue campaigning.

Another very significant departure from the shape of the previous discourse studied is observed in the proportion of content directed to citizens, both by other citizens (indicating a vastly greater internal debate in this cohort), and by all other groups. Crucially, in this case study professional politicians are observed to @-mention citizens in a tenfold greater proportion; activists similarly display greatly more engagement with this public. Finally, the degree to which professional media agents @-mention citizens is also significant, and a very substantial departure from the #le14 case. Here 12.9% of professional media mentions are of citizens (cf. 0.7% in #le14 study). This is likely accounted for by several factors, but notably two: (1) professional media may have been greatly more inclined to solicit reaction, quotation, etc., from 'ordinary' members of society in the case of a historic social change, than in the case of a local election, (2) many 'citizen' coded actors became prominent spokespersons for points of view in favour of and opposing the referendum and adopted positions of issue punditry normally occupied by politicians and professional activists.

Again, a specific examination of the top 50 users to whom tweets were directed provides additional insight on the concentrated activity underlying these patterns. A list, giving the username, referendum stance coding, actor coding, and number of tweets directed to each of these top 50 users is presented in the chapter addenda (Appendix A.6.4). In total, 43,013 tweets were directed to just these 50 users (over 16% of all #marref tweets sent), with proreferendum users accounting for 48% of the recipients, and anti-referendum users 30% (fig. 6.3). Among the most mentioned users, activists and citizens again find prominence, indicating a new space for influential users from both groups in this part of the discourse.

Figure 6.3 Composition of 50 most frequently @-mentioned users in #marref tweets, by referendum stance and by actor category



Finally, a calculation of the External/Internal ratio analysis (*E-I index*) was made for each group, during each period, as was the case in the first case study. The derivation of this calculation (Hanneman 2005) is given in more detail in the previous chapter, with the resultant metric being a decimal value between -1 (indicating all communicative ties are internal to the given group) and +1 (indicating all communicative ties are with modes outside the group).

Table 6.3Group level E-I index, by period, for selected #marref users

	1	2	3	4	5	6
P1	0.741	0.939	0.375	0.599	0.931	-0.012
P2	0.728	0.942	-0.109	0.612	0.898	0.206
P3	0.631	0.969	0.454	0.511	0.936	0.441
P4	0.766	0.940	0.390	0.788	0.955	0.299

Table 6.3 outlines the E-I indices for each category, during each period, and as in the case of the #le14 study, shows positive values for all periods for most actor groups, indicating that the general disposition was toward out-group communication. Several important differences

in actor communicative behaviour may be observed, however. In two cases, the E-I index is negative, indicating mostly in-group communication. These occur for citizens in P1, demonstrating a trend for citizens to engage with one another, more than with professional actors, in the earliest phase of the study. While this value was low in the previous case study (0.216), and is a factor of the numeric dominance of citizens in the user cohort generally (they account for 90% of active users during this period), the negative value in this case indicates a substantial shift toward citizen debate and mutual engagement, from that of the opening period of #le14. It is important to contextualise this study as having a longer duration for its P1 period⁷⁷; data collection started earlier in this case, and may have captured more of an initial context in which professional actors, especially media and politicians, are less engaged, leaving citizens to discuss the topic amongst themselves. Throughout the periods, citizens retain a much lower E-I value than in the previous study (cf. Table 5.4), indicating a sustained level of mutual engagement at all stages, and pointing to a greater degree of citizen debate, with their conversation moved away from the traditionally more influential actors, such as national politicians.

A larger negative value is observed for the E-I index of activists during P2 (the week prior to voting). While this value was greatly higher (0.922) for the local elections, indicating a different communicative strategy by activists for that event, here the level shows that in the days immediately before the referendum, activists made most mention of one another. This significant change is likely accounted for both by heightened debate between pro- and anti-referendum activists, as well as supporting and mobilising @-mentions between those on a particular side.

Other important trends to note in this analysis include the changed strategy of politicians national politicians being more outwardly engaged (likely mentioning activists and prominent citizen campaigners), and local politicians having the highest outward communicative directionality. The latter group is small in number in this study, though here again, their communication on Twitter shows a marked difference to that of their national counterparts, offering continued support for the separate coding of both.

Media too, exhibit a variant pattern of in- and out-group communication to #le14, with both groups more outwardly focussed than previously. The changed focus of professional media (from national politicians to activists and citizens) may account for this in part. Professional media are especially more outwardly communicative in P4 here, compared to #le14, though again this may be a factor of both an international subset @-mentioning users outside the

⁷⁷ P1_{#le14}: 66 days; P1_{#marref}: 105 days

sample, and the engagement by media with citizen campaigners about their reactions to the result.

6.4.4 Descriptive Metrics

Analysis of the hashtags intersecting #marref (co-occurring with it in tweets) provides another useful source of insight on the nature of the discourse captured. A list of the top 50 hashtags intersecting this #marref set is presented in the chapter addenda (Appendix A.6.5) and the three most commonly co-occurring present an indication of the overall nature of the stances exhibited by users discussing the referendum. The most popular co-occurring tag was #VoteYes (present in 29,453 tweets - 11% of all collected), followed by #YesEquality (21,368 tweets) - the name of the main pro-referendum campaign. The third most frequently coincident tag was #VoteNo, but this (and all other coincident tags) were present at a much lower level in the set; #VoteNo appears with #marref only 6,731 times. This overview offers another indication of the overall trend of the contributions to be in favour of the marriage referendum passing. The word "Yes" appears as a hashtag 4,069 times, making it the sixth most popular co-occurrence with #marref, compared with the #No appearing ten times less frequently, ranking it 50th in the same measure.

Similarly to the previous case study, other hashtags in the set offer information on the wider media context of the discourse. The tag #rtept (marking content related to RTÉ's current affairs television program *Prime Time*) occurs in the top 10 coincident tags by frequency, followed closely by #VinB (TV3's *Vincent Browne Show*), which appeared very prominently with the #le14 tag in the previous set. The former helps to account for one of the observed spikes in tweet frequency in the period leading up to the debate, as #rtept co-tagged tweets appear in a cluster on 05 May 2015, corresponding to a prominent marriage referendum debate which took place on the show that evening. As in the case of #le14, the tags for television coverage of the electoral event demonstrate Twitter's place as 'second screen' for commenting audiences.

While most other coincident tags fall into directly relevant categories for the debate - the names of specific voter mobilisation campaigns (e.g. #hometovote, #VoteWithUs), media coverage (e.g. #VinB, #LateLateShow), or contextual terms (e.g. #love #LGBT) - some appearances in the set provide additional context for the discourse. #Eurovision and #Eurovision2015 both prominently appear, and though the televised song contest is unrelated to the referendum, users joking about their occurrence on the same day drives the presence of these.

Finally, the tags for the second referendum on 22 May 2015 - concerning a change to the minimum age requirement for a presidential candidate - appears as #PresRef and #ArasRef, low in the top 50 co-occurring tags. This offers an indication of the relatively secondary

prominence that this other referendum received in the Twitter discourse generally; indeed tracking of both of these terms by the researcher showed that in combination they account for a set only 6% the size of the #marref dataset. The presence of two tags of the same character length, for the same event, and accumulating roughly the same amount of tweets each, also indicates that the critical mass of social discussion required to begin a process of promoting one 'official' tag for the event was likely not reached in the case of this second referendum. For the marriage referendum, social pressure to use only #marref from among the variety of tags initially proposed, was evident from early 2014.

Examination of retweeting in the set also shows some differences between this corpus and that of #le14. Crucially, analysis of which actor cohorts originated retweeted content could necessarily only take place where the originating user was one of those 9,510 coded by the researcher. A first-step analysis determined that 98.4% of all retweets originate in this cohort, however, as the remaining un-coded users (who only used the #marref tag once) are likely more incidental and less prominent in the discussion. However, it is important to note that the following analyses are based only on those RTs, where the originating user is coded.

In the #le14 dataset, most retweets were of content originating from professional media (41.3%) and politicians (26.4%), while in this case study, activists dominate this metric, being the originators of 38.8% of all retweeted content over the duration tracked (cf. Table 5.5). Table 6.4 sets out the specific percentage of retweets that originated in each actor category in each period.

	1	2	3	4	5	6
P1	15.2	0.1	41.1	11.2	3.9	28.5
P 2	11.1	0.0	36.5	33.3	2.6	16.5
P 3	9.6	0.1	38.7	28.4	2.2	21.0
P4	14.9	0.6	39.0	31.6	5.1	8.8
P^*	12.7	0.2	38.8	26.1	3.4	18.7

Table 6.4% of #marref RTs originating in each actor category, by period

Again, the prominence of citizens as influential sources of shared content is readily observed and contrasts strongly with the #le14 dataset. A notable change occurs for citizen influence in the last period, however, when the reporting of count results and the subsequent publication of voting and analysis cause a return to the more usually influential figures (national politicians and professional media) as sources of retweets. This is likely an indication of the sharing of media reporting and statements in the wake of the result.

A relatively unchanged phenomenon throughout the periods, however, is the proportion of retweets originating from activists. Their dominance of the re-shared content within this

discourse is established early and retained throughout, as users retweet them consistently often. Similarly, local politicians remain consistently very low in this measure, potentially indicating less relevance for these as figures of focus and authority in this national discussion than was obviously the case for that of the local elections.

Finally, a similar measure to the previous study was made, computationally assessing the degree to which URLs were included in retweeted messages. Again, a majority (in this case a larger proportion of 82%) of retweets of content originating from professional media agents contained a link. This likely indicates the same processes as previously, wherein journalists and media outlets share links to stories during and after the campaign, and these are widely retweeted. A notable departure from this 'reason for retweet', is seen for activists and citizens, however, who despite being widely retweeted in both cases, showed much lower inclusion of URLs - 29% and 13% respectively. This indicates that the content being retweeted from these agents was much more likely to be declarative statements, reaction, reply, etc., than attempts to direct users toward external sources of information.

Thus, taken together the descriptive metrics for this data show a very different discourse dynamic than that of the previous case study, where conversation is led by activists and citizens, rather than by the more traditionally powerful figures in political discourse seen previously. These analyses show an elevated role for citizen debate and for the influence of (a small number of) highly active citizen actors alongside the more expected agents such as national politicians and professional media.

6.5 Caveats & Limitations

Apart from the known caveats, already outlined in the previous chapter and highlighted in the literature review, that the *representativeness* of hashtag-sampled data is poorly understood by scholars, this chapter has some specific additional limitations to be noted.

First, this case study has highlighted shortfalls in data collection by the various open-source applications tested, during peak frequency of tagged discourse. While the software used in this study (yTK_DCU_v3) was shown to perform well, and specifically to substantially outperform the standard version of yTK used in the previous case study and in the majority of studies in the literature that cite a method, it was still shown to underperform compared to DMI TCAT. The latter ought then to be used in future studies, and has been deployed by the researcher in continued tracking of Twitter discourse, though this too is simply the best available current option; without a known sample of all data with a particular tag from Twitter, it must necessarily be assumed to provide questionable dataset completeness also. The extent to which this completeness is also a factor of the server hardware on which the software is run is also unknown; in this case study a fair comparison may be made between the three concurrent collections, since they ran on identical AWS server instances with the same hardware and bandwidth allowances, though further studies varying these factors may help to clarify the limiting criteria for software performance.

Secondly, this case study adapted the coding strategy for users, in response to the very large number in the collected sample. While this was shown to result in a smaller sub-network of users who had contributed the majority of all captured content, this change makes direct comparison with measures in the previous case study more dubious. The #le14 study contains the contributions of users who only make a single use of that hashtag, while corresponding analyses in this study are necessarily of a slightly more active cohort.

The coding of those 9,510 relevant actors who were included was also a point of some difficulty, in the particular case of separating 'activists' from 'citizens'. Following the definition outlined in this work (chapters 4, 5) this deliberation focused on identifying a professional link with an activist group. This was achieved by studying actor's self-descriptions (and in some cases tweet history) for evidence of either (1) explicit declaration of themselves as activists, (2) direct employment by, or the holding of a position in a political activism organisation, or (3) identification of an account as part of a named campaign of activism around the referendum specifically. This coding helped to maintain consistency with the previous case study, and with the Austrian study in the literature described in chapter 5 (Ausserhofer & Maireder 2013), but left open an ambiguity of whether highly active citizens,

who modified their profile information and very frequently and overtly campaigned for either side of the referendum position, ought to be considered activists also.

This ambiguity could be mitigated somewhat by additional stance coding, such as was undertaken for a small number of the most prominent users (Appendices A.6.3, A.6.4), based on determination of a user's position on the referendum question. A larger systematic coding of this variable across the set, would have allowed for the separation of those users who were professional activists before the referendum campaign, more clearly from those who became active campaigners during it. This coding, however, requires a level of content analysis that is greatly more onerous than the category coding used. Coding of stance is also dependent on the ready availability of indicators of that stance in recent tweets and the Twitter profile, much of which quickly disappeared after the referendum ended. (Indeed, 4 of the top 50 most active users outlined here had deleted their profiles entirely shortly after the referendum, making later coding impossible. 3 of these deletions occurred on the day of the results, in all cases for previously very active users campaigning against the referendum).

6.6 Conclusions

This chapter has presented the second major case study of this research - the collection and analysis of tweets that included the #marref tag. The processes followed by the researcher in identifying this tag as most relevant to the discourse of the 2015 Irish marriage referendum were set out, showing the importance of this phase of investigation for researchers wishing to establish datasets by the means of Twitter data selection currently available. Detail was provided of the collection strategy used for the collation of this #marref dataset, offering critique of the effectiveness of other open-source collection software in use, and upon which many of the studies in the literature depend. The most widely used software, yourTwapperKeeper, was shown to perform poorly in the collection of 'burst' data, where tweets that should be collected are not in situations of very high frequency of contibution. This same analysis showed that the DMI TCAT software, in development at the University of Amsterdam, performed best and is indicated to be a viable choice for researchers seeking more reliable and complete data capture. The developed version of yTK deployed by the researcher for this set was also shown to be more effective than the standard version for collection of tweets during spikes in the usage frequency of the selected-on hashtag, indicating that the previous dataset, #le14, may be incomplete and that analysis based on it should be subject to a recognition of its bias away from periods of very high frequency tagged discussion.

The considerably larger size of this case study dataset exposed some methodological considerations related to the extensive time required to perform manual elements of the coding of data. A strategy for approaching user coding for such large cohorts was adopted via the analysis of frequency of contribution and the reduction of users to be analysed to only those who had more than incidentally used the hashtag for the set. Thus, all users who appeared twice or more as contributors to the corpus of tweets in this study were coded, reducing the number from 79,609 to 9,510. Network and Descriptive Analytics tested on the previous case study were deployed again for this dataset, and its sub-network of coded users. Again, these methods offered insight into the patterns of user behaviour by category, and into the communicative strategies employed - some of which differed significantly from those observed in the analysis of #le14. Importantly, analyses showed that despite the very substantial reduction in the number of users considered for some methods of investigation, those users included accounted for the vast majority of all tweets collected, as well as for the origination of 98% of retweets, presenting this sub-selection method as a viable option for dealing with similarly very large datasets in future research.

188

In addition to confirming the investigative utility of the methods selected for this and the last case study, this chapter has also presented detailed information on the patterns of interaction in political discourse on Twitter, specifically outlining these for the case of a referendum. Taken with the previous chapter, these findings on Irish political discourse on Twitter represent novel investigation for this field, and contribute to knowledge of patterns of participation in this mediated discussion by professionals associated with the field of politics, and by citizens. While the #le14 case demonstrated a role for citizens in a discourse field still largely dominated by a small number of socially influential figures (chiefly national politicians and professional media agents), this chapter's analysis of the #marref discussion showed a significantly increased participation and centrality to the topic for citizen users of Twitter. The participation of activists, too, was shown to be highly influential, with a small number showing disproportionate presence at all periods of the study, and being the focus of directed communication from other groups. This case study, much more than the previous, has shown Twitter as a more transformative and mobilising medium for citizen discourse, consistent with how it (and internet-mediated political discourse more generally) is sometimes portrayed in the literature (e.g. Castells 2000, 2006, 2009; Shirky 2008).

6.6.1 Revisiting the research questions for this case study

Who are the participants in the #marref Twitter discussion and how can they be meaningfully categorised?

Once again, it is clear that in the studied #marref dataset, the enormous majority of users are not professionally associated with politics, though some of the most prominent of these actors have come to blur the line between 'citizen' and 'activist' in their self-presentation and the overt stance adopted on the particular referendum issue. Among coded users professionally associated with politics, national politicians and professional media are still observed to take prominent positions in the discourse, especially close to the voting event, with the closest comparison of their influence and communicative directionality measures to those of #le14 seen here in the period following the vote. Perhaps unsurprisingly given the national level of debate on a social change issue, local politicians exhibited little influence throughout, though their directed communication is again shown to be substantially divergent from national politicians, warranting their separate coding.

An additional coding based on the stance actors took toward the referendum was applied to the 50 most frequent contributors and 50 most frequently mentioned. This coding allowed for some additional insight into the relative prominence of both, and showed that a smaller cohort of campaigners opposing the referendum were responsible for a disproportionately large amount of tweets, as they discussed the issues with the larger segment of proposing users. An expansion of this coding was identified as a potentially insightful further study, however there would likely be a correlation between the number of users to be coded on referendum stance and the difficulty of that coding. Users who were highly involved in the discourse tended to display readily evident data in their profiles to show their political position on the issue, though this was already being removed from profiles in the later coding period after the result, and may not be evident at all for many users less frequently contributing. Content analyses of sufficient samples of tweets from every user to be thus coded would be required, necessitating a great deal of time to complete such a process.

How can analytical methodologies for networks be applied to investigate the patterns of interaction in Twitter discourse, using this case study as an example?

This chapter has again outlined the investigative utility of the methodological combination of approaches from Network and Descriptive Analytics deployed in the #le14 case study. In particular, this #marref case study served to demonstrate (1) the comparative performance and implications of use of various open-source data collection strategies, (2) the application and adaptation of methods to a larger dataset, with a substantially increased number of contributing users, and (3) a comparative analysis between the discourse around a multi-focal, segmented, and party political electoral event (#le14) and that of a single nationally-focussed binary social change issue (#marref).

The methods deployed were shown to have strong explanatory value for the structure and directionality of discourse within the network, and the adaptations to manual user coding were shown to produce useful insight on a sub-set of all users, who could be demonstrated to be the contributors of the majority of the collected tweets. The addition of analysis focussed on the top 50 most active participants in the discourse served to provide additional insight on influence within the discussion, and identified the most prominent users by their participation. Additional coding with respect to their stance on the discourse issue (the 2015 Irish marriage referendum) helped to contextualise these most active contributors and recipients of content.

Finally, the use of descriptive metrics, similar to those used in the toolset of the previous case study, provided more information on the context of the discourse captured, its relation to specific media events which could be correlated with observed spikes in contribution frequency, and the nature of the content originating from actor groups. The latter, in particular helped to differentiate the retweeted contributions of media professionals from

190

those of activists and citizens, highlighting likely reasons for the aggregate sharing of tweets from both.

How do the findings of this case study compare to that of the previous study of #le14 discourse, and what research insights may be drawn to inform the design of further studies?

This chapter has systematically compared the sets of metrics derived from the analysis of the #marref dataset, with that of the #le14 case. While some commonalities occur in the overall network structure and substantial presence of citizen actors, among whom a small number are very highly active and responsible for thousands of the tweets recorded, important anomalies are also noted. Crucially, metrics for the influence of users (% presence, % origination of RTs, % of directed mentions received, etc.) show a different overall 'shape' for the #marref discussion than for that of the local elections in 2014. Activists clearly lead this discourse, adopting a prominent presence from the earliest period and retaining this throughout, directing content to national politicians and professional media, but also interacting with the public. Citizens similarly feature heavily among the most frequent contributors, and in the same measures of influence that highlight activists. The discussion is shown to be closer to theorised paradigms of internet-mediated public spheres than the previous case studies, with citizens engaging one another to a much greater degree, especially at the earlier stages, before the discourse receives a wider media context. Even later in the studied duration, however, citizens are still seen to highly engage with one another and with activists, while politicians and the professional media engage with them - a feature not observed in the #le14 case.

6.6.2 **Opportunities for further study**

The findings of this chapter not only provide additional insight for Irish political discourse on Twitter, and new information on the discourse of #marref in particular, but provide implications for the design of further research. As outlined in the caveats and limitations for this study, the performance of data-collection software has been shown to vary, subject to the rate at which the tracked discourse is occurring, prompting questions as to the completeness of derived sets. This thesis had already outlined the difficulties around using other sources of data for the study of Twitter, and the imperative for social scientists to access a viable open-source tool; thus, future studies focussed on developing knowledge of the extent of completeness of data capture, and the factors affecting it, would benefit the design of future investigation and the interpretation of existing analyses.

Similarly, this study has uncovered a set of communicative patterns among actors that significantly varies from those of the participants in the #le14 discourse. The latter had a useful comparison study in the form of the analysis of actor participation in the 2012 Belgian local elections in Flanders (D'heer & Verdegem 2014), demonstrating some consistencies in interaction and communicative directionality across two similar political locales. Studies of national binary political issues, such as other referendums, would likewise offer a very useful comparison with this analysis and may be informed by some of the methodological approaches at work here.

Finally, some of the questions implied by the evaluation and findings of this work serve to further frame a need for qualitative investigation of the macro social context and micro-level actor deliberation around content contributions. The next chapter, which outlines the interviewing component of this research, begins to respond to these shortfalls in knowledge by addressing questions about participation and conventions in Twitter political discourse to highly active users identified from each of the coding categories used in this thesis.

Chapter 7

Insights from interviews with highly active users in Irish political discourse on Twitter

Insights from interviews with highly active users in Irish political discourse on Twitter

7.1 Introduction

This chapter presents the qualitative findings of the work, outlining brief profiles of each interview participant, and important insights specific to each case. These descriptive profiles foreground thematic analyses later in the chapter, which draw together threads of commonality across interviews, on several key research questions. Chiefly, this investigation seeks to offer contextualising insight to the reader on the social setting of active contributors to political discourse on Twitter, specifically acknowledging and exploring the Irish social context, and whether it implies unique shaping of the discourse in the minds of these participants.

Concepts from Bourdieu (see: Chapter 2) provide a useful framework for focussing the emergent themes across the extensive and rich data collected in the interview responses. Treating the mediated space of Twitter as a field, concepts of its *doxa*, and the *habitus* of individual actors help to coalesce insights on how participants have come to use and understand it as a communications medium. An overview of the social context in which the discourse is situated emerges, and is discussed in this chapter. Concepts of the social and symbolic capital within the field and how this transfers between related (overlapping and overarching fields) provide an approach for understanding particular motivations of participants that were identified.

In keeping with the theoretical approaches outlined earlier in this work (see: Chapter 2), these interviews expose technological agency in the actor-network constituting Irish political discourse on Twitter. Specifically, the degree to which participants depended on technological intermediaries was examined, with responses from participants informing an understanding of the degree of content shaping and the extent of decoupling of content and creator brought about by technological interventions at multiple stages of the Twitter interaction.

Findings based on the specific interrogation of hashtag use among participants form a crucial response to one of the key research goals of this work, and these are explored in this chapter and summarised as a set of insights which are likely to have implications for the interpretation of existing work and which should be accounted for in the design of future

studies. This information directly begins to address acknowledged shortfalls in the literature, and frames further questions that might shape continuing improvements to data collection and analysis. The chapter thereby derives several imperatives for computational social science research on Twitter discourse.

7.2 Participant profiles and overviews

The research interviews were conducted from late 2014, to mid-2015 in Ireland. The interviewing was designed according to the seven-stage process outlined by Kvale (2007) and detailed in the methodology chapter of this thesis. A summary list of interview participants is given in the addenda to chapter 4 (Appendix A.4.7). In all, twenty-three interviews took place following the initial pilot interviews that were used to refine the enquiry.

Participants in this study were selected for interview by (1) their presence in the pilot study for this work, outlined in the Methodology (Chapter 4, Section 4.6.12), and (2) attempts by the researcher to represent all six of the actor categories used in the coding of contributors in the two case studies explored. In addressing the latter, participants from actor category 1 (national politicians) presented an additional potential difficulty in that there existed a strong likelihood that their contributions to Twitter were modulated, at least in part, by the media policy, direction or management of their respective political parties. It was therefore necessary to try to ensure a wider selection of participants from category 1, in order to include government and opposition, ministers as well as backbenchers, and independent as well as party TDs. This resulted in a larger interview cohort of national politicians, but the various insights they provided confirm an effect on their contributions by these specific contextual factors in national professional politics.

As previously outlined, issues of participant anonymity were a crucial concern of the research design, and in some cases the provision of partial or full anonymity was necessary to encourage candidate participation, many of whom clearly expressed that they would not otherwise be comfortable with providing some candid opinions. A notable trend was for national politicians to most likely request some level on anonymity, likely a reflection of the highly managed relationship with media that they maintain. Insights are offered from aggregate themes among this cohort, on their heightened awareness of self-presentation in Twitter, and are discussed in this chapter. In general, about one third of the interviewees requested some level on anonymity for their contributions to the study, with this correlating to those most closely professionally associated with the field of politics - TDs, Councillors, and political correspondents.

While some participants opted to be fully credited for their contributions, and their names and Twitter usernames are provided in the following profiles, this study adopts a policy of using a coded designation for each in all subsequent sections. This coding follows a simple scheme of assigning a letter designation of their actor category (NP: national politician; LP: local politician; A: Activists; PM: Professional Media; OM: Other Media; C: Citizens), followed by a sequential number indicating the order they were interviewed among those in their cohort. Thus, NP4 provides an anonymous designation for the fourth interviewee among the national politician participants, and A3 the third activist interviewed, etc.

This section briefly presents a profile for each of the participants who took part in this research, in order to provide the reader with an overview of their relationship to the discourse, their adoption and usage of Twitter, and important specifics of each case. These profile vignettes are descriptive, drawing on data gathered during the interview process, and they are intended to provide insights on the participants that might usefully inform the later aggregate analysis of their contributions. This chapter then proceeds with examinations, in greater depth, of the themes emerging from the corpus of interviews as a whole.

Where the directions concerning aspects of anonymity required by a given participant conflicts with the provision of profile information about them, this is noted and withheld; otherwise their name, date of joining Twitter, and other contextualising data are provided. In each case, the 'verified' status of their Twitter account (whether specifically named or not) is indicated. This is an official process, carried out by Twitter, to verify the identity of 'noteworthy' individuals (See section 4.3.2.2.3) and is mentioned by several participants as an important indicator of status in the network; "*I'm proud of the blue tick alright*" (Participant PM2, referring to how the status is displayed by default); "*I do think that having a verified account* [...] means you're taken more seriously... by journalists anyway. I've noticed they sometimes include a screenshot of the tweet... showing the profile... I think to show that it's verified" (Participant NP7).

7.2.1 Participant NP1

Participant NP1 is a male opposition TD with the Sinn Féin party, and was interviewed in September 2014 at his office in Leinster House Dublin. He opted not to be identified by name, though indicated during the interview that he was happy for his party and constituency to be identified. Since these pieces of information taken together necessarily identify him, neither his name nor specific constituency will be provided here.

NP1 is a frequent contributor to political discourses on Twitter and could immediately identify a number of topics on which he had recently tweeted. Subsequent analysis of his Twitter contributions showed his recollections to be correct, down to the wording of particular tweets from several days earlier, which might offer some indication of the consideration put into their phrasing. NP1 referred several times to his (largely rural) constituents being unlikely to form any substantial part of the political discussions he is involved in, e.g. "…it's not likely that the people who come back to me about comments on gay marriage or retweets about those kind of issues are the ones in [constituency town] …they're on twitter, some of them, and I hear from them there when it's a local thing, but they're not my audience; that's the wrong word - they're not who I'm in regular contact with on the big issues that matter to me".

Referring to his ownership and authorship of the Twitter account in his name, NP1 was clear about his understanding of the perceived authenticity of his contributions among others in the network, and stated firmly that he is solely responsible for the account. He contextualised his own use of the account in terms of the guidelines and expectations of his party:

"There are policies... party policies, yes... for how social media is supposed to be used. [...] We had training on it. But there's no dictating what I say or anything like that [...] - no, it's more that we'd be expected to retweet certain press releases for the party [...] we would all do that, but the tweets - as in the actual tweets from the account - they're all me. It's me writing those."

"It's important to me that I'm the one who does it. You can tell on some people's - some TDs' [emphasised] - accounts that it couldn't be them saying the stuff. [...] People can tell, so it's important it's... authentic, or why would people engage with you. That's important to me anyway."

NP1 identified several affordances of the Twitter mobile application as central to the management of his presence and interaction in the medium, describing particular settings he used for managing lists of users, custom notifications, and familiarity with particular views within the interface. He was clear that his use of Twitter, which is among the highest in terms of frequency of contributions by politicians on the topics measured in this study, is a source of pride; *"I wouldn't use it if I wasn't going to do it right. It's [...] an important tool. We have it and we*

can use it to organise and to rally around causes... I'm thinking of the Savita⁷⁸ case for example. I don't think that would have been talked about like it was until Twitter made it be talked about. I think we'll see that happening again and again now. [...] I'm proud to be part of something like that."

7.2.2 Participants NP2a and NP2b

The second interview to take place with a national politician was held in Leinster House, in October 2014. In this case, there were two participants, designated NP2a and NP2b⁷⁹, and they indicated that they wished to be fully identified in the work. NP2a is Catherine Murphy (@CathMurphyTD, verified), an independent TD for Kildare North, first elected in March 2005 and re-elected to the Dáil in 2011. She participated in the interview along with her Director of Communications, Anne-Marie McNally (designated participant NP2b in this research), whom she identified as a co-contributor to her Twitter account.

In contrast with other national politicians interviewed, NP2a identified from the outset that her Twitter account is a "*representation*" of herself and her role - one managed by several people, including NP2b. Here again, the importance of Twitter as a sphere of mobilisation was a theme, with discussion particularly of the use being made of it by the political left. Responding to the question on whether political discourse on Twitter translates into action, NP2a cited the water protests⁸⁰ of that week; "*I don't think you'd be seeing [the water protests in Dublin]... at the scale you're seeing it. [...] They were organised because of Twitter to a large extent*".

NP2a and NP2b discussed their adoption of Twitter in September 2010 (this was a jointly managed process), outlining how an important factor in their initial engagement with it was the potential they felt it had for providing a national platform for NP2a, who lacked the political communications machinery of a party TD. Here, they felt that it had particular relevance in bringing stories of NP2a's political action and campaigning on issues to the attention of the national press. Both participants confirmed that they were aware of media agents in their regular discussion network; to whom they do not explicitly address (@-mention) content, but whom they expect to "*pick up on what we're saying there*".

⁷⁸ The death of Savita Halappanavar at University Hospital Galway in October 2012, which led to a national discourse on abortion legislation in Ireland.

⁷⁹ This participant is so designated to show her relationship to participant NP2a, reminding the reader elsewhere that this was a single interview with a pair of participants. It should be noted however that NP2b is not a National Politician, as the NP prefix indicates in most other cases. She does, however, work directly with national politician NP2a, representing her in messages sent on Twitter from her account, and the subject matter of the interview only concerns NP2a.

⁸⁰ Popular protests at the introduction of charges for the domestic use of water and treatment of waste

7.2.3 Participant NP3

Participant NP3 was interviewed in October 2014 opting for full credit, but later withdrew from the study.

7.2.4 Participant NP4

NP4 is a male opposition TD from a political party, who opted for partial anonymity in this research. He is a young TD from a rural constituency and was interviewed via Skype in January 2015, to suit his travel schedule. This interview, as with others carried out over Skype, was shorter than those carried out face-to-face. In this case, NP4 offered to use the video chat capabilities of Skype, in an effort to make the interview more personable, but problems of his connectivity impeded the smooth use of this, and the interview was functionally similar to a simple voice call. While this made for a more difficult interview, in that it curtailed any possibility of the researcher reading visual cues and modulating the interview in response, some important insights were attained.

NP4, though the youngest national politician interviewed, was the least concerned with Twitter and made clear his view that engagement with it as a medium was a chore, required by his party; "I don't have much time for it [...] I have to be there, so I am. I've things I have to retweet... they're signalled to me... I'd often do a few of them together". Like NP1, he sees little relevance for his constituents in his presence on Twitter - "It has nothing to do with [the constituents] – they're not on there" – but in this case, NP4 did not see the same opportunities for issue discourse at a more national level.

While NP4's comments, likely affected by the voice-only nature of the interview, were perfunctory and laconic, his participation was very valuable in uncovering an important shortfall of the pilot study selection of participants to interview; his highly regular (and party-mandated) use of Twitter to retweet press releases, tweets from party colleagues, and points of local information, led to a sufficient frequency of use of the hashtags and keywords associated with Irish political discourse to warrant his inclusion in the cohort of the pilot study. Though only 'broadcasting' into the medium and participating in no meaningful discursive engagement – "*I don't talk to people there at all*" – the trace data selection methods were unable to differentiate his presence as an 'active' actor in the discourse from any others.

7.2.5 Participant NP5

NP5 is a male government TD and minister of state with a verified Twitter account, who opted for partial anonymity. He was interviewed at his office in Leinster House, in February 2015, and spoke of his enthusiasm for Twitter and his frequency of use of it. Like NP1, he was able to identify hashtags he had used, for topics he'd contributed on, though with less accuracy and immediacy. Similarly too, he was very clear about his sole authorship of content sent from his account; "*It's all sent by me, from my phone most of the time, though I do use the website too.* [...] *I'd often comment on things in the evening – nothing to do with politics… football, events I'm at, television*". NP5 was conscious of the authenticity of the personal in his content, citing examples of casual conversations on non-political topics as reasons his Twitter presence would be identifiably "*real*" to others.

NP5 described his initial adoption of Twitter - "I was advised to use it by an election agent of mine in [...] March 2010... I initially used it in a very stiff formal manner, and then I realised this wasn't necessarily the case and began to use it to give people, maybe a sense of my personality and a sense of what my job entails". He describes not having used Facebook before beginning to use Twitter and being initially uncertain how it might be effective for him. Throughout the interview, he discusses the medium in terms of its effectiveness and utility for his political career and demonstrates keen awareness of his own self-presentation within it, and his attempts to modulate that in response to how he wishes the news media to view him; "When you're a councillor and you're running for election, sometimes you want to say something a bit edgy [...] and as a backbencher maybe, similarly, you might want to be the first to tweet that someone has voted against the government [...] to give the insider's view [...] but now I'm not really in the sphere of looking for attention... when you're a minister of state, stuff you have a view about, you now kinda keep to yourself"

NP5 was less certain than other participants on the use of hashtags and their functional value in Twitter as hyperlinks to topics. He described "getting better at it" with time, and indicated that his party had provided training on the use of Twitter specifically. He linked Twitter to current affairs television, and especially to the *Vincent Brown Show*, which has a wellestablished hashtag (#VinB) and identified it as central to the development of media and wider social discourse around "controversial issues", giving '*Pantigate*³⁸¹ as an example.

⁸¹ Comments made by Rory O'Neill (Panti) on RTÉ television in January 2014, alleging specific individuals to be homophobic, which led to legal action, an apology and monetary settlements by RTÉ, and subsequent discussion on social media, coverage in the national press, and debate in parliament.

7.2.6 Participant NP5x

This participant is designated NP5x to show his relationship to participant NP5. Unlike the case of NP2a and NP2b, which use the consecutive letter suffixes to indicate that they were interviewed together, as part of a single session, this participant is designated NP5x to remind the reader that the interview is an appendix to NP5, above.

NP5x is the parliamentary assistant to NP5 and at the suggestion of the latter, was interviewed separately and alone in his own office space following the previous interview. It became apparent at the conclusion of NP5's interview, that he had discussed this research with his assistant and that both were interesting in contributing, though no notification of this was received before the event. This interview and participant are unique among the set in providing a secondary perspective without the presence of the principal subject, and whilst the unusual nature of it compared to the other interviews might have discounted it in other circumstances, it offers particular insights not attainable elsewhere. NP5x opted to be credited fully on his consent form, but to do so would necessarily identify the politician for whom he works (NP5), who did not select the same level of identification. NP5x is therefore not listed by name in this work. He is 27 years old, has worked for NP5 for a number of years and is responsible for some aspects of communication on his behalf, including engagement on national issues related to the ministerial department, as well as constituency issues at a local level. He offered some insights on his own use of Twitter and indicated fluency both with affordances of the medium and with the identification of key actors, naming other politicians and professional media agents whom he considers "use it well, have made a name for themselves there too". He indicated that he was pressed for time, so a full interview did not take place.

Importantly - and unusually, given that NP5 directly suggested this appendix interview with NP5x - NP5x identified *himself* as "the main contributor really; the person who is actually sending the tweets... almost all the time" for the account identified with NP5. This represents a critical insight of this short interview - it is likely the case that national politicians are often represented on Twitter, perhaps in the majority of content sent in their name, by assistants who manage their communications, but that they are sufficiently conscious of this fact being potentially problematic to take steps to obscure it. In the case of NP5, a clear statement was made that he alone was the author of his Twitter output, moments before he invited the researcher to interview his assistant, who revealed the co-authorship of it. It seems less likely that this was the unintentional uncovering of a deception, than that NP5 is sufficiently accustomed to claiming complete personal autonomy over his Twitter account that he maintained this out of habit in his interview, even when he was likely to be clearly

202

contradicted immediately afterward. This serendipitous insight calls into question the assertions made by almost all other national politicians who were interviewed alone, and who similarly stated that they are the only ones tweeting on their accounts. This also likely demonstrates an internalisation on the part of at least one national politician, of a narrative of proclaiming discrete personal agency over an associated Twitter presence to reassure the 'realness' of it.

7.2.7 Participant NP6

In February 2015, NP6 was interviewed via Skype voice call for scheduling reasons and initially opted for partial anonymity, withdrawing entirely from the study several months later.

7.2.8 Participant NP7

NP7 is a former TD and Minister, who was interviewed at his home in Dublin in April 2015. He contextualised his use of Twitter for political discourse by setting it among a range of online and social media communications - "... Facebook, Twitter, various sites, I suppose... online newspapers I read... I'm not sure I could separate them in that they're all interlinked [...] Twitter is talking about something from the Irish Times website, the Irish Times is quoting something from Twitter... none of them are their own thing". Like other participants, however, he identified uniqueness in Twitter's structure, affordances, and patterns of reciprocation that lend it a particular utility for political interaction and debate; "You're in a conversation there all the time. [...] You can't be subtle on Twitter [because] you don't have the room, but you also can't really have a lively debate on Facebook, where there's no culture of back and forth... Twitter has won out there"

In common with most other national politicians, NP7 described his adoption of Twitter (in February 2009) as a "political tool from the start", beginning his use of the medium during his tenure as a minister. He describes some apprehension about how he was "coming across to people who were more experienced in it than me", and acknowledged that he relied on the guidance of his staff in helping to "craft" tweets, gaining greater confidence in his own use of it with time. Despite this context from government, some of his post-ministerial experience of Twitter, as a member of a small party, was similar to that of independent TD NP2a; "when you don't have [party] press officers and all that machinery behind you, getting your name and your quotes out, you have to find opportunities to do that yourself [...] it wasn't so much the case at the start that the media would use [Twitter] as a source... but now it's standard, now you say something there and it will be reported".

7.2.9 Participant LP1

This participant is Green Party Fingal Councillor, Roderic O'Gorman (@RodericOGorman, using Twitter since December 2008), who was interviewed at his office in Dublin, in January 2015, opting to be credited in the study. LP1 was the first local politician interviewed, and expressed enthusiasm for Twitter as a communications tool at the constituency level, describing the particular discourse around the #dubw tag – representing the Dáil constituency of Dublin West, but widely used by local residents to bring up issues for councillors also; "schools, roads, bus routes". As national politicians (NP1, NP2a, NP5, NP7) had similarly identified, Twitter's potential to gain press attention for a politician was a key factor to its adoption by LP1 also; "There are a few stories - local and national coverage [...] - that came about directly because of it. [...] There was a story where the quote from me is a tweet".

LP1's identification of Twitter as important to political discourse at highly local levels - he cites traffic management in Blanchardstown village, as an example - stands in contrast to opinion among national politicians that it is a medium for larger societal issues, and outright rejections by some (especially NP4) of its relevance to constituents and discussion with them on their local concerns.

LP1 described receiving training on the use of Twitter by his party, identifying these as "more guidelines really, than training... acceptable use, what not to say". He reported some unfamiliarity with features of Twitter and stated that his use of it was mostly via the website.

He described his identity on Twitter as very much related to his political role; he is also a university lecturer in law, but neither his job, nor his personal life feature strongly in his contributions, he said.

7.2.10 Participant LP2

LP2 is a city councillor with the Labour Party, who was interviewed at her home in January 2015. She opted to be identified only by her party affiliation and described very similar experiences to national politicians and to LP1 in her use of Twitter as a means of media exposure for profile building. Uniquely among both groups of politicians interviewed, she reported dislike of many aspects of the medium - *"the trolls, the people who are just sitting in it waiting to pick a fight or go on some wild tangent […] yesterday I retweeted [a charity event in the local constituency] and immediately had people ask completely unrelated stuff - do I support [a party colleague] on abortion? What? […] It's as if they're just waiting to see some movement from you to know you're there, then they send whatever controversial agenda is on their mind and complain if you don't respond on that".*
Importantly, LP2 related her negative comments about the nature of the Twitter discourse, and of the interactions from the public, to her selection of partial anonymity; "*I'm doing this* [interview] now, and I'm being honest about how frustrating it is and how just negative a space it can be, but I know that if that was reported back into Twitter, I would be... condemned really, for not 'taking part' or not seeing it as an important democratic thing that I shouldn't complain about."

LP2 began using Twitter on the advice of her party, and her local election campaign manager, whom she identified as *"addicted to it"*. She reported greatly preferring Facebook as a medium, describing it as comparatively *"less time-consuming, less nasty overall [...] probably more effective for me in just getting my message out to people in [the constituency area]"*.

7.2.11 Participant LP3

Participant LP3 is a Councillor and was a Municipal Mayor when interviewed in February 2015. He is among the youngest holders of a local authority mayoral office in Ireland and described his youth as an assistance to his use of Twitter and other social media; "*I'm a digital native, really. [...] I can understand how other politicians would have problems coming to [social networks] and having to learn what's expected of you, what's appropriate"*

Speaking during the course of campaigning ahead of the 2015 marriage referendum, he described Twitter as an important avenue for setting out his and his party's position. He identified the #marref tag as one where he was a frequent contributor, though he qualified its representativeness for the debate; "I don't know that [#marref] is where all of the discussion is happening, obviously. [...] The no side are less inclined to use it. Definitely the no-people who contact me when I tweet something about it, they don't use it... Most of them don't".

LP3 has used Twitter since January 2012, and displayed familiarity with a selection of tools for managing Twitter timelines and output, citing TweetDeck⁸² and Hootsuite⁸³ among the applications he used to coordinate incoming content by topic or user, and to schedule outgoing tweets.

7.2.12 Participant LP4

LP4 is a male councillor residing near a provincial town in a county in the midlands, interviewed in February 2015. Although he opted for partial anonymity, this participant felt that it was important to specify his rural situation and status as an independent local

⁸² https://tweetdeck.twitter.com

⁸³ https://hootsuite.com

politician in contextualising his use of Twitter. Despite his opinion that "[it's] new for me, I don't know - you know - whether I am using it to the right extent or if I'm wasting my time in it at all", he was sufficiently active to place him in the target group for participant selection during this study.

Like other politicians interviewed, he identified Twitter as a useful tool for bringing his work on local issues to the attention of the press, differentiating the recent (2014) local election from previous ones, because local media now paid attention to Twitter in a way that he had not experienced before; "[The local newspaper] had quotes from my tweets about a water quality issue [...] we campaigned on. That was the first time for something like that - you know, because the person writing that story knows my number, would have called me about that in the past and would have gotten their quote that way. It was new that it was Twitter".

Unlike LP1, and perhaps giving an indication of the same attitudes to Twitter in rural constituencies that were highlighted by NP1 and NP4, this participant did not see Twitter as a facilitator of discourse on a local level; "*The [water quality issue] wasn't something where lots of people around the area were chatting to me... I mentioned it on Twitter because I'm now in the habit of mentioning what matters to me on there, and it was picked up, but [...] it didn't start there, it wasn't discussed there really*". Instead, he identified Twitter as a medium of discourse of more national or more widely social issues, citing his contributions on local government reform, local infrastructure, and agriculture as those which were more usual for him; "the things where there's real conversation happening".

7.2.13 Participant A1

Participant A1 is executive director of an international activist NGO, and the founder of a high-profile Irish charity. While also a former appointee to the Seanad (May - July 2007), he is not categorised in this study as a national politician, since this short appointment was linked to his activist work. He was interviewed in February 2015 and contextualised Twitter among a variety of "media channels" that he used as part of his activism; "It has its place, and I do know I use it a lot... a lot! [...] There's an organising value to it, especially around particular campaigns that [Amnesty International] might run and that I want to bring people's attention to. [...] A few years back, you wouldn't think of having a campaign that didn't have a good title to hook people, or a good line to sum it up, now it's all about a good hashtag, isn't it"

Describing his adoption of Twitter (in April 2009) as "a combination of pressure from my children, from my friends, from people who were telling me it was important to politics", he outlined his enthusiastic use of it around electoral events, and issues debates (especially as a "side chat" to televised debates and current affairs programming he identified). He keeps his use "mainly to the phone, sometimes the website... or the iPad, rarely though", and describes his presence in Twitter as "mainly the work-me, I suppose", finding that attempts to include more personal information are sometimes detrimental to that; "people can pick up on something that you're joking about, or a comment on... I don't know, television perhaps and then run with that... 'Why is he joking about X when Y is happening? Why doesn't he care about... whatever it is' - that's frustrating".

7.2.14 Participant A2

A2, Michael Barron (@MichaelNBarron) is the director of *BelongTo*, an organisation supporting young LGBT people in Ireland. He was interviewed in April 2015, during the height of campaigning on the marriage referendum, and this permeated many of his contributions. Having begun to use Twitter in March 2012, he described "[coming] to it late, really, because I knew at that stage that a lot of people were using it [...] were finding it useful".

Contrasting other participants, he describes his presence in Twitter as "a mixture of issues stuff, personal stuff", though he acknowledges that he perceives expectations from others for how he should engage with issues; "There are certain things I'd... be expected to comment on - not that I'm especially the expert on them - but people will ask sometimes what I think of something... I'd be conscious of that".

He described very deliberate actions of addressing politicians and media figures in some of his contributions, matching aggregate patterns for the 'Activist' category of actors in the case studies of this work; "I would [@-mention] certain politicians... especially if I am retweeting something and I'm want to say 'look at this, do something about this' [...] I see other people doing that too and you can tell... it's almost a comment on who should be responding to some issue".

7.2.15 Participant A3

Participant A3 is Rory O'Neill (also known as 'Panti'), an LGBT activist and drag queen who became a central figurehead of the yes campaign for the 2015 Irish marriage referendum, but whose prominence as both a subject of and contributor to political discourse in Ireland began much earlier than this. He was interviewed at his bar in central Dublin, in July 2015, more than two months after the result of the marriage referendum, so many of his responses are reflections on that specific context.

Speaking about his initial adoption of Twitter (in March 2009), A3 said "I heard of it, a few years ago, before it became huge [and] I do try, a little, to keep up with these things. [...] Twitter is one of those ones that kind of suits me. I like the way it passes information around. [...] The links, that's what I'm

interested in". Here, as with other participants, a sharp contrast was made with Facebook in terms of currency, brevity and relevance of information. A3 offered a particularly well-delineated account of issues of the management of presentation of self within Twitter, since he is known both as Rory O'Neill, the person, and Panti, the drag queen character. These 'selves' are represented separately on Facebook, where he maintains separate accounts for both and describes the deliberative processes involved in choosing to which a particular post is made. On Twitter, however, he has one account (@PantiBliss, verified), and feels the need to hold back certain content that doesn't fit his understanding of what the following for that Twitter account expects; *"There are absolutely things that I don't tweet, because they're too Rory; I keep those for Rory's Facebook page*". This highlighting of the deliberation behind the social suitability of contributions made on Twitter echoes sentiments expressed by others, who wish to represent a particular version of themselves from amongst a plurality of self-definitions; LP1 is a good example of this, wishing to maintain a distinct representation of his role as a councillor.

A3 was clear on his general disregard for hashtags in his own contributions on Twitter, despite fully understanding their function as topic markers and navigational affordances within the medium. He was unable to recall any hashtag used recently and could only point to rare occasions where he would include one; "*I guess I include the hashtags when it's something I want to support, in a sense. [...] I'm more likely to want to put it in, if I think it's a considered hashtag -#marref wasn't a bad one; it's simple and short*". When questioned on his actual use of #marref, however, he said that he seldom used it, believing that followers implicitly understood the topical context of his remarks without it.

7.2.16 Participant PM1

Jason Kennedy (@JasonKennedy, verified) is a journalist and videographer with the *Irish Independent* newspaper, having previously been employed in a similar role at the *Irish Times*. Having used Twitter since February 2009, he described his initial experiences with is as "*personal, chatty*" and outlined a process of having to "*clean it up... somewhat*" after his employment at the Irish Times began and it shifted from a personal to a professional representation. "*I would watch what I say*", he said, when asked about whether he has to consider 'appropriateness' of contributions, "*There are definitely things that I've deleted too, and things where I've seen people... politicians... say things in Twitter that become stories in themselves because they're... inappropriate... or hasty*". This modulation of his contributions has continued at the second major newspaper for which he's worked, and he stated that it would "probably affect some of what I would like to say about the [then upcoming] referendum. I'll have to watch that". Like other participants, he described a segmentation of his social media self-presentation, being "more myself on Facebook, which is locked down, than [on Twitter] which people might... go looking through". He described his previous use of the Twitter contributions of politicians in reporting and acknowledged it as an important source to the newspaper, citing recent examples of stories broken from Twitter, and one wholly concerned with comments made by a politician on Twitter.

7.2.17 Participant PM2

David Cochrane, participant PM2, is *Communities Editor* with the *Irish Times* and is the founder of politics.ie, an online forum for the discussion of Irish politics, which he ran until August 2012. His verified Twitter account is @DavidCochrane, which he has used since May 2007 - the earliest adoption among the participants. This provided an interesting technological contextualisation of Twitter adoption, since PM2 initially began using the service during its incarnation as an SMS service, which he found a useful integration with his management of the politics.ie forum.

He describes his role at the Irish Times as "I direct our social media and digital strategy - editorial, and recently commercial as well, across the organisation", providing internal training and direction to staff on their communications with audiences via social media. He also works on "digital strategy and the way the Irish Times uses social channels itself - meaning irishtimes.com - promoting print, engaging with audiences". Likely due to his professional role, PM2 approached his analysis of the use and affordances of Twitter is highly functional terms, outlining what he saw as the "essential value" of hashtags, @-mentions, and the Twitter-verified flag in terms of their specific utility to journalists. His discussion of the commercial aspect to decisions on these was a useful point of view in qualifying remarks from other participants in the media category (see Section 7.4.3).

7.2.18 Participant PM3

PM3 is a political journalist with a verified Twitter account, who was interviewed by Skype following several postponed attempts at a face-to-face meeting. This participant made it clear early in the process of selection that he/she could only provide candid commentary on his/her assessment of Twitter political discourse contributions made by politicians if anonymity is assured, hence this participant's name, employer and gender are withheld in this work.

PM3 described his/her adoption of Twitter as something that "took place early on; it hadn't been around long as a service"; like PM2, the SMS services offered by Twitter were a factor in its use by PM3, who stated that "it had clear potential from the start... from the start of politicians using it, I suppose I mean... to shake things up". PM3 described Twitter's "removal of intermediary layers [...] of PR, and handlers, and spin" as an important consequence of its widespread adoption, for the reporting of politics.

Echoing sentiments of the national politicians who were concerned about 'authenticity', PM3 suggested that Twitter *"has the power to give ordinary people a connection to a politician in a real sense... that could humanise them, increase their respect for them [...] but the corollary is there too... sometimes the mask skips, sometimes you'll see them fuck up completely, and at some level they all must be negotiating that balance*". PM3 offered assessments of specific politicians and political parties use of Twitter in his responses, in some cases drawing conclusions that were independently supported by other participants; his opinion that the political party of NP4 was "foisting Twitter on its members", for example, is borne out by that participant's description of his perfunctory use of the medium.

7.2.19 Participant OM1

This participant opted for partial anonymity and was initially difficult to define in the coding process, falling between categories of *Other Media* and *Activist*, due to the nature of her self-description (Twitter bio) and some of her contributions. On a more detailed examination of a sample of her contributions in the pilot study, she was coded *Other Media*, since her role as blogger and her promotion of that work, most permeates her part in the discourse. OM1's activism is not related to an NGO, charity, or other organised structure and therefore is best categorised as a form of personal lobbying of and awareness building, which further served to warrant her inclusion in this cohort.

She was interviewed at Dublin City University in April 2015, and described her more than seven years on Twitter as "sometimes a struggle to just keep at it when you're worn down by the same institutions there, the same pigeon-holing of you". She described a small network of users, many of whom she readily identified by their screennames, with whom she is most regularly in contact; "and it's not a love-in... I know people say Twitter or forums [...] can just be a collection of everyone congratulating each other on having the same opinion, but I genuinely think there is quite a spread among the people I'd regularly talk to. [...] I remember meeting [a named user] in person - and we'd have knocked heads on Twitter quite a few times by this stage - and thinking, in my mind, 'we have nothing in common; I'd hate you and everything you stand for, if I didn't kind of... come to respect [the person's screenname] as an intelligent debater".

7.2.20 Participant OM2

Participant OM2, (active on Twitter, since January 2011), is a blogger and part-time columnist, contributing to various online news websites. He was interviewed in central Dublin and provided information about his blog on Irish elections, and his presence in Twitter discourse of election issues.

Similarly to professional media agents, he regarded Twitter as a transformative introduction to the political landscape, in its affordance of more direct access by the public to their representatives. He termed this *"a demystification"* of the political elite and discussed what he viewed as profound consequences for journalism that followed from this.

OM2's promotion of his blog content was cited by him as a primary concern in shaping his contributions, overriding any use of other Twitter affordances; "I have to give a link and I have to give a summary - that will make someone click that link [...] hashtags and mentions are useful, sometimes, but not at the expense of those".

7.2.21 Participant C1

Leo O'Shaughnessy (@leoie) is a technology professional at a major multinational company. He opted to be fully credited for the interview given at his home in Dublin, in September 2014. C1 described his early experience of Twitter (from March 2008) in which he says he took several months to fully understand how it worked as a medium; "what I found is that as I've become more and more mature... in Twitter... what I find is that I use it now specifically to get information about a particular subject, i.e. by following a hashtag, or to disseminate information that I have found elsewhere". The use of hashtags was crucial to C1's experience of, and very frequent contributions to, political discussion on Twitter. He regarded them as an essential tool of categorisation and segmentation of discourses, and pointed to their utility (as searchable points of aggregation of other users' content) for finding conversations to join.

However, C1 also offered the most extensive discussion of the more subtle uses of hashtags – humour, irony, sarcasm and even the deliberate positioning of content where it might not otherwise be accessible; "I have used them - it has to be said; aggressively - in that I have created a tweet that I know would be irritating to one political point of view, and then used the hashtag that that political point of view follows, in order to perhaps put it in front of them".

Like other frequent contributors interviewed, C1 could readily identify a cohort of users with whom he regularly interacted, and regarded the network among them as the principle setting

of his own contributions; "Generally I find that the conversations that I join are somebody that I know talking to somebody else that I know".

7.2.21 Participant C2

Ciarán O'Driscoll (@bearaboi) is an Irish marine researcher from west county Cork, resident in Dublin. He was interviewed at the researcher's office and opted to be fully credited in the study. C2's frequency of use of Twitter was among the highest measured in the pilot study, and in the subsequent case studies, and he described hearing about and joining it in (January) 2010; "I'd been on Facebook, and everything else, and wanted to give it a try, and immediately found it useful and enjoyed the community on it".

C2 described his use of Twitter as a resource for political information; "a good source to go to for [...] what's happening, what are people talking about - people in the know, so to speak". Comparing it to Facebook, which he found "a very narrow place to have a debate", C2 expressed his strong preference for Twitter as a medium for political interaction on current issues and - especially for his area of interest and research - on multiregional and transnational discourses like EU policy.

Though this participant was the most giving of praise for Twitter's facilitation of political discourse, he was still keen to contextualise who was actually taking part; "generally middle-class, generally left-leaning people will be on Twitter [...] so while it might seem that everyone's talking about it on Twitter, no – only a certain section of people are there". Interestingly, he cited 'Pantigate' - the same example given by multiple other participants as exemplary of Twitter's ability to translate topical discourse from an internal discussion to a wider one at societal level - as an event which travelled poorly beyond Twitter and Dublin; "The Pantigate issue is one topic where, although it was massive on Twitter and social media, I think beyond that and [beyond] Dublin, it was a debate that most people didn't have an interest in. Speaking to home [west Cork], they heard something about it, but it wasn't a big issue".

7.2.22 Participant C3

Participant C3, Steve Conlon (@SteveConlon), is a PhD researcher in politics and, like other '*Citizen*' categorised interviewees, he opted to be fully credited in this work. He has used Twitter since April 2009, identifying it as a medium of political discourse for him from early on. He was happy to be considered an 'early adopter' and pointed out that he had introduced others to it, including a named acquaintance whom he *"advised [...] to get involved* *in Twitter*" and whom he identified as having strategically used the medium to build a popular profile and, ultimately, a career in television as a commentator on political issues.

C3 described Twitter as "an elitist medium", and the cohort involved in Irish political discourse on Twitter as "in a lot of cases, a politics.ie clique", referring to the use of that site (especially previous to the adoption of Twitter) as a forum for political discussion among a fairly small group of contributors, well-known to one another. Like other participants in his category, C3 demonstrated a fluent knowledge of the affordances of the medium, describing how the character limitation, hashtagging, retweet dynamics, etc. "set the rules" and made editing a tweet before sending it an important consideration in maximising its impact.

Though he was interviewed in July 2015, C3 also cited 'Pantigate' (which happened in January 2014) as an example of the translation of Twitter political discourse in Ireland into tangible political action; "[it] was a hugely important issue, social media had a massive role to play in it, in getting it and disseminating it... across the dinner tables, into pubs, and into newspaper columns. The great thing about it happening on Twitter is that allowed traditional media access to a source that it didn't have previously".

7.3 Understanding the social context of Irish political discourse on Twitter

Drawing on the theories of Bourdieu, outlined earlier in this work, the concept of field may be used to examine the themes emergent from analysis of these interviews. Treating the specific discourse network around issues of Irish politics in the mediated space of Twitter as a *field*, the participants in this study will constitute some the most active actors within it. Investigation of how their habitus has been adapted to it, of how capital operates, and of the field's relatedness to other fields, yields a framework for parsing and presenting some of the findings.

The situatedness of insight from participants' about their use of Twitter cannot be separated from the cultural, social, economic and geo-political domains of contemporary Ireland. Investigation was therefore undertaken of whether participants saw uniquely Irish factors at work in their assessment of the structure and characteristics of the discourse in which they took part. This question served to prompt participants who had not already done to so begin to specifically examine whether the Irish setting might be significant to the phenomena they had described.

Many participants, including national politicians, suggested that Twitter has a role in overturning 'parish pump politics' - the tendency of local community concerns to dominate the agenda for politicians in Ireland, ahead of, or instead of national debates. This was supported in the responses of politicians who indicated that they saw and were involved in very little discussion on Twitter of issues at a constituency level. With the exception of LP1, who cited a very active hashtag for the Dáil constituency in which he is a local councillor, no politicians saw a connection between the political discourse on Twitter and the political discourse taking place at regional or local levels. The specification of what "political discourse on Twitter" means may therefore be informed by this, at least in Ireland; it seems reasonable to conclude that Irish political discourse on Twitter largely concerns national issues.

Another theme to emerge in responses about the Irish context concerned the trust of politicians following the many scandals of corruption among senior politicians and institutions in Ireland in recent years. Several participants, and notably those in the 'other media' and citizen categories, described the damaged perception of politicians in Ireland and the facility of Twitter to be used to *"possibly help [politicians] create an image of themselves as normal people [... rather than] aloof ones that we might be suspicious of"* (OM1, 04/04/2015). The theme of

214

the public's changing perception of politicians, and Twitter's specific role in that, is investigated in the next section.

7.3.1 Habitus, Illusio & Doxa

Since an initial line of enquiry in interviewing investigated how each participant had come to use Twitter, the concept of habitus is one that can be informed by the themes emerging from participants' responses, and the detail they provide on adapting to the medium. While variances in the age of particular participants and their experiences with digital media technology can be seen to affect the fluency with which they adapt their habitus to the medium (cf. LP3's discussion of being a 'digital native', Section 7.2.11), there is evidence across the interviews that those who moved into the space of Twitter from similar online discussion forums - particularly politics.ie in this Irish context - more readily and quickly adapted to it. Participants in many cases viewed Twitter as a natural extension, or accessory to politics.ie, which still exists but which was reported to have become more niche as discourse moved toward Twitter; *"to a certain extent [Twitter] is politics.ie for loads of us now. A lot of the same people you knew were behind the names there... are the ones you talk to... and fight with in Twitter [...] and definitely the issues are the same. Depressingly the same sometimes." (OM2, 11/06/2015)*

The illusio of the space is also frequently referred to in the interviews, with many participants acknowledging a specific set of conventions, using terms like 'the way the game is played', which are very close to Bourdieu's intent for this concept. Broadly, a division exists between the illusio of Twitter, in the medium-specific rules it implied for how users can interact, and in the more specific illusio of political discourse on Twitter, where these are further specified and contextualised, often with reference to outside fields, patterns of historic engagement between actors, and knowledge of the relatedness of particular events. Participants described "coming to terms with the way things are done" (C1, 10/09/2014) for Twitter generally, and a period (sometimes of months and years) of uncertainty of what it meant to have discussion there; "Early on, it wasn't totally clear why I should bother with it $\lceil \ldots \rceil$ it took time to differentiate it from Facebook, but that's really apparent to everyone now" (OM1, 04/04/2015). Users who reported quicker settling into the 'rules' of the space, tended to be those who came to it later (e.g. A3) or via the recommendation of close friends (e.g. C1); in both cases these factors would have helped to pre-establish some of the operating norms of the medium, though these participants too acknowledged some period of adjustment culminating in their present understanding. In relation to the specific context of the Irish political discourse within Twitter, many responses indicated known 'rules of engagement' among participants (and especially among activists, citizens and bloggers). Several identified that the timing of certain

kinds of contributions, so as to maximise the chance of engagement with other actors, was critical.

"If you want something, especially a link of your own, to be seen... then you have to get to know the peak times for the crowd who are likely to be interested. [...] if you time it, and I'd often hold back things, you're better off" (OM1, 04/04/2015)

Similarly, specific social rules for the directing of context to particular actors, and particular categories of actors, were evidenced in some replies. In many cases these were analogised to external social norms.

"I think of it like... would I shout your name every time I was talking about you to someone else, hoping you'd come over. If we were in a room, I mean. Mentioning people in tweets to someone is like that, to me. I hate when I see people do it [...] it's usually a kind of desperate tactic from someone hoping the Taoiseach will notice their chat about water charges or whatever. Ridiculous." (LP2, 22/01/2015)

"It's almost a mark of immaturity that thing of dropping an issue tweet with a link and @-mentions of a load of journalists... like CC-ing the Irish Times on every blog post you make" (OM2, 11/06/2015)

Evidence of how users described their deliberations on the inclusion of hashtags (discussed in more detail in Section 7.4) also highlight mutually negotiated norms for such medium-specific behaviour; considerations of the 'rudeness' of adding a hashtag to one's contribution in a conversational thread of replies already underway, for example, demonstrates nuanced application of external social interaction norms in creating the illusio of this mediated space.

Discussions of poor social etiquette, as it was understood for the field, often mentioned the Twitter facility to block users. A blocking user suspends presentations of all communications from a user they block; they are unable to see any tweets from that person (including those explicitly addressed to them) and receive no notifications about that user. The user being blocked sees the blocker's profile page as if it were a 'locked' profile; the mode of Twitter use chosen by around 10% of accounts, where permission is required to view their content, similarly to the 'friend request' dynamics of Twitter. It is therefore possible, in the case of a public (i.e. not locked) user blocking another, that the owner of the blocked account can log out, see the public profile and tweets of the one who blocked them, and deduce that they have been blocked. Several participants (e.g. NP1, NP5, A3) reported blocking users and receiving complaints from them via other means of communication (e.g. email), or via

intermediaries in Twitter (petitioning on their behalf for them to be unblocked, or complaining about the event). Often these reactions to the use of blocking were linked to discourse around censorship; "One guy recently [...] sent an email about the 'cruelty' of being blocked on Twitter and how it amounted to state censorship - this was to my [government department] email address - I'm North Korea because I blocked someone on Twitter" (NP5, 10/02/2015)

Despite expressing some reservations about blocking users, often contextualised as part of their earlier less experienced use of Twitter, the majority of participants say that they do so, some frequently; "I didn't always block people... I don't like to do it, and I can take it when people are mean to me... I'm used to that, but some people... there's no alternative when it's this persistent abusive trolling. I've blocked dozens recently" (A3, 28/07/2015, specifically referring to tweets directed to @PantiBliss during and immediately after the marriage referendum campaign). The practice of users blocking one another with some regularity, and especially the pattern emerging from amongst participants that blocking is often in the context of specifically political discourse, may have implications for the research, where collection and analysis methods are unable to account for this..

The doxa of this field, meanwhile, is noticeable in part in the many unquestioned assumptions shared by the participants. A sentiment frequently expressed was that Twitter was a "natural setting" (OM1, 04/04/2015) for political discourse, positioned separately to other social media to suit this purpose; "Twitter attracts people who are interested in political discourse, whereas Facebook... people don't use it for that" (C1, 10/09/2014). Many non-politicians interviewed expressed an opinion that political discourse should be happening on Twitter, and several expected the participation in that discussion by actors identifiably powerful in the wider field of politics; "It doesn't surprise me that ambitious politicians like [NP5 is identified here] are involving themselves with it – they must see the opportunity, the gap. [...] Senior people - who you'd expect - have no real presence in it, and they're foolish, because it's part of how political change happens now" (A2, 14/04/2015, referring specifically to #marref discourse within Twitter).

There is also considerable evidence for *taken-for-granted* expectations that a user's presence within Twitter should be an unedited, unmediated and holistic representation of them conversely leading to reports of increased editing and mediation (e.g. NP7's stated previous reliance on staff more experienced on Twitter to assist in formulating his contributions). While a thorough analysis of the issues of self-presentation in social media, and the specific factors related to how political discourse participants present themselves on Twitter, is beyond the scope of the literature and range of interviews made in this work, it is still worthy of note here. In particular, the tension that exists among national politicians around 'authenticity' of Twitter presence was clear. Almost all exhibited some degree of focus on it, most using the term 'authentic' in their descriptions of how they currently engage with the medium, and how important it is for them to reflect aspects of their lives outside politics and to be the only person sending the tweets. Crucially, these sentiments were shared among all other groups, but where suspicion of inauthenticity was expressed, this was directed only at national politicians, with the ordinariness of other users contexts apparently being taken for granted. This likely correlates with the identification of issues of diminished trust for (particularly senior) politicians in contemporary Ireland, already outlined. Many participants specifically discussed named politicians who appeared to be 'real' or 'themselves' within the space, indicating that these expectations for behaviour and self-presentation are internalised among actors in this field, and that perhaps the politicians are right to worry. Responses from professional media participants offered some explanation of why well-known politicians who are perceived to hold back on casual or personally contextual content are labelled 'inauthentic'. PM3, in particular, used specific named politicians - senior government ministers and opposition figures - to illustrate what he saw as a now out-dated use of Twitter by political agents "in much the same way as press releases get sent out [...] one-way broadcasting in of, really, very tightly controlled snippets... the odd slogan-y use of a hashtag to seem relevant". A change in how politics, and especially politicians, are being mediated to the public has come about, according to PM3, and others in his category share this sentiment.

"The politics in this country, for as long as there has been a 'mainstream media', has been a mediated process... very much about control, by politicians, through their handlers... their advisors, their press people [...] Twitter in particular, has completely destroyed that. It is no longer a mediated process; politicians are now closer to the public [...] people are more authentic" (PM2, 14/01/2015)

In this context therefore, it may seem unsurprising that high-profile politicians, whose previous presentation to and interaction with the public were co-constructed by communications specialists, would exhibit greater anxiousness around getting Twitter engagement right. The case of NP5 (and NP5x) outlined earlier, certainly seems to suggest that the narrative of personal agency and personal autonomy is centrally linked to the 'authenticity' of an actor in Twitter discourse, perhaps even a condition for its meaningful use in interacting with the public; *"For those that use it authentically, and genuinely, it is a medium for the public to engage directly with them"* (PM2, 14/01/2015).

7.3.2 Capital in the field, and its accumulation and transfer

This thesis has already outlined how the concept of capital, which represents any accumulable property in a field that interests the actors there, may be differentiated into particular kinds of capital, of which social and symbolic capital are highly relevant to the field of Twitter (Papacharissi & Easton 2013). Social capital, constituting position and influence due to an ability to harness social connections, and symbolic capital, constituting prestige or recognition (Bourdieu 1986), are concepts which map well onto the struggles for power within the field of political discourse on Twitter. Strategies related to each form of capital are evident in Twitter: (1) Actors who are powerful in other fields by virtue of high social capital (e.g. Politicians), may leverage this to attain position in the field of Twitter; and (2) Actors who lack social capital may attempt to accumulate symbolic capital within Twitter in the hope of gaining position. Evidence for the first is readily discernable in the responses from national politicians, and (especially) professional politicians regarding their knowledge of the extent of their following, comparisons of this with other similar actors, and discussion of their power; "I'm aware that if I'm to wade in [on a discussion], people are going to notice... that's part of being a minister" (NP5, 10/02/2015). Overt discussion of the second strategy was more rare, though among activists, bloggers and citizens, several cues to this were available. Some were indirect, such as the discussion by C3 of a named acquaintance that "harnessed" Twitter as a medium in which to gain recognition as a commentator on politics, leveraging this to find employment in television as a pundit. Other participants more directly spoke of the effect of their symbolic capital, as reputation or prestige:

"I would regard myself, at this point, as an authority on some of the issues... so yes, I'd be aware of people @-mentioning me on some topics to seek my view [...] I'd also see screenshots of my tweets quite a bit on the likes of 'the journal'⁸⁴" (OM1, 04/04/2015)

A specific finding from this examination of strategies by participants unveiled another differentiation between local and national politicians; whereas national politicians came to the field of Twitter with power and social capital by virtue of their role in legislation and public discourse, and then acted within the field to develop the symbolic capital of "authentic" self-presentation in response to its norms, local politicians interacted differently.

⁸⁴ An online-only news website, begun as a story aggregator using the websites of major newspapers as sources, but now providing its own journalism and reportage, often with social media inclusions. http://www.thejournal.ie

Those interviewed described strategies to project a specifically professional role-related version of themselves, without the everyday context their national counterparts strived to communicate; "*Tm not there as [name]*, *I'm there to get a job done*" (LP2, 22/01/2015); "*There's the odd tweet about tv or something [...], but it's somewhat important to keep it professional*" (LP1, 21/01/2015). For local politicians, the concern was less about being 'authentically' an everyday person, than demonstrably a political agent. Some explanation therefore, may be offered by these separate imperatives for the observable differences in their contributions to the discourse that were measured in the quantitative pilot study.

7.4 Technological agency investigated

Following the conception of the network of interactions on Twitter, which constitute the Irish political discourse taking place there, as an *actor-network*, an element of the qualitative investigation in this study sought to examine the role of technological agency perceived among the participants. The particular mediation of content by the Twitter network generally, via the limitations and affordances inherent to it, were seen by many participants to create a net effect of suitability for the discussion of political topics specifically. Thus, at a macro level, the content transmission within the Twitter actor-network is mediated to achieve properties of communication appreciated by actors interested in highly current discourse. This was especially evident among participants who contrasted Twitter with other social networks, especially Facebook.

"Facebook is too cluttered up with a lot of... people's personal lives. You see good links on there too, but Twitter seems more focussed on real information, which is probably to do with the brevity of it. There's less faff." (A3, 28/07/2015)

At the micro level, however, themes of technological agency are also uncovered in the study. All users of Twitter necessarily interact within it via specific client applications - the Twitter website and increasingly mobile apps, as outlined in the introduction to this work - but about half of participants in this study additionally describe their use of secondary tools to create or sort content. Software applications such as Tweetdeck and Hootsuite were cited, the former adding a particularly overt technological agency in that it can schedule the sending of tweets at particular times, enabling contributions to the discourse to be untethered from the moment of their creation by the human user. Combined with the notions of there being 'peak times' for content contribution already outlined, it is easy to see how this intermediary scheduling agent could be seen as useful to particular actors wishing to maximise their exposure.

Collation of incoming content is also affected directly by the choice of device and application used to 'read' the incoming stream, especially where affordances are present to sort and search for specific kinds of content. Media category participants here (both professional journalists and bloggers) reported use of the most forms of content-sorting and streaming, with Tweetdeck being used in 3 cases, for its facility to maintain separate streams of content by keyword, hashtag, or other definable search operators; "In my job, there are people who are actively researching stories for me, and they're using all sorts of aggregators and social search tools obviously... but I have several running lists by particular topic or sets of people [...] that I couldn't go without" (PM3,

26/06/2015). Even for users accessing Twitter through the default applications of its website, segregation of content by lists of users is possible, representing another point of technological interjection in the process of interaction on Twitter. About one quarter of participants reported using such lists.

A noticeable theme over the course of the interviewing period was the emergence of negative commentary in response to a change made by Twitter to its timeline algorithms during the period of the study. In August 2014, Twitter began to trial an altered version of the timeline in which tweets were no longer sorted only by chronological recency, but also by 'relevance' to the viewing user. Like similar algorithmic determination of user preference for particular kinds of content in a social network, as in the Facebook timeline, for instance, this technological mediation can profoundly alter the relationship between the creation of content and its viewable presence for other in the medium - previously a highly immediate process on Twitter (Meyer 2014). Participants who were interviewed in early 2015, when this alteration began to be widely introduced for users, commented on them as a loss of an essential value to the Twitter discourse.

"I understand that Twitter has to make money, and that it makes a loss now... but I think they've gone wrong here... they've abandoned the whole point of what it was" (OM2, 11/06/2015)

"Timeline stickiness is not right for Twitter. It's not Twitter anymore when it becomes about saving the popular recent tweets so you can see them, instead of showing you what's actually coming in. [...] I'm not sure where this will lead" (PM2. 14/01/2015)

Like the scheduling of content by intermediary applications, the preservation of popular content in others' timelines for hours or days, represents a decoupling of content and author which is somewhat at odds from the original design of the medium, and has implications for researchers, as it does for users perceptions of conversational interaction on Twitter; *"I see the 'While you were away⁸⁵' tweets when I log in now, and I do sometimes think 'what's the point in replying?'* ... *if they're from hours ago, the kind of conversation is different*" (A2, 14/04/2015). This disruption of real-time conversation should be considered by researchers attempting to correlate user influence with a measurement of the conversational interactions stemming from their contributions, since this is no longer an 'organic' process where users see and reply to a

⁸⁵ This is the header Twitter places above that section of tweets first presented when a user logs in, following the changes to the timeline. These tweets are the ones algorithmically derived for the user, based on unrevealed metrics of popularity and relevance.

tweet; instead, an unknown algorithmic determination of popularity for either a user or their specific tweet may cause it to be retained, much more widely seen, and disproportionately responded to.

Finally, some participants touched on the existence of bots and aggregators within Twitter. The former are automated software-based robots designed to carry out specific tasks on Twitter, operating there in violation of Twitter's terms of use. Bots may be used to tweet or retweet particular content, usually in an effort to promote a link or hashtag. C1 mentioned that his knowledge of the existence of bots within the space of Twitter has caused him to occasionally doubt the humanity of some users he considers interacting with; "*I have had that experience…* [laughing] it's almost a sci-fi cliché… of having to stop and wonder if a tweet I'm seeing was really written by a person, or just a bot. [...] I've checked the profile to see if it looks real" (C1, 10/09/2014).

Similarly, this user and other 'citizen' users describe the role of *aggregators* in the discourse essentially websites and applications that are collecting and archiving aggregate content on a topic - based on keyword inclusion or hashtags, and similar to the collection methods of researchers. C1 cites their existence as a relevant factor in his use of certain hashtags; "*If I know there's an aggregator for some topic, say the US congressional elections, I'd make extra sure to include it so that [the tweet] becomes part of that*" (C1, 10/09/2015). This notion of users performing for known technological agents is an interesting insight, especially where it affects their use of hashtags - another key area of investigation in this work.

7.4 How hashtags are really being used

A specific research goal of this thesis is to begin a process of qualitative investigation to address the gaps in knowledge about how hashtags are used, and thus how the selection of data upon them is, or is not, a representative process. Among the issues highlighted in the literature is the identification of a body of conversation happening prior to or subsequent to a tagged tweet, which appears alone in a set collected using that hashtag, and without that full conversational context around it. Behaviours implying why this might occur were readily detectible in the responses of participants, who discussed specific social pressures and the etiquette of Twitter reply that explain this.

The suitability of a hashtag dataset to represent the totality of conversation on a topic has also been questioned in several studies (e.g. Bruns & Burgess 2011; D'heer & Verdegem 2014), and this investigation of how consistently and deliberatively they are used, provided some insights. Other revelations of the findings on hashtag usage, which are not predicted in the literature but were uncovered in the interviewing, are also presented in this section.

7.4.1 Conversation 'hanging off' the hashtag

This thesis has already outlined the problem of the existence of conversational components 'hanging off' the hashtag (Chapter 1; See also: Bruns & Burgess 2011; Bruns & Stieglitz 2012). These contributions on a topic, forming constituent parts of a conversation thread that emerges out of an initial hashtagged tweet, or is on-going before some subsequent reply is tagged and enters the conventional collection parameters, signal an intriguing unknown for researchers. Revelations from those participants who were asked about their own use of hashtags in this study emerge to offer some insight on the extent of these uncollected elements; (1) It is likely that the majority of conversation resultant from a hashtagged tweet, especially one stating a position or offering information, will go untagged, and (2) the act of hashtagging a tweet made in reply, especially one made 'further down' a thread of conversation on a topic, may be unwelcome and subject to social pressure not to occur.

These insights are grounded in findings from across several participant categories. The first is evidenced in the comments of five participants who variously stated that there are diminishing returns for the search value of including hashtags in a conversation, e.g. "when I'm replying - that's obviously a case where I wouldn't use one. If my reply, or even if it's after several people's replies [is to] a tweet with the hashtag, then people can already find it… why include it again?" (OM1,

04/04/2015). Indeed, Twitter's default 'detail' view for a single tweet presents the replies to it underneath it⁸⁶, such that any tagged contribution found by other users, who searched using the tag it contained, will necessarily show all contributions stemming from it as replies. Knowledge of this presentation of results prompted some to further comment that repeated use of a hashtag in an already tagged thread of discussion might look foolish to others, e.g. *"it's almost stupid to think of every tweet, back and forth, all with the same hashtags… it's ludicrous, who would do it*" (Participant A3, 28/07/2015). These sentiments were expressed exclusively among non-politician participants, though such users tended to be the longer established users in the medium, and the more frequent contributors, who demonstrate the most knowledge of its affordances.

The second insight related to 'hanging off' conversations - that tagging in mid-conversation is unlikely and may be frowned upon within the field - comes from a observances among blogger and citizen participants that there were an increasing number of their acquaintances whom they noticed purposely abstaining from using hashtags, and that to add one in conversation with them might be *"intrusive [...] certainly not what they want"* (Participant OM2, 11/06/2015). Participants, again those longer experienced with Twitter, saw the flagging of a hitherto-untagged conversation as potentially rude, given that the content it is in reply to may have deliberately not been tagged.

"I've noticed a growing number of people not using hashtags, even though they're talking about events [which have hashtags established], just to stay out of the bigger conversation record [...] people who've pulled out of using hashtags even though they're talking about the same topic you're talking about"

(C2, 21/09/2014)

7.4.2 Factors influencing inclusion and exclusion

The most commonly cited reason for deliberately excluding a hashtag, known to be established and relevant to a Twitter contribution, was space. The 140-character constraint on a tweet leads in some cases to a dilemma for users over whether to truncate content to fit it, or to forego its use, and this specific question was put to participants who did not volunteer the answer as part of another response. An interesting dichotomy emerged between those actors who hold position in other fields (national politicians, professional media, and wellknown activists) and those who wish to build position within Twitter. The former, believing that their standing already implies that others will find their tweets without the aid of tags,

⁸⁶ See Fig 4.1 in chapter 4 of this work, for an example

were almost universally unwilling to sacrifice the quality of their content in order to fit the tag; "I don't mean to willy-wave, but some concerns are a factor of how many followers you have [...] I have 19,100 followers, so I deal with some of this differently" (PM2, 14/01/2015). All participants whose accounts were Twitter-verified, indicating a superior position in terms of status in the field, reported a similar preference for their own content over any hashtags that would necessitate shortening it.

"They're not top of my agenda and I'm always forgetting to put them in, or I get to the 140 limit and it's either make this sound nicer, or stick a hashtag on it - and I will always opt to make it sound nicer" (A3, 28/07/2015)

In contrast, the citizen and 'other media' participants, displayed greater reverence for the 'rules' of the field they're trying to gain power in, and uniformly choose to include the hashtag whenever possible, modifying text to do so.

"I feel as though I have failed in some way - that it's incomplete if I have left the hashtag out [...] you're not contributing, if you leave out the hashtag" (C1, 10/09/2014)

This divergence between users with larger followings and greater social capital, and those who do not, may have profound consequences if this pattern holds more widely for Twitter, since it implies that contributions from more influential users, with already established profiles external to Twitter, are less likely to be hashtagged and would thus go uncollected by the current research methodologies.

7.4.2.1 Economic factors for curtailment

An unexpected additional insight of the interviews with 'professional media' and 'other media' participants was the uncovering of economic factors driving behaviour not to include hashtags. PM1 tangentially related that his employer, a major newspaper, had advised him to cut back on their use and speculated that this might have been because of his own overuse. This point of information was not pressed further by the researcher, but PM3 later also reported limiting his use of hashtags "for work reasons" and context for both cases may be offered in the detailed discussion of the commercial imperatives among contemporary media outlets, from PM2 who is responsible for 'digital strategy' at his newspaper.

He stated that he advises his colleagues at the *Irish Times* to be judicious and sparing in their inclusion of hashtags. They have to be somewhat used, he said, because "*[they're] a good way of*

doing things, it's the way the game is played for Twitter", but he saw important economic imperatives for ignoring them in some cases: "I have to be slightly commercial about this - if you're growing a large social media audience - and that's part of my job - if you include a hashtag in your tweet, you're actually sending people away from your tweet, and not to your website. There are traffic implications". This same economically motivated desire to ensure that the contributions of professional journalists on Twitter drive web traffic to their news website is likely to be a factor in the advice received by the other journalists interviewed. A technological solution to assist journalists in accurate and consistent hashtagging was also rejected for this reason.

"There was a tool we were being asked to use that would have put the relevant hashtags in all of our tweets, but I had to say no - what's the benefit for us? It just wasn't something that worked for us" (PM2, 14/01/2015)

Similarly in the case of non-professional media agents, effective enticement of users to visit their website must be a factor in their construction of Twitter contributions. OM1, who revealed that she derives income from advertising on her blog, which she promotes on Twitter, may feel similar imperatives not to add a hashtag, which would become a competing link in her tweet. While this insight requires substantial further research to contextualise, it represents an interesting finding from the qualitative data collected in this study, and is not anticipated by the literature.

7.5 Implications for research

In addition to providing a substantial amount of new qualitative data on the social situatedness of Irish political discourse on Twitter, which may guide the goals and strategies of further investigations, this chapter has also set out insights that may be used to directly inform interpretation of previous research and the design of future studies.

7.5.1 Hashtags

This thesis has already set out some of the potential shortcomings of selection on hashtags in the creation of a dataset for computational analysis of a Twitter discourse. In doing so it supports the theory that abounds, albeit in a small amount of the existing literature and studies, that a degree of bias must necessarily exist in the representativeness of the hashtag set to the total topical discourse, though the direction or shaping effect of that bias is unknown. Insights from this depth study, however, show that:

- The notion that certain tweets might result in a conversational thread related to them, in which none of the content is tagged, is well supported and explained by the development of social conventions in the field for the use of hashtags. Such content must exist, to a currently unknowable degree within the methods available, and is explained by these norms for Twitter.
- 2. The incidence of hashtags being applied on a tweet contributing to a conversation already underway may be low, and is likely to be subject to growing social pressure to fall lower, as this practice is viewed as intrusive and unacceptable by some. Thus conversations on a topic that starts untagged may continue thus and not be possible to capture using the current collection methods.
- 3. Actors in the discourse with higher social capital (greater following, status by virtue of standing in other fields, etc.) may be less likely to hashtag their contributions at all, exhibiting a preference for the honing of their content over the inclusion of a tag, in situations where the character constraint is a limiting factor. Taken with points 1 and 2 above, this is likely to result in some situations where contributions on a topic from an influential participant go untagged and are the impetus of threads of relevant conversation that similarly go untagged⁸⁷.

⁸⁷ A further small-scale investigative study, to examine this phenomenon, took place following the interview analysis. Using the archive of all tweets from the pilot study users, whose total contributions on Twitter were tracked throughout, the 3 most active national politician, activist and professional media users' tweets from the 3-day period 21-13 May 2015 (the most active period for #marref

- 4. Commercial imperatives of online news sources may incentivise the deliberative exclusion of hashtags by journalists and bloggers, where their tweets contain a link to a website dependent on web traffic for its economic viability. Since these links, especially from professional media agents, constitute information offerings likely to result in subsequent topical discussion, a similar situation to point 3 may arise.
- 5. Highly active 'citizen' users of Twitter for political discourse, who value the conventions of the space and view participation in it as important to their status are likely to make greater effort to tag their contributions in order to 'play the game' and to maximise the chance of their content being seen. A potential exists that this behaviour, taken in combination with points 3 and 4, may lead to a hashtag-based dataset being biased towards 'ordinary users' rather than those professionally associated with politics, and that subsequently derived analyses may overstate the role of citizens in topic discourse.

These insights are based, of course, on a small sample focussed on depth analysis. Investigation to better ascertain the macro effect of these findings, implied from micro level interrogation of participants, would be beneficial to the field.

These findings nonetheless begin to contextualise the extent of hashtag selection bias, and may usefully inform any review of findings resulting from analysis of datasets predicated on hashtags. The next chapter applies these insights in the interpretation of the case study data in this thesis, and findings derived from them.

7.5.2 Support for actor sub-categorisations

The insights set out in this chapter on the social factors influencing actions within Twitter also support the quantitative findings of the pilot study for this work, which suggested (based on observations of trace data) that national and local politicians, and professional and nonprofessional media agents, exhibited different patterns of interaction on Twitter. This quantitative insight led to the separate coding of these actors, and ultimately to the coding

discussion) were examined. Of these tweets that were not replies, 31% are declarative or information sharing statements directly pertaining to the marriage referendum, but not hashtagged. In more than half of cases, these relevant but untagged tweets led to further threads of @-reply conversation with other users. While this finding comes from a narrow scope of enquiry limited by the pilot study user set, and to the #marref topic, it does nonetheless indicate some support for the notion that influential actors may be less likely to tag their tweets, trusting that their large following and the context of their timing may make the action unnecessary. This insight warrants further study, using similar means to this test case to establish the degree of tagging in the tweet output of influential actors who have large followings on Twitter.

schema used throughout the case studies of this work. By selecting participants for interview from the various actor categories, qualitative investigation of social factors underlying these behaviour patterns was possible.

The findings set out in this chapter show that national and local politicians were subject to different motivations for their interaction, separate strategies for the use of Twitter, etc. Similarly, professional and other media agents were differentiated by their actions in response to their perceived status and following. Overall, the schema derived in the pilot study has resulted not only in a useful framework for the analysis of Twitter interactions by quantitative methods of examining the trace data they created, but also in a useful categorisation of participants for interviewing, which aided in identifying commonalities and anomalies amongst the themes that emerged from their responses.

7.5.3 False positive "contributors"

The specific case of NP4 in this study highlighted a significant failing of computational methods that focus only on trace data. By the regular tweeting and retweeting of content containing keywords and hashtags of relevance to this study, NP4 was included among the set of active participants in the discourse. Analysis following his interview, in which he made clear that his contributions were a necessity directed by his political party and constituted a one-way interaction with the medium, showed that indeed his presence was limited to the routine perfunctory actions mandated. While it may be of value to some studies to capture these behaviours, constituent as they are in the totality of how users are interacting with Twitter, research seeking to gather a set of highly active actors, as the pilot study in this thesis intended, may need to account for these kind of 'false positives' by parsing their results using measurements of reciprocity to eliminate users who are not engaged in two-way communication.

7.6 Limitations of this qualitative research

As in the case of many studies incorporating qualitative research, particularly where this research is secondary to a qualitative investigation in a mixed methods model, the sample included in the phase of the empirical research is restricted to a small size, and designed to represent only some degree of the large amount of individuals constituting the six actor categories in the Irish context of Twitter political discourse. Accordingly, the results cannot be generalised to a broader population, though the nuance and depth of analyses achieved in this portion of the investigation complements and informs the quantitative research here, and similar studies in the field of computational social investigation of Twitter. Insights from these interviews also serve to inform the broader thesis goal of developing methodological imperatives to account for factors uncovered in analysis of the social context of Twitter interactions, though here too this study can only offer a small and specific glimpse at the processes involved.

7.6.1 Opportunities for future research

It follows from the limits of this study, outlined above, that a number of opportunities are open for further research in this area. Aside from the obvious comparative research benefit inherent in a similar study situated in another (political and geographically distinct) locale, which might uncover commonalities between discourses and the effects of their settings, and hone in on those insights from this study that might be unique to Ireland, various further research questions are implied.

The first is a greater qualitative investigation of the behaviour of politicians, national and local, with regard to Twitter. The information on their concerns around authenticity of presentation, their perceptions of the medium and its value for their agendas and career, and their reconciliation of their positions within Twitter to those in other media and in society more widely, all offer intriguing threads that demand detailed further study.

Another opportunity arises in addressing the very small sample of 'citizen' actors interviewed in this study, through the creation of a specific qualitative investigation into those participants in Twitter discourse who are not professionally associated with the field of politics in any way, but who are actively involved in such discussions. The responses from this category in particular tended to be detailed, passionate, and motivated by a view of Twitter as potentially transformative for the political status quo. Further interrogation of the deep context of the participation of such actors would provide very valuable data to those social science researchers who are interested in the 'normalisation vs. mobilisation' debate within social networks where politics is discussed.

Similarly, a longitudinal study of actors - of all categories surveyed here, but especially of both citizens and of politicians - could provide invaluable insight on their changing relationship with the medium. This study, which offered snapshots of particular actors' attitudes to and impression of interaction within Twitter, has highlighted important temporally situated concerns present at different moments of development in both the issue discourses and the medium of Twitter itself. Since both of these dynamics are important to understand when attempting to more holistically place an analysis of political discourse on Twitter, qualitative enquiry with fixed participants, staged over a period of time, could uncover crucial insights on how perceptions of group dynamics change, how technological development of the medium itself impacts its usage, and how the habitus of individuals continues to adapt to the norms and conventions of this mediated space in which they interact.

7.7 Conclusions

This chapter presented the qualitative findings of this work, outlining the individual profiles of participants interviewed, with information and insights specific to each, based on their responses. An analysis followed of the themes emerging from these interviews, which may be summarised as a set of findings (specific to the Irish context) on the adoption of and adaptation to Twitter, the extent of technological agency in the discourse network, and the uncovering of specific attitudes towards affordances of the medium and their use. These findings may be used to inform the design of new investigation and to support and critically assess current studies in the Irish context, such as those of #le14 and #marref in this thesis - a process undertaken as part of the conclusions of this work, in the next chapter.

Interrogation of the factors underlying the inclusion or exclusion of hashtags, in particular, was shown in this chapter to reveal social contexts underpinning the issue of 'hanging off' conversations identified in some qualitative studies in the literature. This investigation of hashtags also resulted in findings unanticipated by the literature, such as a tendency among more influential actors to forego their use entirely, and a corresponding emphasis on their inclusion by 'citizen' participants. Similarly, commercial imperatives of large professional media organisations and of individual bloggers were also revealed to drive deliberate exclusion of hashtags. These findings were summarised as a set of five related insights.

These insights, as well as the support for the coding schemata derived in this chapter, were discussed in terms of their implications for research; both for the interpretation of previous analyses, and more widely as considerations for research design in the field generally. The limits of this qualitative branch of the research were outlined, and opportunities for further studies implied by it were discussed.

The next chapter concludes this thesis by revisiting the main strands of the theoretical approach and analysis of the field's literature, reviewing the findings of the empirical research, and drawing together these themes in the mixed methods *Explanatory Design* model to arrive at an informed final engagement with the research questions and interrogation of the primary hypothesis.

Chapter 8

Conclusions

8. Conclusions

This final chapter provides concluding remarks on the findings of this research, which has explored the emergent field of the study of political discourse on Twitter within computational social science. The thesis has introduced and described the dominant methodologies of the field and set out findings derived from the review of a large corpus of studies, which indicate the extent of use of particular methods of data collection and analysis in the cross-disciplinary literature on this topic. This document has discussed these methods and identified combinations of these, which were applied in two large-dataset case studies of Twitter discourse around major political events in Ireland in 2014 and 2015. The collection and analysis of the tweets thus collected, and the investigation of the interactions between the users who created these were set out and informed, in part, a qualitative stage of research that sought deep contextualising insight from some of the most active users involved. This chapter first summarises these findings from the quantitative and qualitative phases of investigation, and drawing on the *Explanatory Design* model of mixed methods research in use in this thesis, offers a set of conclusions on the effectiveness of the methodologies and the interpretation of specific results on these Irish case studies.

The chapter addresses the identified research themes of this work, implied by the primary hypothesis and set out in the introductory chapter. With reference to these, a reflection on the thesis is outlined, which draws conclusions across the chapters of this work. A summary of the methodological imperatives for future studies of political discourse on Twitter follows, informed by the approaches and findings of this thesis and intended as a derived set of guidelines that will shape the future work of the researcher, and may be useful to others in the field. The extent of this and other contributions to the field of computational Twitter discourse research is explored in the chapter, which closes with a discussion of the limitations of this thesis and the potential avenues for future research that it has opened.

8.1 Summary and Interpretation of Findings

This thesis draws on the work of Latour and others on Actor-Network Theory, conceptualising Twitter as an *actor-network* of interconnected nodes in the form of human users and points of technological agency, sometimes via software and algorithmic mediation and transformation of exchanged content, and sometimes more literally as 'bots' and aggregators taking on user roles of creating and compiling content. In choosing this approach, the researcher sought a balance between the various and diverging emphases of technological determinism and social shaping approaches to understanding how media technologies and society interact, retaining elements of both in a holistic consideration of the complex network of interactions facilitated by technology and contextualised by social, cultural, and economic factors. Some of these factors were unpicked in the work using the conceptual framework of Bourdieu's field theory, in which the concepts of *habitus, doxa, illusio* and struggles for and accumulation of *capital* were mapped onto the mediated social and technological space of Twitter. These concepts focused questioning of actors in the qualitative phase of the research and provided a framework to relate insights derived there to the quantitative findings on specific Twitter political discourses as structured networks.

Chapter 3 of this work set out a comprehensive review of the cross-disciplinary research literature on the study of political discourse on Twitter and identified commonalities across studies in that corpus of work. While theoretical framing was explicitly set out in only a minority of the studies, those that did so were most commonly established within a theoretical approach from political science of normalisation versus mobilisation - these studies sought to reconcile power and position in Twitter discourses with hegemonic societal structures outside the medium, as well as identifying if and how Twitter challenged these norms or enabled new modes of communicative participation, especially among citizens. Theories from communications and psychology were less prevalent, though appeared also, the latter often providing a theoretical framing for studies reliant on Sentiment Analysis of tweet content. This literature review identified Network Analysis as the dominant investigative approach in use in the field, disparate methods from which were observed to have been deployed for a range of investigations of meso-level aggregate user interactions. The focus of these studies tends to be on measures of network cohesion, density, and reciprocity, and many were specifically related to single electoral events. In the vast majority of studies reviewed, reporting by authors of the data collection strategy, and in many cases of any specifics of the size, selection variable, or period for the dataset was poor or absent. Those studies providing this information variously reported use of Twitter's API for data collection, in some cases specifically citing software used; here yourTwapperKeeper (discussed in detail in

chapter 4 of this work) was observed to be the most often used. In studies not stating specifics for the data collection method (which API and software were used), identification of one or more hashtags as the basis for the set analysed was stated in most cases. Overall, a broad picture emerges of a large set of studies from differing disciplines and geopolitical locales, which tend to include little explicit detail on their data selection and collection methods, few references to other studies outside their immediate research and topic areas, and therefore various *ad hoc* approaches in their analyses. While many studies offer intriguing insights based on their findings, this lack of overall methodological cohesion for the field makes cases difficult to compare and results in a promising but fragmented body of work, without a sense of best practice in research design. Thus the conclusions of chapter 3 of this thesis foreground an opportunity for original research to begin to systematically address some of the field's methodological issues.

Chapter 4 of this work outlined a mixed methods approach to the research agenda of the thesis, specifically outlining a phased set of studies - two quantitative case studies and a qualitative interviewing stage - set out following an explanatory design model. The aim of this research design was to investigate the methodological utility of specific methods chosen from the extant studies (and in some cases newly adapted or developed) by applying these to discussion on Twitter of major Irish political events in 2014 and 2015. These case studies had goals both of testing applied analytical methods for their efficacy in returning useful and comparative knowledge on the network of discourse, and of revealing novel insights on the discourse of Irish politics on Twitter. The thesis methodology outlined the design of a pilot study, used for the continued testing of methods, for revealing intelligence on actor interaction of value to a coding process, and for selection of participants for interview in the qualitative phase of the work. This interviewing stage, described in chapter 7, was identified in the methodology as a source of explanatory insight which may be applied to the quantitative findings of the work, and inform methodological considerations for research design. The next subsection of this chapter sets out this reconciliation of quantitative and qualitative data, describing how insights derived from interview may explain, contextualise, and bound the case study data, and imply necessary new approaches in similar future studies.

8.1.1 Results and conclusions of the case studies

In both case studies - the first of discussion around the Irish local elections in 2014 tagged #le14, and the second of discussion around the Irish marriage referendum in 2015 tagged #marref - metrics were computationally derived from the trace data associated with collected tweets in each set. From this metadata, and from computational descriptive analyses that identified features of inclusion in the tweets, patterns were revealed of interaction and

communicative directionality for the network of contributing users in each case, as well as for sub-groups identified by their professional association with politics.

8.1.1.1 The #le14 and #marref case studies

Analysis of the Twitter discourse of the 2014 local elections in Ireland, based on the set of #le14-tagged tweets collected, showed that the majority of users responsible for these contributions were 'citizens', professionally unaffiliated with the field of politics. In most cases, these users contributed only incidentally to the discussion captured, with a small but highly active minority responsible for most of the contribution and directed communication from their cohort. National and local politicians were observed to interact differently by all metrics derived from the data, and though the context of the event as a local election was acknowledged, their computationally-identified discourse dynamics previous to the case study had shown some of these differences also, and suggested a utility in separately analysing these two sets of professional politicians. In the case of this particular discourse, national politicians were observed to be less communicative with citizens, activists or bloggers (and other non-professional media), than with established professional media agents such as broadcast and print journalists. Descriptive metrics were outlined for the network of discourse between actor groups, and for the corpus of tweets, identifying a loosely-knit and decentralised network of users, in which identifiable patterns of inter-group directed communication mirrored that of a similar study carried out on the 2012 local elections in Belgium (D'heer & Verdegem 2014) and demonstrated likely commonality between similar Western European political settings.

Analysis of the collected Twitter discourse tagged #marref, pertaining to the 2015 referendum on the *Thirty-Fourth Amendment to the Constitution of Ireland*, showed a similar dominance of citizen users by number and percentage presence at all periods of the study analysed. In this case, the citizens featured much more strongly than in #le14 by measures of communicative directionality between groups, which showed them being engaged with by politicians and professional media in a manner not observed in the first case study. The #marref discourse was observed to be led by Activist users, whose centrality to the network was established early in the studied duration and was maintained throughout the period, affording this small group disproportionately large influence by several measures of directed discussion and descriptive analysis of their content.

In both cases, these studies demonstrated the efficacy of the combination of network and descriptive analytics deployed for providing computable metrics to describe the patterns of interaction among groups of users. Similarly, the use of periodisation in these analyses

provided a framework for the examination of discourse dynamics temporally related to the political event discussed. In the second case study in particular, methodological practices by the researcher uncovered new insights on the data-gathering process and on limitations for the dataset deriving from it. Both case studies provided an overall picture of a large, loosely connected discourse network in which a vast numerical majority of citizen users belies a very small but highly active cohort amongst them, who are responsible for most of their Twitter interactions. Thus, the very large sets of tweets in both studies were shown to mostly derive from a small community of only several hundred users, with many users only incidentally involved in that portion of the discourse collected based on the hashtag.

8.1.1.2 Insights and conclusions based on interview responses

The interviewing phase of this research drew on the responses of 22 participants (in 21 interviews from late 2014 to mid-2015), from which themes were compiled based on commonalities identified across the set. Insights revealed in this interviewing and thematic analysis were presented in chapter 7 and may be summarised under three broad categories: (1) knowledge revealed on the social context of participants' use of Twitter for political discussion, (2) knowledge revealed on technological agency in the network of interactions, and (3) knowledge revealed on the inclusion (and exclusion) of hashtags in the discourse.

While participants were specifically questioned on whether they could identify any uniquely Irish contextualising factors for the Twitter political discourse in which they are involved, most cited perceptions about Irish politics that are common in most contemporary Western nations. The issue of trust of politicians was identified both explicitly and implicitly in responses about the perceived 'authenticity' of political actors in the medium.

All participants described periods of adjustment to Twitter when they first adopted its use, many very early on in its development, and many having come to it from similar online political discourse in other media, chiefly the *politics.ie* forum. Most could readily identify other users with whom they regularly interact, corresponding to the quantitative findings of a smaller highly active core among the very large sets of users recorded overall. The participants described social conventions unique to Twitter, which shape conversational interaction within it and which affect the inclusion of hashtagged contributions to topical discourse, especially where these occur in threads of reply. Trends for more influential users with larger followings to be less inclined to include tags, especially when constrained by the character limit were identified from responses and shown to bear out in a small-scale followup study. Previously undocumented factors unveiled for the exclusion of hashtags from content where they might otherwise be relevant included commercial imperatives of media

239

wishing to drive web traffic, and social concerns over 'exposing' content from users deemed to have begun to intentionally forego hashtag usage.

8.1.2 Interpretation

Chapter 7 set out implications for the interpretation of existing research (including the case studies in this thesis), and for the design of future research. In keeping with the development of this identified knowledge gap in the literature, these implications focused substantially on how revealed insights on user deliberations around hashtag inclusion might augment researchers' understanding of the representativeness and completeness of hashtag-based datasets.

Trends revealed - showing how users understand their own place in the discourse, based on signalling factors such as their following, their topic knowledge, and their history of interaction with others - lent support to the category coding of actors derived from their professional association to the wider field of politics - from which users interviewed described struggles to transfer position and equate power, or in which some wished to gain new position. Interviewing of the small sample of 'local politicians' and 'other media' category participants revealed alternative usage practice and agendas at work for these, especially with regard to their relationship with professional media agents outside of Twitter. These insights offered qualitative support for the separate coding of these actor categories, which had been shown in the major case studies to produce useful segmentation of quantitative findings on user interaction by group.

Finally, the use of a qualitative phase of investigation was notable for its ability to uncover avenues of further enquiry and revelations about existing data that might not otherwise be attainable within the computational quantitative analyses more general to the field. The example of the discovery of one 'highly active' participant to have been a frequent contributor to hashtagged datasets via a mandated and perfunctory set of behaviours, rather than by any meaningful intent to participate in the discourse, served to highlight this unanticipated value of mixing methods. This insight, which would not have been otherwise uncovered pointed to the potential masking of other 'false positive' identifications of active contribution, and suggested methodological additions to account for these in future studies.
8.2 Key Research Themes and Primary Hypothesis Addressed

The introduction to this thesis outlined the primary hypothesis on which it is predicated, and set out six research themes that are derived from it, informing the particular sub-questions of this work across chapters. These themes are broad and interrelated and thus revisiting them in this concluding chapter may aid in drawing together specific conclusions from particular phases of the thesis as a whole.

8.2.1 Addressing the networked structure of Twitter political discourse

This work, and its primary hypothesis, posits an understanding of Twitter as a network of discourse(s), structurally arranged as a series of connections between users. The research has shown that this is not a single network, but variously understandable as constituent networks constructed of both the more permanent ties between users and those following them, and the more ephemeral ties of user mentions of one another in directed discourse. The latter, from which *mention networks* at macro-, meso- and micro-levels may be derived and analysed, offers a powerful conception of Twitter discourse, and a framework in which to investigate who is taking part in it and how the aggregate interactions between cohorts of users shape the discussion and vary over time.

8.2.2 Understanding the social context of Twitter political discourse

This thesis has presented contextualising knowledge on the social situatedness of the political discourses investigated, by both quantitative and qualitative means in its mixed methods approach. Computationally derived descriptive metrics for tweets showed how those collected often 'intersect' with other hashtags, many of which are established markers for wider media events and discourses, or associated with user behaviours, or other tagging of contextual value. This method of analysing co-occurring hashtags for a set has been shown in the research to effectively contextualise spikes in tweet frequency over the course of a tracked discourse, correlating these with other events such as televised debates, scandals, or even unrelated co-temporal events in society. The deeper, and substantially greater insight on the social context of participation in political discourse on Twitter in this work has derived from the inclusion of a qualitative stage of investigation in the research. Interviewing of highly active users has offered new knowledge on the adaptation of their habitus to Twitter, their

understanding of and assumptions about the medium, and their struggles within it to accumulate and reconcile social and symbolic capital. The Irish context of Twitter political discourse is uncovered in part in this work, offering new contributions to the field on deliberative behaviours by Irish users contributing to a hashtagged discourse, and on previously undocumented factors shaping their patterns of contribution.

8.2.3 Establishing the extant methodologies for Twitter discourse analysis

Chapter 3 of this work has set out an extensive review of those studies in the broad interdisciplinary literature that focus on the analysis of political discourse on Twitter. Commonalities and notable anomalies were identified across this corpus, providing an overview of the extant theoretical framings, methods, and data collection strategies. A literature in which poor reporting of methodological practice was observed, was set out by categorical divisions of approach to data selection, data sources, disciplinary framing, and analytics. Network Analysis was shown to be the dominant strategy for the investigation of Twitter discourse, and this thesis identified and combined useful metrics and tools from this largely computer science approach, deploying these in the analyses of the major case studies of this work. Where reported in other studies reviewed, data collection was shown to heavily rely on the Twitter API and on open-source software that queried it and parsed and stored results. Of this software, yourTwapperKeeper - developed in 2010 - was shown to be the dominant application in use.

8.2.4 Testing the effectiveness of collection and analytical methods

Following from refinement in a pilot study, the quantitative case studies for this work offered large-scale Twitter datasets from an Irish political context with which to test methods. The collection of these sets was itself an important methodological consideration, and this thesis provides substantial detail on the affordances of Twitter's API for available content and trace data, the structure of this data, and the limitations of current approaches to its collection. Original research throughout the thesis contributed newly developed tools, and comparative findings derived from concurrent data collection using various means. Combined methods from Network Analysis were shown to provide useful and comparable data with explanatory value for the shape of discourse, and manual coding of actors by their professional association with politics yielded a system of cohort-based analysis of interaction that provides valuable knowledge on dynamics of participation and influence. Periodisation of these analyses further strengthened their explanatory potential, by enabling segmented investigation of different stages of a discourse, with respect to a political event. The deployment of both data collection and analysis methods in the two case studies resulted in a set of new insights about political discourse on Twitter in the Irish context, and a derived set of research imperatives that may contribute to the field by informing the design of future studies - these are set out in the next section.

8.2.5 Informing methodology with qualitative investigation of actors

This important theme of the research was manifest in the inclusion of a qualitative stage of investigation, from which new insights derive, based on the interview responses of selected highly active participants. The social context of these participants was explored, revealing how they had come to use Twitter and some of the conventions they had developed for that use, much of which was shown to materially effect how their contributions are structured and thus have implications for the capture of the discourse using current methods. Analysis of the extent of technological agency was performed, situating the current modes of participation in a temporal set of affordances and technological mediations. New and relevant knowledge for the field was uncovered in these processes, specifically with regard to how hashtags are used by highly active contributors to a topical discourse, and thus what implications of bias this may cause for resultant datasets selected upon these tags.

8.2.6 Developing the understanding of how hashtags are used in political discourse

A key consideration of the research, and particularly of the development of the explanatory phase of qualitative investigation, was the degree of selection bias stemming from the use of hashtags as determiners of which data objects are collected in studies. The literature review has shown that the dominant mode of data collection for the field is the selection of tweets based on one or more topic-defining hashtags. This research has shown that those (few) studies that described this selection paradigm as potentially limiting for the representativeness of the resulting data, are correct in that assumption. Interviewing of highly active participants in the case study discourses has shown that using hashtagged data to represent Twitter discussion on a given political topic or event biases the sample toward 'citizen' actors, who are not professionally associated with the field of politics. Similarly, the potential for underrepresentation of professional media actors due to the economic imperatives for their non-inclusion of hashtags was a novel and important finding. Finally, the widely agreed

conventions apparent among interviewed actors that hashtags should not be added to or used in later conversational interaction has important implications for the degree to which captured data represents such discourse, biasing it instead toward more general declarative or information-sharing statements.

8.2.7 The Primary Hypothesis

Political discourse on Twitter can be understood as mediated interactions between and within dynamic networked publics, analysis of which may be derived from combined computational methods, and meaningfully informed by contextual qualitative knowledge of the actors involved.

The researcher contends that while the methods of analysis of political discourse on Twitter presented here are a small collection of those possible, and provide specifically for computationally attainable insights about the structure of the network of interactions revealed in tweet datasets, these have nonetheless been shown to result in informative findings and to increase knowledge of the discourse, proving this hypothesis within their limitations. The social contextual knowledge accumulated by the qualitative stage of this research, though situated in a specifically Irish context, and limited by the narrow extent of participants selected, has also been demonstrated to inform understanding of the discourse network, its dynamics, and the validity and limitations of data collection and subsequent analysis deriving from it.

8.3 Methodological Imperatives for the Study of Political Discourse on Twitter

A goal of this research was to identify salient methods for collection and analysis of Twitter trace data, to test these in case studies, to examine derived findings informed by qualitative investigation of the social factors shaping participation, and thus to arrive at a set of methodological imperatives for future research. These conclusions on effective practice for computational social science researchers wishing to collate and examine political discourse on Twitter are made within the many limitations identified for this work (discussed in preceding chapters and summarised here, see Section 8.5), but nonetheless constitute an important contribution of this thesis to the field. While chapters 5 and 6 of this thesis offer greater detail on the specific individual methods combined in their analyses of #le14 and #marref, this section sets out broader imperatives which are derived from this research and which may helpfully inform future studies.

8.3.1 Reporting

This research has identified a widespread lack of methodological reporting in studies in the literature, which results in often-subjective interpretation by readers of results derived without clearly set-out parameters, and a lack of comparative value of metrics presented without complementary contextualising information on the dataset and its derivation. Thus an obvious imperative for future research in this field is the inclusion of several specific descriptions of factors relevant to the interpretation of results, namely:

- 1. Metrics for the dataset, including the number of tweets, number of users contributing these, period of the study, and selection criteria (including specific hashtag, keywords or tracked users and why/how these were chosen).
- 2. Information on the data collection strategy, including the software used (giving version, where applicable), detail of any gaps in collection, and (in the case of bespoke approaches) the specific API used.
- 3. Information on modifications to the data, including any strategies for sub-sampling, criteria for discarding of tweets, and detail of any approaches to data storage or processing which affect the data objects (e.g. discarding or combining fields when inserting into a database, etc.)

Although this reporting strategy may seem simple and self-evident, the majority of this information is not provided for most extant studies, and therefore limits the extent to which these may be used in meaningful comparative settings.

8.3.2 Digital Ethnographic Practices

This thesis has demonstrated the research utility of contextual knowledge of the conventions and deliberative practices among users around the selection of a dominant hashtag for a particular discourse topic. It is a recommendation of this work, therefore, that future studies incorporate investigative practice focused on clearly understanding these processes, in order to authoritatively describe and contextualise hashtags chosen as selection criteria and account for any potential biases toward particular users, political stances, etc.

Such attention to the community of a studied discourse is also likely to meaningfully inform interview participant selection, either by directly suggesting users of note to interview, or as a source of collation of keywords, hashtags, and tracked users to implement a pilot study, as in this work, from which users may be computationally selected for given criteria.

8.3.4 Security of data

In addition to the ethical imperatives for the safe-guarding of data intrinsic to research in the social sciences, computational methods such as were deployed in this work are subject to non-obvious considerations for the integrity and security of data. The incident during which the webserver used in data collection between the two major case studies was hacked highlights concerns around best practice for the security of software and hardware that is necessarily exposed to the internet. While open-source collection software does not actively prompt this, it may be advisable for researchers to regularly archive and remove data from an active survey, both backing-up on-going datasets and preventing their exposure (which constitutes a breach in Twitters *Terms of Use* for the data.

8.3.5 Actor coding

Where coding of actors takes place as part of analytical methods, this thesis has shown that pilot studies and applied computational descriptive analyses can help to posit and test viable and effective coding categories. Following the demonstration in this work of the consistent differences in interaction patterns between local and national politicians, and between professional and other media, it is a recommendation of this research that these categories be utilised, tested, and critiqued in other work. Similar derivation of new actor coding strategies is possible, based on ethnographic observations in a discourse cohort, leading to postulated differences between identifiable actor attributes, confirmed by computational analyses of aggregate interactions or descriptive tweet metrics.

Based on the processes of coding for political stance in the #marref case study, this thesis has identified an imperative for such coding to take place during, or immediately following the tracked political event, as overt evidence of users' stances with regard to a topic is most easily detected then, and subject to sudden change. The software developed for the coding of actors in this work (yUC) facilitates concurrent coding of users during the course of continued discourse tracking and data collection, which may substantially improve the detection of temporally specific indicators on which to code. This software will likely become part of the DMI TCAT open source project and continue to be developed with this concurrent deployment in mind. Other strategies for coding may also be adjusted to facilitate retrieval of contextualising information specific to the period of a study.

8.3.6 Specific Analytics

The case studies in this work have demonstrated the insights obtainable from the deployment of particular methods of network and descriptive analysis. While this set may be expanded, as new explanatory efficacies are demonstrated, these analytics represent a coherent set chosen from disparate approaches in evidence in the literature and systematically deployed to provide an overview of the network structure of a discourse, identification of aggregate communicative patterns within it, measurements of the directionality of addressed content, and computable metrics for describing the types of tweets sent and indicating how these may relate to a more generalised political and communicative context outside of Twitter (e.g. URL inclusion, hashtag intersection analysis).

Of particular note here is the use of the *E-I index* (see Chapters 5, 6) as a simple, effective and comparable metric of the tendency of a group to be inwardly or outwardly communicative. This value may be derived for user sets as a whole, for actor categories, or for specific defined groups of any size, thus making it a useful method for macro- and meso-level analyses of communication.

8.3.7 Periodic Analyses

Finally this thesis has identified an explanatory utility in segmenting the discourse around a particular political event into stages related to the modes of discursive participation and mapping well onto observed trends in the frequency distribution of tweets. In these studies, a 4-part periodisation was used to examine discourse prior, proximally prior, during, and post a particular electoral event, and this strategy was shown to reveal changes in the patterns of interaction between actors which had strong explanatory value, and which would be obfuscated in an aggregate analysis of the period generally. Further study is required to determine whether these periods should be set to specific durations, though these case studies have demonstrated that relative values for the period, independent of the absolute value of duration and tweets, are useful metrics in understanding the discourse dynamics. It may be that case that regardless of some period lengths, divisions based on relative position with regard to a political event offer insight on how influence, presence and interaction changes in the on-going discussion. Naturally this periodic analysis can only be applied in cases where a studied discourse had a definable event around which it can be tracked. For studies predicated on hashtags about general social change, political parties, lobbying, etc., segmentation of analysis may prove more difficult, though it is likely that this strategy would provide valuable research insight when applied with regard to a particular event in the ongoing survey, showing how it had shaped and changed participation and interaction.

8.4 Contributions to the Field

This research contributes to knowledge in the field of analysis of Twitter political discourse through its systematic engagement with the themes set out in this chapter and the derivation of imperatives for future studies based on the specific analyses, findings and insights of the original research constituting this work. Following the prioritised research questions, outlined in the Introduction to this work, this thesis had demonstrated a viable and useful methodological approach for the capture and analysis of Twitter data, which yields metrics by computational means that offer useful comparison across studies. The methodological toolset derived in this work combines disparate extant methods from existing studies and demonstrates their combined utility for the analysis of political discourse on Twitter, using large-scale case studies. The introductory chapter identified this field as a new methodological frontier in computational social science, where the opportunities and challenges of working effectively with 'big data' inherent to the social web are as yet without methodological consistency and best practices. Chapter 3 of this thesis has set out a thorough review of literature related to the field extant across several disciplines, identifying commonalities, anomalies and shortfalls in theoretical framing, methods, and reporting. While this thesis cannot hope to institute or fix best practices in a field so interdisciplinary and emergent, substantial contributions have been made here to several areas identified in the literature as necessitating study, and insights have been offered in new study contexts, and deriving from new research designs.

The methodologies explored for network analysis in particular have prompted the researcher to develop new categories for actor coding, derived from observation of user interactions, which were then tested, verified and demonstrated to provide useful explanatory value across the pilot and case studies. The coding of actors as a process has also been addressed by the researcher through the derivation of imperatives for such practice and its adaptation in future studies, and the development of a software solution to aid efficiency of the process. Similarly, the area of data collection is an important focus of this thesis and one where novel software development by the user, and testing of various open-source collection strategies has yielded new insight on factors limiting completeness of datasets derived using this dominant method in the field. A more complete understanding of the totality of issues around the representativeness and completeness of hashtag-based datasets in the study of topical discourses on Twitter will require considerable further study, some avenues of which are identified in the final section of this chapter, but this area is one where this research has sought to directly address documented issues and uncover unknowns.

249

The mixed methods approach to this research has resulted in a qualitative phase of study from which interview responses have revealed important insights related to the social context of contributions to Twitter. This interview-based research is also novel to the field and has been shown here to have value in confirming and contextualising a priori assumptions about specific user practices with regard to their interactions on Twitter, uncovering unforeseen social factors related to the deliberative processes of users creating tweets, and providing a source from which to postulate about new factors shaping the discourse. In the particular area of hashtag-use by contributors to Twitter political discourse in Ireland (which has obvious and significant implications for the selection, collection and analysis of datasets) the interviewing phase of this research was shown to be a crucial new avenue of investigation. Findings set out in Chapter 7 show that social factors are likely to directly shape the extent of conversational discourse on a topic that goes untagged, and that the likelihood of users choosing to forego hashtags may be related to their status and following within the Twitter. Finally, this component of the research has unveiled new economic factors that affect the inclusion of hashtags in content originating from media agents in Ireland, warranting further study on whether this trend is more widespread.

Finally, this work has provided complementary sets of findings on the Twitter discourse around two recent major Irish political events. These findings have comparative value for the field, being closely documented in their methodological procedures to allow for meaningful interpretation of the evaluations and conclusions presented. In both cases, the scale of dataset used for these analyses is novel to the literature and represents new knowledge on Twitter political discourse participation in Ireland, and the patterns of interaction and communicative strategies constituting it. The second of these studies, on the discussion on Twitter of #marref, applies an identified methodological toolset to a single national binary issue, providing knowledge both of its specific discourse and of how such methods may be adapted to this setting and to the very large dataset associated with it. The methods chosen and deployed in these case studies have demonstrated efficacy in identifying commonalities and anomalies in actor behaviour over periods of the discourse, offering new insights into the dynamics of Twitter-mediated discussion of national political events.

8.5 Limitations of the Research

A primary limitation of this research is that the testing of methodologies was evaluated on only two case studies of Twitter political discourse. Furthermore, data collected for these two cases was selected upon only one hashtag in each case; #le14 and #marref. While consistent with studies set out in the literature, the specificity of the discourse captured by the data gathering strategies in this work results in subsequent analyses and findings relating only to those tweets which formed a part of a demonstrably wider discussion on Twitter of both the 2014 Irish local elections and 2015 Irish marriage referendum; insights on the patterns of interaction and communicative strategies of users grouped by their professional relationship to the field of politics, derive from and apply to only those users who included a hashtag in their Twitter contributions. In both case studies chapters, further caveats have also been outlined for each case, but the narrowness of focus of this research on two hashtag sets is a crucial constraint and implies the need for further studies to compare the insights observed for the specific cases and the efficacy of the methods deployed to deliver these.

The use of hashtag-selected data has been identified from the outset in this work as applying limitations of completeness, conversational context, and unknown biases extending from poorly understood user conventions and deliberative processes around hashtag inclusion. While this work has offered some insights on the potential direction and extent of these biases, and identified factors previously unmentioned in the extant literature, the process of understanding how a hashtagged dataset relates to and is representative of a wider topical discourse is one only beginning. Caveats of representativeness for hashtagged data aside, the methodological testing processes of this thesis have also identified shortfalls in data collection by means of open-source software retrieving content from Twitter's API. The insight that those sections of topical discussion in which the most sudden short spikes in use of the selected-upon hashtag occur are subject to some extent of incompleteness must be acknowledged as another limiting factor for both case studies in this work, and for previous studies in the literature. However, the demonstrably greater efficiency of data collection observed in the use of software developed by the researcher and of tools developed by the Digital Methods Initiative does show that the open source approach upon which much of the data collection in the field is predicated, can lead to iterative improvements and may continue to increase the precision of dataset collation. Collaboration by the researcher with others working on the open source projects around Twitter analysis, and the sharing of devised tools for its processes, is hoped to benefit the field and enable improved practices among other scholars.

The methods of analysis deployed in the two Irish political discourse case studies are predicated upon computational analysis of Twitter trace data, and apart from coding by the researcher of users by category there is no component of content analysis. Aside then from the removal of those tweets from users who were identified during coding processes as irrelevant, no aspects of the methodologies used in this work objectively or subjectively assess the relevance of tweet content to the topic of discourse implied by the hashtag inclusion. While any attempt as such assessment would be greatly onerous, and constitutes an approach not associated with the computational analysis in the field, it must be acknowledged that without such oversight the datasets investigated are likely to contain some degree of spurious data. Analysis of user interaction based on these datasets must therefore be subject to some unknown degree of error based on the inclusion in them of irrelevant content, malicious or intentionally unrelated contributions from 'troll' users and spam bots, etc. Such analyses rely on the size of the dataset to capture aggregate patterns of communication and interaction to compensate for this shortcoming.

Coding processes used in this research are also an identifiable limitation. Firstly, since the manual coding of users depends both on the rigour of decisions by the researcher and the specificity of the code categories, this aspect of the research is always open to revision and improvement. In particular, the nature of the 'citizen' category for users is here, as in other studies where it is used, a default into which users are filed when they do not fit other categorisations. Hence, the development of criteria to better define what a 'citizen' is, rather than defining these by what they are not, may help to give greater meaning to this and similar schema. Secondly, the application of user coding, while shown to be an effective variable for group-based investigations of communicative behaviour in the discourse, is subject to the limits of how many users may be coded by a researcher or team. The use of the very large datasets possible to derive from Twitter by computational means leads to correspondingly enormous sets of contributing users to be coded. Strategies are required to sub-select a sample of users, as was the case in the second case study of this work, but this refocusing of analytic methods on a partial set necessarily affects the representative applicability of findings and diminishes the value of derived metrics for comparative studies where sampling differs. Work is required in the field to establish a best practice for how this sub-sample should be selected and obtained, in order that a consistent process emerge for the coding of subsets of large user cohorts.

As with any study where findings are derived from interview responses, limitations of the associated processes must be acknowledged. In this case, the selection of participants for interview is an obvious limitation, both because the sample was of a small size, varied in number across the actor categories from which actors were chosen, and was subject to factors

252

of feasibility in many cases. Participants initially selected often declined, particularly in the case of national politicians and professional media, and in several cases scheduled interviews were cancelled or repeatedly postponed, sometimes resulting in compromise interviewing via Skype, which diminished rapport and access to potential insights. At various stages, participants withdrew or increased their level of anonymity in the study, necessitating the removal of data, or the restructuring of its reporting to the detriment of some knowledge and context. The timing of the interviews, which took place over an extended period in order to suit imperatives of the overall research management, as well as to suit the availability of some participants, was often something of a constraint on the potential intelligence to be gathered, since participants declined to or were unable to comment on concurrent issues that they might have spoken more freely about at a later time. Similarly, distance from some issues, especially specific political events that were discussed, led to poor recollection in other cases. Finally, a crucial limitation of the interviewing was in its focus on highly active contributors to the measured topics only - thus no insights are available on the contexts, habitus, understanding of convention, deliberative practice, etc., of users who are more incidentally involved in this discourse.

Lastly, it must be acknowledged that the selection, collection, and analysis of Twitter trace data in this thesis is directly dependent on several levels of current technological states, which are subject to often significant change in the short to medium term. While the approach taken by the researcher has sought to exclude identifiably 'volatile' affordances of Twitter as mechanisms of selection or analysis (e.g. measuring tweet influence via the 'likes/favorites' count, see also Appendix A.4.2), the methodologies defined and deployed in this work are necessarily grounded in a current set of limits, which include the particulars of Twitter's API, the current efficiencies of webserver hardware and software used in collection, the technological mediation of Twitter affordances and presentation of content to users, etc. Thus, methodological developments for this field must be temporal, though should be guided by theoretical framing and identification of meaningful characteristics of political communication that are well established and transcend specific technical settings.

8.6 **Possibilities for Further Research**

As outlined in the chapters describing each of the two major case studies (5 & 6), the novelty to the field of the large datasets and combinations of data collection and analysis methods make obvious a benefit of further similar studies. Such studies situated in other locales and based upon discussion of other political topics, would facilitate comparison and critique of these case study findings and of the efficacy of their constituent methods. There are several specific opportunities for valuable contribution to the research in this field, opened up by these cases:

- 1. Similar strategies of data selection, collection and analysis applied to (a) other Irish political issues, and (b) other internationally situated political issues, would begin to provide comparative data on discourse patterns across events and locales, and would be likely to uncover shortfalls in the methods deployed here, leading to the sort of iterative improvement that this thesis has made to extant methods. This is especially relevant in the case of actor coding practice, and the findings and conclusions to be derived from network analytics dependent on such researcher-contributed coding.
- 2. Further studies on binary political issues, in particular, may serve to advance the coding practice for the actors and would help to contextualise the findings of the second case study in this work, which currently lacks a comparison in the literature, and which displays significantly different influence and interaction patterns for citizen users than does either the first case study, or existing studies similar to it.
- Studies that vary the scale of dataset would be useful for critiquing the efficacy of these methods at smaller sample sizes, and for contributing to the necessary adaptation of some to larger sizes.
- 4. Longitudinal analysis of Twitter discourse around similar political events (e.g. subsequent local or general elections) would be particularly effective at uncovering societal dynamics for participation in such discussions, and the temporal relationship of patterns revealed in specific studies to known data for Twitter and internet usage at the time. Similarly, such research would help to chart the development of patterns particular to current affordances, conventions or technological mediations within Twitter. Analysis of the extent of continued participation of particular users, especially by category, would be a feasible and likely insightful potential investigation in such repeated studies.

This thesis has particularly focussed on hashtag-based selection of data, and while alternative strategies may become feasible in the short or medium term, depending on new API

affordances etc., continued research on the limitations of hashtag sets is highly relevant both to the design of future research and the meaningful interpretation of extant studies. Therefore research focussed on particular issues related to data collection would strongly benefit the field: (1) further studies to establish discrepancies between data collected via different software applications, (2) new studies to assess the effect of hardware and bandwidth limitations on data collection, and (3) studies combining data from several related hashtags to develop intelligence on how particular hashtags come to be dominant markers for a topic.

Since this work has already shown the frequency of usage spikes for a hashtag to be a limiting factor for the completeness of a collection, further work to develop computational analyses and comparative metrics for how 'spiky' a particular dataset is would be a useful reporting consideration for researchers who wish to aid others in understanding the potential limits of their data.

This thesis has included qualitative study and demonstrated that insights obtained from it are capable of informing the interpretation of data and providing important considerations for improving on-going research. Thus, an obvious avenue for further benefit to this field lies in the continued use of qualitative data-gathering to provide explanatory context and to uncover otherwise unknowable modifying factors on how participation in Twitter discourse is socially, culturally, politically, and economically shaped. Particular possibilities of note in this regard include:

- Expanded participant selection to provide insights from both a wider variety of highly active users, as well as those who are less frequently involved though nonetheless contribute to the discourse.
- 2. Interviewing of Twitter users on the particular topics of *trust and reputation* among users in the space who discuss and opine on politics, and *self-presentation* in the specific context of Twitter, both of which were shown to influence deliberation over contributions and conversational interaction in this research, but which necessitated too great a degree of tangential research to properly investigate.
- 3. Longitudinal analyses of particular Twitter users, to establish the temporal situatedness of their patterns of communicative behaviour, the processes of adaptation of their habitus to the medium, the effect on them of technological change and dynamics of available affordances, and their individual viewpoints on participation and its context over time.

Finally, during the course of this work, other opportunities to contribute to the field of computational Twitter discourse analysis in the social science were considered by the

researcher and had to be excluded because of the limitations of a timely completion of this thesis. Nonetheless, several are viable strategies for obtaining further insight, particularly in two areas: (1) extension beyond the hashtag to examine conversational threads of discourse 'hanging off' it, and (2) alternative social scientific methods for collecting qualitative insights that might inform methodologies and findings.

The first of these requires considerable algorithmic development of new methods of selection and analysis, but is possible within the current constraints of the Twitter API, albeit using a highly 'lossy' strategy of substantial data collection and computational post-analysis. It is possible at present to use Twitter trace data to identify whether a given tweet was made in reply to another (identified by a unique id). Thus, any strategy which intelligently collects tweets subsequent to a given potential 'trigger tweet' could assess whether each new tweet was made in reply to it, comparing the *in_reply_to_id* field to known ids already collected. Such an approach would require this process of checking against existing tweets to happen outside of the cycle of API and database transactions that constitute the current datagathering paradigm, since it is time-intensive. However, such a strategy would be able to reveal currently unattained knowledge on how threads of conversation are arranged and flow from particular tweets, at what rates, and subject to what dynamics of inclusion, etc. Since the goal of such a process is to extend beyond hashtag selection, and since this thesis has shown strong implication based on interview responses that conversations made in reply to a given hashtag are highly unlikely to be tagged, a strategy is required for the selection and collection of tweets based on other likely indicators of conversational contribution, which would form a set to be checked for replies made to known starting points. Developing such a strategy is a second programmatically onerous task, though should be possible via the collection of a large set of tweets from specific users demonstrably likely to converse with others identified within the set. Thus, an over-collecting of tweets and a subsequent parsing of these for the components of conversational threads could yield potentially very useful new data on the conversational component of Twitter discourses that currently goes uncollected and therefore unanalysed.

Finally, alternative methodologies for the collection of qualitative data, which were not chosen for this research, but which are viable and feasible possibilities to contribute to knowledge of factors shaping participation are: (1) surveys and (2) media diaries. Both methods have been effectively deployed in studies of media use, though the latter does not appear in the reviewed corpus of research literature on the topic of Twitter political discourse analysis. While surveys have been used, these have tended to be deployed in the collection of coding data (e.g. political party affiliation, indication of having voted or not in a recent election, etc.) and not for social contextual data that may help researchers to understand

256

issues like why hashtags are or are not used. Both methods could be effective in prompting users to reflect on their own Twitter use, and to provide related insights, from a potentially much wider cohort than might be possible with in-depth interviewing. Thus, such strategies may help to address the problems of narrowness of participant selection identified in the discussion of the limitations of interview design in this work, by including insights from a greater number of users at a wider level of activity within political discourse on Twitter.

Bibliography

Bibliography

- Abroms, L. C., & Craig Lefebvre, R. (2009). 'Obama's Wired Campaign: Lessons for Public Health Communication', *Journal of Health Communication*, 14/5: 415–23. DOI: 10.1080/10810730903033000
- Adamic, L. A., & Glance, N. (2005). 'The political blogosphere and the 2004 US election: divided they blog'. *Proceedings of the 3rd international workshop on Link discovery*, pp. 36–43. ACM.
- Adrianus. (2013). 'Twopblog: The last 100 million Twitter accounts'. Retrieved October 29, 2015, from http://www.twopblog.com/2012/05/last-100-million-twitter-accounts.html
- Ampofo, L., Anstead, N., & O'Loughlin, B. (2011). 'TRUST, CONFIDENCE, AND CREDIBILITY: Citizen responses on Twitter to opinion polls during the 2010 UK General Election', *Information, Communication & Society*, 14/6: 850–71. DOI: 10.1080/1369118X.2011.587882
- Anstead, N., & O'Loughlin, B. (2012). 'Semantic polling: the ethics of online public opinion',.
- Ausserhofer, J., & Maireder, A. (2013). 'NATIONAL POLITICS ON TWITTER: Structures and topics of a networked public sphere', *Information, Communication & Society*, 16/3: 291–314. DOI: 10.1080/1369118X.2012.756050
- Barrett, G. (2011). A Road Less Travelled: Reflections On The Supreme Court Rulings In Crotty, Coughlan and McKenna (No.2). Dublin, Ireland: The Institute of International and European Affairs. Retrieved November 11, 2015, from <http://www.iiea.com/publications/a-road-less-travelled-reflections-on-thesupreme-court-rulings-in-crotty-coughlan-and-mckenna-no2>
- Beevolve.com. (2015). 'An Exhaustive Study of Twitter Users Across the World'. Social Media Analytics | Beevolve. Retrieved October 29, 2015, from <http://www.beevolve.com/twitter-statistics/>
- Bekafigo, M. A., & McBride, A. (2013). 'Who Tweets About Politics?: Political Participation of Twitter Users During the 2011Gubernatorial Elections', *Social Science Computer Review*, 31/5: 625–43. DOI: 10.1177/0894439313490405
- Benson, R. D., & Neveu, E. (2005). *Bourdieu and the journalistic field*. Cambridge ; Malden, MA: Polity.
- Bermingham, A., & Smeaton, A. F. (2011). 'On using twitter to monitor political sentiment and predict election results',.
- Bijker, W. E. (1995). Of bicycles, bakelites, and bulbs: toward a theory of sociotechnical change. Inside technology. Cambridge, Mass: MIT Press.
- Bijker, W. E. (2010). 'How is technology made?--That is the question!', *Cambridge Journal of Economics*, 34/1: 63–76. DOI: 10.1093/cje/bep068
- Bijker, W. E., Hughes, T. P., & Pinch, T. J. (1987). The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology. MIT Press.

- Bode, L., Hanna, A., Sayre, B., Yang, J., & Shah, D. V. (2011). 'Mapping the political twitterverse: Finding connections between political elites',.
- Borra, E., & Rieder, B. (2014). 'Programmed method: developing a toolset for capturing and analyzing tweets', (D. Axel Bruns and Dr Katrin Weller, Ed.)Aslib Journal of Information Management, 66/3: 262–78. DOI: 10.1108/AJIM-09-2013-0094
- Bourdieu, P. (1977). *Outline of a theory of practice*. Cambridge studies in social anthropology; 16. Cambridge; New York: Cambridge University Press.
- ——. (1986). 'The Forms of Capital'. Richardson J. (ed.) Handbook of Theory and Research for the Sociology of Education, pp. 241–58. Greenwood: New York, NY.
- -----. (1990). The logic of practice. Stanford, Calif: Stanford University Press.
- -----. (1993). *The field of cultural production: essays on art and literature*. New York: Columbia University Press.
- (2003). 'Participant Objectivation*', Journal of the Royal Anthropological Institute, 9/2: 281– 94. DOI: 10.1111/1467-9655.00150
- Bourdieu, P., & Thompson, J. B. (1991). *Language and symbolic power*. Cambridge, Mass.: Harvard University Press.
- Boutet, A., Hyoungshick Kim, & Yoneki, E. (2012). 'What's in Twitter: I Know What Parties are Popular and Who You are Supporting Now!'., pp. 132–9. IEEE. DOI: 10.1109/ASONAM.2012.32
- boyd, danah. (2010). 'Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications'. Papacharissi Z. (ed.) A Networked Self: Identity, Community, and Culture on Social Network Sites, pp. 39–58. Routledge: Chicago.
- ——. (2015). 'Research on Twitter and Microblogging'. http://www.danah.org/. Bibliography, . Retrieved November 3, 2015, from <http://www.danah.org/researchBibs/twitter.php>
- boyd, danah, & Crawford, K. (2012). 'CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society*, 15/5: 662–79. DOI: 10.1080/1369118X.2012.678878
- boyd, danah, Golder, S., & Lotan, G. (2010). 'Tweet, tweet, retweet: Conversational aspects of retweeting on twitter'. System Sciences (HICSS), 2010 43rd Hawaii International Conference on, pp. 1–10. IEEE.
- Boynton, G. R., Cook, J., Daniels, K., Dawkins, M., Kopish, J., Makar, M., McDavid, W., et al. (2014). 'The Political Domain Goes to Twitter: Hashtags, Retweets and URLs', *Open Journal of Political Science*, 04/01: 8–15. DOI: 10.4236/ojps.2014.41002
- Brabazon, T. (2008). 'Review Here Comes Everybody: The Power of Organizing without Organizations'. *Times Higher Education (THE)*. Retrieved December 28, 2015, from https://www.timeshighereducation.com/books/here-comes-everybody-the-power-of-organizing-without-organizations/401300.article
- Broersma, M., & Graham, T. (2012). 'SOCIAL MEDIA AS BEAT: Tweets as a news source during the 2010 British and Dutch elections', *Journalism Practice*, 6/3: 403–19. DOI: 10.1080/17512786.2012.663626

- Bruns, A. (2012a). 'How long is a tweet? Mapping dynamic conversation networks on Twitter using GAWK and GELPHI', *Information, Communication & Society*, 15/9: 1323–51. DOI: 10.1080/1369118X.2011.635214
- ——. (2012b). 'Ad Hoc innovation by users of social networks: The case of Twitter', ZSI Discussion Paper, 16/2012: 1–13.
- ——. (2015). 'USING GAWK TO PREPARE TCAT DATA FOR TABLEAU'. Mapping Online Publics. Research, . Retrieved November 30, 2015, from http://mappingonlinepublics.net/2015/03/02/using-gawk-to-prepare-tcat-data-for-tableau-part-1/>
- Bruns, A., & Burgess, J. (2013). 'Cumulative and comparative social media analytics for crisis communication', *Selected Papers of Internet Research 14.0*.
- Bruns, A., & Burgess, J. E. (2011). 'The use of Twitter hashtags in the formation of ad hoc publics'. ARC Centre of Excellence for Creative Industries and Innovation; Creative Industries Faculty; Institute for Creative Industries and Innovation. Presented at the 6th European Consortium for Political Research General Conference, August 27, University of Iceland, Reykjavik.
- Bruns, A., Burgess, J. E., Crawford, K., & Shaw, F. (2012). '# qldfloods and@ QPSMedia: Crisis communication on Twitter in the 2011 south east Queensland floods',.
- Bruns, A., Burgess, J., & others. (2011). '# ausvotes: How Twitter covered the 2010 Australian federal election',.
- Bruns, A., & Highfield, T. (2013). 'POLITICAL NETWORKS ON TWITTER: Tweeting the Queensland state election', *Information, Communication & Society*, 16/5: 667–91. DOI: 10.1080/1369118X.2013.782328
- Bruns, A., & Stieglitz, S. (2012). 'Quantitative Approaches to Comparing Communication Patterns on Twitter', *Journal of Technology in Human Services*, 30/3-4: 160–85. DOI: 10.1080/15228835.2012.744249
- ——. (2013). 'Towards more systematic *Twitter* analysis: metrics for tweeting activities', *International Journal of Social Research Methodology*, 16/2: 91–108. DOI: 10.1080/13645579.2012.756095
- Burgess, J. E., & Crawford, K. (2011). 'Social media and the theory of the acute event', *Performance and Participation, Seattle*, Paper presented at Internet Research 12.0.
- Burt, R. S., Kilduff, M., & Tasselli, S. (2013). 'Social Network Analysis: Foundations and Frontiers on Advantage', *Annual Review of Psychology*, 64/1: 527–47. DOI: 10.1146/annurev-psych-113011-143828
- Castells, M. (2000). *The rise of the network society*. Information age, 2nd ed. Oxford; Malden, Mass: Blackwell Publishers.
- -----. (2006). *The network society: from knowledge to policy*. Washington, DC: Johns Hopkins Center for Transatlantic Relations.
- -----. (2009). Communication power. Oxford ; New York: Oxford University Press.
- Chadwick, A. (2006). Internet politics: states, citizens, and new communication technologies. New York: Oxford University Press.

- Chadwick, A. (2011). 'Britain's First Live Televised Party Leaders' Debate: From the News Cycle to the Political Information Cycle', *Parliamentary Affairs*, 64/1: 24–44. DOI: 10.1093/pa/gsq045
- Chadwick, A. (2013). *The hybrid media system: politics and power*. Oxford studies in digital politics. Oxford; New York: Oxford University Press.
- Chae, B. (Kevin). (2015). 'Insights from hashtag #supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research', *International Journal of Production Economics*. DOI: 10.1016/j.ijpe.2014.12.037
- Coase, R. H. (1937). 'The Nature of the Firm', *Economica*, 4/16: 386–405. DOI: 10.1111/j.1468-0335.1937.tb00002.x
- Collins, R. (1986). *Media, culture, and society: a critical reader*. London; Beverly Hills: Sage Publications.
- Conover, M. D., Gonçalves, B., Flammini, A., & Menczer, F. (2012). 'Partisan asymmetries in online political activity', *EPJ Data Science*, 1/1. DOI: 10.1140/epjds6
- Couldry, N. (2013). *Media, society, world social theory and digital media practice.* [Place of publication not identified]: Polity Press.
- Creswell, J. W. (2013). Qualitative inquiry and research design: choosing among five approaches., 3rd ed. Los Angeles: SAGE Publications.

-----. (2014). A concise introduction to mixed methods research. Los Angeles: SAGE.

- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research.*, 2nd ed. Los Angeles: SAGE Publications.
- Dang-Xuan, L., Stieglitz, S., Wladarsch, J., & Neuberger, C. (2013). 'AN INVESTIGATION OF INFLUENTIALS AND THE ROLE OF SENTIMENT IN POLITICAL COMMUNICATION ON TWITTER DURING ELECTION PERIODS', Information, Communication & Society, 16/5: 795–825. DOI: 10.1080/1369118X.2013.783608
- Davis, R. (1999). The Web of Politics: The Internet's Impact on the American Political System. New York: Oxford University Press.
- Denzin, N. K., & Lincoln, Y. S. (2005). The SAGE handbook of qualitative research., 3rd ed. Thousand Oaks: Sage Publications.
- D'heer, E., & Verdegem, P. (2014). 'Conversations about the elections on Twitter: Towards a structural understanding of Twitter's relation with the political and the media field', *European Journal of Communication*, 29/6: 720–34. DOI: 10.1177/0267323114544866
- Diakopoulos, N. A., & Shamma, D. A. (2010). 'Characterizing debate performance via aggregated twitter sentiment'., p. 1195. ACM Press. DOI: 10.1145/1753326.1753504
- Dimitrova, D. V., Shehata, A., Stromback, J., & Nord, L. W. (2014). 'The Effects of Digital Media on Political Knowledge and Participation in Election Campaigns: Evidence From Panel Data', *Communication Research*, 41/1: 95–118. DOI: 10.1177/0093650211426004

- Dixon, D. (2012). 'Analysis Tool or Research Methodology: Is there an epistemology for patterns?'. Berry D. M. (ed.) Understanding digital humanities. Palgrave Macmillan: Houndmills, Basingstoke, Hampshire; New York.
- Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). 'Social Media Update 2014 (Pew)'. *Pew Research Center: Internet, Science & Tech.* Retrieved December 30, 2015, from http://www.pewinternet.org/2015/01/09/social-media-update-2014/>
- Dutton, W. H. (2009). 'The Fifth Estate Emerging through the Network of Networks', *Prometheus*, 27/1: 1–15. DOI: 10.1080/08109020802657453
- Earle, P. S., Bowden, D. C., & Guy, M. (2011). 'Twitter earthquake detection: earthquake monitoring in a social world', *Annals of Geophysics*, 54/6: 708–15.
- Ellison, N. B., & boyd, d. (2013). 'Sociality through Social Network Sites'. Dutton W. H. (ed.) *The Oxford Handbook of Internet Studies*, pp. 151–72. Oxford University Press: Oxford.
- Elmer, G. (2013). 'Live research: Twittering an election debate', *New Media & Society*, 15/1: 18–30. DOI: 10.1177/1461444812457328
- Enli, G. S., & Skogerbø, E. (2013). 'PERSONALIZED CAMPAIGNS IN PARTY-CENTRED POLITICS: Twitter and Facebook as arenas for political communication', *Information, Communication & Society*, 16/5: 757–74. DOI: 10.1080/1369118X.2013.782330
- Ess, C. (2010). *Digital media ethics*. Digital media and society series, reprinted. Cambridge: Polity.
- Fan, W., & Gordon, M. D. (2014). 'The power of social media analytics', Communications of the ACM, 57/6: 74–81. DOI: 10.1145/2602574
- Feller, A., Kuhnert, Sprenger, T. O., & Welpe, I. M. (2011). 'Divided They Tweet: The Network Structure of Political Microbloggers and Discussion Topics', International AAAI Conference on Web and Social Media; Fifth International AAAI Conference on Weblogs and Social Media.
- Freelon, D. (2014). 'On the Interpretation of Digital Trace Data in Communication and Social Computing Research', *Journal of Broadcasting & Electronic Media*, 58/1: 59–75. DOI: 10.1080/08838151.2013.875018
- Fuchs, C. (2011). Foundations of critical media and information studies. Routledge advances in sociology. Milton Park, Abingdon, Oxon; New York: Routledge.
- Gainous, J., & Wagner, K. M. (2014). Tweeting to power: the social media revolution in American politics. New York: Oxford University Press.
- Gilmore, J. (2012). 'Ditching the pack: Digital media in the 2010 Brazilian congressional campaigns', New Media & Society, 14/4: 617–33. DOI: 10.1177/1461444811422429
- Godnov, U., & Redek, T. (2014). 'The Use of Twitter for Political Purposes in Slovenia', Romanian Journal of Political Science, 14/1.

- Golbeck, J., Grimes, J. M., & Rogers, A. (2010). 'Twitter use by the U.S. Congress', *Journal of the American Society for Information Science and Technology*, n/a n/a. DOI: 10.1002/asi.21344
- Golder, S. A. (2006). 'Usage patterns of collaborative tagging systems', *Journal of Information* Science, 32/2: 198–208. DOI: 10.1177/0165551506062337
- Graham, T., Broersma, M., & Hazelhoff, K. (2013). 'Twitter as an instrument for connected representation',.
- Groshek, J., & Al-Rawi, A. (2013). 'Public Sentiment and Critical Framing in Social Media Content During the 2012 U.S. Presidential Campaign', *Social Science Computer Review*, 31/5: 563–76. DOI: 10.1177/0894439313490401
- Grossman, L. (2009). 'Iran Protests: Twitter, the Medium of the Movement'. Time.
- Grusell, M., & Nord, L. (2012). 'Three attitudes to 140 characters the use and views of twitter in political party communications in Sweden', *Public Communication Review*, 2/2.
- Gupta, A., & Kumaraguru, P. (2012). 'Credibility ranking of tweets during high impact events'., pp. 2–8. ACM Press. DOI: 10.1145/2185354.2185356
- Halstead, N. (2015). 'Twitter Ends its Partnership with DataSift Firehose Access Expires on August 13, 2015'. *DataSift Blog*. Retrieved November 3, 2015, from <http://blog.datasift.com/2015/04/11/twitter-ends-its-partnership-with-datasiftfirehose-access-expires-on-august-13-2015/>
- Hanna, A., Sayre, B., Bode, L., Yang, J., & Shah, D. V. (2011). 'Mapping the Political Twitterverse: Candidates and Their Followers in the Midterms.' *ICWSM*, pp. 510–3.
- Hanneman, R. A. (2005). 'Introduction to Social Network Methods: Chapter 8: More Properties of Networks and Actors'. Introduction to social network methods. Retrieved August 17, 2015, from http://www.faculty.ucr.edu/~hanneman/nettext/C8_Embedding.html
- Harrington, S., Highfield, T., & Bruns, A. (2012). 'More than a backchannel: Twitter and television', Audience interactivity and participation, 13–7.
- Hawthorne, J., Houston, J. B., & McKinney, M. S. (2013). 'Live-Tweeting a Presidential Primary Debate: Exploring New Political Conversations', *Social Science Computer Review*, 31/5: 552–62. DOI: 10.1177/0894439313490643
- Hemphill, L., Otterbacher, J., & Shapiro, M. (2013). 'What's congress doing on twitter?'., p. 877. ACM Press. DOI: 10.1145/2441776.2441876
- Highfield, T. (2012). 'Talking of Many Things: Using Topical Networks to Study Discussions in Social Media', *Journal of Technology in Human Services*, 30/3-4: 204–18. DOI: 10.1080/15228835.2012.746894
- Highfield, T., Harrington, S., & Bruns, A. (2013). 'Twitter as a technology for audiencing and fandom: The# Eurovision phenomenon', *Information, Communication & Society*, 16/3: 315–39.

- Highfield, T., Kirchhoff, L., & Nicolai, T. (2011). 'Challenges of Tracking Topical Discussion Networks Online', *Social Science Computer Review*, 29/3: 340–53. DOI: 10.1177/0894439310382514
- Honeycutt, C., & Herring, S. C. (2009). 'Beyond microblogging: Conversation and collaboration via Twitter'. System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on, pp. 1–10. IEEE.
- Hong, S. (2013). 'Who benefits from Twitter? Social media and political competition in the U.S. House of Representatives', *Government Information Quarterly*, 30/4: 464–72. DOI: 10.1016/j.giq.2013.05.009
- Howison, J., Wiggans, A., & Crowston, J. (2011). 'Validity issues in the use of social network analysis with digital trace data', *Journal of the Association for Information Systems*, 12: 767– 97.
- Huang, C. (2011). 'Facebook and Twitter key to Arab Spring uprisings: report | The National'. *The National (UAE)*. Retrieved October 29, 2015, from http://www.thenational.ae/news/uae-news/facebook-and-twitter-key-to-arabspring-uprisings-report>
- Ipsos MRBI. (2015). Social Networking Quarterly (Ireland) August 2015 (Social Networking Usage Infographic)., p. 1. Dublin, Ireland: IPSOS. Retrieved December 21, 2015, from http://ipsosmrbi.com/wp-content/uploads/2015/11/SN_Aug15.pdf>
- Iyengar, S., & Kinder, D. R. (1987). News that matters: television and American opinion. American politics and political economy. Chicago: Univ. of Chicago Press.
- Jackson, N., & Lilleker, D. (2011). 'Microblogging, Constituency Service and Impression Management: UK MPs and the Use of Twitter', *The Journal of Legislative Studies*, 17/1: 86–105. DOI: 10.1080/13572334.2011.545181
- Jenkins, R. (2007). Pierre Bourdieu. Key sociologists, Rev. ed., [reprinted]. London: Routledge.
- Jungherr, A. (2014). 'The Logic of Political Coverage on Twitter: Temporal Dynamics and Content: Coverage on Twitter', *Journal of Communication*, 64/2: 239–59. DOI: 10.1111/jcom.12087
- Jungherr, A., & Jurgens, P. (2014). 'Through a Glass, Darkly: Tactical Support and Symbolic Association in Twitter Messages Commenting on Stuttgart 21', Social Science Computer Review, 32/1: 74–89. DOI: 10.1177/0894439313500022
- Jürgens, P., Jungherr, A., & Schoen, H. (2011). 'Small worlds with a difference: new gatekeepers and the filtering of political information on Twitter'., pp. 1–5. ACM Press. DOI: 10.1145/2527031.2527034
- Juris, J. S. (2012). 'Reflections on #Occupy Everywhere: Social media, public space, and emerging logics of aggregation', *American Ethnologist*, 39/2: 259–79.
- Kim, M., & Park, H. W. (2012). 'Measuring Twitter-based political participation and deliberation in the South Korean context by using social network and Triple Helix indicators', *Scientometrics*, 90/1: 121–40. DOI: 10.1007/s11192-011-0508-5

King, N., & Horrocks, C. (2010). Interviews in qualitative research. Los Angeles: SAGE.

Kitchin, R. (2014). The data revolution. Thousand Oaks, CA: SAGE Publications Ltd.

- Klinger, U. (2013). 'MASTERING THE ART OF SOCIAL MEDIA: Swiss parties, the 2011 national election and digital challenges', *Information, Communication & Society*, 16/5: 717–36. DOI: 10.1080/1369118X.2013.782329
- Krackhardt, D., & Stern, R. N. (1988). 'Informal Networks and Organizational Crises: An Experimental Simulation', *Social Psychology Quarterly*, 51/2: 123. DOI: 10.2307/2786835
- Kvale, S. (1994). 'Ten standard objections to qualitative research interviews', *Journal of phenomenological psychology*, 25/2: 147–73.
- -----. (2007). *Doing interviews*. The sage qualitative research kit. Thousand Oaks, CA: Sage Publications.
- Larsson, A. O., & Moe, H. (2011). 'Studying political microblogging: Twitter users in the 2010 Swedish election campaign', *New Media & Society*, 14/5: 729–47. DOI: 10.1177/1461444811422894
- ——. (2014). 'Triumph of the Underdogs? Comparing Twitter Use by Political Actors During Two Norwegian Election Campaigns', SAGE Open, 4/4. DOI: 10.1177/2158244014559015
- Latour, B. (1996). Aramis, or, The love of technology. Cambridge, Mass: Harvard University Press.

-----. (2007). Reassembling the social: an introduction to Actor-Network-Theory. Clarendon lectures in management studies. Oxford: Oxford Univ. Press.

- LaValle, S., Lesser, E., Shockley, R., Hopkins, M., & Kruschwitz, N. (2010). 'Big Data, Analytics and the Path From Insights to Value'. *MIT Sloan Management Review*. Retrieved December 16, 2015, from http://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/>
- Law, J. (1992). 'Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity', *Systems practice*, 5/4: 379–93.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., et al. (2009). 'SOCIAL SCIENCE: Computational Social Science', *Science*, 323/5915: 721–3. DOI: 10.1126/science.1167742
- Lee, E.-J., & Jang, J. -w. (2013). 'Not So Imaginary Interpersonal Contact With Public Figures on Social Network Sites: How Affiliative Tendency Moderates Its Effects', *Communication Research*, 40/1: 27–51. DOI: 10.1177/0093650211431579
- Lee, E.-J., & Oh, S. Y. (2013). 'Seek and You Shall Find? How Need for Orientation Moderates Knowledge Gain from Twitter Use: Twitter and News Knowledge', *Journal of Communication*, 63/4: 745–65. DOI: 10.1111/jcom.12041
- Lee, E.-J., & Shin, S. Y. (2014). 'When the Medium Is the Message: How Transportability Moderates the Effects of Politicians' Twitter Communication', *Communication Research*, 41/8: 1088–110. DOI: 10.1177/0093650212466407
- Lewis, S. C., Zamith, R., & Hermida, A. (2013). 'Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods', *Journal of Broadcasting & Electronic Media*, 57/1: 34–52. DOI: 10.1080/08838151.2012.761702

- Lilleker, D., & Vedel, T. (2013). 'The internet in campaigns and elections.' *The Oxford Handbook of Internet Studies*, p. 632. Oxford University Press: Oxford.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, Calif: Sage Publications.
- Lin, Y.-R., Keegan, B., Margolin, D., & Lazer, D. (2014). 'Rising Tides or Rising Stars?: Dynamics of Shared Attention on Twitter during Media Events', (P. Holme, Ed.)PLoS ONE, 9/5: e94093. DOI: 10.1371/journal.pone.0094093
- Lipman, V. (2014). 'Top Twitter Trends: What Countries Are Most Active? Who's Most Popular? - Forbes'. *Forbes*. Retrieved December 30, 2015, from <http://www.forbes.com/sites/victorlipman/2014/05/24/top-twitter-trends-whatcountries-are-most-active-whos-most-popular/>
- Lister, M., Dovey, J., Giddings, S., Grant, I., & Kelly, K. (2009). New media: A Critical Introduction., 2nd ed. Milton Park, Abingdon, Oxon; New York, N.Y: Routledge.
- MacKenzie, D. A., & Wajcman, J. (1985). *The Social shaping of technology: how the refrigerator got its hum*. Milton Keynes; Philadelphia: Open University Press.
- -----. (1999). The social shaping of technology., 2nd ed. Buckingham [Eng.]; Philadelphia: Open University Press.
- Maireder, A., & Ausserhofer, J. (2013). 'Political discourses on Twitter: networking topics, objects and people', *Twitter and Society, New York, Peter Lang*, 291–341.
- Manovich. (2011). 'Trending: the promises and the challenges of big social data.' *Debates in the Digital Humanities*, p. 460-476. Univ Of Minnesota Press.: Minnesota, USA.
- Mascaro, C., & Goggins, S. P. (2012). 'Twitter as virtual town square: Citizen engagement during a nationally televised republican primary debate'. APSA 2012 Annual Meeting Paper.
- Mayer-Schönberger, V., & Cukier, K. (2014). *Big data: a revolution that will transform how we live, work, and think.*, First Mariner Books edition. Boston: Mariner Books, Houghton Mifflin Harcourt.
- Maykut, P. S., & Morehouse, R. (1994). Beginning qualitative research: a philosophic and practical guide. The Falmer Press teachers' library. London; Washington, D.C: Falmer Press.
- McLuhan, M. (1964). Understanding media: the extensions of man., 1st MIT Press ed. Cambridge, Mass: MIT Press.
- McMillan, S. J. (2000). 'The Microscope and the Moving Target: The Challenge of Applying Content Analysis to the World Wide Web', *Journalism & Mass Communication Quarterly*, 77/1: 80–98. DOI: 10.1177/107769900007700107
- Meier, F., Elsweiler, D., & Wilson, M. L. (2014). 'More than liking and bookmarking? towards understanding twitter favouriting behaviour'. Proceedings of the International AAAI Conference on Weblogs and Social Media.
- Messina, C. (2007a). 'Groups for Twitter; or A Proposal for Twitter Tag Channels'. FactoryCity. Retrieved February 17, 2014, from http://factoryjoe.com/blog/2007/08/25/groups-for-twitter-or-a-proposal-for-twitter-tag-channels/>

- ——. (2007b). 'Twitter hashtags for emergency coordination and disaster relief'. FactoryCity. Retrieved February 17, 2014, from <http://factoryjoe.com/blog/2007/10/22/twitter-hashtags-for-emergencycoordination-and-disaster-relief/>
- Metaxas, P. T., & Mustafaraj, E. (2012). 'Social Media and the Elections', *Science*, 338/6106: 472–3. DOI: 10.1126/science.1230456
- Meyer, R. (2014). 'Twitter Starts to Change the Central Logic of Its Service'. *The Atlantic*. Retrieved January 3, 2016, from <http://www.theatlantic.com/technology/archive/2014/08/twitter-starts-tochange-the-central-logic-of-its-service/378650/>
- Mirer, M. L., & Bode, L. (2015). 'Tweeting in defeat: How candidates concede and claim victory in 140 characters', New Media & Society, 17/3: 453–69. DOI: 10.1177/1461444813505364
- Moe, H., & Larsson, A. O. (2013). 'UNTANGLING A COMPLEX MEDIA SYSTEM: A comparative study of Twitter-linking practices during three Scandinavian election campaigns', *Information, Communication & Society*, 16/5: 775–94. DOI: 10.1080/1369118X.2013.783607
- Mohammad, S. M., & Kiritchenko, S. (2014). 'Using Hashtags to Capture Fine Emotion Categories from Tweets', *Computational Intelligence*, n/a – n/a. DOI: 10.1111/coin.12024
- Moore, H. (2013). 'Twitter heads for stock market debut by filing for IPO'. The Guardian.
- Morozov, E. (2011). The net delusion: the dark side of internet freedom., 1st ed. New York: Public Affairs.
- -----. (2013). To save everything, click here: the folly of technological solutionism. New York: Public Affairs.
- Murthy, D. (2011). 'Emergent digital ethnographic methods for social research'. *Handbook of Emergent Technologies in Social Research*, pp. 158–79. Oxford University Press: New York, NY.
- Mustafaraj, E., Finn, S., Whitlock, C., & Metaxas, P. T. (2011). 'Vocal minority versus silent majority: Discovering the opionions of the long tail'. *Privacy, Security, Risk and Trust* (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on, pp. 103–10. IEEE.
- Newton, C. (2015). 'Twitter officially kills off favorites and replaces them with likes'. *The Verge*. Retrieved December 17, 2015, from http://www.theverge.com/2015/11/3/9661180/twitter-vine-favorite-fav-likes-hearts
- Nilsson, B., & Carlsson, E. (2014). 'Swedish politicians and new media: Democracy, identity and populism in a digital discourse', *New Media & Society*, 16/4: 655–71. DOI: 10.1177/1461444813487964
- O'Brien, J. (2013). 'Github: yourTwapperKeeper'. *GitHub*. Reference, . Retrieved March 25, 2014, from <https://github.com/540co/yourTwapperKeeper>

- Osterburg, G. (2013). Update on the Twitter Archive At the Library of Congress (White Paper)., p. 5. Washington DC, USA: Library of Congress. Retrieved from <But more than five years later, the project is in limbo. The library is still grappling with how to manage an archive that amounts to something like half a trillion tweets. And the researchers are still waiting. Read more: http://www.politico.com/story/2015/07/library-ofcongress-twitter-archive-119698.html#ixz3utJcbQSF>
- Otterbacher, J., Shapiro, M. A., & Hemphill, L. (2012). 'Tweeting vertically? Elected officials' interactions with citizens on Twitter',.
- Oudshoorn, N., & Pinch, T. J. (2003). *How users matter: the co-construction of users and technologies.* Inside technology. Cambridge, Mass: MIT Press.
- Papacharissi, Z., & Easton, E. (2013). 'In the Habitus of the New: Structure, Agency, and the Social Media Habitus'. Hartley J., Burgess J., & Bruns A. (eds) A Companion to New Media Dynamics, pp. 167–84. Wiley-Blackwell: Oxford, UK.
- Papacharissi, Z., & de Fatima Oliveira, M. (2012). 'Affective News and Networked Publics: The Rhythms of News Storytelling on #Egypt', *Journal of Communication*, 62/2: 266–82. DOI: 10.1111/j.1460-2466.2012.01630.x
- Parmelee, J. H. (2014). 'The agenda-building function of political tweets', *New Media & Society*, 16/3: 434–50. DOI: 10.1177/1461444813487955
- Parmelee, J. H., & Bichard, S. L. (2013). Politics and the Twitter revolution: how tweets influence the relationship between political leaders and the public. Lexington studies in political communication, 1. paperback ed. Lanham, Md.: Lexington Books.
- Patton, M. Q. (1990). Qualitative research and evaluation methods., 3 ed. Thousand Oaks, Calif: Sage Publications.
- Pew Research Center. (2013). 'Twitter and the Campaign'. Pew Research Center's Journalism Project. Retrieved November 2, 2015, from <http://www.journalism.org/2011/12/08/twitter-and-campaign/>
- Plotkowiak, T., & Stanoevska-Slabeva, K. (2013). 'German politicians and their Twitter networks in the Bundestag Election 2009', *First Monday*, 18/5. DOI: 10.5210/fm.v18i5.3816
- Preston, P. (2001). *Reshaping communications: technology, information and social change*. London; Thousand Oaks, Calif: SAGE.
- Prior, M. (2013). 'Media and Political Polarization', Annual Review of Political Science, 16/1: 101–27. DOI: 10.1146/annurev-polisci-100711-135242
- Procter, R., Vis, F., & Voss, A. (2013). 'Reading the riots on Twitter: methodological innovation for the analysis of big data', *International Journal of Social Research Methodology*, 16/3: 197–214. DOI: 10.1080/13645579.2013.774172
- Putnam, R. D. (2001). Bowling alone: the collapse and revival of American community., 1. touchstone ed. New York, NY: Simon & Schuster.
- Quercia, D., Capra, L., & Crowcroft, J. (2012). 'The Social World of Twitter: Topics, Geography, and Emotions.', *ICWSM*, 12: 298–305.

- Rainie, L., Smith, A., Schlozman, K. L., Brady, H., & Verba, S. (2012). 'Social Media and Political Engagement'. *Pew Research Center: Internet, Science & Tech.* Retrieved November 2, 2015, from http://www.pewinternet.org/2012/10/19/social-mediaand-political-engagement/>
- Rash, W. (1997). Politics on the nets: wiring the political process. New York: W.H. Freeman.
- Raymond, M. (2010). 'How Tweet It Is!: Library Acquires Entire Twitter Archive'. *Library of Congress Blog.* Government, . Retrieved December 20, 2015, from <http://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitterarchive/>
- Richardson, L., & Ruby, S. (2007). RESTful web services. Farnham: O'Reilly.
- Roback, A., & Hemphill, L. (2013). 'I'd have to vote against you: issue campaigning via twitter'. Proceedings of the 2013 conference on Computer supported cooperative work companion, pp. 259–62. ACM.
- Rogers, R. (2009). End of the virtual digital methods: inaugural lecture delivered on the appointment to the Chair of New Media & Digital Culture at the University of Amsterdam on 8 May 2009. Amsterdam: Amsterdam University Press.
- ——. (2010). 'Internet Research: The Question of Method—A Keynote Address from the YouTube and the 2008 Election Cycle in the United States Conference', *Journal of Information Technology & Politics*, 7/2-3: 241–60. DOI: 10.1080/19331681003753438
- Scola, N. (2015). 'Library of Congress: Twitter archive project limbo'. POLITICO. Retrieved December 20, 2015, from http://www.politico.com/story/2015/07/library-of-congress-twitter-archive-119698.html>
- Scott, J., & Carrington, P. (2014). The SAGE Handbook of Social Network Analysis. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd.
- Seidman, I. (2013). Interviewing as qualitative research: a guide for researchers in education and the social sciences., 4th ed. New York: Teachers College Press.
- Shamma, D. A., Kennedy, L., & Churchill, E. F. (2010). 'Conversational Shadows: Describing Live Media Events Using Short Messages.' *ICWSM*.
- Shirky, C. (2008). *Here comes everybody: the power of organizing without organizations*. New York: Penguin Press.
- Smith, M. A., Rainie, L., Himelboim, I., & Shneiderman, B. (2014). Smith, M. A., Rainie, L., Himelboim, I., & Shneiderman, B. (2014). Selected by the probability of the probability
- Spinello, R. A., & Tavani, H. T. (Eds). (2004). Readings in cyberethics., 2nd ed. Sudbury, Mass: Jones and Bartlett Publishers.
- Statistica. (2015). 'Global social networks by users 2015 | Statistic'. Statista. Social Media Analytics, . Retrieved October 29, 2015, from http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

- Stieglitz, S., & Dang-Xuan, L. (2012). 'Political Communication and Influence through Microblogging--An Empirical Analysis of Sentiment in Twitter Messages and Retweet Behavior'., pp. 3500–9. IEEE. DOI: 10.1109/HICSS.2012.476
- Stone, B. (2006). 'Six More Twitter Updates!'. *Twitter Blogs*. Retrieved December 17, 2015, from <https://blog.twitter.com/2006/six-more-twitter-updates>
- Strandberg, K. (2013). 'A social media revolution or just a case of history repeating itself? The use of social media in the 2011 Finnish parliamentary elections', *New Media & Society*, 15/8: 1329–47. DOI: 10.1177/1461444812470612
- Swartz, D. (1997). Culture & power: the sociology of Pierre Bourdieu. Chicago: University of Chicago Press.
- Tashakkori, A., & Teddlie, C. (2009). Foundations of mixed methods research: integrating quantitative and qualitative approaches in the social and behavioral sciences. Los Angeles: SAGE.

-----. (2010). Sage handbook of mixed methods in social & behavioral research., 2nd ed. Los Angeles: SAGE Publications.

- Thelwall, M. (2010). 'Researching the public web'. *eResearch Ethics*. Academic, . Retrieved from <http://eresearch-ethics.org/position/researching-the-public-web/>
- Tonkin, E., Pfeiffer, H. D., & Tourte, G. (2012). 'Twitter, information sharing and the London riots?', Bulletin of the American Society for Information Science and Technology, 38/2: 49–57.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). 'Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment.', *ICWSM*, 10: 178–85.
- Twitter. (2015a). 'History of the REST & Search API'. *History of the REST & Search API*. Reference, . Retrieved February 17, 2012, from <https://dev.twitter.com/docs/history-rest-search-api>
- ——. (2015b). 'Posting photos on Twitter'. Retrieved December 17, 2015, from <https://support.twitter.com/articles/20156423#>
- ——. (2015c). 'FAQs about verified accounts'. Twitter Help Center. Retrieved December 19, 2015, from https://support.twitter.com/groups/31-twitter-basics/topics/111-features/articles/119135-about-verified-accounts#
- Twitter Developers. (2015). 'Tweets'. *Twitter Developer Documentation: API Tweet Objects*. Retrieved December 16, 2015, from <https://dev.twitter.com/overview/api/tweets>
- Twitter Inc. (2013). 'Form 10K: Twitter Inc. Annual Report to SEC'. UNITED STATES SECURITIES AND EXCHANGE COMMISSION.
- Twitter Investor Relations. (2014). 'Twitter Reach: Q4'13 avg Monthly Active Users 241 million, up 30% y/y. #TWTRearnings pic.twitter.com/74Eva0gRqc'. @*TwitterIR*. microblog, . Retrieved February 11, 2014, from <https://twitter.com/TwitterIR/status/431175668122935297>
- ——. (2015). Twitter Q3 2015 Selected Company Metrics and Financials (Financial Metrics). Retrieved December 20, 2015, from

<http://files.shareholder.com/downloads/AMDA-2F526X/1254804449x0x856828/5EE94FC8-9A5A-4587-ACBD-D536A9027A62/Q315_Selected_Company_Metrics_and_Financials.pdf>

- Uprichard, E. (2013a). 'Focus: Big Data, Little Questions?'. *Discover Society*. Retrieved December 16, 2015, from http://discoversociety.org/2013/10/01/focus-big-data-little-questions/
- ——. (2013b). 'Sampling: bridging probability and non-probability designs', International Journal of Social Research Methodology, 16/1: 1–11. DOI: 10.1080/13645579.2011.633391
- Vaccari, C., & Valeriani, A. (2015). 'Follow the leader! Direct and indirect flows of political communication during the 2013 Italian general election campaign', *New Media & Society*, 17/7: 1025–42. DOI: 10.1177/1461444813511038
- Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J. T., Nagler, J., & Tucker, J. (2013). 'Social media and political communication. A survey of Twitter users during the 2013 Italian general election', *Rivista italiana di scienza politica*, 3: 381–410. DOI: 10.1426/75245
- Vergeer, M., & Hermans, L. (2013). 'Campaigning on Twitter: Microblogging and Online Social Networking as Campaign Tools in the 2010 General Elections in the Netherlands', *Journal of Computer-Mediated Communication*, 18/4: 399–419. DOI: 10.1111/jcc4.12023
- Vergeer, M., Hermans, L., & Sams, S. (2011). 'Is the voter only a tweet away? Micro blogging during the 2009 European Parliament election campaign in the Netherlands', *First Monday*, 16/8. DOI: 10.5210/fm.v16i8.3540
- Vergeer, M., Hermans, L., & Sams, S. (2013). 'Online social networks and micro-blogging in political campaigning: The exploration of a new campaign tool and a new campaign style', *Party Politics*, 19/3: 477–501. DOI: 10.1177/1354068811407580
- Vis, F. (2013). 'TWITTER AS A REPORTING TOOL FOR BREAKING NEWS: Journalists tweeting the 2011 UK riots', *Digital Journalism*, 1/1: 27–47. DOI: 10.1080/21670811.2012.741316
- Wagner, C., Liao, V., Pirolli, P., Nelson, L., & Strohmaier, M. (2012). 'It's Not in Their Tweets: Modeling Topical Expertise of Twitter Users'., pp. 91–100. IEEE. DOI: 10.1109/SocialCom-PASSAT.2012.30
- Wallsten, K. (2014). 'Microblogging and the News: Political Elites and the Ultimate Retweet'. Solo A. M. G. (ed.) *Political Campaigning in the Information Age:*, pp. 128–47. IGI Global.
- Wasserman, S., & Faust, K. (1994). Social network analysis: methods and applications. Structural analysis in the social sciences. Cambridge; New York: Cambridge University Press.
- Weare, C., & Lin, W.-Y. (2000). 'Content Analysis of the World Wide Web: Opportunities and Challenges', Social Science Computer Review, 18/3: 272–92. DOI: 10.1177/089443930001800304
- Weber, I., Popescu, A.-M., & Pennacchiotti, M. (2013). PLEAD'13: proceedings of the Workshop on Politics, Elections and Data: October 28, 2013, San Francisco, CA, USA.

- Wilhelm, A. G. (2000). *Democracy in the Digital Age: challenges to political life in cyberspace*. New York: Routledge.
- Williams, R. (1974). *Television: technology and cultural form*. Routledge classics. London; New York: Routledge.
- Williams, S. A., Terras, M. M., & Warwick, C. (2013). 'What do people study when they study Twitter? Classifying Twitter related academic papers', *Journal of Documentation*, 69/3: 384–410. DOI: 10.1108/JD-03-2012-0027
- Winner, L. (1980). 'Do artifacts have politics?', Daedalus, 121-36.
- Woolcock, M. (2010). 'The Rise and Routinization of Social Capital, 1988–2008', Annual Review of Political Science, 13/1: 469–87. DOI: 10.1146/annurev.polisci.031108.094151
- Wu, T. (2011). The master switch: the rise and fall of information empire. New York: Knopf.
- Zeng, D., Chen, H., Lusch, R., & Li, S.-H. (2010). 'Social media analytics and intelligence', Intelligent Systems, IEEE, 25/6: 13-6.
- Zimmer, M. (2010). 'Is it Ethical to Harvest Public Twitter Accounts without Consent?'. MichaelZimmer.org. Retrieved November 3, 2015, from <http://www.michaelzimmer.org/2010/02/12/is-it-ethical-to-harvest-publictwitter-accounts-without-consent/>
- Zinko, C. (2009). 'What is Biz Stone doing?'. *SFGate*. Retrieved February 17, 2014, from <http://www.sfgate.com/living/article/What-is-Biz-Stone-doing-3165704.php#src=fb>

Appendices

A.3 Addenda to Chapter 4

A.3.1 Corpus of academic literature related to political

discourse on Twitter (1/3)

Year Author(s)		Title	Item Type
2008 Abroms, Lorien C.; Craig Lefebvre, R.		Obama's Wired Campaign: Lessons for Public Health Communication	Journal Article
2009 Garrett, R. Kelly		Politically Motivated Reinforcement Seeking: Reframing the Selective Exposure Debate	Journal Article
2009 Lazer, D.; Pentland, A.; Adamic, L.; Aral, S.; 2009 L.: Gutmann, M.: Lebara, T.: Kine, G.: Maco.	; Barabasi, AL.; Brewer, D.; Christakis, N.; Contractor, N.; Fowler, M.: Rov. D.: Van Alstvne, M.	SOCIAL SCIENCE: Computational Social Science	Journal Article
2010 Ammann, Sky L.	ere fan hen en e	Why Do They Tweet? The Use of Twitter by U.S. Senate Candidates in 2010	Journal Article
2010 Boyd, Danah; Golder, Scott; Lotan, Gilad		Tweet, tweet, retweet: Conversational aspects of retweeting on twitter	Conference Paper
2010 Chi, Feng; Yang, Nathan; others		Twitter in congress: Outreach vs transparency	Journal Article
2010 Diakopoulos, Nicholas A.; Shamma, David A	А.	Characterizing debate performance via aggregated twitter sentiment	Conference Paper
2010 Golbeck, Jennifer; Grimes, Justin M.; Roger	rs, Anthony	Twitter use by the U.S. Congress	Journal Article
2010 Grant, Will J.; Moon, Brenda; Busby Grant,	Janie	Digital Dialogue? Australian Politicians' use of the Social Network Tool Twitter	Journal Article
2010 Gulati, Jeff; Williams, Christine B.		Communicating with Constituents in 140 Characters or Less: Twitter and the Diffusion of Technology Innovation in the United States Congress	Journal Article
2010 Lilleker, Darren G; Jackson, Nigel A.		Towards a More Participatory Style of Election Campaigning: The Impact of Web 2.0 on the UK 2010 General Election	Journal Article
2010 Shamma, David A.; Kennedy, Lyndon; Chur	rchill, Elizabeth F.	Conversational Shadows: Describing Live Media Events Using Short Messages.	Conference Paper
2010 Shogan, Colleen J.		Blackberries, Tweets, and YouTube: Technology and the Future of Communicating with Congress	Journal Article
2010 Small, Tamara		Canadian politics in 140 characters. Party politics in the Twitterverse	Journal Article
2011 Ampofo, Lawrence; Anstead, Nick; O'Lough	hlin, Ben	TRUST, CONFIDENCE, AND CREDIBILITY: Chrizen responses on Twitter to opinion polls during the 2010 UK General Election	Journal Article
2011 Anstead, N.; O'Loughlin, B.		The Emerging Viewertariat and BBC Question Time: Television Debate and Real-Time Commenting Online	Journal Article
2011 Bode, Leticia; Hanna, Alexander; Sayre, Ber	:n; Yang, JungHwan; Shah, Dhavan V.	Mapping the political twitterverse: Finding connections between political elites	Journal Article
2011 Bruns, Axel; Burgess, Jean E.		The use of Twitter hashtags in the formation of ad hoc publics	Conference Paper
2011 Bruns, Axel; Burgess, Jean; others		# ausvotes: How Twitter covered the 2010 Australian federal election	Journal Article
2011 Chadwick, A.		Britain's First Live Televised Party Leaders' Debate: From the News Cycle to the Political Information Cycle	Journal Article
2011 Chadwick, A.		The Political Information Cycle in a Hybrid News System: The British Prime Minister and the "Bullygate" Affair	Journal Article
2011 Chi, Feng, Yang, Nathan		Twitter Adoption in Congress	Journal Article
2011 Feller, Albert; Kuhnert; Sprenger, Timm O.;	; Welpe, Isabell M.	Divided They Tweet: The Network Structure of Political Microbloggers and Discussion Topics	Journal Article
2011 Hanna, Alexander; Sayre, Ben; Bode, Letici	ia; Yang, JungHwan; Shah, Dhavan V.	Mapping the Political Twitterverse: Candidates and Their Foltowers in the Midterms.	Conference Paper
2011 Jackson, Nigel; Lilleker, Darren		Microblogging. Constituency Service and Impression Management: UK MPs and the Use of Twitter	Journal Article
2011 Jürgens, Pascal; Jungherr, Andreas; Schoen	n, Harald	Small worlds with a difference: new gatekeepers and the filtering of political information on Twitter	Conference Paper
2011 King, G.		Ensuring the Data-Rich Future of the Social Sciences	Journal Article
2011 Larsson, A. O.; Moe, H.		Studying political microblogging: Twitter users in the 2010 Swedish election campaign	Journal Article
2011 Lassen, D. S.; Brown, A. R.		Twitter: The Electoral Connection?	Journal Article
2011 Macnamara, Jim		Pre and post-election 2010 online: What happened to the conversation?	Journal Article
2011 Mustafaraj, Eni; Finn, Samantha; Whitlock,	, Carolyn; Metaxas, Panagiotis T.	Vocal minority versus silent majority: Discovering the opionions of the long tail	Conference Paper
2011 Sæbø		Understanding Twitter Use among Parliament Representatives: A Genre Analysis	Chapter
2011 Shamma, David A.; Kennedy, Lyndon; Churr	rchill, Elizabeth F.	Peaks and persistence: modeling the shape of microblog conversations	Conference Paper
2011 Vergeer, Maurice; Hermans, Liesbeth; Sam	ns, Steven	Is the voter only a tweet away? Micro blogging during the 2009 European Parliament election campaign in the Netherlands	Journal Article
2012 Aharony, Noa		Twitter use by three political leaders: an exploratory analysis	Journal Article
2012 Anstead, Nick; O'Loughlin, Ben		Semantic polling: the ethics of online public opinion	Journal Article
2012 Baxter, Graeme; Marcella, Rita		boes Scotland å@"likeå@" This? Social Media Use by Political Parties and Candidates in Scotland during the 2010 UK General Election Campaign	Journal Article
2012 Boutet, A.; Hyoungshick Kim; Yoneki, E.		What's in Twitter: I know What Parties are Popular and Who You are Supporting Now!	Conference Paper
2012 Broersma, Marcel; Graham, Todd		SOCIAL MEDIA AS BEAT: Tweets as a news source during the 2010 British and Dutch elections	Journal Article
2012 Bruns, Axel; Burgess, Jean E.; Crawford, Kat	ate; Shaw, Frances	# qldfloods and@ QPSMedia: Crisis communication on Twitter in the 2011 south east Queensland floods	Journal Article
2012 Bruns, Axel; Liang, Yuxian Eugene		Tools and methods for capturing Twitter data during natural disasters	Journal Article
2012 Bruns, Axel; Stieglitz, Stefan		Quantitative Approaches to Comparing Communication Patterns on Twitter	Journal Article
2012 Burgess, Jean; Bruns, Axel		(Not) the Twitter election: the dynamics of the# ausvotes conversation in relation to the Australian media ecology	Journal Article
2012 Conover, Michael D; Gonçalves, Bruno; Fl	lammini, Alessandro; Menczer, Filippo	Partisan asymmetries in online political activity	Journal Article
2012 Gilmore, J.		Ditching the pack: Digital media in the 2010 Brazilian congressional campaigns	Journal Article
2012 Grusell, Marie; Nord, Lars		Three attitudes to 140 characters the use and views of twitter in political party communications in Sweden	Journal Article

2012	Hong, Sounman; Nadler, Daniel	Which candidates do the public discuss online in an election campaign?: The use of social media by 2012 presidential candidates and its impact on candidate salience	Journal Article
2012	Jungherr, Andreas	Online Campaigning in Germany: The CDU Online Campaign for the General Election 2009 in Germany	Journal Article
2012	Kim, Minjeong: Park, Han Woo	Measuring Twitter-based political participation and deliberation in the South Korean context by using social network and Triple Helix indicators	Journal Article
2012	Lee, Eun-Ju; Oh, Soo Youn	To Personalize or Depersonalize? When and How Politicians' Personalized Tweets Affect the Public's Reactions: Politicians' Twitter Communication	Journal Article
2012	Mascaro, Christopher; Black, Alan; Goggins, Sean	Tweet recall: examining real-time civic discourse on twitter	Conference Paper
2012	Mascaro, Christopher; Goggins, Sean Patrick	Twitter as virtual town square: Citizen engagement during a nationally televised republican primary debate	Conference Paper
2012	Metaxas, P. T.; Mustafaraj, E.	Social Media and the Elections	Journal Article
2012	Otterbacher, Jahna; Shapiro, Matthew A.; Hemphill, Libby	Tweeting vertically? Elected officials' interactions with citizens on Twitter	Journal Article
2012	Papacharissi, Zizi; de Fatima Oliveira, Maria	Affective News and Networked Publics: The Rhythms of News Storytelling on #Egypt	Journal Article
2012	Peterson, Rolfe Daus	To tweet or not to tweet: Exploring the determinants of early adoption of Twitter by House members in the 111th Congress	Journal Article
2012	Pew Research Center	Twitter and the Campaign	Blog Post
2012	Rainie, Lee; Smith, Aaron; Schlozman, Kay Lehman; Brady, Henry; Verba, Sidney	Social Media and Political Engagement	Blog Post
2012	Stieglitz, Stefan; Dang-Xuan, Linh	Political Communication and Influence through Microblogging-An Empirical Analysis of Sentiment in Twitter Messages and Retweet Behavior	Conference Paper
2013	Ausserhofer, Julian; Maireder, Axel	NATIONAL POLITICS ON TWITTER: Structures and topics of a networked public sphere	Journal Article
2013	Baxter, Graeme; Marcella, Rita	Do Online Election Campaigns Sway the Vote? A Study of the 2011 Scottish Parliamentary Election	Journal Article
2013	Bekafigo, M. A.; McBride, A.	Who Tweets About Politics?: Political Participation of Twitter Users During the 2011Gubernatorial Elections	Journal Article
2013	Bruns, Axel; Burgess, Jean	Cumulative and comparative social media analytics for crisis communication	Journal Article
2013	Bruns, Axel; Highfield, Tim	POLITICAL NETWORKS ON TWITTER: Tweeting the Queersland state election	Journal Article
2013	Bruns, Axel; Stieglitz, Stefan	Towards more systematic <i>Towards more systemat</i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i></i>	Journal Article
2013	Chadwick, Andrew	The hybrid media system: politics and power	Book
2013	Christensen, Christian	WAVE-RIDING AND HASHTAG-JUMPING: Twitter, minority å© third partieså@" and the 2012 US elections	Journal Article
2013	Conway, B. A.; Kenski, K.; Wang, D.	Twitter Use by Presidential Primary Candidates During the 2012 Campaign	Journal Article
2013	Dang-Xuan, Linh; Stieglitz, Stefan; Wladarsch, Jennifer; Neuberger, Christoph	AN INVESTIGATION OF INFLUENTIALS AND THE ROLE OF SENTIMENT IN POLITICAL COMMUNICATION ON TWITTER DURING ELECTION PERIODS	Journal Article
2013	Elmer, G.	Live research: Twittering an election debate	Journal Article
2013	Enli, Gunn Sara; Skogerbø, Eli	PERSONALIZED CAMPAIGNS IN PARTY-CENTRED POLITICS: Twitter and FaceBook as arenas for political communication	Journal Article
2013	Gayo-Avello, D.	A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data	Journal Article
2013	Graham, Todd; Broersma, Marcel; Hazelhoff, Karin	Twitter as an instrument for connected representation	Journal Article
2013	Graham, Todd; Broersma, Marcel; Hazelhoff, Karin; van 't Haar, Guido	BETWEEN BROADCASTING POLITICAL MESSAGES AND INTERACTING WITH VOTERS: The use of Twitter during the 2010 UK general election cannaion	Journal Article
2013	Groshek, J.; Al-Rawi, A.	Public Sentiment and Critical Framing in Social Media Content During the 2012 U.S. Presidential Campaign	Journal Article
2013	Hawthorne. J.: Houston. J. B.: McKinnev. M. S.	Live-Tweeting a Presidential Primary Debate: Exploring New Political Conversations	Journal Article
2013	Hemphill, Libby: Otterbacher, Jahna: Shapiro, Matthew	What's congress doing on twitter?	Conference Paper
2013	Hong Southman	Who henefits from Twitter? Social media and nollifical commetition in the U.S. House of Renresentatives	lournal Article
2013	Jungherr, Andreas	Tweets and votes, a special relationship: the 2009 federal election in germany	Conference Paper
2013	Klinger, Ulrike	MASTERING THE ART OF SOCIAL MEDIA: Swiss parties, the 2011 national election and digital challenges	Journal Article
2013	Lee, E-J.; Jang, JW.	Not So Imaginary Interpersonal Contact With Public Figures on Social Network Sites: How Affiliative Tendency Moderates Its Effects	Journal Article
2013	Lee, Eun-Ju	Effectiveness of Politicians' Soft Campaign on Twitter Versus TV: Cognitive and Experiential Routes: Twitter Vs. TV	Journal Article
2013	Lee, Eun-Ju; Oh, Soo Youn	Seek and You Shall Find? How Need for Orientation Moderates knowledge Gain from Twitter Use: Twitter and News knowledge	Journal Article
2013	Maireder, Axel; Ausserhofer, Julian	Political discourses on T witter: networking topics, objects and people	Journal Article
2013	Moe, Hallvard; Larsson, Anders Olof	UNTANGLING A COMPLEX MEDIA SYSTEM: A comparative study of Twitter-linking practices during three Scandinavian election campaigns	Journal Article
2013	Otterbacher, Jahna; Shapiro, Matthew A.; Hemphill, Libby	Interacting or Just Acting?	Journal Article
2013	Parmelee, John H.; Bichard, Shannon L.	Politics and the Twitter revolution: how tweets influence the relationship between political leaders and the public	Book
2013	Plotkowiak, Thomas; Stanoevska-Slabeva, Katarina	German politicians and their Twitter networks in the Bundestag Election 2009	Journal Article
2013	Prior, Markus	Media and Political Polarization	Journal Article
2013	Roback, Andrew: Hemphill, Libby	l'd have to vote against you: issue campaigning via twitter	Conference Paper
2013	Strandberg, K.	A social media revolution or just a case of history repeating itself? The use of social media in the 2011 Finnish parliamentary elections	Journal Article
2013	Vaccari, Cristian; Nielsen, Rasmus Kleis	What Drives Politicians' Online Popularity? An Analysis of the 2010 U.S. Midterm Elections	Journal Article
2013	Vaccari, Cristian; Valeriani, Augusto; Barberá, Pablo; Bonneau, Richard; Jost, John T.; Nagler, Jonathan; Tucker, Joshua	Social media and political communication. A survey of Twitter users during the 2013 Italian general election	Journal Article

A.3.1 Corpus of academic literature related to political discourse on Twitter (2/3)
201	.3 Vergeer, M.; Hermans, L.; Sams, S.	nline social networks and micro-blogging in political campaigning: The exploration of a new campaign tool and a new campaign style	Journal Article
201	.3 Vergeer, Maurice; Hermans, Liesbeth	ampaigning on Twitter: Microblogging and Online Social Networking as Campaign Tools in the 2010 General Elections in the Netherlands	Journal Article
201	.3 Weber, Ingmar; Popescu, Ana-Maria; Pennacchiotti, Marco	LEAD'13: proceedings of the Workshop on Politics, Elections and Data : October 28, 2013, San Francisco, CA, USA	Book
201	Boynton, George Robert, Cook, James, Daniels, Kelly, Dawkins, Melissa: Kopish, Jory; Malkar, Maria; Ma McCavdi, Willam, Munty, Margaret; Osmundson, John; Steenblock, Taylor; Sudarmawan, Anthony; Mese, Philip, Socia, Alpariain	he Political Domain Goes to Twitter: Hashtrags, Retweets and URLs	Journal Article
201	(4 Dimitrova, D. V.; Shehata, A.; Stromback, J.; Nord, L. W.	he Effects of Digital Media on Political Knowledge and Participation in Election Campaigns: Evidence From Panel Data	Journal Article
201	.4 Gainous, Jason; Wagner, Kevin M.	weeting to power: the social media revolution in American politics	Book
201	.4 Godnov, Uros; Redek, Tjasa	he Use of Twitter for Political Purposes in Slovenia	Journal Article
201	.4 Jungherr, A.; Jurgens, P.	hrough a Glass, Darkly: Tactical Support and Symbolic Association in Twitter Messages Commenting on Stuttgart 21	Journal Article
201	.4 Jungherr, Andreas	he Logic of Political Coverage on Twitter: Temporal Dynamics and Content: Coverage on Twitter	Journal Article
201	.4 Larsson, A. O.; Moe, H.	riumph of the Underdogs? Comparing Twitter Use by Political Actors During Two Norwegian Election Campaigns	Journal Article
201	.4 Lee, EJ.; Shin, S. Y.	when the Medium Is the Message: How Transportability Moderates the Effects of Politicians' Twitter Communication	Journal Article
201	.4 Lin, Yu-Ru; Keegan, Brian; Margolin, Drew; Lazer, David	ising Tides or Rising Stars?: Dynamics of Shared Attention on Twitter during Media Events	Journal Article
201	.4 Maruyama, Misa T.; Robertson, Scott P.; Douglas, Sara K.; Semaan, Bryan C.; Faucett, Heather A.	ybrid media consumption: how tweeting during a televised political debate influences the vote decision	Conference Paper
201	4 Nilsson, B.; Carlsson, E.	wedish politicians and new media: Democracy, identity and populism in a digital discourse	Journal Article
201	(4 Parmelee, J. H.	he agenda-building function of political tweets	Journal Article
201	.4 Smith, Marc A.; Rainie, Lee; Himelboim, Itai; Shneiderman, Ben	fapping Twitter topic networks: From polarized crowds to community clusters	report
201	.4 Solo, Ashu M.G	olitical Campaigning in the Information Age	Book
201	.4 Wallsten, Kevin	ficroblogging and the News: Political Elites and the Ultimate Retweet	Chapter
201	15 D'heer, Evelien; Verdegem, Pieter	what social media data mean for audience studies: a multidimensional investigation of Twitter use during a current affairs TV programme	Journal Article
201	Larsson, A. O.	ioing viral? Comparing parties on social media during the 2014 Swedish election	Journal Article
201	L5 Mirer, M. L.; Bode, L.	weeting in defeat: How candidates concede and claim victory in 140 characters	Journal Article
201	.5 Sobaci, Mehmet Zahid; Eryiğit Kadir Yasin; Hatipoğlu Ibrahim	he Net Effect of Social Media on Election Results: The Case of Twitter in 2014 Turkish Local Elections	Chapter
201	.5 Vaccari, C.; Valeriani, A.	ollow the leader! Direct and indirect flows of political communication during the 2013 Italian general election campaign	Journal Article

A.3.1 Corpus of academic literature related to political discourse on Twitter (3/3)

A.4 Addenda to Chapter 4

A.4.1 JSON Formatted Tweet Object Example

Tweet object returned via Twitter REST API v1.1, September 2014. (Key fields highlighted in bold). Note that informed consent was sought and received from this user for the inclusion of this tweet as an example of the format, since doing so necessarily exposed the user's full name, location, image, bio, etc.

Query: GET http://api.twitter.com:80/1.1/statuses/show.json?id=323850827544555520

HTTP/1.1 {

```
contributors = null,
text = Anti-gay folks on the radio all day calling #ccven skewed but they
       forget there just isn't an actual divide among experts on equal
       marriage,
geo = null,
retweeted = false,
in reply to screen name = null,
truncated = false,
lang = en,
entities =
                       symbols =
                                       [],
                       urls =
                                       [],
                       hashtags =
                                       [
                                       text = ccven,
                                        indices = [44, 50]
                                        }
                                      ],
                       user mentions =[ ]
               },
in reply to status id str = null.
id = 323850827544555520.
source = web,
in_reply_to_user_id_str = null,
favorite = false,
in_reply_to_status_id = null,
retweet count = 0,
created at = Mon Apr 15 17:30:28 +0000 2013,
in reply to user id = null,
favorite_count = 2,
id str = 323850827544555520,
place = null,
user =
               location = Thurles, Co. Tipperary,
               default profile = true,
               profile background tile = false,
               statuses count = 19621,
               lang = en,
               profile_link_color = 0084B4,
profile banner url =
               https://pbs.twimg.com/profile banners/18290913/1363918687,
               id=18290913,
               following = false,
               protected = false,
               favourites count = 3213,
               profile text color = 333333,
               description = 23 year old business graduate. Leftist, likes pop
               culture, feminism and politics,
               verified = false,
               contributors enabled = false,
               profile sidebar border color = CODEED,
               name = David Gormley,
               profile_background_color = CODEED,
               created at = Sun Dec 21 19:14:15 +0000 2008,
               default profile image = false,
               followers count = 1150,
               profile_image_url_https =
               https://si0.twimg.com/profile images/1410309122/230305 normal.j
               pq,
               geo enabled = false,
```

```
profile background image url =
                  http://a0.twimg.com/image_url_https =
profile_background_image_url_https =
                  https://si0.twimg.com/images/themes/theme1/bg.png,
                  follow_request_sent = false,
                  entities =
                                     {
                                    description=
                                                       {
                                                       urls=[]
                                                       },
                                    url=
                                                       {
                                                       urls=
                                                               [
                                                                 {
                                                                 expanded_url =
                                                                 http://www.davidjgormley
                                                                 .wordpress.com,
                                                                 indices = [0, 22],
                                                                 display_url =
                                                                 davidjgormley.wordpress.
                                                                 com,
url =
                                                                 http://t.co/lIS11wHOBm
                                                                 }
                                                                ]
                                                       }
                                    },
                  url = http://t.co/lIS1lwHOBm,
                  utc_offset = 3600,
time_zone = Dublin,
notifications = null,
                  profile_use_background_image = true,
friends_count = 449,
profile_sidebar_fill_color = DDEEF6,
                  screen_name = DayvyG,
                  id_str = 18290913,
                  profile_image_url =
                  http://a0.twimg.com/profile images/1410309122/230305 normal.jpg
                  listed_count =19,
                  is_translator = false
                  },
coordinates = null
```

}

A.4.2 Stars and Hearts – Implications of changing Twitter affordances

The "favorite" [*sic* – *US spelling was used, regardless of locale*] option on Twitter was introduced in November 2006, intended as a bookmarking mechanism and shown as a star button with the word "favorite", under each tweet. Users who "come across a Twitter update that [they] really like [could] save it by clicking the little star" (Stone 2006). Within Twitter's web and mobile clients, the number of favorites accumulated was not shown in the default timeline view; users could only access this information upon opening the more detailed view of the tweet (e.g. Fig 4.1).

Users referred to the use of the feature as *favouriting* or *starring* a tweet, though it was initially difficult to establish whether this behaviour was equated with *liking* a Facebook post, whether it was indicative of a specifically positive sentiment about the tweet, or what the total connotations of the feature were for the user (Ellison & boyd 2013). Similarly, without the context of the number of favorites on display in the default view, it is unlikely that the feature was associated with a weighting value for the merit or worth of the tweet when it was first introduced. Changes to the mediation of this information in 2014 led to the number of favorites (and retweets) being directly visible for each tweet in the default timeline view, offering these as metrics of the tweets success for users. There is little discussion in the literature about the action and perceived purpose of favouriting a tweet, though one study found that the feature was being "overly repurposed", with users making use of the button to show like, agreement, appreciation of informational value, approval, and for bookmarking (Meier et al. 2014). This study settled on a number of 'most identified' connotations amongst these various usages, which included indicating agreement, and showing appreciation. The authors of the study, as well as others in the literature (e.g. boyd 2015) have therefore suggested that the *favorite count* could be used as a metric for the influence of a tweet, since many users have been observed using the favorite, rather than retweet, option on content they agree with or appreciate. The use of this variable could therefore potentially capture a hither-to unmeasured set of more passive user interactions around content.

However, in November 2015, Twitter again changed the mediation of this affordance, modifying the symbolic button for its use from a star to a heart and dubbing the process a *like* rather than a *favorite* (Newton 2015). While this change may help to bring homogenise the action with that of a Facebook like, and result in more consistent intent by users employing it, these repeated changes make it difficult for researchers to reliably associate the metric with a particular action or intention through time, and would require any use of this metadata in studies to be subject to segmented analyses by the different periods in which the feature was variously presented, as well as significant caveats about its interpretation. For this reason, the metadata associated with favorites have tended to go largely ignored in the literature, and may continue to be so for some time.

A.4.3 Data collection software screenshots

A.4.3.1 Single archive default view in yourTwapperKeeper



A.4.3.2 Archive list view in yTK_DCU_v3

cf. Fig 4.3, Chapter 4

A.4.4 Sample Participants' Guide

PhD study on the context of political discourse on Twitter Dónal Mulligan (donal.mulligan2@mail.dcu.ie)

A Guide for Interview Participants

This guide will provide you with some detail of the interview process and structure, outlining the topics I would like to address and the ethical considerations given to data collection involved and your privacy.

Please feel free to contact me by email if there is any further issue that I can clarify for you.

Why is this research happening?

These interviews form part of the research being undertaken by Dónal Mulligan toward a PhD investigating factors related to political discourse that takes place on Twitter. It is funded by, and takes place in, the School of Communications at Dublin City University.

Specifically, these interviews aim to assess your attitudes to, opinions on, and actions related to contributions you make on Twitter as part of discussions of contemporary politics in Ireland. The outline of topics below will help you to assess the range of this component of the research.

The project also involves automated archiving and analysis of public Twitter content related to political discourse in Ireland, including content from you that included certain hashtags or keywords. That component of the research is intended to form big-picture findings related to the discussion generally, which your interview can provide specific and personal context for. You will not be asked about specific tweets you've sent or comments you've made, though your identification as a potential participant in this research means that this automated analysis included you and highlighted you as a user of interest because of the frequency of your interactions, the number of your followers, etc.

Interview Topics

- Twitter as a medium for political communication
- Political communication online more generally
- How groups form around topics on Twitter
- Twitter's role in political change
- The specific case of Ireland and politics on Twitter here
- Your adoption and use of Twitter for political discourse
- Your opinion on the benefits and limitations of it
- Your own use of Twitter features (e.g. retweets, hashtags)

Your interview will be loosely structured around these topics, allowing you time to provide the level of information you're comfortable with for each, and having scope for you to discuss related issues.

School of Communications, Dublin City University (http://dcu.ie/communications) Supervisor: Dr. Miriam Judge (miriam.judge@dcu.ie) PhD study on the context of political discourse on Twitter Dónal Mulligan (donal.mulligan2@mail.dcu.ie)

Interview Structure, Location & Duration

I wish to conduct the interview with you in a loosely structured format, which will permit me to cover those broad topics stated, while allowing you the flexibility to take our discussion outside of these questions to issues you might feel are important to include. This format also allows us to deal with further questions that might arise in the course of your responses, including questions you may have for me. It is my intention to format the interview so as to allow you to be confident that your opinions can be fully expressed, while retaining focus on the central topics of interest to my research.

I have access to private meeting rooms at Dublin City University (located in Glasnevin, Dublin 9), which are available to you should you chose to be interviewed there. I will, however, offer you the choice of location, as I understand that another location may be more comfortable or convenient to you.

The interview will last around 45-60 minutes.

Data Collection & Privacy

So as to allow you to be fully informed of the treatment of data collected during the interview and to help you to feel comfortable about fully and accurately giving your opinion on the topics, I would like to offer the following information:

Your interview will be audio-recorded. To allow me to give you my full attention and to accurately assess our discussion during later analysis, I would like to record our interview. This recording will be kept for a period of 24 months after the submission of the study for examination, on a password-protected computer in my office at DCU, after which it will be deleted. The conversation will be transcribed as part of the thesis, and will be available in full only to my supervisor and those examiners involved with the assessment of this research. This transcript, your signed consent form, and any other information which may identify you will be kept securely by my supervisor (below) and subject to the data storage policies outlined by DCU's Research Ethics Committee. No part of this information will be reproduced in any other publication without your prior consent (see "Reference or Quotation in Other Publications", below).

Your involvement, for which I am very grateful, is voluntary and you may choose to withdraw from the interview at any point, or to decline to answer any question asked.

The privacy of contributions made by you will be determined by your instruction in the consent form, and I offer varying degrees of anonymity for you to select from prior to interview. Some people feel most comfortable to fully express their opinions when they are assured in advance of full anonymity while others prefer to be fully credited in the work and might for the valuable contribution of their time and information. It is certainly an aid to the authenticity and richness of the research to provide some detail of the sources, so a choice of several levels is available to you as follows:

School of Communications, Dublin City University (http://dcu.ie/communications) Supervisor: Dr. Miriam Judge (miriam.judge@dcu.ie) PhD study on the context of political discourse on Twitter Dónal Mulligan (donal.mulligan2@mail.dcu.ie)

Full anonymity	You prefer that no information about your identity be included in my final thesis
Partial anonymity	You prefer to be referred to by profile information, including political affiliation, or professional association with politics. e.g. "A political blogger" or "An Independent opposition TD"
Full Credit	You prefer to be credited by name, including reference to your Twitter username e.g "Dónal Mulligan (@donalmulligan)"

You will be asked to select from one of these options on the Consent Form.

Reference or Quotation in Other Publications

Since my research is on a highly contemporary subject, contributions made by me prior to the completion of my thesis may be of interest to other scholars and I may write one or more academic papers for publication or conference discussion. I would like, where relevant, to include details from interviews which can add context and rich nuance to such papers. I am therefore asking you separately to allow me to make reference to your interview responses in such work, with your prior permission. Before the publication of any such work, I will provide you with a copy for approval and you may withdraw your consent for that specific publication, in which case I will remove any such references from that paper.

School of Communications, Dublin City University (http://dcu.ie/communications) Supervisor: Dr. Miriam Judge (miriam.judge@dcu.ie)

A.4.5 Sample Consent Form

PhD study on the context of political discourse on Twitter Dónal Mulligan (donal.mulligan2@mail.dcu.ie)
CONSENT FORM
I wish to state my consent to be interviewed by Dónal Mulligan, <i>PhD researcher in</i> Online Communications at Dublin City University School of Communications.
I understand that this interview forms part of Dónal Mulligan's School of Communications funded doctoral research project related to political discourse on Twitter in Ireland.
I agree to the interview bring conducted, recorded, transcribed, archived and destroyed according to the details supplied to me in the Participants' Guide
I understand the choices with regard to the anonymity of my contributions, which are detailed in the Guide, and I agree and declare my choice by checking one of the boxes below.
[] Full Anonymity
[] Partial Anonymity
[] Full Credit
I consent to reference and quotation of my interview responses in other publications, with my prior review.
[] Yes
[]No
Signed: Print Name:
Date:
School of Communications, Dublin City University (http://dcu.ie/communications)
Supervisor: Dr. Miriam Judge (miriam.judge@dcu.ie)

A.4.6 Interview Questions

Interview questions for participants identified algorithmically as frequent contributors to the discourse on politics in Ireland. These questions are given in the order used and formed the basis of a semi-structured interview specifically designed to allow the participant to develop their responses on particular topics and themes (Kvale 2007). Corresponding research questions and themes are outlined in italics after each broad group.

Q 1: My research has shown that there is an identifiable community of people in Ireland who regularly take part in discussions about current political issues, and that you are one such person. Tell me a little about how you came to use Twitter...

Informs the participant about the study, eases their first response by allowing for a personal narrative point of view, and provides an opportunity for participants to pose questions about this research or clarify anything not understood from the guide.

[Follow up, where relevant] How did you come to use Twitter specifically for discussion of political issues?

Q 2: Could you identify other users with whom you regularly interact when discussing political issues on Twitter? I don't need specific usernames, but I'd like to know whether you're aware of regularly communicating with the same people on these topics.

Prompts a participant to examine the social aspect of the medium, to situate their contributions on these topics, and to evaluate the degree of awareness of an identifiable group of regular correspondents.

[Follow-up] What associations do these users have to politics? Are they professionally involved?

Elicits responses related to key research sub-questions: Who are the participants and what are the limits to and conventions of participation? What is the level of awareness among participants of the link between their correspondents' standing on Twitter and in the fields of Politics or Media.

Q 3: Could you give an example of a recent discussion that comes to mind, letting me know who was involved and how it unfolded?

Develops (Q2), but contextualizes this with regard to a specific interaction. Provides an opportunity to evaluate whether the topic fits the political discourse theme, and to re-prompt if necessary.

Q 4: Are there other ways you discuss politics online and do these involve the same people?

Elicits responses related to key research sub-questions: What is the nature of online political discourse and how has it come about? How does Twitter mediated political discourse compare with elsewhere online?

Q 6: Do you think you adapted to Twitter by taking previous ways of communicating about politics and using them there? Did you talk to the same people about the same sorts of things with the same frequency before, for example?

Evaluates comparative mediation of this discourse and develops the theme of interaction between fields. Elicits responses related to key research sub-questions: *How do highly active users engaged in political discourse on Twitter view its relationship as a technology to the discourse it mediates and the people taking part? How much do factors of normalisation play a part here, with users porting over pre-existing behaviours and networks?*

Q 6: Thinking still about your own use of Twitter to take part in these kinds of discussion, what benefits do you think it affords you and those you interact with?

Prompts consideration of Twitter as medium and the relationship of specific affordances to discourse and social power.

Q 7: How do you organise the discussion you have on Twitter, where so much is happening at any given time? How does Twitter itself help you do this?

Assesses participant's knowledge of organizational and information-streaming affordances in Twitter (hashtags, lists, favourites, etc.)

[Prompt for specific information about the client used for access (e.g. Twitter website, TweetDeck, iPhone app, etc., if not given]

Q 8: Do you regularly use hashtags in your tweets? How?

Directly introduces the research-critical area of hashtags, and understanding of their function and established conventions. Establishes perceived frequency of own use, comparable with actual computed value.

[Follow-up] What does it signify to you when someone else's tweet contains a hashtag?

Assesses participant understanding of conventions of use.

[Follow-up] Are there situations where you would <u>not</u> use a hashtag, despite what you're saying belonging to a particular topic? Why?

Prompts, if not already volunteered, discussion of specific incidences in which hashtags are and are not used. Develops a key research sub-question: *Among active participants in political discourse on Twitter, are there identifiable conditions in which specific hashtags might be expected but are not used?* This has strong implications for selection bias in current collection strategies.

Q 9: Can you remember a recent hashtag you used and tell me a little about it – who else is using it and in what ways?

Contextualises hashtag use by examination of a specific example; allows for prompting back toward previous question and follow-ups if these were difficult in the abstract.

Q8 & 9 share key research sub-questions: How does a multiplicity of publics form around political issues on Twitter and how does the medium of their formation affect their function? What are the affordances and conventions of Twitter hashtags and what limits their use? Does the studied behaviour, motivation, and action of such actors match the rational and deliberative role ascribed to them by the assumptions of quantitative study?

Q10: Do you think political discussions taking place on Twitter are being translated into action? How?

Elicits social and political contextualisation of the discourse. Addresses the research theme of transfer of power between fields.

Q 11: Thinking about Twitter's role in bringing about political change, are there any examples you could give me of such changes globally, or in Ireland?

Develops Q10 with context of a specific event.

Q 12: Thinking specifically of the Arab Spring⁸⁸ and the reporting of Twitter's involvement in the organisation of protest, how do you assess that involvement and Twitter's role in it?

Specifically addresses a frequently cited example of political action linked to Twitter discourse and mobilisation.

Key research sub-questions: How have concepts of the radical potential of Twitter to mobilise publics been borne out? How can we understand Twitter as a medium enabling both radical revolutionary mobilisation and conservative debate? To what extent has the technological nature of Twitter shaped and been shaped by the discourse it mediates?

Q 13: Are there any factors related to how we use Twitter for political discussion or translate that discussion into change or action that you think are specific to Ireland?

Addresses a sub-question: What is the awareness, among highly engaged participants, of the social and political contextual factors for their Twitter discourse. Prompts participant consideration of transfer of power structures, and normalised conventions, from related fields.

Q 14: As I've been reading a lot about the effect of technology on society and vice versa, I've seen that some scholars think that technology affects us and how we communicate and organise ourselves without it being influenced itself. Others think that when we use communications technologies we shape it by finding ways to apply it and make it suit our needs.

Thinking specifically about Twitter and your experience of it, what would you have to say about that relationship between society and technology in this case?

Introduces a primary philosophical underpinning of the research – *the technology-society* relationship – and elicits comment on it from participants. This question allows a re-framing of the previous discussion via a stepping back and prompts a holistic review of Twitter as medium and technology and the implications of this for the discourse within it.

⁸⁸ The Arab Spring was used as the example for the earliest interviews in this study. During the development of the research, other Ireland-specific examples were used, and the 2015 Marriage Referendum was specifically used in later interviews, especially with agents who were substantially involved in it.

A.4.7 Overview of Interviews Conducted

Interviewees in the study are coded using a designation of their research category, and a number corresponding to the order of the interview in the series for that category. Categories are (NP: National Politicians, LP: Local Politicians, A: Activists, PM: Professional Media, OM: Other Media, C: Citizens). Thus, *OM2* would indicate the second interviewee among those participants associated with Irish politics through their role in Other Media, such as blogs.

Code:	Info:	Date:	Location:	Duration:
Pilot1	Post-doc researcher	29/06/2015	Leinster House, Dublin	36m 10s
Pilot2	Senator	29/06/2015	Leinster House, Dublin	21m 19s
NP1	Opposition TD (SF)	30/09/2014	Leinster House, Dublin	34m 09s
$NP2^{89}$	Opposition TD (Ind ⁹⁰)	09/10/2014	Leinster House, Dublin	44m 19s
NP3	[Withdrew]	12/10/2014	Leinster House, Dublin	31m 22s
NP4	Opposition TD (FF)	14/01/2015	via Skype	19m 34s
NP5	Government TD	10/02/2015	Leinster House, Dublin	32m 58s
NP5x91	Parliamentary Aide	10/02/2015	Leinster House, Dublin	05m 03s
NP6	[Withdrew]	10/02/2015	via Skype	22m 08s
NP7	Former TD (GP)	13/04/2015	Interviewee's home, Dul	olin 30m 59s
LP1	Dublin Councillor (GP)	21/01/2015	Interviewee's office, Dub	olin 22m 39s
LP2	City Councillor	22/01/2015	Interviewee's home	26m 31s
LP3	Dublin Councillor (SF)	17/02/2015	SF Party HO, Dublin	43m 16s
LP4	Councillor (Ind)	25/02/2015	Rural midlands town	33m 11s
Al	NGO leader	11/02/2015	Interviewee's home, Wid	cklow 44m 09s
A2	LGBT Activist	14/04/2015	Dublin City University	22m 02s
A3	LGBT Activist	28/07/2015	Interviewee's office, Dub	olin 31m 43s
PM1	National Print Journo	03/11/2014	Dublin City University	31m 02s
PM2	National Online Editor	14/01/2015	Dublin City University	49m 12s
PM3	Political Correspondent	26/06/2015	via Skype	18m 18s
OM1	Political Blogger	04/04/2015	Dublin City University	40m 00s
OM2	Political Blogger	11/06/2015	Westbury Hotel, Dublin	35m 11s
C1	Tech professional	10/09/2014	Participant's home, Dub	lin 54m 34s
C2	Marine Researcher	21/09/2014	Dublin City University	37m 32s
C3	Politics PhD Student	07/07/2015	Dublin City University	44m 56s
			, , ,	

⁸⁹ There were two participants for this interview and both opted for full credit. They are Catherine Murphy TD and her Communications Director Anne-Marie MacNally. In this study, these participants are identified as NP2a and NP2b

⁹⁰ This participant was an independent member of the opposition at the time of interview, and her responses are contextualised by this, rather than by her later affiliation with a political party.

⁹¹ This interview is a special case, as it was with the parliamentary assistant to the participant identified as NP5, and carried out at NP5's suggestion. The interview is short and concerns the Twitter output of NP5 rather than that of the interviewee in question.

A.6 Addenda to Chapter 6

A.6.1 Example conversation with a user (via Twitter Direct Message, 04 February 2014), about the use of the #ssmref hashtag as an alternative to #marref



Note that this image is altered; the user's profile picture has been pixelated, at their request, to preserve their anonymity.

A.6.2 Tweets, tagged #marref, captured per minute by yTK_DCU_v3, during Referendum Count (23 May 2015)



A.6.3 Table of the 50 most frequent contributors to

#marref-tagged tweets, coded

Note: Stance coding corresponds to the users declared position on the marriage referendum, as evidenced in their tweets and Twitter profile page. Codes: 0 = neutral - user has an officially neutral stance, such as was the case with most media outlets; 1 = in favour – user has declared an intention to vote yes, or actively campaigns for others to do so; 2 = opposes – user has declared an intention to vote no, or actively campaigns for others to do so.

Twitter	Number of	Stance	Actor
Username:	Tweets	Coding	Coding
@Martin17773	2382	2	6
@TheScarletElf	1963	2	6
@AntoniaPDon	1612	2	6
@IzzyKamikaze	1332	1	6
@MFM_Ireland	1276	2	3
@pmbarrins	1222	2	6
@david75donovan	1211	1	6
@HangBlaa	1087	1	6
@HelenORahilly	1068	1	6
@Cunionsandphey	946	1	6
@Colmogorman	905	1	3
@CaroleLeger	887	1	6
@Paul71	785	2	6
@timelesstotty	734	1	6
@kencurtin	720	1	6
@YesEquality2015	690	1	3
@Lamhfada	689	1	6
@Omaniblog	661	1	6
@daithigorman	635	1	6
@YesEqualityCork	619	1	3
@Independent_ie	614	0	4
@Dodslaw	565	2	6
@Blackhall99	550	1	6
@TheEvertBopp	518	2	6
@RebeccahLouise	508	1	6
@Cgoraghallaigh	495	1	6
@Khayden2412	485	1	6
@alanearly	484	1	6
@pauljoneill	477	2	6
@mrs_bopp	465	2	6
@hometime_music	435	1	6
@thomasJKW	433	2	6
@spiller2	424	1	6
@Donal_OKeeffe	423	1	6
@NewstalkFM	418	0	4

@BPDCarey	410	1	6
@KeithM	408	2	6
@kehbayhoody	405	1	6
@NORTHANNE	402	2	6
@BabyMonsterGoon	385	1	6
@YewtreeGirl	377	1	5
@anita_furlong	373	1	6
@GallMandy	372	2	6
@pmcchambers	364	1	6
@MPTIreland	362	2	3
@MaryFitzgibbon	362	2	2
@LaobhaoiseNihE	360	1	6
@BeLonG_To	359	1	3
@brian_dneen	359	1	6
@Dailigh	358	1	6

A.6.4 Table of the 50 most frequently @-mentioned

contributors in #marref-tagged tweets, coded

Note: Stance coding corresponds to the users declared position on the marriage referendum, as evidenced in their tweets and Twitter profile page. Codes: 0 = neutral - user has an officially neutral stance, such as was the case with most media outlets; 1 = in favour – user has declared an intention to vote yes, or actively campaigns for others to do so; 2 = opposes – user has declared an intention to vote no, or actively campaigns for others to do so.

Username	Number of	Stance	Actor
@-mentioned:	Tweets	Coding	Coding
@YesEquality2015	6742	1	3
@Colmogorman	2504	1	3
@DavQuinn	2293	2	3
@IrishTimes	2081	0	4
@MFM_Ireland	1776	2	3
@KeithM	1545	2	6
@UnaMullally	1350	1	4
@labour	1337	1	1
@PantiBliss	1161	1	3
@ireland	1144	0	4
@Independent_ie	938	0	4
@EndaKennyTD	890	1	1
@YesEqualityCork	807	1	3
@mrs_bopp	769	2	6
@FineGael	698	1	1
@YouTube	679	0	4
@RonanMullen	670	2	1
@NewstalkFM	613	0	4
@MPTIreland	608	2	3
@PaddyJManning	606	2	6
@marriagequality	591	1	3
@noelwhelan	586	1	1
@MariaLaoise	560	2	6
@TheEvertBopp	555	2	6
@BeLonG_To	545	1	3
@jerrybuttimer	543	1	1
@GallMandy	524	2	6
@thejournal_ie	523	0	4
@morningireland	519	0	4
@SenatorKZappone	518	1	1
@simoncoveney	518	1	1
@Paul71	515	2	6
@TheScarletElf	509	2	6
@campaignforleo	497	1	1
@AodhanORiordain	486	1	1

@rte	485	0	4
@MaryFitzgibbon	471	2	2
@HelenORahilly	468	2	6
@fiannafailparty	453	0	1
@HangBlaa	436	1	6
@markohalloran	425	1	6
@rtenews	424	0	4
@IzzyKamikaze	382	1	5
@RefCom_ie	370	0	1
@SenDavidNorris	341	1	1
@daithigorman	334	1	6
@johndwalsh	317	2	6
@AmnestyIreland	312	1	3
@kencurtin	302	1	6

A.6.5 Table of the 50 hashtags most frequently co-

occurring with #marrer in the dataset				
Hashtag	Occurrence	Hashtag	Occurrence	
#VoteYes	30274	#YesEquality2015	880	
#YesEquality	21444	#MakeItHappen	876	
#VoteNo	6799	#VoterMotor	848	
#Ireland	5579	#LetsMakeHistory	798	
#hometovote	4349	#Eurovision	763	
#Yes	4128	#LateLateShow	743	
#MakeGraTheLaw	4048	#love	714	
#equality	3986	#Eurovision2015	675	
#MarriageEquality	3305	#WeMadeHistory	647	
#rtept	3004	#IrelandSaysYes	637	
#VoteYesForEquality	2873	#Cork	631	
#vinb	2407	#BeLonGToYES	622	
#cblive	2279	#Dublin	615	
#EqualityForAll	1982	#equalmarriage	604	
#latelate	1520	#marriageeqaulity	577	
#MarRef2015	1462	#marian	569	
#IrelandVoteYes	1422	#YesToEquality	562	
#LGBT	1366	#ArasRef	548	
#IVoted	1354	#VoteWithUs	543	
#marriagequality	1221	#IrelandSaidYES	529	
#proudtobeirish	1202	#BeMyYes	529	
#PresRef	1119	#IRL	497	
#YesVote	1074	#ntfm	495	
#LoveIsLove	972	#Irish	466	
#YestoLove	938	#No	457	

occurring with #marref in the dataset